Connections between biodiversity and ecosystem functioning in large-scale natural ecosystems

James G. Hagan
Doctoral thesis

UNIVERSITY OF GOTHENBURG

Department of Marine Sciences
Faculty of Science
2023

SVANENMÄRKET

Trycksak
3041 0234

Front cover: A biodiverse landscape illustrating different types of species interactions that can generate biodiversity effects on ecosystem functioning: i. local niche partitioning as illustrated by a moose and deer feeding on different plants ii. dominance by a single species shown by a single barnacle species covering the intertidal zone iii. spatial niche partitioning where different algae species occupy different depth zones and different snail species feed on different algae species and iv. temporal niche partitioning where birds and bats feed on insects at different times of the day.

Abstract

Billions of years of evolution have given us a planet that supports a remarkable diversity of life. Estimates for the number of Eukaryotic species frequently number in the millions and the Prokaryotes are much more diverse than that. This *biodiversity* makes up the ecosystems that we, as humans, rely on to sustain almost every aspect of our lives. But, despite our reliance on these biodiverse ecosystems, we are eroding them at an alarming rate through habitat destruction, overexploitation and our transformation of the climate. Indeed, some estimates suggest that the rate at which species are going extinct is as high as previous mass extinction events that have sporadically occurred throughout earth's history. How will this loss of biodiversity affect the functioning of ecosystems that we rely on? How much biodiversity do we need for healthy ecosystems? These are some of the questions that researchers began to address in the early 1990's. Based on hundreds of experimental manipulations of biodiversity, there is a general scientific consensus that biodiverse ecosystems tend to be more stable and more productive than depauperate ones. However, much of this work has taken place in artificial, experimental systems and at small scales of space and time. Thus, several questions remain. For example, if small-scale experiments show that biodiversity is important for ecosystem functioning, will the effects be the same at large scales? If ten species are required to maximise ecosystem functioning in a one square meter experimental grass patch, how many are required in a whole meadow, or in a landscape with many meadows? In my thesis, I attempt to extend our knowledge so we can better understand the consequences of biodiversity loss in natural systems and at larger scales of space and time.

In **Paper I**, I re-examined experimental work on biodiversity and ecosystem functioning from the last 30 years through the lens of community assembly theory. The aim was to understand what these experiments may tell us about how biodiversity loss will impact ecosystem functioning in natural ecosystems. My analysis showed that there are probably many cases where the results of experiments will not easily transfer to natural ecosystems. Rather than studying the community of species present in a local place, as is done in the experiments, I argue that we should instead focus on the pool of species present in the whole landscape, and the processes that govern the composition of local communities.

Many experiments performed over the last 30 years have shown that a high-diversity community of species is only rarely higher functioning than the highest functioning single species (i.e. monoculture). In **Paper II**, I used a set of theoretical simulations, an experiment in a bacteria-based model system, and a synthesis of previously published experiments to show that this may be because experiments have been performed in relatively homogeneous environments. When environmental heterogeneity increased, we found that the functioning of diverse species mixtures increased relative to the highest functioning monocultures.

But, despite the general trend observed in **Paper II**, there were many experiments in the synthesis where a single species in monoculture was highest functioning across the range of environmental conditions. This contradicted many theoretical models for the effect of biodiversity on ecosystem functioning. Thus, in **Paper III**, I wanted to study species along an environmental gradient to see if we would obtain similar results. I did this using a transplant experiment with four common species of marine seaweeds on Swedish rocky shores. These species occupy relatively distinct depth zones on the shores, which are characterised by different environmental conditions. I thus hypothesised that the four species would grow best at the depth where they are most common.

Counter to my predictions, the experiment showed that only one species responded strongly to being transplanted to a different depth zone.

For **Paper IV**, I took the results obtained from **Paper III** and attempted to model what would happen to the biomass production of the seaweed communities if each of the four species went extinct. I found that the biomass production of these rocky shore communities would probably only be strongly affected if one of the seaweed species (*Fucus vesiculosus*) went extinct. This is because the four species showed high productivity outside of the depth zones where they are naturally found and, therefore, may be able to compensate for the loss of any of the other species.

Arguably the most direct way to calculate an effect of biodiversity on ecosystem functioning is to compare a mixture of interacting species to a null expectation where species do not interact based on species' functioning in monoculture. However, in natural systems, this is generally not possible because we rarely have natural monocultures. In **Paper V**, I developed a Bayesian analytical pipeline to impute missing monoculture data which enables comparisons of mixtures and monocultures in natural ecosystems. Combined with a previously developed statistical partition, I was able to show that a combination of local-scale species interactions, local-scale dominance by a few high functioning species and spatial niche partitioning all contributed to a positive effect of biodiversity on ecosystem functioning in two, natural marine ecosystems.

Based on these five papers, I conclude that the hundreds of experiments that have been done to date provide useful but imprecise information about how biodiversity loss may affect the functioning of natural ecosystems. To understand the ecosystem-level effects of biodiversity loss more thoroughly, we will need to carefully study how biodiversity is changing across multiple scales of space and time and use methods that can detect the consequences of these changes. **Papers IV** and **V** suggest avenues for how this may be done.

Populärvetenskaplig sammanfattning

Vårt hem, som vi kallar planeten jorden, är så vitt vi vet unik i världsrymden. Till skillnad från grannarna i vårt solsystem myllrar jorden av liv. Hav och sjöar är fyllda med fisk och plankton, och på land har vi frodiga skogar och bördiga gräsmarker. Till och med på de mest ogästvänliga platserna finns det liv. Denna mångfald är fascinerande i sig: färgglada korallrev, savanner med zebror, lejon och elefanter, eller varför inte en havsörn som majestätiskt seglar på varma uppåtvindar. Men den biologiska mångfalden är också av största betydelse för mänskligheten. Vi skulle inte klara oss utan alla de tjänster och varor som naturen tillhandahåller.

Men trots vårt beroende av den biologiska mångfalden, utarmar vi den i en alarmerande takt genom förstörelse av livsmiljöer, överexploatering och vår påverkan på klimatet. Mycket tyder på att hastigheten med vilken arter dör ut är lika hög som tidigare massutrotningshändelser under jordens historia. Hur påverkar denna förlust av biologisk mångfald funktionen i de ekosystem som vi är beroende av? Hur många arter behövs för ett väl fungerande ekosystem? Det här är några av de frågor som forskare började studera i början av 1990-talet. Baserat på hundratals experimentella manipulationer av biologisk mångfald råder det idag vetenskaplig konsensus om att intakta ekosystem med hög biologisk mångfald tenderar att vara mer stabila och mer produktiva jämfört med utarmade ekosystem. Mycket av detta arbete har dock skett i artificiella, experimentella miljöer på liten skala i både rum och tid. Det finns därför många obesvarade frågor. Om till exempel småskaliga experiment visar att biologisk mångfald är viktigt, är effekterna desamma eller större på andra skalor? Om tio arter krävs för att maximera produktionen i en gräsplätt på en kvadratmeter, hur många krävs på en hel äng eller i ett landskap med många ängar? I min avhandling försöker jag öka vår kunskap i dessa frågor så att vi bättre kan förstå konsekvenserna av en utarmad biologisk mångfald på större skalor i naturliga ekosystem.

Min avhandling innehåller fem artiklar. I **Artikel I** utvärderade jag 30 år av experiment som studerat hur variation i biologisk mångfald kopplar till olika ekosystemfunktioner. Syftet var att förstå vad dessa experiment kan säga oss om hur förlust av biologisk mångfald påverkar ekosystemens funktion i naturliga ekosystem. Min analys visade att det förmodligen finns många fall där resultaten i experimenten inte lätt kan översättas till naturliga ekosystem. I stället för att studera de arter som finns på en enskild plats, som man gör i experimenten, argumenterar jag därför för att man bör fokusera på den pool av arter som finns i hela landskapet, samt på de processer som styr den lokala artsammansättningen.

Många experiment som gjorts de senaste 30 åren har visat att ett samhälle med hög biologisk mångfald sällan har högre produktion jämfört med den enskilt mest produktiva arten (i monokultur). I **Artikel II** använde jag en uppsättning av teoretiska simuleringar, ett experiment med bakterier, och en syntes av tidigare publicerade experiment för att visa att detta kan bero på att tidigare experiment har utförts i relativt homogena miljöer. När jag ökade miljövariationen fann jag att produktionen hos mer diversa artsamhällen ökade i förhållande till de mest produktiva monokulturerna.

Men trots den allmänna trenden i Artikel II, hittade jag många experiment i min genomgång i vilka en enskild art i monokultur uppvisade högst funktion oberoende av vilken typ av miljö de befann sig i. Detta står i kontrast mot vad jag förväntade mig. Enligt väl etablerad ekologisk teori bör betydelsen av biologisk mångfald öka med ökad miljövariation. I **Artikel III** ville jag studera arter längs en miljögradient för att se om jag skulle få liknande resultat. Jag gjorde detta med hjälp av ett transplantationsexperiment med fyra vanliga arter av marina brunalger på svenska klippstränder.

Dessa arter återfinns i havet i relativt distinkta djupzoner på stränderna, zoner som kännetecknas av olika miljöförhållanden. Till exempel är de arter som lever i den grundaste zonen utsatta för regelbunden uttorkning och kraftiga temperaturväxlingar, medan de arter som lever djupare måste klara av konkurrens från andra arter och betning av olika snäckor. Jag antog därför att de fyra arterna skulle växa bäst på det djup där de är vanligast. I motsats till vad jag förväntade mig visade experimentet att endast en art uppvisade tydlig respons på att bli transplanterad till en annan djupzon.

I **Artikel IV** tog jag resultaten från Artikel III och modellerade vad som skulle hända med biomassaproduktionen i algsamhällena om någon av de fyra arterna skulle utrotas. Jag fann att biomassaproduktionen av alger på dessa klippstränder förmodligen bara skulle påverkas negativt om en av arterna (*Fucus vesiculosus*) dör ut. Detta beror på att de fyra arterna uppvisade hög produktivitet utanför de djupzoner där de återfinns naturligt, varför de skulle kunna kompensera för förlusten av någon av de andra arterna.

Det mest direkta sättet att kvantifiera betydelsen av biologisk mångfald för ett ekosystems funktion är förmodligen att jämföra funktionen i ett samhälle bestående av många interagerande arter med vad vi kan förvänta oss baserat på arternas funktion i monokultur (där de inte interagerar med andra arter). I naturliga system är detta dock ofta inte möjligt eftersom vi sällan har naturliga monokulturer. I **Artikel V** utvecklade jag ett Bayesianskt analytiskt tillvägagångssätt för att skatta saknade monokulturdata, vilket möjliggör jämförelser mellan mer diversa artsamhällen och monokulturer i naturliga ekosystem. I kombination med en tidigare utvecklad statistisk metod för att räkna på biodiversitetseffekter kunde jag visa att en kombination av lokala artinteraktioner, lokal dominans av ett fåtal högproduktiva arter, och rumslig nischuppdelning, alla bidrog till en positiv effekt av biologisk mångfald på ekosystemens funktion i två naturliga marina ekosystem.

Baserat på dessa fem artiklar drar jag slutsatsen att de hundratals experiment som har gjorts hittills har gett användbar men vag information om hur en utarmad biologisk mångfald kan påverka funktionen hos naturliga ekosystem. För att på djupet förstå effekterna på ekosystemnivå måste vi mer noggrant studera hur den biologiska mångfalden förändras över större skalor i både rum och tid, och vi bör använda metoder som faktiskt kan upptäcka konsekvenserna av dessa förändringar. Artikel IV och V föreslår alternativ för hur detta kan uppnås.

Table of contents

Acknowledgements

It has been quite a journey since I first touched down in Sweden four and a half years ago. It is not possible to mention everyone who has influenced this thesis in these acknowledgements. But, I will try. Starting with one of the most important people, I would like to thank my supervisor, Lars Gamfeldt (also known to me as Gammy), for giving me the opportunity to pursue my PhD studies in Göteborg. We've had our ups and downs but you always supported and encouraged me to take whatever path I thought was best for me. I grew a lot over these four and a bit years both as a person and a researcher and you always gave as much as you could. I am thankful for that. To Jon Havenhand (my favourite and only co-supervisor) for always telling me exactly what you thought. Lars is very Swedish. He doesn't like conflict and is always polite. Despite living in Sweden a long time, you're not and I really appreciate that. Fabian Roger, thanks for your advice and guidance along the way. To Helle Plough, my examiner, for supporting me and making sure I stayed on track throughout these four years.

Anne Bjorkman (emphasis on *Bjork* without the ö) and Per Jonsson. You are both truly inspiring researchers. Anne, thank you for inviting me to spend time with your lab group, especially when I needed it most. Per, thank you for always being available for a chat and to read my drafts despite your retirement (sorry Eva-Marie).

Bram Vanschoenwinkel from the Vrije Universiteit Brussels. Thank you for your guidance and constant willingness to hack it out in front of a whiteboard with me. Then, to Bram's lab group (the *ShoeBois*: Joren, Eliza, Lisa, Vincent, Tim, Jasper etc.), thanks for hosting me so often and with such hospitality. I promise I will endeavour to finally try a "Bicky Burger".

My PhD colleagues. First, Rickard. You, my friend, have shown me that sneezes can be deafening. Birte, thanks for taking care of my plants while I was away (which was most of the time). Adele and Alex, thanks for giving me the 'good coffee' when I needed it. Milad, I really know a lot about the *Remarkable* now. Steffington, thanks for all the Molly pictures and warm dinners. Benny-Whenny. Thanks for the espressos and for your enthusiasm in the field. I will never forget that dive from the pier (you know what I'm talking about). Diego Moreno. Who would've thought that a Mexican and a South African could reminisce so delightfully about the music of our teenage years. There can't be many other people in Sweden that appreciate *The Red Jumpsuit Apparatus* as much as we do. Also, thanks for the hot-sauce. Simon, thanks for the sports! Also, to the Tjärnö volleyball crew Ida (långa pojkarna) and Louise. Luisa and Lucas, I'm still traumatised by *Klopfers*.

To the many MSc students that I taught on BIO451 and MAR440 who, rather frighteningly, probably number in the 200's. You all taught me a lot and it was one of the great joys of my PhD to try and share my knowledge and enthusiasm of statistics and R with you.

In the category of miscellaneous. Gunilla Johansson. Thanks for your warmth and helpfulness with managing my chaotic life at Tjärnö. Lovisa Thoursie, thanks for all your hugs and infectious enthusiasm. Seb Swart, my fellow countryman, thanks for being there to talk some rugby from time to time.

To my dearest Swedish friends and housemates: Ashina, Sandra, Tova, Tilde, Wilhelm, Christian, Shiro (dog #1), Shika (dog #2) and especially Elin. You made Sweden feel like home for me. I feel that I will always have a place to visit if I ever travel back up north and, I hope you all know that you are always welcome wherever I may be. Tack så mycket. Jag kommer sakna er alla.

My family: Dad, Mom, Kate and Warren. Thank you for supporting me and for being okay with me spending much of my life on the other side of the world. It has been and still is challenging but I know you all just want the best for me. And, also, to my family-in-law: Mechtilde, Toon, Zeenmeister (aka Xenia) and Laurien (not sure that Michiel makes the cut just yet). Bedankt dat jullie het familievormige gat in mijn hart mee hebben opgevuld de afgelopen vier jaar.

To Iliana, my now wife and illustrator of my thesis cover. Jouw steun voor mij in de laaste vier jaren was wel tof. Amai. Ik ben zo blij dat ik sta langs u in dit leven. Ik kijk uit naar wat de toekomst ons nu gaat brengen. Ik hou van u.

Finally, to the Swedish people. Thank you for hosting me in your beautiful country. I will deeply miss your wonderful nature, Bregott and Prästost. A midsummer will surely never pass without me reminiscing about the time that I spent here.

List of publications included in this thesis

This thesis is based on the following papers which are referred to by their roman numerals:

Paper I: Hagan JG, Gamfeldt L, Vanschoenwinkel B. (2021). We should not necessarily expect positive relationships between biodiversity and ecosystem functioning in observational field data. *Ecology Letters* 24: 2537-2548. https://doi.org/10.1111/ele.13874

- I conceived the idea for the research with input from BV and LG. I designed the conceptual framework and performed the analyses. I wrote the manuscript with input from BV and LG.

Paper II: Gamfeldt L[†], **Hagan JG**[†], Farewell A, Palm M, Warringer J, Roger F. (2023). Scaling-up the biodiversity–ecosystem functioning relationship: the effect of environmental heterogeneity on transgressive overyielding. *Oikos* 2023: e09652. https://doi.org/10.1111/oik.09652

- I collaborated with LG and FR on the design of the study. I collected approximately half of the data and performed approximately half of the analysis. I collaborated with LG and FR to write the manuscript and I led the revision process.

Paper III: Scrofner-Brunner B[†], **Hagan JG**[†], Cappelatti L, Hassellöv J, Wißmann M, Gamfeldt L (2023). Macroalgae maintain growth outside their observed distributions: Implications for biodiversity-ecosystem functioning at landscape scales. *Journal of Ecology* 111: 1362-1373. https://doi.org/10.1111/1365-2745.14103

- I collaborated with BSB and LG to design the study. I collected the data with BSB, LG, MW, JH, and LC. I analysed the data and wrote the first draft in collaboration with BSB. BSB and I led the revisions.

Paper IV: Hagan JG (*in review*). Compensation alters estimates of the number of species required to maintain ecosystem functioning as spatial scale increases: A case study with intertidal macroalgae. *Functional Ecology*.

- I conceived the idea, collated the data, analysed the data and wrote the manuscript.

Paper V: Hagan JG, Schrofner-Brunner B, Gamfeldt L (*in review*). Quantifying biodiversity effects on ecosystem functioning across space and time in natural and semi-natural ecosystems. *PNAS*. Preprint available on *ResearchSquare*: https://doi.org/10.21203/rs.3.rs-3249429/v1

- I conceived the idea and developed the analytical workflow that we used in the study. I simulated data to test the workflow. I analysed two empirical datasets using the analytical workflow with input from BSB. I collaborated with BSB and LG to design the data collection for the first empirical dataset, but I had a minor role in collecting the data. Data for the second case-study was collected in 2011 by LG. I wrote the first draft of the manuscript with input from BSB and LG.

[†]Shared first authorship

Publications related to but not included in this thesis

Hagan JG, Henn JJ, Osterman WHA (2022). *Matters Arising:* Plant traits alone are good predictors of ecosystem properties when used carefully. *Nature Ecology and Evolution* 7. Response to: van der Plas et al. (2020, *Nature Ecology and Evolution*). https://doi.org/10.1038/s41559-020-01316-9

Hagan JG, Bergmann R, Snoeks J, Dolmans V, Vanschoenwinkel B. *InvTraitR*: Improving biomass estimations for freshwater invertebrates using taxonomic backbones and habitat similarity. *Manuscript and R-package*. https://github.com/haganjam/FW_invert_biomass_allometry.

Luypaert T, **Hagan JG**, McCarthy ML, Poti M. (2020). Status of marine biodiversity in the Anthropocene. In: *YOUMARES9 – A Conference from and for YOUng MARine RESearchers*. Leibich V, Bode M, Jungblut S (eds). Springer International Publishing. https://doi.org/10.1007/978-3-030-20389-4_4

Olofsson M[†], **Hagan JG**[†], Karlson B, Gamfeldt L (2020). Large seasonal and spatial variation in nano- and microphytoplankton diversity along a Baltic Sea—North Sea salinity gradient. *Scientific Reports* 10: 17666. https://doi.org/10.1038/s41598-020-74428-8

Bushke FT, **Hagan JG**, Santini LS, Coetzee BWT (2021). Random population fluctuations bias the Living Planet Index. *Nature Ecology and Evolution* 5: 1145–1152. https://doi.org/10.1038/s41559-021-01494-0

Gould E, … **Hagan JG**, *et al.* (2023). Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *EcoEvoRxiv*. https://doi.org/10.32942/X2GG62

[†]Shared first authorship

Additional analyses presented in this thesis

Unless otherwise stated, all data and code presented in this thesis that was not included in the publications listed previously can be found in the following Github repository: https://github.com/haganjam/BEF_thesis.

Section 1: Introduction

Biodiversity – Ecosystem Function (BEF). It's quite the mouthful to say out loud as I have discovered over the last few years. But what is it exactly? I would define BEF as the subfield of ecology that concerns itself with understanding the ecosystem-level consequences of declines in biodiversity. This definition is necessarily broad, and the idea is operationalised in several different ways. Nevertheless, BEF is currently firmly rooted as a sub-field of ecology. In 2009, Solan *et al.* systematically reviewed the BEF field using 1000 publications. Today, a simple *ISI Web of Science* search with the same search terms produces almost 7000 publications.

But how did BEF come to be a sub-field of ecology? Can we trace its intellectual roots through time and make sense of the current mass of literature? In the first section of the introduction (*1.1: Historical context and intellectual roots*), I attempt to trace the origins of the BEF field in the ecological literature and show how it fits into ecology more generally. There are several excellent reviews of this topic. Notably, Morin's (2011) and Hector and Wilby's (2009) chapters, Tilman *et al.*'s (2014) literature review along with deLaplante and Picasso's (2011) and Frank's (2022) historical reviews. I cannot hope to be completely comprehensive in this introduction but interested readers will find these chapters and reviews useful starting points.

1.1: Historical context and intellectual roots

"It has been experimentally proved that if a plot of ground be sown with one species of grass, and a similar plot be sown with several distinct genera of grasses, a greater number of plants and a greater weight of dry herbage can thus be raised." – Darwin (1859)

In the early 2000's, Hector and Hooper (2002) traced the origins of the BEF field to agricultural experiments in the 19th century which were referred to by Darwin in *The Origin of Species* (see quote above). These agriculturalists experimented with mixtures of different numbers of species and examined the amount of hay and herbage that they produced. From this experiment and, presumably other experiments at the time, they concluded that more herbage and hay were produced when more species were grown in a plot. Indeed, in Darwin's unfinished book, *Natural Selection*, he remarked that: "A greater absolute amount of life can be supported…when life is developed under many and widely different forms,…the fairest measure of the amount of life being probably the amount of chemical composition and decomposition within a given period." (quoted from Hector and Hooper 2002). Similar concepts are at the heart of the current BEF field.

Whilst these ideas were present in the ecological literature in the 19th and 20th centuries, they never really gained momentum. Hector and Wilby (2009) cite an example from Carlander (1955) who studied the biomass of fish present in reservoirs in the Midwest of the United States (Fig. 1). Like Darwin in *Natural Selection*, Carlander (1955) hypothesised that: "Presumably fish production will increase as the number of niches increases ... [and] probably the proportion of occupied niches increases as the number of species of fishes increases.'' (quoted from Hector and Wilby 2009). Other than this and potentially a few other examples, these and other, similar questions were largely ignored until the 1990's. Instead, the BEF field as we know it today probably had its origins in the *diversity-stability debate*.

The diversity-stability debate

Several authors have argued that the BEF field traces its origins to the diversity-stability in ecology that occurred between the 1950's and 1990's (deLaplante & Picasso 2011; Morin 2011). During the 1950's, several prominent ecologists worked on how diversity and complexity in ecological communities affected stability. Odum (1953, cited from MacArthur 1955) argued from the perspective of energy flow through an ecological network. He reasoned that high diversity ecological networks should have more redundancy which means that energy flow is less likely to be disrupted. MacArthur (1955) made similar claims to Odum (1953) using consumer-resource models. Elton (1957) was the most wide-ranging and used a variety of evidence to assert that biodiversity leads to more stable ecosystems. For example, Elton (1957) noted that simple food webs are more prone to invasion than more complex ones and that species-poor Lotka-Volterra theoretical systems were prone to instabilities. By modern standards, this level of evidence is relatively weak. However, the idea that diversity begets stability remained popular until May (1972) began his mathematical analyses of the problem.
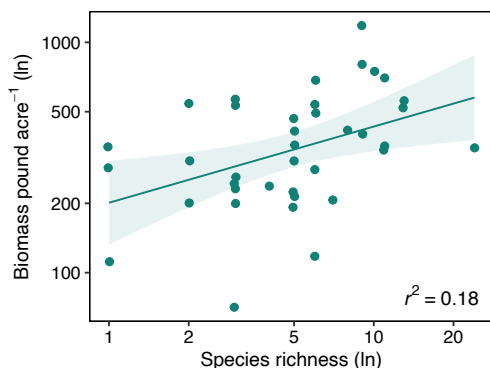

Fig. 1: One of the first studies to examine the relationship between biodiversity and ecosystem functioning. Carlander (1955) regressed the number of fish species (ln-scale) with the standing crop of fish biomass (pounds acre^{-1}, ln-scale) in reservoirs in the Midwest of the United States. Figure replotted using data extracted from Hector and Wilby (2009) in *Princeton Guide to Ecology*. Regression line and 95% confidence interval are overlain along with the coefficient of determination (r^2).

Despite widespread acceptance of the idea that diversity begets stability, May (1972) approached the problem in a novel way. May (1972) used Lotka-Volterra models to show that, in randomly assembled communities governed by the Lotka-Volterra competition, both the diversity and the number of connections among species decreased stability. However, May (1972) defined stability as the probability that the populations of all species would return to equilibrium if one species experienced an arbitrarily small population perturbation. Pimm (1980) built on May's (1972) analysis. Pimm's (1980) criticism of May (1972) was that the arbitrarily small perturbations were not ecologically meaningful. Therefore, instead, Pimm (1980) used similar Lotka-Volterra models but defined stability as the species-deletion stability which is the probability that a local extinction will not lead to any other local extinctions. He came to similar conclusions: diversity begets instability.

Other authors (e.g. De Angelis 1975) used the same kinds of models but with some more realistic assumptions (e.g. fixed trophic links such as a predator always needs a prey) and came to completely different conclusions. The debate was ongoing.
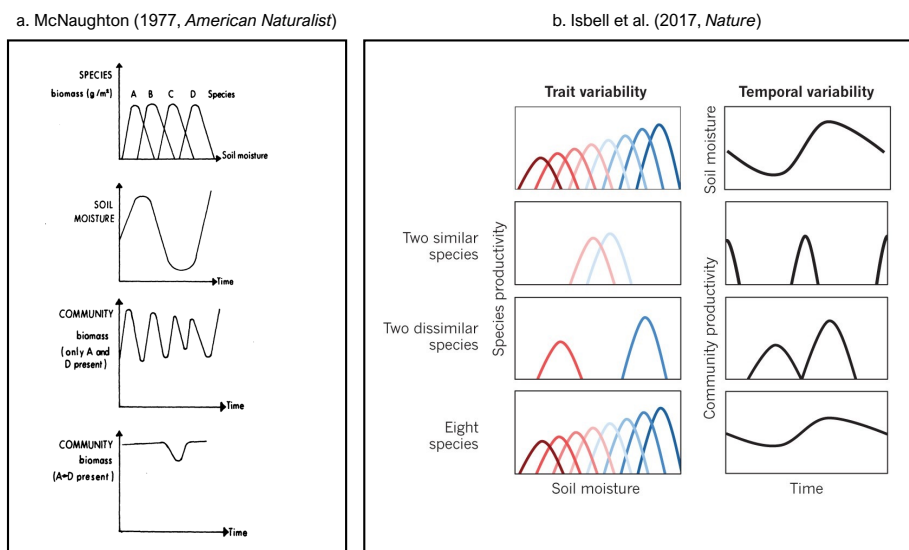

Fig. 2: (a) Conceptual diagram from McNaughton (1977) showing how species diversity and the consequent increase in trait diversity can buffer ecosystem properties (i.e. biomass) from environmental fluctuations like soil moisture. **(b)** Isbell *et al.* published a very similar figure in *Nature* in 2017.

In the 1970's, the diversity-stability had been largely mathematical. McNaughton (1977), however, sought to change that. In a controversial paper, McNaughton (1977) argued that the diversity-stability debate could not be settled without reference to empirical ecological systems. Moreover, he argued that the debate, in its current form was unfounded. Original ideas around stability (e.g. Elton 1957) were based around the stability of aggregate ecosystem properties (e.g. the total biomass of a community of species) and not the fluctuations of individual populations that had been the focus of De Angelis (1975), May (1972) and Pimm (1980). Rather, McNaughton (1977) proposed a conceptual model for how a diverse set of species could increase the stability of aggregate ecosystem properties like community-level biomass (Fig. 2a). This conceptual model is the basis of many of the most exciting hypotheses in the BEF field today (**Paper V**; Loreau *et al.* 2021) and is eerily similar to some recent conceptual models (Isbell *et al.* 2017; Fig. 2b). Moreover, in this and other works (e.g. Mellinger & McNaughton 1975), McNaughton (1977) provided a range of empirical support mostly from observational and experimental work on grassland ecosystems. Thus, not only was McNaughton (1977) able to show that there were definitional problems with how different researchers were studying stability, he also provided evidence from empirical ecological systems to support his claims.

Pimm (1984), in a highly influential review, was able to provide clarity on the issues raised by McNaughton (1977). In brief, Pimm (1984) recognised three response variables that were studied in relation to stability: i. individual species abundances (May 1972; Pimm 1980), ii. species composition

(e.g. McNaughton 1977) and iii. total community biomass (e.g. McNaughton 1977). Pimm (1984) also defined different types of 'stability'. For example, *stability* in the sense of individual species returning to equilibrium when perturbed (May 1972) or *resistance* as the tendency of a variable like biomass to resist changes through time (McNaughton 1977). Pimm (1984) then reviewed the predictions of how diversity affects stability considering different response variables and definitions of stability. Most importantly, he showed that studies like May (1972) and Pimm (1980) do not make predictions about what empirical ecological systems should look like. Rather, the point that these models make is that empirical ecological systems probably cannot have too many species that are too highly connected because, in nature, we only observe the systems that persist and are, therefore, stable. In contrast, McNaughton's (1977) conceptual model does make a prediction that is testable in natural systems: the total community biomass should be more resistant and resilient in more species-rich communities. In my opinion, the ideas presented in Pimm (1984) are still highly relevant today as I frequently come across papers that continue to cite May (1972) in the wrong context (e.g. Ratzke *et al.* 2020).

The rise of the biodiversity-ecosystem functioning field

Whilst questions regarding diversity and stability were important ecological topics in the 20th century (see previous section), these ideas did not become mainstream until the 1990's. As deLaplante and Picasso (2011) argued, this may be because ecologists, conservationists and other nature-conscious groups were becoming more aware and concerned about environmental degradation. Therefore, the socio-political backdrop meant that there was a strong incentive to find causal relationships between biodiversity and ecosystem properties like stability that may confer human benefit as this would provide a strong justification for conserving nature and biodiversity. Whatever the reasons, in my opinion, three important publications brought diversity-stability and BEF questions into mainstream ecological and scientific discourse.

The first important publication was Schulze and Mooney's (1993) edited volume: *Biodiversity and Ecosystem Functioning* from which the BEF field probably gets its name. The idea for the volume came from studies during the 1980's that had begun integrating insights from population biology and ecosystem ecology which, at the time, were largely separate fields. Studies like Carpenter *et al.* (1987) and Vitousek *et al.* (1987) had shown how the addition or removal of certain species could have dramatic effects on ecosystem properties like primary productivity and nutrient cycling. Therefore, with this as a starting point, Schulze and Mooney (1993) wanted a volume that could summarise whether biodiversity in complex communities could affect ecosystem functioning.

The second important publication was Tilman and Downing's (1994) publication in *Nature*. Tilman and Downing (1994) created a species richness gradient of between one and 25 species through a long-term nitrogen-addition experiment. Using this gradient and an unexpected drought year, they showed that the primary productivity of more species-rich communities was more resistant to the drought and was able to recover more completely following the drought. Tilman and Downing (1994) used this result to suggest that conserving nature and biodiversity was, therefore, critical for maintaining the stability of primary productivity. However, this conclusion was heavily criticised.

The biggest critic was Givnish (1994) who doubted Tilman and Downing's (1994) conclusions. Givnish's (1994) main point was that the species richness gradient was created by adding nitrogen which should select for plant communities that are dominated by species with low root:shoot ratios

due to reduced competition for nutrients. Such low root:shoot ratios would make species more susceptible to drought. Therefore, Givnish (1994) argued that Tilman and Downing's (1994) conclusion was more based on different nitrogen regimes selecting for different traits rather than a causal effect of biodiversity.

The third and final important publication was Naeem *et al.*'s (1994) publication in *Nature*. Unlike Tilman and Downing (1994), Naeem *et al.* (1994) directly manipulated species composition and biodiversity in experimental chambers and monitored several ecosystem processes including decomposition rates, community-level respiration and plant productivity (three biodiversity levels: 9, 15, and 31 species). They found that, on average, productivity increased with biodiversity. Thus, whilst Tilman and Downing (1994) showed that the resistance and resilience of primary productivity increased with biodiversity, here Naeem *et al.* (1994) were arguing that biodiversity could also increase the overall *magnitude* of productivity. However, even though Naeem *et al.* (1994) kept the environmental conditions constant among experimental units (unlike Tilman and Downing 1994), their biodiversity gradient was fully nested so that specific species in the high biodiversity treatment (31 species) were not present in the lower biodiversity treatments (9 and 15 species). Andre *et al.* (1994) criticised these results because large plant species were absent from the lower biodiversity treatments which, in their view, could explain Naeem *et al.*'s (1994) results.
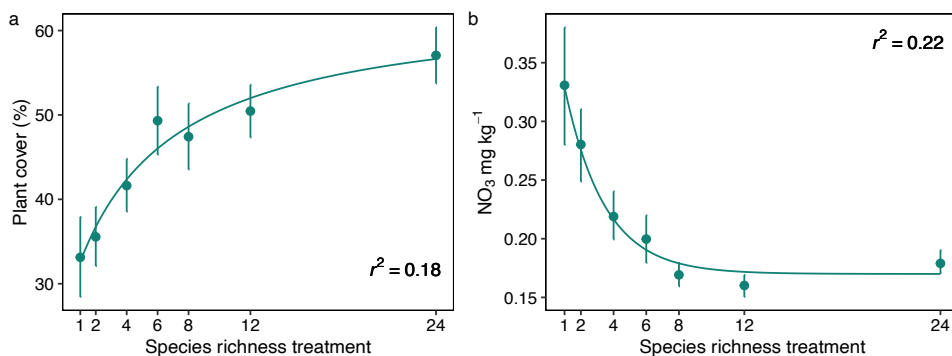

Fig. 3: The relationship between the species richness treatment and **(a)** total plant cover (%) and **(b)** total nitrate (NO₃) in the rooting zone in Tilman *et al.*'s (1996) BEF experiment. Points and error bars indicate the mean ± SE. Fitted curve in **(a)** is: $y \sim 27 + (36.4*x)/(5.48+x)$, ($r^2 = 0.18$, n = 147, P < 0.001) and in **(b)** is: $y \sim 0.17 + 0.24e^{-0.41x}$, ($r^2 = 0.22$, n = 147, P < 0.001). Replotted using data extracted from Tilman *et al.* (1996).

In my opinion, both Givnish's (1994) and Andre *et al.*'s (1994) criticisms were important because they prompted researchers to think more carefully about how to test whether biodiversity causally relates to ecosystem properties. This likely culminated in Tilman *et al.*'s (1996) experiment where 147 plots (3 x 3 m) were cleared of all vegetation and then sown with seven different levels of species richness (1, 2, 4, 6, 8, 14 and 24 species). All patches were sown with 10 g of seeds which were split evenly between species and the species were chosen randomly from a pool of 24 species. This experimental design aims to simultaneously address the criticism of Naeem *et al.*'s (1994) experiment by decoupling species composition (i.e. the identity of species) from species diversity and the criticism of Tilman and Downing's (1994) study where environmental conditions (i.e. nitrogen levels) were confounded with species diversity. As with Naeem *et al.* (1994), Tilman *et al.* (1996)

found that species richness led to increased plant cover (a surrogate for productivity), (Fig. 3a). They hypothesised that this result was due to different species having slightly different nutrient uptake patterns as they also found that more nitrogen was used up in the higher species richness treatments (Fig. 3b). They concluded that their experiment added to the growing body of evidence that any anthropogenic actions that reduced biodiversity would impact the magnitude and stability ecosystem functions.

Tilman *et al.* (1996) was, as far as I can tell, the first BEF experiment as it is recognised in the field today. What do I mean by a BEF experiment? This is an experiment that keeps the environment constant whilst manipulating some aspect of biodiversity (mostly species richness) and randomising species composition. It is worth pausing this historical overview of the BEF field to explain, in detail, what a BEF experiment is because it is central to many further discussions.

The BEF experiment: The workhorse of BEF research

What is a BEF experiment? It is an experiment that aims to directly manipulate biodiversity (e.g. species richness, genetic diversity etc.) whilst controlling for species composition (i.e. the identity of species) and environmental conditions. It turns out that this is rather difficult (Naeem *et al.* 1994; Tilman & Downing 1994). However, Tilman *et al.* (1996) provided a satisfying experimental design solution to the problem (see also previous section) which has arguably defined the field since. The general solution, originally proposed by Tilman *et al.* (1996), was to consider a pool of S species and a set of R different species richness treatments. Species are then randomly inoculated into different experimental units from the pool of S species depending on the species richness treatment. For example, if there were three replicate experimental units of a two-species treatment, then two species from the pool of S species would be randomly assigned to each of the three replicates. Thus, each of the three replicates would be inoculated with a different two-species combination. In all cases, the total abundance of inoculated individuals, J (e.g. seeds, juveniles, adults etc.), is kept constant irrespective of the species richness treatment. The communities inoculated into the experimental units are then allowed to develop and after an arbitrary period of time, some measure of ecosystem functioning (usually biomass productivity) is measured. Finally, the species richness treatment is related to ecosystem functioning (see Fig. 4 for an overview).

Within this general framework, there are two slightly different experimental designs (Fig. 4). The first design replicates all S species as monocultures. Thus, for a species richness treatment of one, there are at least S replicates ($2S$ if each monoculture is replicated twice etc.). In addition, in the first design, the highest species richness treatment is S. This means that, unlike the other species richness treatments, any replicates of the highest species richness treatments are identical in species composition (Fig. 4a). The problem with this design is that there is much more variation in species composition among replicates in the low species richness treatments than at the higher species richness treatments which can cause statistical problems like unequal variance (Fukami *et al.* 2001).

The second experimental design is similar but aims to solve this problem of the species richness treatments differing in variation in species composition among replicates. In the second design, each species richness treatment has N replicates, and the highest species richness treatment is less than S species. Thus, for each replicate for species richness treatment, species are randomly drawn from the pool of S species. This design means that not all S species are replicated in monoculture and there is variation in species composition among the highest species richness treatments. Practically, the results

of these different experimental designs generally do not differ very much, but it is important to bear in mind when examining a BEF experiment. Other experimental designs are also used. Specifically, some designs do not hold total abundance of inoculated individuals constant (e.g. Cardinale *et al.* 2003) and others are short term and do not allow changes in the abundance of species (see *Section 4* for further discussion). But, in my reading of the literature, these are considerably less common.
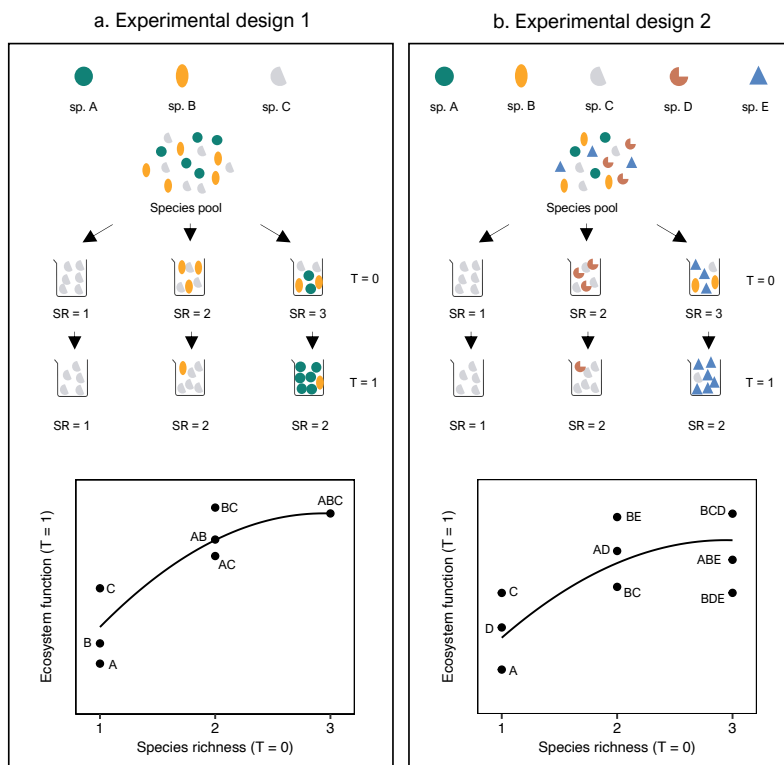

Fig. 4: (a, b) The two different experimental designs of a typical BEF experiment that are probably most commonly used. I only illustrate one replicate for each species richness treatment (SR). Moreover, the curves are hypothetical and meant to illustrate the kind of data obtained from such an experiment.

Keen readers may have noticed that we test the effect of the *species richness treatment* on ecosystem functioning at some later time-point (T = 1 in Fig. 4). Experimentally, this makes sense given that we are directly manipulating the species richness treatment. However, practically it means that a community that develops from a species richness treatment of three can be comprised of one or two species when ecosystem functioning is measured (Fig. 4b). This can occur if species are outcompeted or if they cannot tolerate the environmental conditions. In my opinion, this is a critical point when interpreting BEF experiments which I will discuss in detail later. Moreover, it has some important implications for thinking about how BEF experiments relate to natural ecosystems, something that I elaborated on in **Paper I**.

Controversies surrounding the emerging BEF field

The design of the BEF experiment proposed by Tilman *et al.* (1996) addressed the problems brought up by Givnish (1994) and Andre *et al.* (1994). However, Huston (1997), along with Aarssen (1997), levelled a further criticism which is described by Huston (1997) as "one of the most subtle hidden treatments in experimental ecology". Huston (1997) pointed out that in Tilman *et al.*'s (1996) design (and the design of the typical BEF experiment, see previous section), the higher species richness treatments have a higher probability of selecting a species with a specific property like high biomass or large maximum height. Huston (1997) referred to this as the "selection probability effect". This same mechanism leads to lower variation in species composition among replicates in high species richness treatments compared to low species richness treatments (Fig. 4a, previous section). The consequence of the selection probability effect is that there is a correlation between the species richness treatment and the maximum possible productivity because the 24 species in Tilman *et al.*'s (1996) experiment differed considerably in their ability to grow to a large size (Aarssen 1997). Indeed, Huston (1997) provides several lines of evidence that the biomass observed in Tilman *et al.*'s (1996) experiment was dominated by a few species which suggests that the positive effect of the species richness treatment on ecosystem functioning was an artefact of the experimental design.

In addition to the criticisms around the selection probability effect as a cause of the positive effect of biodiversity on ecosystem functioning, Aarssen (1997), Grime (1997) and Huston (1997) all questioned the relevance of the experiments. In their reading of the literature, even if biodiversity does causally affect ecosystem functioning, the species composition and, specifically, whether a community includes one or several highly productive species will be a much more important determinant of functioning than biodiversity. This is most easily observed by some of the least biodiverse regions having the highest productivity (e.g. temperate forests, Liang *et al.* 2016) or the concomitant decrease in biodiversity and increase in productivity following eutrophication (Hautier *et al.* 2009; Silvertown *et al.* 2006). Moreover, the results of Tilman *et al.*'s (1996) experiment conflicted with a common model at the time whereby the relationship between biodiversity and plant productivity was hump-shaped (Grime 1973; Huston 1979).

Tilman (1997) responded to these criticisms by attempting to explain the claims that the Tilman *et al.* (1996) experiment can and cannot make. Specifically, Tilman (1997) argued that the classic BEF experiment does allow the claim that the species richness treatment and, by extension, the number of species originally sown, causally affects productivity. What Tilman *et al.* (1996) cannot and do not show is why this effect occurs nor do they argue that biodiversity is more or less important than species composition. However, Tilman *et al.* (1997) used simple mathematical theory based on resource competition to show that the observed results in the 1996 experiment could be due to i. the selection probability effect suggested by Aarssen (1997) and Huston (1997), ii. differential resource use among species or iii. a combination of these effects. This was an explicit acknowledgement from Tilman (1997) and Tilman *et al.* (1997) that the selection probability effect is a valid process by which biodiversity can affect functioning. This final claim was, however, disputed (Aarssen 1997; Huston 1997)

Whether the selection probability effect could be considered a valid biological mechanism was tackled by Wardle (1999). The main argument made by Wardle (1999) was that if the selection probability effect is a valid process by which biodiversity can affect ecosystem functioning in nature, two assumptions must be met. First, communities would need to assemble at random with respect to

their effects on ecosystem functioning. Second, there would need to be variation in the pools of species that randomly colonise different communities. Wardle (1999) noted that there may be some specific systems where these assumptions are met (e.g. when fungi colonise wood) but that this is probably unlikely in most ecological systems. Rather, Wardle (1999) suggested that having all possible monocultures that are present in mixtures or by using removal experiments might make these experiments more ecologically relevant.

These issues were not fully resolved in subsequent years. However, there were two important developments. First, there was an acknowledgement by some researchers in the field that, even if the selection probability effect is a valid biological mechanism, it might still be desirable to separate it from other effects such as complementary resource use. Indeed, Hector (1998), Loreau (1998) and Wardle *et al.* (1997) all proposed methods to test the null hypothesis that ecosystem functioning in mixtures of species could be accounted for by changes in the proportional contributions of individual species in that mixture. Secondly, there was flurry of theoretical contributions aiming to examine the mechanistic basis of biodiversity effects on ecosystem functioning, notably Loreau (1998a; 2000) and Lehman and Tilman (2000).

The theoretical contributions used a combination of resource competition theory (Loreau 1998a; Tilman 1999) and statistical simulations (Doak *et al.* 1998; Tilman *et al.* 1998) to explore the mechanistic basis of biodiversity effects on ecosystem functioning. The combination of these theoretical treatments led to several hypotheses. First, biodiversity tends to increase the stability of aggregate ecosystem functions like biomass productivity, and it can do this if all species in a community fluctuate randomly (Doak *et al.* 1998) or if competition leads to some species increasing at the expense of others (Tilman *et al.* 1998). Second, resource competition theory showed that biodiversity can increase the magnitude of ecosystem functions like productivity both by selection probability effects (*sensu* Huston 1997) and through resource complementarity effects (Loreau 1998a; Tilman 1999). Third, environmental conditions affect both species composition and functioning (Bengtsson 1998). This means that the relationship between biodiversity and functioning across different ecosystems can be masked by the environmental variation (Loreau 1998a). This third hypothesis was subsequently tested in the pan-European BIODEPTH experiment where BEF experiments were performed in different grasslands across Europe. These experiments showed that, within most of the BIODEPTH sites, biodiversity did increase productivity and that across sites overall productivity differed strongly due to environmental conditions (Hector *et al.* 1999). These theoretical papers along with large experiments like BIODEPTH being done despite substantial criticisms of the experimental design (Aarssen 1997; Huston 1997; Wardle 1999) led to tensions in the literature.

The tensions eventually boiled over when Naeem *et al.* (1999) published a policy report for a lay audience which summarised the evidence for the importance of biodiversity for ecosystem processes relevant to people and society. Their conclusions were that biodiversity and species composition were of near-equal importance for maintaining ecosystem functioning and that policy makers should preserve biodiversity to protect these important processes. Wardle *et al.* (2000) were highly critical of this report stating that Naeem *et al.* (1999) were not presenting a balanced view of the evidence given that there were many unresolved issues regarding how biodiversity effects functioning. Wardle *et al.* (2000) also argued that, given these unresolved issues, Naeem *et al.* (1999) risked setting a dangerous precedent for conservation if the link between biodiversity and ecosystem functioning turned out to be weak (deLaplante & Picasso 2011). Rather, Wardle *et al.* (2000)

suggested that Naeem *et al.* (1999) should have focused on points of agreement between ecologists. Specifically, that conserving species is important because they provide ecosystem functions and services and that this would be true regardless of whether a causal link between biodiversity and ecosystem function could be demonstrated (Wardle *et al.* 2000; deLaplante & Picasso 2011).

Resolutions

As Frank (2022) noted, there was an uneasy resolution that came following the 1999 controversy. This came about after discussions between authors from both sides at a BEF conference in Paris in 2000. The discussions led to a joint paper titled: *Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges* published in *Science* in 2001 that aimed to summarise the state of knowledge surrounding BEF and point to directions for future research.

The first resolution concerned the interpretation of the selection probability effect which all authors now referred to as simply the *selection effect*. Loreau *et al.* (2001) identified two high-level mechanisms by which biodiversity can increase ecosystem functioning corresponding to theory. The first mechanism relates to processes of niche differentiation or facilitation (or positive interactions more generally) among species that tends to increase the functioning of species in mixture relative to species in monoculture. This was called the *complementarity effect*. The second mechanism is the selection effect which combines random sampling of species from a species pool and then dominance of high functioning species as it was originally conceptualised by Aarssen (1997) and Huston (1997). Loreau *et al.* (2001) recognised that both effects were operating in BEF experiments. They also noted that they probably could operate to some degree in natural ecosystems with the importance of the selection effect depending on the strength of random sampling processes during community assembly but that this remained to be demonstrated.

This resolution was strengthened by Loreau and Hector (2001) who proposed a statistical partition to compare selection and complementarity effects in BEF experiments when monoculture data for all species were available (experimental design 1, Fig. 4a). This partition built on previous work from both authors (Hector 1998; Loreau 1998b) which had tried to separate out the influence of the selection effect. The method is based on calculating a net biodiversity effect (*NBE*) which is defined for some mixture of *S* species as:

Equation 1-3

$$NBE = Mix_{obs} - Mix_{exp}$$

$$Mix_{obs} = \sum_{i=1}^{S} Y_i = \sum_{i=1}^{S} RY_{obs,i} \, M_{obs,i}$$

$$Mix_{exp} = \sum_{i=1}^{S} RY_{exp,i} \, M_{obs,i}$$

Therefore, the net biodiversity effect is the difference between the observed functioning of a mixture (*Mix*obs) and the expected function of the mixture (*Mix*exp). The observed functioning of species *i* in mixture is given as Y_i and, thus, the sum of Y_i to *S* is *Mix*obs (i.e. the observed mixture functioning). However, Y_i can also be expressed as the product of the *RY*obs,i and monoculture

functioning (M_{obs}) where $RY_{obs,i} = Y_i/M_{obs,i}$. The $RY_{exp,i}$ is then the expected relative yield which, in a mixture where each species is inoculated with a $1/S$ proportion, is simply $1/S$ for all species. We can then generalise this expression to the net biodiversity effect as:

Equation 4-5

$$NBE = \sum_{i=1}^{S} \Delta RY_i M_i$$

$$\Delta RY_i = RY_{obs,i} - RY_{exp,i}$$

This expression quantifies the extent to which the observed functioning of some mixture deviates from the expected mixture value based on species functioning in monoculture and the proportion that the species were initially inoculated with. Loreau and Hector (2001) then showed that the *NBE* could be partitioned into the contributions from the selection effect and the complementarity effect as follows:

Equation 6

$$NBE = \sum_{i=1}^{S} \Delta RY_i M_i = S \overline{\Delta RY}\ \overline{M} + S\ cov(\Delta RY_i, M_i)$$

In the expression above, the first term measures the complementarity effect, and the second term measures the selection effect. The complementarity effect quantifies the extent to which the increase in mixture functioning relative to the expectation is above a zero-sum expectation (i.e. the average of the ΔRY_i values are greater than one). This occurs when positive species interactions such as niche partitioning and facilitation outweigh any negative species interactions. In contrast, the selection effect quantifies the extent to which the increase in mixture yield relative to the expectation is due to species increasing their relative yield at the expense of others and those increases being correlated with functioning in monoculture (but see Fox 2005; discussed further in *Section 4*).

The quantification of the complementarity and selection effects can be most easily understood by examining a few hypothetical examples. Let us consider a hypothetical case with two species in monoculture and a 50:50 mixture of the two species. In monoculture, species 2 is higher functioning than species 1 (500 *versus* 200 ecosystem function units). As a first example, if, in the mixture, species 2 outcompetes species 1 and, therefore, completely dominates the mixture, the functioning of the mixture would be equal to the functioning of species 2 in monoculture (i.e. 500 ecosystem function units, Fig. 5a). This would lead to a net biodiversity effect of 150 which is completely due to the selection effect (Fig. 5d). Moreover, the species richness of the mixture is one (i.e. species richness declined in the mixture) and the functioning of the mixture is the same as the functioning of the highest functioning species in monoculture. This means there is no *transgressive overyielding* which occurs when the mixture functioning exceeds that of the highest functioning monoculture.

As a second example, consider the case where, in the mixture, species 1 and 2 coexist and both increase their functioning 1.4-fold beyond their monoculture expectations (i.e. $RY_{exp,i}$, $M_{obs,i}$), (Fig. 5b). In this case, we also have a net biodiversity effect of 150 but the effect is completely due to the complementarity effect (Fig. 5d). In contrast to the first example, the species richness of the mixture

remains at two because both species coexist but there is still no transgressive overyielding (i.e. mixture functioning is the same as that of the highest functioning monoculture, Fig. 5b). As a third and final example, consider a case where species 1 and 2 coexist and both increase their functioning beyond their monoculture expectations. However, now also consider that species 2 increases its mixture functioning more than species 1 (1.7-fold *versus* 1.3-fold), (Fig. 5c). In this example, we have a net biodiversity effect of 200 and it is due to a combination of the complementarity and the selection effect (Fig. 5d). Moreover, there is transgressive overyielding because the functioning of the mixture exceeds the functioning of species 2 in monoculture (Fig. 5c).

These examples are not meant to be comprehensive but are rather meant to give an intuition of how Loreau and Hector's (2001) partition works. This is important because it is widely used in the BEF field and, in **Paper V**, I apply an extension of this partition developed by Isbell *et al.* (2018) to data from natural ecosystems. It is also important to note that these examples are not necessarily generalisable. For example, a combination of complementarity and selection effects are not necessarily required for transgressive overyielding to occur. In addition, these examples served to introduce the concept of transgressive overyielding which is the focus of **Paper II**. Transgressive overyielding is an important concept to understand because many BEF researchers and sceptics view transgressive overyielding as the only valid way to demonstrate a biodiversity effect (for some discussion, see Schmid *et al.* 2008).

Loreau and Hector (2001) used this statistical partition to analyse data from the BIODEPTH experiment where BEF experiments in grasslands were performed in eight countries across Europe (see previously and Hector *et al.* 1999). Generalising across sites, Loreau and Hector (2001) showed that both selection effects and complementarity effects were operating in these experiments. However, complementarity effects tended to be stronger, more consistently positive and increased with species richness. Selection effects, in contrast, were slightly weaker and were frequently negative. Thus, Loreau and Hector's (2001) partition showed that both Aarssen (1997) and Huston (1997) were correct in pointing out that the selection effect operates in BEF experiments but also that there were positive interactions among species that increased functioning in accord with resource competition theory (Loreau 1998a; Tilman 1999). However, at least in the BIODEPTH experiments, there was limited evidence of transgressive overyielding (Hector *et al.* 1999) in contrast to the resource competition models (Loreau 1998a; Tilman 1999). Therefore, the complementarity effects were weaker than the theoretical models suggested. Nevertheless, Loreau and Hector's (2001) statistical partition was an important advance in being able to tease out selection effects (which many researchers did not consider a valid biodiversity effect) from complementarity effects.

The second resolution was about why the results of BEF experiments did not match patterns in natural ecosystems. Specifically, one of the main points of contention was that, in grasslands and herbaceous ecosystems, the relationship between biodiversity and productivity was often hump-shaped (Fraser *et al.* 2015; Grime 1973; Huston 1979). However, as was shown by Loreau (1998a) theoretically, these kinds of studies are simply asking different questions. Productivity varies with environmental conditions and these factors also alter biodiversity. However, BEF experiments are examining the effect of a biodiversity treatment not the observed biodiversity (**Paper I**) and environmental conditions are held constant. Therefore, the results are not mutually exclusive. Across different sites with different environmental conditions, productivity can have a hump-shaped relationship to biodiversity and biodiversity can still increase productivity within a site (Fig. 6, see also Schmid (2002) for a thorough overview).
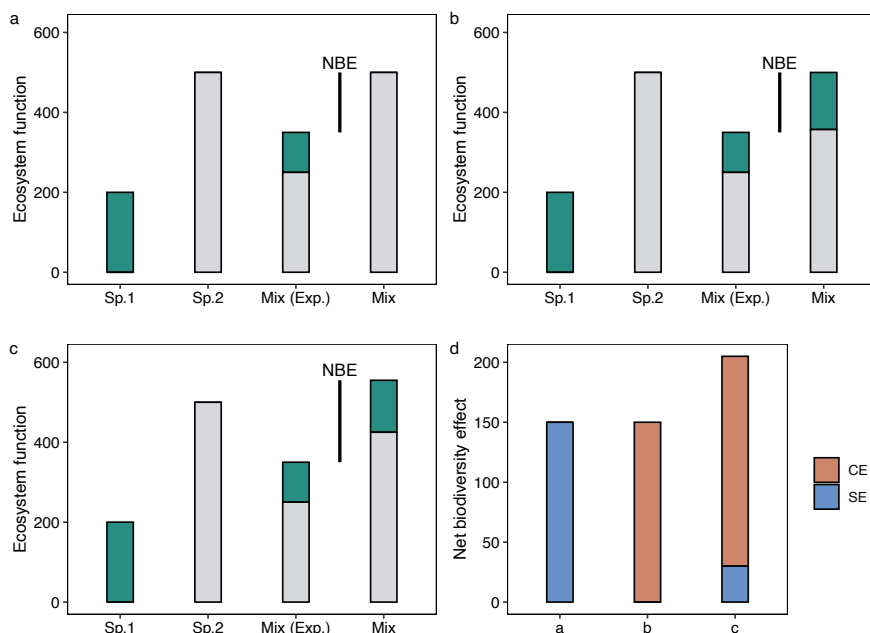
Fig. 5: (a-c) The effect on ecosystem functioning of combining two species (1 and 2, different colours) in a 50:50 mixture (Mix) compared to their respective functioning in monoculture (Sp.1 and Sp.2) and the expected mixture functioning (Mix (Exp.)) under the null hypothesis of no species interactions. The difference between Mix (Exp.) and Mix is the net biodiversity effect (*NBE*, vertical black line). In **(a)** the *NBE* is fully accounted for by a selection effect (*SE*) as species two dominates the mixture. This leads to a positive NBE but no transgressive overyielding (i.e. the mixture functioning does not exceed that of the highest functioning monoculture). However, in **(b)**, the magnitude of the *NBE* is the same as in **(a)** but it is fully accounted for by the complementarity effect (*CE*) as both species are higher functioning in mixture than expected based on their monoculture functioning. In **(c)** a combination of *SE* and *CE* is operating and there is transgressive overyielding. **(d)** Summary of the *NBE*, *CE* and *SE* for the three examples **(a-c)**. Ecosystem function is an arbitrary quantity but could represent, for example, primary productivity (g m^{-2} day^{-1}).

The third resolution based on Loreau *et al.*'s (2001) joint publication but also previous studies (Bengtsson 1998; Loreau 2000) was that any hypothesised effects of biodiversity are necessarily based on phenotypic trait diversity. Both BEF theory and experiments focused on species richness for convenience with the assumption that species richness is a proxy for phenotypic trait diversity. Selection effects are due to dominance of species with particular traits, and complementarity effects are, in theory, driven by species with different traits (Loreau 2000). Thus, when thinking about biodiversity effects on ecosystem function, the hypothesis is that biodiversity, whether it be species richness, number of families or genetic diversity, affects functioning is based on how these different measures of biodiversity map onto phenotypic trait variation.

The final resolution concerned how the conclusions from the recent BEF work generalised to different ecosystems and trophic levels. Loreau *et al.* (2001) acknowledged that the theoretical and empirical work in the BEF field had, so far, mostly studied how biodiversity affects biomass

productivity of primary producers (e.g. grassland plants). Other research strands concerned how top predators, ecosystem engineers and other, so-called, keystone species (i.e. species that disproportionately affect ecosystem processes given their abundance) affected ecosystem functioning in a range of different ecosystems (Power *et al.* 1996). However, these studies did not isolate the effects of biodiversity as separate from species composition as the recent BEF field had done. Therefore, Loreau *et al.* (2001) suggested that much more work needed to be done to test whether the current results from BEF research generalised to other ecosystem types and trophic levels.
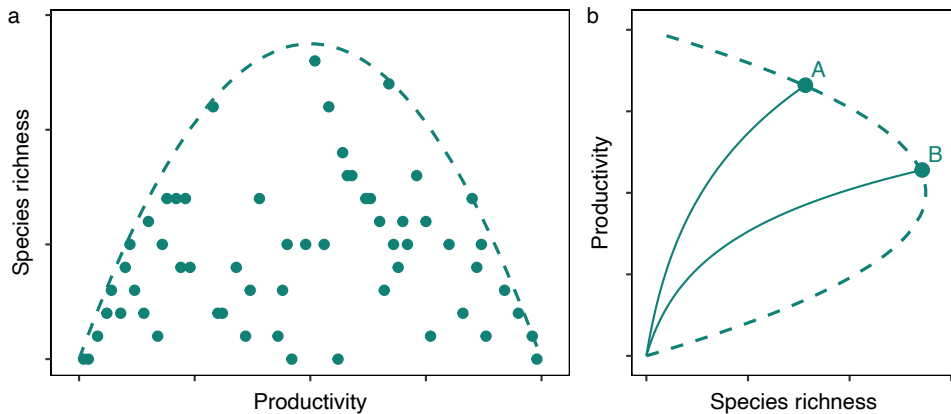

Fig. 6: The conflict between the hump-shaped relationship between productivity and species richness in observational data and the positive relationship between species richness and productivity in BEF experiments was resolved in the following way. **(a)** First, it was recognised that the hump-shaped relationship generally corresponded to the maximum species richness for a given level of productivity across different ecosystems (e.g. a mesic grassland or an arid grassland). **(b)** Second, within, for example, a mesic grassland (point A), productivity is typically higher and species richness is typically lower than in a more arid grassland (point B). However, within both mesic and arid grasslands, BEF experiments suggest that species richness increases productivity (curves labelled A and B). Figures redrawn after Loreau *et al.* (2001) and Schmid (2002).

Consolidation

With the 2001 consensus paper, the BEF field began consolidating itself. This happened mainly through more and more BEF experiments being carried out in different ecosystems using different study organisms. As the evidence from these experiments built-up in the literature, the first quantitative (Balvanera *et al.* 2006) and qualitative (Hooper *et al.* 2005) evidence reviews were published that summarised the state of the BEF field.

In Balvanera *et al.*'s (2006) meta-analysis, they analysed 446 experimental effect sizes for the effect of biodiversity on ecosystem functioning (e.g. productivity, stability, invasion resistance etc.). This meta-analysis affirmed the conclusions from the original BEF theory and experiments (e.g. Tilman *et al.* 1996): biodiversity tends to positively affect both the magnitude and stability ecosystem functioning. These effects were strongest when the environment was tightly controlled. However, Balvanera *et al.* (2006) also noted that there was considerable variation in the effect sizes which was driven by slight differences in experimental design, different measures of ecosystem function (e.g. productivity *versus* decomposition), scale of ecological organisation (e.g. community versus

ecosystem) along with several other factors. Thus, much of the variation was due to how researchers defined and measured biodiversity and ecosystem functioning which made generalisation difficult.

Hooper *et al.*'s (2005) qualitative review was broader and emphasised areas of consensus among researchers working in the BEF field. Many of these points were emphasised previously by Loreau *et al.* (2001) and, here, I only summarise the points that are most relevant to the background for this thesis. First, Hooper *et al.* (2005) agreed that in BEF experiments, biodiversity tends to increase both the magnitude and stability of ecosystem functioning but, as reported by Balvanera *et al.* (2006), these effects can vary depending on the ecosystem-type, organism group and ecosystem function that is measured. Secondly, these positive effects result from a combination of complementarity and selection effects but that it is unclear whether selection effects are valid mechanisms in natural systems. Thirdly, Hooper *et al.* (2005) also acknowledged the importance of the functional traits of organisms in driving ecosystem function. Thus, any effects of some component of biodiversity (e.g. species richness) occurs through the functional traits of the organism whether that be through dominant species having specific traits (drivers of selection effects) or trait differences among species (drivers of complementarity effects). Finally, Hooper *et al.* (2005) emphasised that any effects of biodiversity on ecosystem function in BEF experiments are smaller or equal to the effects of species composition as evidenced by the considerable variation within any given species richness treatment (e.g. Hector *et al.* 1999). This means that, for example, knowing which species is lost can be as informative as knowing that, on average, 10 species have been lost. This final point allowed the BEF field to match-up with research in ecosystem ecology that had determined the importance of individual species' functional traits for ecosystem functions.

These evidence reviews both suggested in some form that the evidence supporting the positive effects of biodiversity on the magnitude and stability of ecosystem functioning should at least warrant a precautionary approach whereby biodiversity is preserved in order to safeguard ecosystem functions essential for people. However, despite the increased evidence (mainly based on BEF experiments), many fundamental issues that had previously been raised were never satisfactorily dealt with. For example, is the selection effect (*sensu* Huston 1997) a valid mechanism for how biodiversity might affect ecosystem functioning in natural systems? Can we logically and philosophically view species composition and a biodiversity variable like species richness as having independent effects on ecosystem functioning when they are inextricably linked (Bengtsson 1998)? Can a biodiversity effect be considered valid when there is no transgressive overyielding? Without transgressive overyielding, a mixture will still have lower functioning than at least one species in monoculture. Most experiments took place at small spatial and temporal scales (Balvanera *et al.* 2006). Are these effects relevant at larger scales of space and time? None of these issues were fully resolved.

Perhaps even more importantly, there was (and still is) considerable disagreement in the literature about whether relationships between biodiversity and ecosystem functioning can be used to justify biodiversity conservation (Schwartz *et al.* 2000; Srivastava & Vellend 2005). For example, Schwartz *et al.* (2000) argued that BEF experiments would only justify conservation if all species were required for maintaining ecosystem function. However, in many experiments, the relationship was saturating (e.g. Fig. 2a). Srivastava and Vellend (2005) argued more from the perspective of how species loss happens in natural systems. They pointed out that the type of species loss simulated in BEF experiments (i.e. random draws from a species pool) does not occur in natural systems. Rather, species with certain traits, like top-predators, are more extinction prone. The extinction of these

species will affect the ecosystem in ways that are not predicted by a BEF experiment. Both publications concluded that the conservation justification based on such studies was, at best, weak.

To conclude, by the early- to mid-2000's, the BEF field had consolidated itself. However, there were still a wide variety of unresolved issues; both technical (e.g. the selection effect) and philosophical (e.g. species composition *versus* species richness, conservation justification arguments). Nonetheless, between 2000 and 2010, the BEF field firmly established itself as a sub-field of ecology. In the next section (*1.2: Background: The "hard core" of the BEF research program*), I go through the definitions, theories and empirical evidence that underlie the field today.

1.2: Background: The "hard core" of the BEF research program

It may seem odd to have a *Historical context and intellectual roots* section along with a section on *Background*. So, why have I structured the introduction to my thesis like this? In my view, a sense of history is important. The BEF field garnered much controversy after the first few experiments and many of these controversies were never fully resolved. For example, I can say with some confidence that very few new entrants to the BEF field really question whether the selection effect is a valid biodiversity effect that operates in nature. Rather, when one reads the more recent BEF literature, statements like: "biodiversity affects ecosystem functioning through a combination of complementarity and selection effects" are commonplace and unquestioned (Maureaud *et al.* 2019). I wrote the previous section to give space to these controversies as understanding the history was instrumental to my understanding of the BEF field.

In this section, rather than review the history, I go through the fundamental principles that define the field today. These fundamental principles include the definitions, theories, and empirical evidence that, currently, most of the workers in the field agree on and upon which the current BEF research seeks to build upon. I am not much of a philosopher but these fundamental principles form what Imre Lakatos may have called the "*hard core*" of the BEF research program (Chalmers 1999). These fundamental principles are rarely questioned and are typically supported by a large body of theory and empirical evidence. Because they are rarely questioned does not make them correct but it makes them critical to understand the BEF research program as it operates today.

Definitions

What do we mean by biodiversity and what do we mean by ecosystem function? In the BEF field, neither term is particularly well-defined. If we look at the standard definition of biodiversity, it typically refers to the variety of life at multiple hierarchical levels i.e. genes, species, traits and ecosystems (e.g. see Convention on Biological Diversity). Other definitions can include aspects of biological communities that are not necessarily linked to variation *per se*. For example, Hooper *et al.* (2005) include things like relative abundance and McGill *et al.* (2015) suggest including measures like population size. In the BEF field, however, we are typically focused on biodiversity as variation *per se*. Most commonly, this has been operationalised as species richness (i.e. the number of species). But, as discussed, theory is based on phenotypic trait diversity and, recently, more quantitative measures of biodiversity such as functional and phylogenetic diversity have been used.

Ecosystem functioning is trickier to define. Most authors would agree with the broad idea that an ecosystem function relates to processes that affect the fluxes or stocks of energy, nutrients or

matter in an ecosystem. These include primary productivity, decomposition rates etc. However, some authors include fluxes (e.g. productivity) and stocks (e.g. standing biomass), (Hooper *et al.* 2005) whilst others only consider fluxes as ecosystem functions (Cardinale *et al.* 2012). Either way, if one considers stocks, fluxes or both, the definitions are still extremely broad. Thus, the decision regarding what an ecosystem function is or is not generally falls on the individual researcher to define in their study system. But in saying that the definitions are extremely broad, the vast majority of BEF studies measure the magnitude of either biomass productivity or some aspect of nutrient uptake (e.g. inorganic uptake in plants, grazing rates in herbivores etc., Cardinale *et al.* 2012). Most researchers are comfortable with these measures of ecosystem functioning because they relate to the fundamental goal (in a non-teleological sense) of life which is to extract resources and reproduce. In addition, theoretical treatments of BEF almost exclusively focus on biomass productivity or nutrient uptake.

There are two other aspects of ecosystems that are commonly incorporated under the banner of BEF: stability and invasion-resistance (i.e. invasibility). As described, the BEF field grew out of the diversity-stability debate in ecology. Now, however, stability is generally seen as an aspect of ecosystem functioning (Cardinale *et al.* 2012). Thus, ecosystem functions are described in terms of their magnitude and their stability which is typically measured as temporal variability (e.g. the coefficient of variation through time). Other aspects of stability like the resistance and resilience of ecosystem functions are also studied (Donohue *et al.* 2013). Moreover, the resistance of communities to invasion by other species (both native and non-native) is also studied within the BEF field (Kennedy *et al.* 2002; Levine 2000). Invasion-resistance is generally studied by examining how many new species establish in a community compared to its biodiversity (Levine *et al.* 2004). In this thesis, I focus on the magnitude of ecosystem functioning and do not consider stability or invasion resistance. Thus, I do not treat these topics further.

Theoretical basis for biodiversity effects on ecosystem functioning

Classic theory that relates biodiversity to the magnitude of ecosystem functioning is largely based on competition theory. There are several models that have been used to explore how biodiversity effects ecosystem functioning (e.g. Loreau 1998a; Tilman *et al.* 1997; Turnbull *et al.* 2013; Yachi & Loreau 1999). Here, I will first illustrate the main results that are common to all these models using the classic Lotka-Volterra competition model for two species following Loreau (2004). For two species, the Lotka-Volterra competition model takes the following form:

Equation 7

$$\frac{dN_i}{dt} = \frac{r_i N_i}{K_i}\left(K_i - N_i - \alpha_{ij}N_j\right) \quad i,j = 1,2$$

In this equation, N_i is the biomass of species i. The fixed parameters in the model are: r_i – the intrinsic growth rate of species i, K_i – the carrying capacity of species i and α_{ij} which measures the competitive effect of species j on species i. In this model, the intraspecific competition coefficient is implicitly one. The α_{ij} parameters thus relate the effect of species j on species i in units of species i. In this model, differences in K represent differences in species' competitive abilities whilst the α parameters indicate niche differentiation (Vellend 2016).

For this model with two species, it can be shown mathematically that both species can stably coexist (i.e. coexist indefinitely in the absence of demographic stochasticity) if the following inequality is satisfied:

Equation 8

$$\alpha_{21} < \frac{K_2}{K_1} < \frac{1}{\alpha_{12}}$$

Intuitively, the inequality shows the competitive difference (K_2/K_1) cannot be too large relative to the competitive effects. For example, if α_{12} is 0.8, then K_2 can only be 25% greater than K_1. Thus, there is a balance between the species niche differences (α-values) and the competitive ability differences (K-values), (Chesson 2000; Barabás *et al.* 2018). If the inequality is met, then we can calculate the equilibrium abundances of species 1 (N^*_1) and species 2 (N^*_2) as follows:

Equation 9-10

$$N_1^* = \frac{K_1(1 - \frac{K_2}{K_1}\alpha_{12})}{1 - \alpha_{12}\alpha_{21}}$$

$$N_2^* = \frac{K_1(\frac{K_2}{K_1} - \alpha_{21})}{1 - \alpha_{12}\alpha_{21}}$$

Using these coexistence conditions, the equilibrium abundances of species 1 and 2 (N_1^* and N_2^*) and their carrying capacities (K_1 and K_2 which are equivalent to monoculture functioning), we can explore the key theoretical results that form the basis of the BEF field. First, under conditions of stable coexistence in the Lotka-Volterra model, an initial 50:50 mixture of species 1 and 2 will lead to a mixture where both species persist indefinitely. Moreover, the mixture functioning will almost always be higher functioning than the average monoculture functioning of the species (i.e. the carrying capacities, Fig. 7a) and this can be due to a combination of selection and complementarity effects. However, transgressive overyielding only occurs when niche partitioning is strong enough (i.e. low enough α_{ij} values, Fig. 7a).

When the conditions for stable coexistence are not met, one species will inevitably go extinct if a 50:50 mixture of species 1 and 2 is created. However, if the species differ in their monoculture functioning and the species with the higher monoculture functioning outcompetes the other species, then the mixture functioning will be higher than the functioning of the average monoculture (Fig. 7b). In this case, the positive effect of biodiversity on ecosystem functioning is due to the selection effect. However, without stable coexistence, no transgressive overyielding is possible and the mixture, over time, becomes a single species monoculture (Fig. 7b).

These basic theoretical results have been reproduced in multiple models that include more mechanistic resource competition (Loreau 1998a; Turnbull *et al.* 2013). But the general results tend to be the same. In the framing of a BEF experiment, we tend to observe positive relationships between biodiversity and ecosystem functioning. This is because there can be some niche partitioning or facilitation leading to complementarity effects or because of selection effects when species differ in their monoculture functioning and high functioning species in monoculture tend to dominate the

mixtures. However, when only selection effects are operating, transgressive overyielding is not possible and the species richness of the mixture will decrease through time.
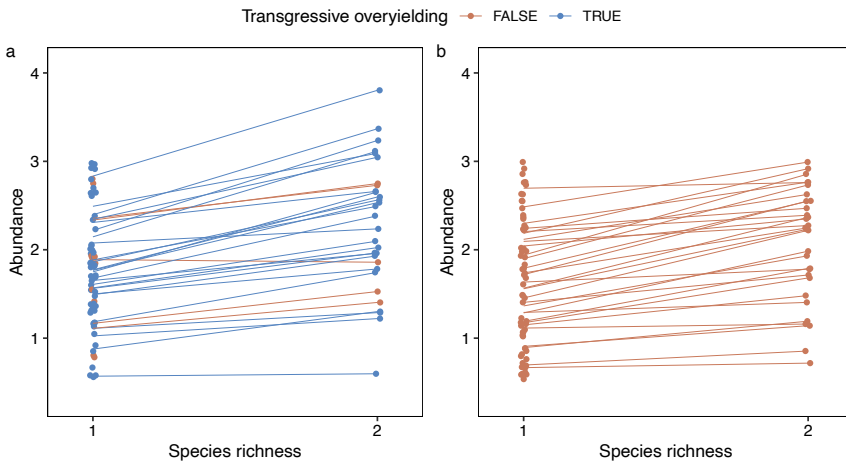

Fig. 7: Simulated equilibrium abundances of 30 BEF experiments using a Lotka-Volterra competition model (Loreau 2004). In **(a)** parameter values lead to stable coexistence of the two-species mixture and in **(b)** one species always outcompetes the other species in mixture. For simplicity, if stable coexistence was not possible, the species with the highest monoculture functioning (K values) dominated the mixture. K_i values were drawn from Uniform(0.5, 3) and α_{ij} values were drawn from Uniform(0.25, 1.25). Parameters were assessed for stable coexistence as described in the text. Lines connect the average functioning of the two monocultures and the mixture for a given simulated experiment.

Based on these and a host of other theoretical results, there is little doubt about the interpretation of a BEF experiment. Moreover, Loreau and Hector's (2001) partition was an effective way to calculate complementarity and selection effects and, therefore, to try to separate out which idealised competition scenario was more consistent with the data (Fig. 7a *versus* Fig. 7b). However, these theoretical models were geared towards interpreting a typical BEF experiment. They do not necessarily make predictions about what the effect of biodiversity on ecosystem functioning should look like in natural ecosystems, something I address in **Paper I**.

Empirical evidence: BEF experiments

In my view, the evidence-base that underlies the BEF field is overwhelmingly based on classic BEF experiments (i.e. Fig. 4). Indeed, several hundred of these experiments have been done using a variety of organisms from terrestrial, marine and freshwater ecosystems worldwide (Cardinale *et al.* 2012). There have been several excellent quantitative reviews of these experiments which I will use to provide an overview of the general patterns. In addition, there are two grassland BEF experiments that have been going for more than 20 years and have had an arguably disproportionate impact on the field (BioDiv: https://www.cedarcreek.umn.edu/ and the Jena Experiment: http://the-jena-experiment.de/). Thus, I will also put some special focus on these experiments as well.

24

As far as I can tell, the five most recent meta-analyses of biodiversity-ecosystem functioning experiments are: Cardinale *et al.* (2011) for primary producers, Griffin *et al.* (2013) for the effect of predator biodiversity on prey abundance, Gamfeldt *et al.* (2015) for marine ecosystems, Raffard *et al.* (2019) for intraspecific biodiversity and Hong *et al.* (2022) for ecosystems experiencing global change. Some publications overlap in these meta-analyses but, given that they all address different trophic levels, ecosystems etc., independently reviewing their results is still useful.

Cardinale *et al.*'s (2011) analysis was the most comprehensive and summarised the results from 574 independent manipulations of species richness (as the measure of biodiversity) from 192 publications. Their main results can be summarised as follows. Across the experiments, the effect of species richness on primary producer biomass and nutrient uptake tended to be positive. These positive effects were due to a combination of complementarity and selection effects (as quantified as per Loreau and Hector 2001). On average, the magnitudes of complementarity and selection effects were similar for terrestrial ecosystems, but selection effects were mostly negative in aquatic ecosystems. Despite average positive effects of species richness, the most species richness mixture was only higher functioning than the highest functioning monoculture (i.e. transgressive overyielding) in around 30% of experiments. Finally, the most common shape of the relationship between species richness and ecosystem functioning was positive but decelerating.

Both Griffin *et al.* (2013) and Gamfeldt *et al.* (2015) came to similar conclusions. Species richness tended to increase average ecosystem functioning defined as prey suppression by Griffin *et al.* (2013) and either production, consumption or biogeochemical fluxes by Gamfeldt *et al.* (2015) but, on average, there was no transgressive overyielding. Gamfeldt *et al.* (2015) also tested for the shape of the relationship between species richness and ecosystem functioning. They found that the shape of the relationship differed based on the ecosystem function measured: linear for production, positive but decelerating for consumption and inconclusive for biogeochemical fluxes. Neither study applied Loreau and Hector's (2001) partition but, together, the results are generally consistent with Cardinale *et al.* (2011).

Hong *et al.* (2022) analysed 46 experiments that tested the effects of both species richness and an anthropogenic stressor (e.g. warming, eutrophication etc.) on ecosystem functioning across microbes, phytoplankton and terrestrial plants. On average, Hong *et al.* (2022) showed that biodiversity positively affected ecosystem functioning under control and stressed conditions which is consistent with the previously cited meta-analyses (Cardinale *et al.* 2011; Gamfeldt *et al.* 2015; Griffin *et al.* 2013). But, Hong *et al.* (2022) also found that the effects of species richness on ecosystem functioning tended to be stronger under stressful conditions and that complementarity effects were stronger than selection effects (*sensu* Loreau & Hector 2001). They did not, however, analyse transgressive overyielding.

The four meta-analyses cited only analysed the effect of species richness manipulations on ecosystem functioning. However, Raffard *et al.* (2019) analysed experiments that examined the effect of the number of genotypes or phenotypes of a single species on ecosystem functioning. The experimental designs match that of typical BEF experiments (e.g. Fig. 4). Interestingly, Raffard *et al.* (2019) also report an average increase in ecosystem functioning with the number of genotypes or phenotypes. Moreover, the shape tended to be positive and decelerating as found by Cardinale *et al.* (2011). Therefore, the results from these manipulations of intraspecific biodiversity are similar to studies manipulating species richness.

In addition to these meta-analyses, there are two grassland BEF experiments that have both been running for than 20 years now: BioDiv (species richness: 1, 2, 4, 8 and 16) and the Jena experiment (species richness: 1, 2, 4, 8, 16 and 60). Both experiments show positive, decelerating relationships between species richness and biomass productivity (Jochum *et al.* 2020; Reich *et al.* 2012; Wagg *et al.* 2022) driven by a combination of complementarity and selection effects (*sensu* Loreau and Hector 2001). These results match with the five meta-analyses reviewed. However, given that these experiments have been running for more than 10 years, researchers have also been able to test whether the effect of species richness on biomass productivity has changed through time. Interestingly, in both experiments, the effect of species richness on biomass productivity strengthened through time (Reich *et al.* 2012; Wagg *et al.* 2022). In the case of the BioDiv experiment, this was due to the complementarity effect increasing in strength through time (Reich *et al.* 2012). For the Jena experiment, however, this was due to monoculture functioning declining through time faster than functioning in mixtures (Wagg *et al.* 2022). Nonetheless, in both these cases biodiversity effects on functioning increased through time, a conclusion supported by a recent synthesis (Qiu & Cardinale 2020).

The results of these meta-analyses and long-term experiments indicate that, in a BEF experiment, the increase in some aspect of ecosystem functioning (e.g. biomass production, consumption etc.) with biodiversity (usually species richness but also the number of genotypes/phenotypes) is an empirical generality. Based on the evidence, the most common relationship is a positive, decelerating curve. Moreover, when calculated, complementarity effects tend to be slightly stronger than selection effects, but this varies considerably between experiments. Finally, for experiments that have been ongoing for several years, positive effects of biodiversity on ecosystem functioning tend to increase through time.

Given the weight of empirical evidence, few ecologists would doubt that biodiversity (usually species richness) does frequently increase ecosystem functioning in the context of a BEF experiment. However, whether the results from BEF experiments are relevant to understanding how biodiversity affects ecosystem functioning in natural ecosystems is a different matter. In the next section, I review other approaches to BEF that aim to determine the effect of biodiversity in natural ecosystems.

Empirical approaches: Beyond the BEF experiment

Many have argued that the typical BEF experiment is too unrealistic to be useful for understanding how biodiversity loss in natural systems will affect ecosystem functioning (Lepš 2004; Wardle 2016). One of the main arguments is that BEF experiments simulate random species loss whilst species loss in natural ecosystems occurs due to environmental changes, dispersal limitation or competitive exclusion (De Laender *et al.* 2016; Díaz *et al.* 2003; Lepš 2004). Another common argument is that the communities assembled in BEF experiments are unrealistic (Buchmann *et al.* 2018). For example, communities in BEF experiments tend to have a much more even relative abundance distribution than natural communities (Hillebrand *et al.* 2008). Several authors have attempted to address these claims (Duffy 2009; Jochum *et al.* 2020) but the realism of BEF experiments is still frequently questioned.

In my reading of the literature, there have been four main approaches to trying to understand how biodiversity affects ecosystem functioning in realistic, natural ecosystems: i. removal experiments, ii. realistic biodiversity loss experiments, iii. correlative approaches that relate

biodiversity to ecosystem functioning in field data and iv. statistical partitions based on the Price equation. Removal experiments are based on removing certain species from intact natural communities and examining how those communities respond in terms of both biodiversity and ecosystem functioning (Díaz *et al.* 2003). Such experiments can be customised to remove certain species that, for example, are particularly extinction prone or have specific traits that may disproportionately affect ecosystem functioning. In the case of single trophic levels, removal experiments tend to reduce ecosystem functioning and the results appear similar to BEF experiments (see Kardol *et al.* (2018) for one of the longest running removal experiments).

Realistic biodiversity loss experiments are another popular alternative to the classic BEF experiment. The idea is to examine how biodiversity might be expected to change in the future due to, for example, anthropogenic stressors or stochastic extinction, and examine how ecosystem functioning might change as a result. This can be done by artificially creating communities like in a BEF experiment but with non-random species compositions (Bracken *et al.* 2008) or it can be done by parameterising models of ecosystem functioning and simulating changes in biodiversity according to specific scenarios of biodiversity loss (Solan *et al.* 2004; Thomsen *et al.* 2019; **Paper IV**). Either way, these experiments tend to show that the effect of realistic species loss on ecosystem functioning differs from that of random species loss (Bracken & Williams 2013; Smith *et al.* 2020; Smith & Knapp 2003; Thomsen *et al.* 2017). Whether the effect is stronger or weaker depends on the species loss scenario such as whether rare species or dominant species are lost (Lisner *et al.* 2023; Smith & Knapp 2003) or whether species show compensatory responses or not (Thomsen *et al.* 2017).

The third approach is to study empirical relationships between measures of biodiversity and measures of ecosystem functioning in observational field data from natural systems. If you were brave enough to read the whole introduction, you will remember the hump-shaped relationship between biomass productivity and biodiversity which contradicted the results from BEF experiments (Fig. 6a; Fraser *et al.* 2015; Schmid 2002). The controversy was, however, resolved by recognising that BEF experiments kept the environmental conditions constant whilst observational field studies reporting the hump-shaped relationships were examining these relationships across ecosystems. Of course, in observational studies we cannot directly control for environmental conditions, but we can measure the environment and attempt to statistically control it (Duffy *et al.* 2017). This is the approach that is typically taken: biodiversity is related to some measure of ecosystem functioning in a multiple regression framework in order to try and statistically control for environmental variation (Duffy *et al.* 2017; Gamfeldt *et al.* 2013; Grace *et al.* 2016).

In the last 10 or so years, this approach has become exceedingly popular (reviewed in van der Plas 2019) which is likely due to the availability of monitoring data for a wide variety of ecosystems and taxon groups. Recently, there have been two reviews of these studies. First, Duffy *et al.* (2017) quantitatively reviewed 133 estimates of the effect of biodiversity on ecosystem functioning after statistically controlling for environmental variation and found that 75% of the estimates were positive. Second, van der Plas (2019) systematically reviewed 726 estimates of the effect of biodiversity on ecosystem functioning across a wide variety of taxa and ecosystem functions. On balance, there were more positive effects than negative effects, but the majority of effects were neutral. van der Plas' (2019) database was more comprehensive and Duffy *et al.* (2017) used a vote-counting approach to quantitative meta-analysis based on a variety of different tests (e.g. significant regression coefficients, variables retained after model selection etc.) which is generally not recommended. Thus, van der Plas' (2019) conclusions are, in my opinion, more reliable. I interpret these studies as showing that positive

relationships between biodiversity and ecosystem functioning are observed in observational field data but not at the same frequency as in BEF experiments (some may disagree with this conclusion) (**Paper I**).

The fourth and final approach to studying the effects of biodiversity on ecosystem functioning in natural systems is to use statistical partitions based on the Price equation pioneered by Fox (2006). The approach partitions a difference in ecosystem function between two communities (in space or time) into additive components that can be attributed to different effects. Originally, Fox (2006) partitioned the difference into three effects: i. random loss of species, ii. a change in species composition and iii. a change in functioning of species present in both communities. However, Fox's (2006) original partition has been expanded to other effects as well (Genung *et al.* 2020; Lefcheck *et al.* 2021; Winfree *et al.* 2015). There have been no syntheses of these results but, in my reading of the literature, some studies find a strong role for species richness in explaining changes in ecosystem functioning across communities (Albrecht *et al.* 2021; Lefcheck *et al.* 2021), others find weaker effects (Genung *et al.* 2020; Winfree *et al.* 2015). A problem with this approach is that it requires species-specific estimates of functioning in communities (Fox 2006) which can be very difficult to obtain for certain ecosystem functions like nutrient uptake. Moreover, it is purely descriptive in determining what drives a difference in functioning between communities and, therefore, does not attempt to establish any causal relationships which is often the goal of BEF research.

My view is that all four of these approaches show some evidence that biodiversity affects ecosystem functioning. However, the effects are not as consistent as in BEF experiments, especially when correlating biodiversity with ecosystem functioning in field data. Moreover, both removal experiments along with realistic biodiversity loss experiments show that the effects on functioning of biodiversity loss can differ from what we observe in a BEF experiment.

Multiple facets of biodiversity

In BEF research, biodiversity has overwhelmingly been operationalised as species richness. However, most researchers would probably agree that species richness is not a very good variable in the context of BEF research. The reason for this was pointed out very early on by Bengtsson (1998) and has been consistently repeated: relating species richness to ecosystem functioning assumes that all species are equivalent in their effects on functioning (Roger 2017). This assumption is easily rejected based on the differences among species in monoculture functioning (Hector *et al.* 1999; Tilman *et al.* 1996; Vile *et al.* 2006), removal experiments (Díaz *et al.* 2003) and realistic species loss experiments (Bracken & Williams 2013) and is because species differ in their traits. Moreover, BEF theory is based on phenotypic trait diversity among species. Therefore, species richness is only useful based on how it relates to phenotypic trait variation.

The problem, however, is that species richness will only be a good proxy for phenotypic trait variation under two scenarios: i. species randomly occupy niche space or ii. species are evenly spread out in niche space (Díaz & Cabido 2001). However, this is often not the case as species are frequently non-randomly clumped in niche space (Díaz & Cabido 2001; Scheffer & van Nes 2006). Given the limitations of species richness as a variable, a large body of BEF work has attempted to use better measures of biodiversity that are more directly linked to phenotypic trait variation.

Biodiversity metrics that directly measure trait variation are typically known as *functional diversity metrics* (or simply *functional diversity*). Early work trying to measure functional diversity

lumped species into functional groups based on prior knowledge and used the number of functional groups as a proxy for functional diversity (Naeem & Li 1997; Tilman *et al.* 2001). However, as Petchey and Gaston (2002) noted, traits are generally continuous variables and, as a result, any lumping of species into functional groups is going to be arbitrary at some level. Moreover, the trait differences between any two functional groups are considered equal which is highly unlikely given a set of species with continuous trait values. Thus, many authors have proposed continuous measures of functional diversity that explicitly account for differences among species in multivariate trait space. A full overview of functional diversity metrics is well-beyond the scope of this thesis (for an excellent overview, see Roger 2017). However, the basic principle is to place species in multivariate trait space and calculate measures of variation. For example, Petchey and Gaston (2002) define functional diversity as the sum of branch lengths in a dendrogram created by clustering species based on their traits. Functional diversity metrics are, currently, a standard tool when conducting BEF research.

The problem with functional diversity metrics is that the choice of traits is usually relatively arbitrary. Moreover, the choice of traits will undoubtedly influence the value of functional diversity which makes it an important consideration to which there is no satisfactory solution (Petchey & Gaston 2006). The issue of choosing traits can be somewhat circumvented by using *phylogenetic diversity* among species as a proxy for functional diversity. Phylogenetic diversity is typically defined as the sum of branch lengths on a molecular phylogeny that connects species present in a community (Tucker *et al.* 2017). The idea is that the relatedness among species summarises overall trait differences among species (Cadotte 2013). However, this is not always the case as closely related species can have large trait differences if, for example, they have evolved via adaptive radiation (Losos 2008). Moreover, phylogenetic diversity among species in any given community cannot always be assumed to represent functional diversity depending on the community assembly mechanisms (Mayfield & Levine 2010; Srivastava *et al.* 2012).

Given these different metrics of functional diversity, many studies have compared the predictive value of species richness, functional diversity and phylogenetic diversity for ecosystem functioning (Cadotte 2013; Venail *et al.* 2015). Early studies showed that quantitative functional diversity metrics were better predictors of aboveground biomass than species richness in sites from the BIODEPTH experiment (Petchey *et al.* 2004). Similarly, Cadotte (2013) showed that phylogenetic diversity was a better predictor of biomass productivity than species richness in a grassland biodiversity experiment. Cardinale *et al.* (2015), in contrast, showed that species richness was a better predictor of biomass productivity than phylogenetic diversity across 16 grassland biodiversity experiments. It must, however, be noted that the differences in predictive power of species richness, functional diversity and phylogenetic diversity differ by a maximum of 10% (but often less). These studies (and many others) indicate that neither species richness, functional diversity nor phylogenetic diversity are consistently superior when trying to predict ecosystem functioning.

This body of work is, whilst interesting, largely trivial in my opinion. Whether or not species richness is a slightly better predictor of functioning than phylogenetic diversity or functional diversity (or *vice versa*) is irrelevant. All three metrics are incomplete descriptors of the level of phenotypic trait diversity relevant to a given ecosystem function in a community (Roger 2017) and there are good reasons as to why this is the case (Mayfield & Levine 2010; Petchey & Gaston 2006; Srivastava *et al.* 2012). Functional diversity and phylogenetic diversity are more directly linked to phenotypic trait diversity on which BEF theory is based. However, in my view if these limitations are acknowledged, any of the three metrics can be useful variables. In this thesis, I generally talk about biodiversity more

generally and this is often measured as species richness. However, this is purely for practical reasons. Any effects of species richness on functioning can only logically act through phenotypic trait variation (Bengtsson 1998; Loreau 2000).

Multifunctionality

Much of the recent work in the BEF field has studied more than one function (Byrnes *et al.* 2014). This is because, during the mid- to late-2000's, BEF research was criticised for being too focused on single functions (e.g. consumption, biomass productivity). As Gamfeldt *et al.* (2008) pointed out, this may miss a very important aspect of biodiversity: different species may be important for different functions in a given place and at a given time. Thus, many functions may be more susceptible to the loss of biodiversity than single functions (Gamfeldt *et al.* 2008). Moreover, it has been argued that the lack of transgressive overyielding seen in BEF experiments is due to single functions being studied (Byrnes *et al.* 2014). The argument is that single species may maximise single functions but are unlikely to maximise multiple functions simultaneously. However, as far as I can tell, there has never been a proper test of this hypothesis. Rather, studies have shown that the presence/abundance of different species correlates with different functions suggesting that species are, to some extent, functionally unique (Hector & Bagchi 2007; Isbell *et al.* 2011). In addition, BEF experiments, and observational field studies have studied the relationship between biodiversity and metrics of multifunctionality, usually the average value among a set of functions known as average multifunctionality (e.g. (function 1 + … + function N)/N). A recent meta-analysis of BEF experiments that measured multiple functions found that biodiversity positively affected average multifunctionality as observed for single functions in BEF experiments (Lefcheck *et al.* 2015). Although I do not treat multifunctionality specifically in this thesis, two papers that I worked on during my PhD address aspects of ecosystem multifunctionality.

Summary

This section (*Background: The "hard core" of the BEF research program*) was designed to give readers an overview of the major theoretical and empirical results that underpin the BEF field. Moreover, I tried to highlight certain topics that are and have been popular in the BEF research in recent times (e.g. functional/phylogenetic diversity and multifunctionality). The next section attempts to describe unsolved problems in the BEF field and how this thesis aims to address some of them.

Section 2: Current issues in the BEF field

In this section, I describe two of the major unsolved problems in the current BEF field, both of which I address in my thesis. The first is what the results of BEF experiments tell us about the consequences of biodiversity loss in natural ecosystems (**Paper I**). The second unsolved problem is how biodiversity affects functioning at large scales where the environment varies in multiple dimensions of space and time (**Papers II, III**). The two problems are highly linked, which is why **Papers IV** and **V** are relevant for both issues.

2.1: What do experiments tell us about the consequences of biodiversity loss in natural ecosystems?

A major topic in the BEF literature over the last 10-15 years has centred on whether the results from BEF experiments (like those described in Fig. 4) can inform us about the consequences of biodiversity loss in natural ecosystems. The idea can be summarised as follows. We know that biodiversity is declining at a range of different spatial and temporal scales (Barnosky *et al.* 2011; McGill *et al.* 2015; Newbold *et al.* 2015; Pimm *et al.* 2014). We also know, based on BEF experiments, that experimentally reducing biodiversity, on average, causes a decrease in many different ecosystem functions (Cardinale *et al.* 2012). Therefore, we need to halt and reverse biodiversity loss if we want to maintain high functioning ecosystems. Whilst this seems reasonable enough, there are major issues with this simple deduction which I will attempt to unpack in this section.

The first issue is that the relationship between biodiversity and ecosystem functioning in BEF experiments is based on unrealistic communities of species (Lepš 2004). Specifically, experimental communities are created by drawing species randomly from a species pool (Fig. 4). Therefore, for these communities to represent natural ecosystems, the natural communities with few species must be random subsets of the natural communities with more species (Wardle 2016). As has been shown many times, this is not the case. Rather, in natural ecosystems, an assemblage is made up of species that are adapted to the environment and can tolerate the local species interactions (Germain *et al.* 2017, 2018). This means that, as Lepš (2004) notes, one would never find a community made up only of unproductive species in a "productive environment" (i.e. an environment with high resource availability) yet this is exactly the type of unrealistic communities often created by BEF experiments.

A second and related issue is that, when species are lost from communities in natural ecosystems, they are generally not lost randomly (Wardle 2016). Rather, species with certain traits are more or less likely to be lost from communities and this probability of loss can covary with their effects on functioning. For example, species that are locally rare tend to have a higher extinction risk (Hubbell 2011). Moreover, removal experiments have demonstrated that removing rare species from a community has minor effects on ecosystem functions like productivity (Lisner *et al.* 2023; Smith & Knapp 2003). This point becomes even more important when we consider environmental change.

Most changes in species composition occur because of some kind of environmental change (Harpole *et al.* 2016; Newbold *et al.* 2015; Zhou *et al.* 2020). And, I would argue that this is the type of species loss that most researchers are concerned with. For example, many researchers are interested in questions like: what will happen to biodiversity and ecosystem functioning if we continue polluting

this stream or if heatwaves become more common due to climate change? As De Laender *et al.* (2016) showed theoretically, the relationship between biodiversity and ecosystem functioning following environmental change will depend critically on which species are most affected by the environmental change. More specifically, De Laender *et al.* (2016) showed that, when dominant species respond most positively to an environmental change, functioning increases whilst species richness decreases due to competitive exclusion, resulting in a negative BEF relationship. Probably the most studied example of this is the increase in productivity and decrease in biodiversity following nutrient enrichment in many plant communities (Hautier *et al.* 2009; Silvertown *et al.* 2006) but there are other, similar examples (Baert *et al.* 2018; Spaak *et al.* 2017).

A third important issue is that, despite mostly positive relationships between biodiversity and functioning in experiments, relationships between biodiversity and functioning in field data from natural ecosystems are more variable (reviewed in van der Plas 2019; **Paper I**). Studies using field data from natural ecosystems typically use some sort of linear regression model to examine the relationship between a measure of biodiversity (e.g. species richness, functional diversity etc.) and a measure of functioning (e.g. nutrient fluxes, productivity etc.) whilst adjusting for various environmental covariates (e.g. climate, soil type etc.). As discussed previously, there have been two syntheses of these kinds of studies. Duffy *et al.* (2017) found mostly positive relationships. However, van der Plas (2019) found that most relationships were neutral and that there were more positive than negative relationships.

Finally, most of the experimental BEF work has been done on small spatial scales with limited environmental heterogeneity (Cardinale *et al.* 2011; Gonzalez *et al.* 2020). Thus, the observed effect of biodiversity on ecosystem functioning, it is argued, refers to an effect at these small spatial scales. Vellend *et al.* (2013) pointed out that, for BEF experiments to be relevant for understanding the effects of biodiversity loss in natural systems, the declines in biodiversity in natural systems must be happening at similar spatial scales. However, in a meta-analysis of repeated vegetation surveys, Vellend *et al.* (2013) found no general decline in plant species richness through time. Instead, increases in plant species richness were as common as decreases in plant species richness. Similar results have now been found for a range of taxonomic groups in a variety of habitats (Blowes *et al.* 2019; Dornelas *et al.* 2014; Elahi *et al.* 2015). These studies have, however, shown that species composition is changing through time (Blowes *et al.* 2019; Dornelas *et al.* 2014).

Undoubtedly, Vellend *et al.* (2013) and others (Blowes *et al.* 2019; Dornelas *et al.* 2014; Elahi *et al.* 2015) raised an interesting point. However, in my view, they are wrong in their assertion that this "calls into question the use of BEF research to justify conserving biodiversity" (paraphrasing from Vellend *et al.* 2013). Firstly, even if biodiversity in only half of all communities is declining (as they state), then BEF research is still useful for that half, a non-trivial amount. Secondly, the idea that local biodiversity is, on average, stable is, in my opinion, a very optimistic way to look at the data (see also Gonzalez *et al.* 2016). All these studies use time-series data (i.e. communities that have been monitored through time). That, by definition, excludes many land-use changes that go from high biodiversity to zero biodiversity such as a grassland being converted to agricultural fields, parking lots or highways in which, for example, *all plant species* are lost (see Cardinale & Loreau's 2020 and Baum *et al.*'s 2020 replies to Blowes *et al.* 2019). This kind of land-use change is arguably the most pervasive driver of biodiversity loss and is not accounted for in these studies. When land-use is specifically modelled as a driver of biodiversity loss, analyses suggest that local biodiversity is

declining, on average (Newbold *et al.* 2015). Thus, in my view, local biodiversity probably is declining but accurately quantifying this across the whole globe is exceedingly challenging.
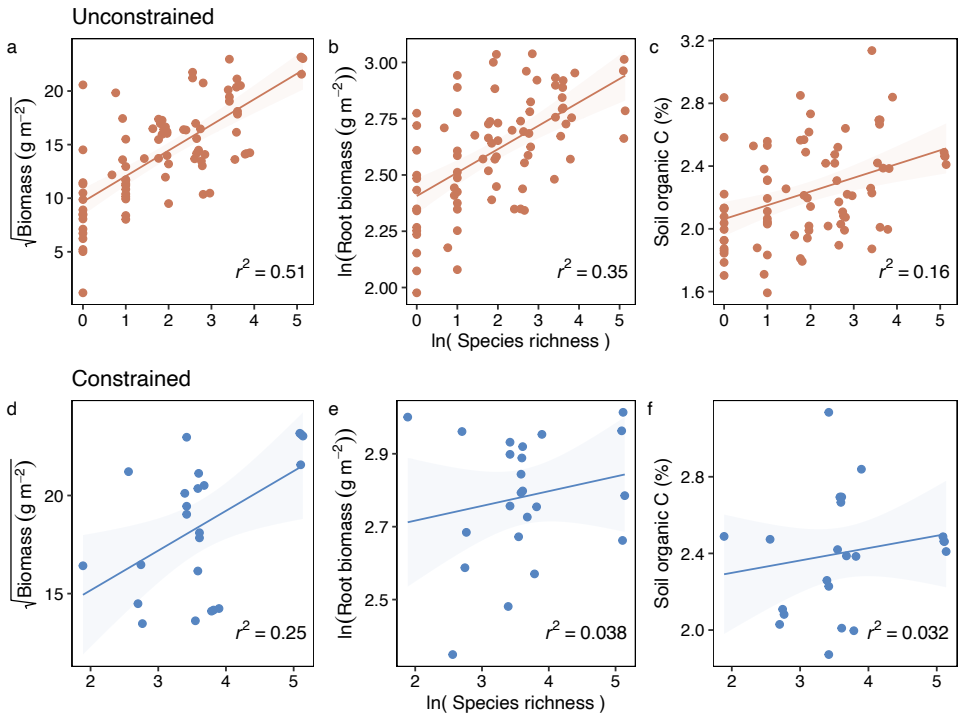

Fig. 8: The relationship between biodiversity (measured as ln-transformed species richness) and three measures of ecosystem functioning (aboveground biomass, root biomass and soil organic carbon) for **(a-c)** all experimental communities in the Jena experiment and **(d-f)** only communities with realistic species composition based on field data. Line is a linear regression line with 95% confidence interval. The model's coefficient of determination (r^2) is also reported. It is clear that that the relationships are considerably weaker when the communities are constrained to communities with realistic species compositions. A subset of the data are replotted from Jochum et al. (2020).

I do not want to paint a biased view of the literature here. Indeed, many researchers have argued that BEF experiments do tell us about the consequences of biodiversity loss in natural ecosystems (Duffy 2009; Eisenhauer *et al.* 2016; Jochum *et al.* 2020). Specifically, Duffy (2009) argued that there are many BEF experiments that have in fact mimicked more realistic species loss scenarios, such as those observed over ecological gradients, and that they tend to find stronger effects of biodiversity on ecosystem functioning than the random assembly experiments. Indeed, there are examples of this (Bracken *et al.* 2008; Bracken & Williams 2013; Solan *et al.* 2004) but there are also many counterexamples (Lisner *et al.* 2023; Smith *et al.* 2020; Smith & Knapp 2003) and probably not enough to make such a generalisation. Moreover, in my view, whether or not realistic species loss BEF experiments find stronger effects than random loss BEF experiments is a moot point. Critics are not necessarily arguing about the strength of the effect. Rather, critics are trying to point out that, for ecosystem functioning, which species are lost is considerably more important than whether

biodiversity, in general, declines (Bengtsson 1998). In my reading of the literature, this latter point is undeniable.

The only empirical analysis of which I am aware that tried to tackle this question directly was Jochum *et al.* (2020). Jochum *et al.* (2020) re-analysed data from two grassland biodiversity experiments (cited previously: Jena experiment and BioDiv). However, they used extensive plant community composition data from natural grasslands on which the experiments were based to exclude communities in the experiments that were not found in the natural grasslands. For example, many species in monoculture were not found in the data from the natural grasslands and thus were excluded from the analyses. The results were mixed. For some functions, excluding unrealistic assemblages changed the slope of the relationship between biodiversity and functioning and for other functions, it did not (Fig. 8). Whilst the approach is innovative, the level of the function values of the experimental plots and the compositionally similar plots in the natural grassland could be completely different due to differences in community assembly (Schmid *et al.* 2022). Thus, in my view, the evidence is not particularly strong.

Based on the balance of evidence, as I see it, there are major issues with the idea that a loss of local biodiversity in natural ecosystems will lead to reductions in ecosystem functioning. However, this does not mean that BEF experiments do not tell us anything about natural ecosystems (Eisenhauer *et al.* 2016). In **Paper I**, I use ecological theory to try and think carefully about what a BEF experiment does and does not tell us about the consequences of biodiversity loss in natural ecosystems. In **Papers IV** and **V**, I try and more robustly quantify biodiversity effects on ecosystem functioning using field data from natural systems.

2.2: Biodiversity and ecosystem functioning at large spatial scales

Most experimental BEF work has been done on small spatial scales with limited environmental heterogeneity (Gonzalez *et al.* 2020). In Cardinale *et al.*'s (2011) meta-analysis of 574 BEF experiments, the median plot size of the experiments done on terrestrial plants was 3 m^2 and 0.1 L for experiments done on aquatic organisms. However, when it comes to understanding and managing natural ecosystems, we are typically interested in much larger spatial scales (e.g. landscapes, regions etc.) (Isbell *et al.* 2017). This raises some important questions. For example, if experimental BEF work has shown that ecosystem functioning decreases with biodiversity loss at small scales, would biodiversity loss also decrease ecosystem functioning at larger scales? And, if so, would the effect be stronger or weaker? If *N* species are required to maximise ecosystem functioning in an experimental plot, how many species are required at the scale of a landscape or a region composed of 100's or 1000's of such plots? Questions like this are, currently, at the frontier of the BEF literature (reviewed in Gonzalez *et al.* 2020).

Although there is currently a lot of interest in these questions, Tilman (1999) wrote about the BEF relationship and spatial scale 24 years ago in his opinion piece about the BIODEPTH project (Hector *et al.* 1999). The BIODEPTH project showed that approximately 16 species were required in a 1 m^2 plot to have high levels of productivity (the exact number at this scale could be debated). Tilman (1999) then asked: If 16 species are required in 1 m^2, how many would be required in 100 ha? To answer this question, he employed the species-area relationship (Rosenzweig 1995) along with the theory that regional diversity is a determinant of local diversity (Cornell & Harrison 2014). Tilman (1999) then defined the local species-area relationship as follows:

Equation 11

$$S_L = cA_L{}^z$$

Where S_L is the local species richness (i.e. 16 in this example), c is a constant, A_L is the local plot area (i.e. 1 m² in this example) and z is the power-law exponent which typically ranges from 0.15 to 0.3 depending on the ecosystem (Rosenzweig 1995). At the regional scale, the species-area relationship is defined in the same way:

Equation 12

$$S_R = cA_R{}^z$$

But, now S_R is the quantity we want and A_R is the regional area which, in this example, is 100 ha. To solve for S_R, all that we need is a simple substitution:

Equation 13-14

$$c = \left(\frac{S_L}{A_L{}^z}\right)$$

$$S_R = \left(\frac{S_L}{A_L{}^z}\right)A_R{}^z = S_L\left(\frac{A_R}{A_L}\right)^z$$

Using this expression and calculating S_R for all possible values of z between 0.15 and 0.3 (0.01 intervals), we arrive at a required regional diversity (i.e. S_R) of between 127 and 1010 species. Thus, for a typical 1 m² patch of European grassland to contain 16 species and assuming that regional diversity is important for maintaining local diversity (Cornell & Harrison 2014), a 100 ha patch of grassland would need between 127 and 1010 species based on the species-area relationship (see also Tilman *et al.* (2014) for a similar analysis).

This approach is undoubtedly creative but it has problems. First, whilst it can be argued that, across all sites in the BIODEPTH project, 16 species are required to maintain high levels of productivity (Hector *et al.* 1999), one could just as easily argue the opposite. In some sites there was no discernible relationship between biodiversity and productivity (i.e. Silwood, Greece and Ireland) and, in other sites, some monocultures were as productive as the average highest diversity mixture (Hector *et al.* 1999). Thus, the idea that 16 species are *required* is spurious. Second, whether between 127 and 1010 species are needed in a 100 ha region to make sure that a typical 1 m² plot will have 16 species depends on the mechanisms of species coexistence (Hart *et al.* 2017). If a given 1 m² plot has sufficient environmental variation to allow the long-term coexistence of 16 species, there is no reason why having more species in the region should be required because this suggests that local species richness is fully or partially maintained at the regional scale. Thus, the BEF relationship at large spatial scales is likely considerably more complicated than suggested by Tilman's (1999) simple extrapolation.

Since Tilman (1999), there have been several studies about the BEF relationship and spatial scale and these studies have typically focused on two questions:

1. Does the effect of biodiversity on ecosystem functioning strengthen with spatial scale?

2. Are more species required to maintain ecosystem functioning at large compared to small spatial scales?

Here, I review both theoretical and empirical studies that have attempted to answer these questions.

Does the effect of biodiversity on ecosystem function strengthen with spatial scale?

As discussed in previous sections, small scale BEF studies have shown that ecosystem function can be maximised by many coexisting species (if complementarity effects dominate) or a few competitively dominant and high functioning species (if selection effects dominate) (Cardinale *et al.* 2012). And, in the majority of BEF experiments, the highest functioning species in monoculture has higher functioning than the most diverse mixture (i.e. there is no transgressive overyielding) (Cardinale *et al.* 2011; Chisholm & Dutta Gupta 2023; Gamfeldt *et al.* 2015). However, many authors have noted that, even if one or a few competitively dominant species maximise ecosystem functioning at small spatial scales, many species may still be required to maximise functioning at large spatial scales if species respond to spatial environmental variation (Chesson *et al.* 2013; Isbell *et al.* 2018) because spatial and temporal environmental variation tends to increase with scale (Hart *et al.* 2017; Levin 1992). As a result, several authors have hypothesised that the effect of biodiversity on ecosystem function should increase with spatial scale (Duffy 2009).

As far as I can tell, the first theoretical paper to specifically address whether the effect of biodiversity on ecosystem function changes with spatial scale was Cardinale *et al.* (2004). In the paper, Cardinale *et al.* (2004) used a simple Lotka-Volterra competition model to describe the biomass of different species (b_i) at the scale of a single patch through time as:

Equation 15

$$b_i(t + 1) = b_i(t) \times \exp\left(r_i \left(1 - \frac{b_i(t) + \alpha \sum_{j \neq i}^{N} b_j(t)}{K_i}\right)\right)$$

Where r_i is the intrinsic rate of increase in biomass of species i in the absence of other species, K_i is the carrying capacity of species i and α is the interspecific interaction coefficient. In this formulation of the Lotka-Volterra model, the intraspecific interaction coefficient is assumed to be 1. For all models, r_i varied between species such that the geometric mean was equal to 0.2 (Cardinale *et al.* 2004). Using this model, I simulated typical BEF experiments (i.e. Fig. 4) using four different scenarios (Table 1) which largely match the simulations performed by Cardinale *et al.* (2004). In each scenario, I simulated 600 regions composed of 20 patches. All patches in a region were seeded with between 1 and 20 different species (i.e. initial species richness level). Therefore, each level of initial species richness had 30 replicates with a randomly drawn composition of species (as per Fig. 4b). All patches were seeded with the same number of initial colonists (B_0) and these were equally divided between species in a given patch (B_0/N) in the line with the replacement series design. Each patch was simulated for 300 time-steps as exploratory simulations showed that this was a sufficient number of time-steps for the model to reach an equilibrium.

In the scenarios 1 and 3, all 12000 patches across all 600 regions are identical and all 20 species have the same carrying capacity (i.e. $K = 200$ for all species). However, the interspecific interaction

coefficient (α) is set to 0.2 which means that species can partition their niches within each patch. In scenario 1, I evaluate the BEF relationship where the individual patch is the experimental unit. This means I examine the relationship between patch-scale species richness and community biomass. In scenario 3, region is the experimental unit and, as a result, I examine the relationship between region-scale species richness and the average community biomass across the 20 patches. In scenarios 2 and 4, in each region, there are 20 different patch types and species differ in their carrying capacities. More specifically, each species has one patch type where it has the maximum carrying capacity ($K_{\max}$) of 1000 as determined by a Guassian function (Thompson *et al.* 2020, Table 1). In addition, the interspecific interaction coefficient is set to 1. This means that species partition niches between patch types but there is no niche partitioning within a patch type. In scenario 2, I evaluate the BEF relationship at the patch-scale and, in scenario 4, I evaluate the BEF relationship at the region-scale (see Table 1 for an overview of these scenarios). In all four scenarios, to measure the effect of initial species richness on community biomass (a proxy of ecosystem functioning), I fitted a power function using non-linear least squares to the data. I then measured the effect of biodiversity on ecosystem functioning using the slope parameter (i.e. b-parameter, $y \sim a \times (SR^b)$).

Table 1: Four different scenarios simulated using Cardinale et al.'s (2004) model to determine whether the effect of species richness on biomass productivity changes with spatial scale.

Scenario	K_i	α	Experimental units	Fig. 9
1	200	0.2	12000 individual patches	a
2	$K_{max}e^{-\left(\frac{j-\{1,2,...,P\}}{2\sigma}\right)^2}$ $for\ j = 1, ... , N$	1	12000 individual patches	b
3	200	0.2	600 regions composed of 20 patches	c
4	$K_{max}e^{-\left(\frac{j-\{1,2,...,P\}}{2\sigma}\right)^2}$ $for\ j = 1, ... , N$	1	600 regions composed of 20 patches	d

The question that Cardinale *et al.* (2004) and I, here, have tried to address with these four scenarios is whether the effect of biodiversity on ecosystem function changes with spatial scale. In scenarios 1 and 3, species could partition their niches within a given patch and, under these assumptions, measuring the BEF relationship at the patch-scale or the regional-scale has no effect on the slope (b-parameter, Table 2 and compare Fig. 9a and c). Moreover, in both scenarios, we observe transgressive overyielding because of the weak interspecific interactions within a patch (i.e. $\alpha = 0.2$). Similar results are found when we look at scenarios 2 and 4: measuring the BEF at relationship at the patch- or region-scale also does not affect the slope (b-parameter, Table 2 and compare Fig. 9b and d). However, there are important differences with scenarios 2 and 4. First, at the patch-scale, the species richness after 300 time-points tends to a species richness of one as the species with the highest carrying capacity (K) outcompetes the others. This also means that, at the patch-scale, there is no transgressive overyielding. But, this changes at the regional-scale (scenario 4) where species coexist by partitioning their niches between patches and this causes transgressive overyielding to emerge (a prediction that I tested directly in **Paper II**). Therefore, even though a single species can maximise

functioning at the patch-scale (Fig. 9b), no single species can maximise functioning in a region composed of different patch types (Fig. 9d).

Table 2: Estimated coefficients and standard errors from a power function fit to the data from each of the four scenarios (Fig. 9) for the effect of initial species richness on total community biomass (a proxy for ecosystem functioning).

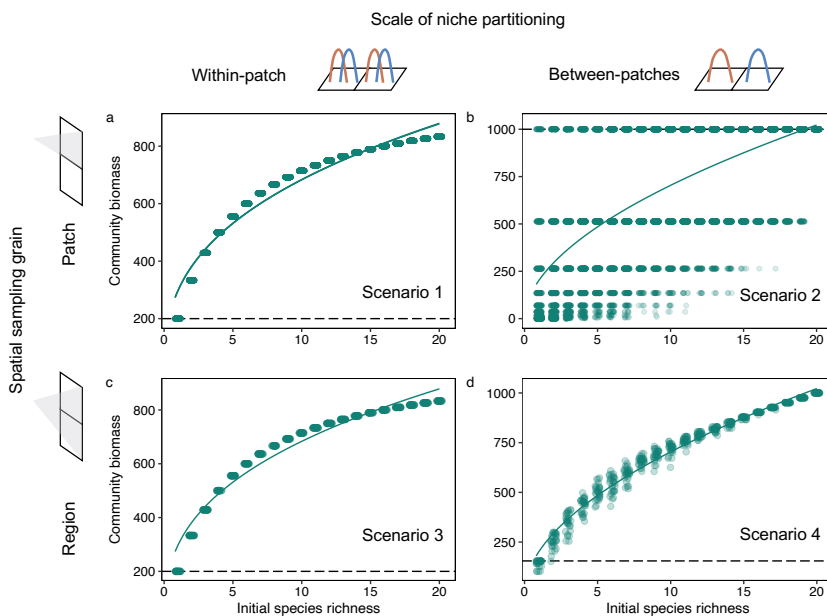| Parameter: $y \sim a \times (SR^b)$	Scenario 1		Scenario 2	
	Estimate	SE	Estimate	SE
a	297	0.62	204	3.7
b	0.36	0.001	0.54	0.007
	Scenario 3		Scenario 4	
	Estimate	SE	Estimate	SE
a	297	2.77	204	2.0
b	0.36	0.004	0.54	0.004

Fig. 9: The relationship between initial species richness and total community biomass for the four simulated scenarios **(a)** 1, **(b)** 2, **(c)** 3 and **(d)** 4. In scenarios 1 and 2, the relationship is evaluated at the patch-scale whilst in scenarios 3 and 4, the relationship is evaluated at the region-scale. In all cases, a power-function of the form: $y \sim a \times (SR^b)$ is plotted which was fit using non-linear least squares. Estimated model parameters can be found in Table 2. Different coloured bell curves represent different species.

The results of this simple model indicate that, in the context of a typical BEF experiment, simply changing our scale of observation is unlikely to have an impact on the slope between initial species richness and ecosystem functioning (in this case, biomass). Similar results have been found

using a completely different modelling approach. Thompson *et al.* (2018) modelled a BEF relationship at the patch scale as a simple power function:

Equation 16

$$Y_i = a_i S_i^{b_i}$$

Where Y_i is the ecosystem functioning of patch i, S_i is the species richness of patch i, a_i is the ecosystem functioning in patch i when S_i is one and b_i is the effect of species richness on ecosystem functioning. Using this simple equation at the patch-scale, Thompson *et al.* (2018) scaled ecosystem functioning to larger regions comprised of A patches as:

Equation 17

$$Y_A = \sum_{i=1}^{A} Y_i$$

This means that, for regions simulated to comprise A different patches, the BEF relationship can be estimated as:

Equation 18

$$Y_A = a_A S_A^{b_A}$$

In this equation, Y_A is the total ecosystem functioning in a region comprised of A patches, S_A is the number of species in the region and a_A and b_A are the estimated coefficients of the BEF relationship. It should be noted that in this model, neither environmental heterogeneity nor species interactions are explicitly considered. Thompson *et al.* (2018) simulated many different scenarios. However, only one of their scenarios (case I in their paper) are relevant in the context of a typical BEF experiment. In this scenario, all patches in a region have the same a_i, S_i and b_i values. Therefore, in a given region, all patches have the same number of species and the BEF relationship is the same. The number of species in the region are varied as: Normal(10, 3). To vary spatial scale, in a given simulation, a region has between 1 and 50 patches. For regions with different numbers of patches, 2000 replicates are simulated (i.e. 2000 regions with 1 patch, 2000 regions with 2 patches etc.). Across the 2000 regions for a region with a certain number of patches, the slope (b_A) of the relationship between species richness (S_A) and functioning is measured (Y_A). In line with Cardinale *et al.* (2004), Thompson *et al.* (2018) found that the slope of the BEF relationship (i.e. b_A) did not vary with spatial scale in this scenario.

In my view, these models indicate that, in the context of a typical BEF experiment where regions are seeded with different numbers of species and the environmental conditions between regions are the same, the slope of the relationship between biodiversity and ecosystem functioning is unlikely to change (Cardinale *et al.* 2004; Thompson *et al.* 2018). Interestingly, this was true when all patch types in a region were the same and species partitioned niches within patches or when all patch types in a region were different and species partitioned niches between patch types (Fig. 9b and d). However, when patch types in a region differed and species partitioned niches between patches, we observed transgressive overyielding at the region but not at the patch scale. This latter prediction was

largely untested (but see Cardinale 2011; Griffin *et al.* 2009; Wacker *et al.* 2008) but I found support for it in **Paper II**.

For those who know the BEF literature well, the fact that the BEF slope does not change with spatial scale may come as a surprise. Recently, Qiu and Cardinale (2020) examined whether the effect of biodiversity on ecosystem functioning changed with spatial scale across 374 BEF experiments performed on terrestrial plants. The effect of biodiversity on functioning was measured as the net biodiversity effect (see Fig. 5 for details) and spatial scale was measured as the plot size relative to the average size of the focal plant species. They showed that the net biodiversity effect did increase with spatial scale across experiments. Can the conflict between the theoretical predictions and these empirical results be explained? In my view, it is not straightforward to compare the theoretical predictions from models like those of Cardinale *et al.* (2004) and Thompson *et al.* (2018) to the empirical results from a study like Qiu and Cardinale (2020). The models show that, if we observe a system at both patch- and region-scales, the slope of the BEF relationship remains constant. However, the experiments studied by Qiu and Cardinale (2020) differ in a number of ways and it is very difficult to rule out possible confounding variables. For example, do the studies at the low spatial scale spectrum differ fundamentally from those at the high end? I think this is very likely and thus cannot, in my opinion, be used to reject the theoretical predictions.

The fact that, in the context of a typical BEF experiment, spatial scale does not necessarily affect the slope of the BEF relationship does not mean that spatial scale is unimportant. The models I presented are designed in the context of a BEF experiment where everything is held constant but initial species richness is varied between patches and/or regions in random draws. In my opinion, these models do not necessarily make predictions about what we should see in natural systems where many factors (e.g. environment, dispersal, competition etc.) determine both the biodiversity and ecosystem functioning of a system (see also **Paper I**). In this latter context, the BEF relationship becomes an emergent property of an ecological system and there is another body of theory that has examined how spatial scale affects the BEF relationship in this context.

Theory that has examined the effect of spatial scale on the BEF relationship outside of the confines of a BEF experiment comes from the metacommunity literature. Without going into too much detail, metacommunity theory views ecological systems as local communities that are linked by the dispersal of multiple interacting species (Leibold *et al.* 2004). The metacommunity models that have been used in the BEF literature are mostly models of the so-called *insurance hypothesis* (Bond & Chase 2002; Loreau *et al.* 2003; Shanafelt *et al.* 2015; Yachi & Loreau 1999). These models all have a very similar structure. Local patches vary in some environmental variable in space and time and species respond differently to that environmental variable. The differences in species responses are such that all species are highest functioning in the absence of competition (i.e. in monoculture) under some value of the environmental variable (i.e. species are specialised). In any given local patch at a given time, ecosystem functioning is maximised by either the highest functioning species under that environmental condition (Loreau *et al.* 2003; Shanafelt *et al.* 2015) or by a few species if local niche partitioning at the patch-scale is incorporated (Bond & Chase 2002; Leibold *et al.* 2017; Thompson *et al.* 2021). A general prediction that all these models tend to make is that biodiversity will increase functioning because dispersal allows high functioning species to dominate at specific times and in specific places (Isbell *et al.* 2017, 2018).
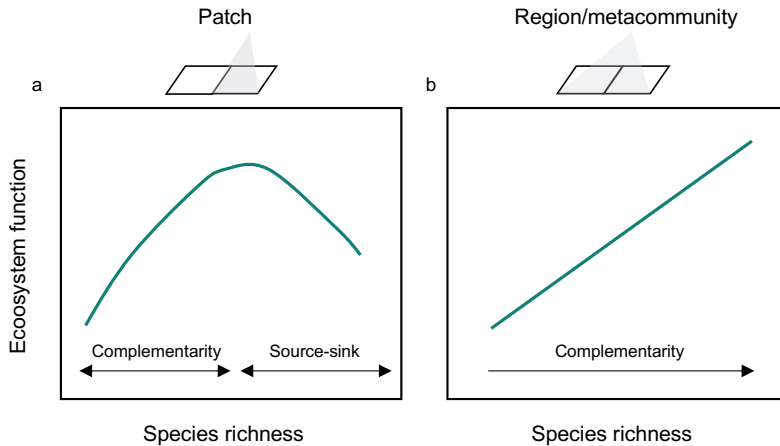
Fig. 10: Metacommunity theory predicts a **(a)** hump-shaped relationship between species richness and ecosystem functioning at the patch-scale. This occurs because, at high levels of dispersal between patches, source-sink dynamics occur whereby poorly adapted species increase species richness but decrease functioning. However, at the **(b)** region/metacommunity-scale a positive relationship is predicted because there is greater environmental heterogeneity between patches. Figure redrawn after Bond and Chase (2002).

A key difference between these models and those based on BEF experiments like Cardinale *et al.* (2004) is that all metacommunities are initialised with the same number of species. Variation in patch- or metacommunity-scale biodiversity or functioning is then determined by the features of the patches and the metacommunity as a whole (e.g. dispersal, environmental heterogeneity, species specialisation etc.). In my view, Bond and Chase (2002) most clearly summarised the predictions that these metacommunity models tend to make. First, adding species at a patch-scale that are adapted to the local conditions and/or are complementary in their use of resources (i.e. patch-scale niche partitioning) leads to increased ecosystem functioning. This tends to occur when the dispersal rate is intermediate which allows species to adaptively track a varying environment (Loreau *et al.* 2003). However, at high dispersal rates, source-sink dynamics develop where species that are poorly adapted to the local environmental conditions persist in patches through continued dispersal (Pulliam 1988). This leads to an increase in patch-scale species richness but a decrease in ecosystem functioning. Thus, metacommunity theory predicts a hump-shaped relationship between biodiversity and functioning when dispersal is a strong determinant of biodiversity at the patch-scale (Bond & Chase 2002; Mouquet & Loreau 2002) (Fig. 10a).

At regional scales, however, the predicted pattern is different. If patches differ in their environmental conditions and species respond differently to that environmental variation (i.e. insurance hypothesis), then the relationship between biodiversity and functioning measured at the regional-scale should be positive (Bond & Chase 2002) (Fig. 10b). However, this occurs when, within regions, the rate of dispersal between patches differs. If dispersal is too low, species cannot adaptively track the spatially and temporally varying environment which can cause a reduction in species richness. Similarly, if dispersal is too high, one or a few species may dominate the whole region (Loreau *et al.* 2003). There are many different variations on these metacommunity models. But, the prediction that the patch-scale BEF relationship is hump-shaped and the region-scale BEF relationship

is positive linear (or potentially log-linear) is relatively robust across models (Loreau *et al.* 2003; Mouquet & Loreau 2002; Thompson *et al.* 2021).

In my view, even though these metacommunity models are incomplete (e.g. many do not allow within-patch niche partitioning, Loreau *et al.* 2003), the predictions they make are more likely to be relevant to understanding what we see in natural ecosystems than theory that is based on BEF experiments (Cardinale *et al.* 2004). This is because they model how properties of an ecological system (e.g. dispersal, environmental heterogeneity etc.) lead to emergent patterns in biodiversity and ecosystem functioning at different scales rather than specifically varying biodiversity experimentally. Indeed, there is more and more support for the basic predictions that these metacommunity models make.

At a patch-scale, one of the key predictions that these models make is that increasing the dispersal rate should increase patch-scale biodiversity but have limited or negative effects on functioning (Fig. 10a, if dispersal goes from intermediate to high). Recently, Ladouceur *et al.* (2020) summarised 12 different grassland experiments that reduced dispersal limitation by seed addition and found an increase in patch-scale species richness but no effect on biomass productivity (their proxy for functioning). Freitag *et al.* (2023) also performed a seed addition experiment across 73 grassland sites with varying land-use intensity and found similar results. Moreover, Dee *et al.* (2023) in what is arguably the most robust analysis of the relationship between biodiversity and functioning in observational field data, found that adding rare species increased patch-scale species richness but decreased productivity in global grasslands. This is exactly the pattern expected under source-sink dynamics (Bond & Chase 2002; Pulliam 1988). Moreover, the converse is also well-supported. When rare species are removed from the patch-scale, there is often no effect or a positive effect on ecosystem functioning (Lisner *et al.* 2023; Smith *et al.* 2020; Smith & Knapp 2003).

Whether the patch-scale BEF relationship is hump-shaped and the region-scale BEF relationship is positive linear as predicted by metacommunity theory has not been robustly tested. The best example is probably Venail *et al.* (2010) who created experimental metacommunities with different levels of dispersal. They found, in line with predictions, that there was no clear BEF relationship at the patch-scale but a strong, positive linear BEF relationship at the region scale. Other studies have mostly evaluated the BEF relationship at different spatial grains in observational data. For example, Craven *et al.* (2020) found a negative BEF relationship at small spatial grains but a positive BEF relationship at large spatial grains in North American forest data. In contrast, Chisholm *et al.* (2013) analysed global forest data and found mostly positive BEF relationships at small spatial grains and more variable relationships at larger grains. Many other, similar studies have been done (e.g. Mao *et al.* 2023) but, together, in my view, the evidence from these observational studies is largely equivocal with respect to this metacommunity prediction. Moreover, many observational studies have major issues in terms of their ability to infer causes due to confounding generally not being treated robustly (see Dee *et al.* 2023 for further discussion).

There is a third set of theoretical models that addresses how the relationship between biodiversity and ecosystem functioning may change with spatial scale. These models rely on scaling theory to generate empirical expectations of the BEF relationship at different spatial scales (e.g. Barry et al. 2021; Gonzalez et al. 2020). Specifically, these models take the observation that species richness and area scale non-linearly (Rosenzweig 1995) whilst ecosystem functioning and area scale linearly (Barry *et al.* 2021) to show that, as a result, the BEF relationship changes when observed at different scales in natural ecosystems (Barry et al. 2021; Gonzalez et al. 2020). I find this approach rather

tautological. I suppose these approaches indicate that we should probably expect the BEF relationship to change with changes in spatial scale. However, I have never seen another study attempt to scale-up as Barry *et al.* (2020) do. Therefore, I do not think this approach has much to offer beyond the first sets of models presented (i.e. those based on BEF experiments and those based on metacommunity theory). As a result, I do not treat them further in this thesis.

Are more species required to maintain ecosystem functioning at large compared to small spatial scales?

A related question is whether more species are required to maintain or maximise ecosystem functioning at large compared to small spatial scales. Indeed, this is the question that Tilman's (1999) analysis examined. However, since Tilman (1999) initially posed the question, most studies have approached this question by examining whether different species drive ecosystem functioning in different sites. The logic followed by these studies is that, if different species drive an ecosystem function in different sites, then more species should be required in a landscape comprising multiple sites.

As far as I can tell, the first study to take this general approach was Isbell *et al.* (2011). Isbell *et al.* (2011) examined which plant species best predicted plant productivity at different sites. Winfree *et al.* (2018) took a similar approach. Specifically, Winfree *et al.* (2018) examined the set of species required to provide a threshold amount of pollination at a given site. If the set of species differed between sites, the number of species required to meet some threshold amount of pollination increased as more sites were considered. A range of papers from Rachel Winfree's group using the same basic method have come to similar conclusions (Genung *et al.* 2020, 2023; Simpson *et al.* 2022) as have other studies in different ecosystems (e.g. Schiettekatte *et al.* 2022 using coral reef fish). Although these studies use slightly different methods, they all come to the same general conclusion: more species are required to maintain ecosystem functioning at large compared to small spatial scales because different species perform most of a given function in different sites.

Recent theoretical work also supports these findings. Thompson *et al.* (2021) used simulations based on Lotka-Volterra competitive communities to determine whether more species are required to maintain functioning as spatial scale increases. Specifically, in the model, there are different patch types that vary in a single environmental dimension. Species differ in their environmental optima (i.e. conditions under which their growth rates are maximized) and these optima are distributed evenly along the range of conditions in the patch types (i.e. species are specialised to different patch types). At the patch-scale, communities follow Lotka-Volterra dynamics (Thompson *et al.* 2021). Within this framework, the approach is very similar to Winfree *et al.* (2018). At a given scale (i.e. one patch, two patches etc.), they examine the relationship between species richness (S_i) and functioning (in this case, biomass: B_i) by fitting a Michaelis-Menten function:

Equation 19

$$B_i = \frac{(S_i a_i)}{(S_i + b_i)}$$

In the Michaelis-Menten function, the b_i parameter (also known as the half-saturation constant) is the number of species required to sustain half the asymptotic level of biomass (Thompson *et al.*

2021). This can be interpreted as the number of species required to reach half the maximum functioning at a given scale. Thompson *et al.* (2021) found that, at larger spatial scales (i.e. more patches with different environmental conditions), this half saturation constant (b_i) increased. Thus, in line with the empirical analyses presented, at cumulatively larger spatial scales, more species are required to maintain a given level of ecosystem functioning.

Both the empirical analyses cited previously and Thompson *et al.*'s (2021) theoretical model make an important assumption. They assume that if a species that is performing most of a given function at one site goes extinct, it will not be compensated by a different, potentially less abundant species in the local community or a species that colonises from a different site. This assumption is explicit in Thompson *et al.*'s (2021) model because each species is assigned a unique environmental condition under which their growth rate is highest (although some compensation is possible because there is limited niche overlap). And it is implicit in the observational approaches because they cannot rule out the counterfactual scenario where species compensate fully or partially for the loss of other species. This is despite the fact that there are well-known examples where such strong compensation following species extinction has been observed (Leibold *et al.* 2017; Ernest & Brown 2001).

There are, of course, arguments as to why compensation of the type I describe may be unlikely in some of the observational studies I cited previously. For example, Winfree *et al.* (2018) do provide some arguments as to why they believe their results would not be overturned by high levels of compensation. For example, Winfree and Kremen (2008) found limited evidence for spatial density compensation among pollinators in the crop systems studied by Winfree *et al.* (2018) and Simpson *et al.* (2022) provide evidence that many of the pollinators studied are highly specialised. However, whilst these arguments are valid, there is also considerable evidence to suggest that, at least in many systems, high levels of compensation may be expected.

The first point that I find important is that it is a well-known in ecology that species are often absent from a site because of competition and not necessarily due to an inability to grow, survive and reproduce (Germain *et al.* 2018; Kraft *et al.* 2015b). Indeed, some of the most famous ecological stories (e.g. Connell's barnacles) are evidence of this. Second, the analysis of many competition models has shown that species do not necessarily perform best (i.e. grow, survive and reproduce) at sites where they are most abundant (Fox 2012; Rosenzweig 1981). Rather, as explored theoretically by Fox (2012) species abundance and performance can be decoupled due to complex interactions among multiple species. Indeed, studies of North American trees have shown that growth rates are often decoupled from local abundance (McGill 2012) or probability of occurrence metrics derived from species distribution models (Bohner & Diez 2020; Midolo *et al.* 2021). Third, in natural systems, we often observe direct evidence of compensation from asynchrony in species' population fluctuations even though population synchrony seems to be more common (Gonzalez & Loreau 2009; Houlahan *et al.* 2007; Lamy *et al.* 2019; Vasseur *et al.* 2014). Together, in my opinion, this suggests that we should at least take a critical view of the extent to which species may be able to compensate for the loss of others and how this may affect observed patterns (Schiettekatte *et al.* 2022; Winfree *et al.* 2018). This is something that I attempted in **Paper III** and **Paper IV**.

Section 3: Summary of papers

This thesis is made up of five papers that attempt to address both of the issues described in the previous section (*Section 2: Current issues in the BEF field*). **Paper I** relates to how the effect of biodiversity on ecosystem functioning is studied in natural ecosystems whilst **Papers II** and **III** are about how the effects of biodiversity on ecosystem functioning are expected to respond to changes in spatial scale. **Papers IV** and **V** are relevant for both of these issues. In these five papers, I focus on single ecosystem functions and not ecosystem multifunctionality.
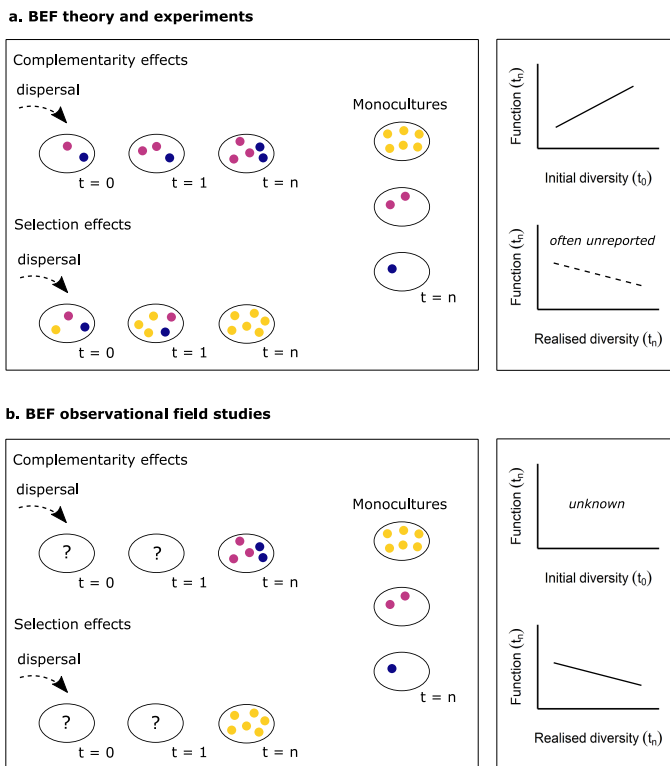

Fig. 11: (a) In BEF theory and experiments, a set of species (akin to a regional species pool) is used to inoculate environmentally similar patches with different levels of initial diversity. In each patch, the inoculated species coexist or go locally extinct. After some arbitrary length of time, the ecosystem function is measured (i.e. function t_n). At this time point (t_n), the realised diversity (i.e. the number of remaining species) might be the same or different than was initially inoculated. Usually, in BEF experiments, the relationship between initial diversity (t_0) and function at t_n is then examined and this relationship is frequently positive as predicted by theory. The relationship between realised diversity and ecosystem function is often not reported (but see e.g. Tilman et al. 2001; Reich et al. 2012). This is different from BEF studies using observational field data **(b)**. In observational field data, we can only measure realised diversity and ecosystem function at some arbitrary time-point (t_n). Unless detailed time-series data are available, the history of community assembly remains unknown. Thus, field data may suggest a strong effect of individual species when positive effects of initial diversity on function are unobserved.

3.1: Paper I

In **Paper I**, I examined how the results of BEF experiments are applied to observational data from field studies. The paper is based on the observation that, in BEF experiments, there is generally a positive effect of biodiversity on ecosystem functioning (Cardinale *et al.* 2012; Tilman *et al.* 2014). However, in observational field data, we tend to find more variability in the relationship between biodiversity and functioning (van der Plas 2019). My main goal in writing this paper was to point out that BEF experiments test for the effect of the species richness treatment (i.e. number of species initially inoculated into a habitat patch) on ecosystem functioning. Within the context of such an experiment, a positive effect of biodiversity can arise when species partition niches or facilitate one another (i.e. complementarity effects) or if certain high functioning species dominate (i.e. selection effects). However, in observational data from field studies, we do not observe the number of species initially inoculated into a habitat patch. Instead, we observe biodiversity after a period of biotic interactions where some species are competitively excluded (Fig. 11). Therefore, in observational data from field studies, there may have been positive selection effects, but the result is a low biodiversity community. Therefore, I argued, that we should not necessarily expect to find the same type of relationship between biodiversity and ecosystem functioning in observational field data as we do in BEF experiments.
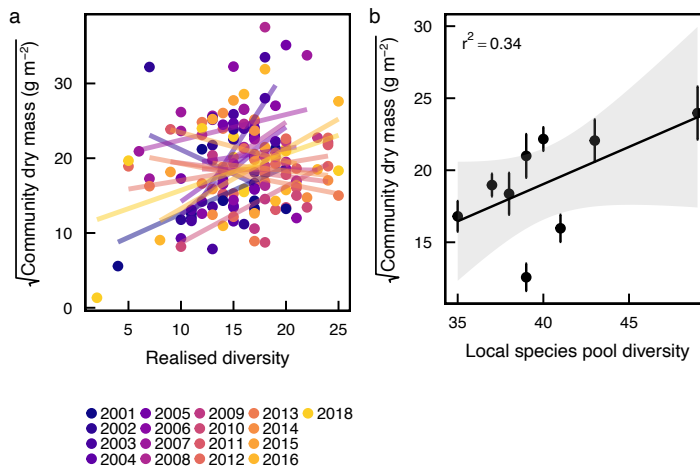

Fig. 12: (a) The relationship between realised diversity and community dry mass (g m^{-2}, square root transformed) of macroalgae observed at nine sites in 17 different years between 2001 and 2018 (colours are different years). As predicted, the realised diversity–function relationships are highly variable between years. All fitted lines are simple linear regressions. **(b)** The relationship between local species pool diversity of algae (all species observed over 17 years) and mean ± SE community dry mass (g m^{-2}, square root transformed) for the nine sites across the 17 years. Despite considerable variation, the relationship is positive as would be expected based on BEF theory and experiments. Data are from kelp forests off the coast of California (see **Paper I** for details). The fitted line is a simple linear regression (intercept CI$_{95\%}$ = [−27.7 to 24.5]; slope CI$_{95\%}$ = [−0.13 to 1.2].

So what should we expect? One of my key conclusions was that linking predictions from biodiversity-ecosystem function theory and experiments to observational field data requires considering the pool of species available during colonisation: the *local species pool* as this is the variable that is most analogous to how the effects of biodiversity on functioning are studied in BEF experiments. I supported this and the other assertions made in the manuscript using a combination of ecological theory and a re-analysis of several biodiversity datasets. For example, I used 17 years of monitoring data of macroalgae from kelp forests off the coast of California to test whether the diversity of the local species pool had a consistent positive effect on community dry mass (Fig. 12b). Thus, I was able to provide some evidence for the ideas presented.

Interestingly, since publication, two papers have directly tested the hypotheses I developed in **Paper I**. Schmid *et al.* (2022) created communities of species richness S that were derived from an initial species richness of $2S$. This was done by experimentally removing the subordinate species in the $2S$ communities to mimic strong biotic filtering. These communities with species richness S derived from $2S$ communities were compared to communities with species richness S that were derived from an initial species richness of S. Thus, the communities have equal S, but the species pool richness differed two-fold (i.e. $2S$ *versus* S). Schmid *et al.* (2022) found that, on average, communities with species richness S that were derived by biotic filtering from $2S$ communities had 32% more aboveground biomass than communities with species richness S that were derived from an initial species richness of S. This result supports one of the key hypotheses from **Paper I** that the biodiversity of the local species pool can have important effects on ecosystem functioning even when we observe a low biodiversity community. The second publication, Le Provost *et al.* (2022) took a more correlative approach where they compared the effects of local biodiversity to that of local species pool biodiversity on ecosystem functions in grassland ecosystems. Le Provost *et al.* (2022) found that both scales of biodiversity affected ecosystem functioning.

A secondary goal of the paper was to address a criticism regarding the relevance of BEF research to conservation and arguments for nature preservation in general. Up until now, BEF research has largely focused on how biodiversity affects ecosystem functioning measured on small spatial scales. As mentioned previously, both Vellend *et al.* (2013) and Wardle (2016) point to research showing that local-scale biodiversity is not necessarily declining in natural ecosystems (Blowes *et al.* 2019; Dornelas *et al.* 2014). Therefore, they argue that BEF research is largely irrelevant to conservation. But taking a species pool view shows that, even without local biodiversity declines, biodiversity loss at regional scales—which determines local species pools—may still negatively affect ecosystem functioning.

3.2: Paper II

As discussed in *Section 2*, one of the most topical issues in the BEF field today is spatial scale. The reason for this is that BEF experiments have generally taken place at small spatial scales with limited environmental heterogeneity and the evidence underpinning the BEF field relies heavily on these experiments. The small spatial scales of BEF experiments mostly reflect practicality. Manipulating species richness (or some other aspect of biodiversity) is not trivial and generally requires high levels of replication to decouple species richness and species composition (Fig. 4). Therefore, doing these experiments at different spatial scales or with different levels of environmental heterogeneity is challenging.

Spatial scale and environmental heterogeneity are also important topics in the BEF field because they may explain why so few BEF experiments observe transgressive overyielding (i.e. mixtures having higher functioning than the highest functioning monoculture) (e.g. Cardinale *et al.* 2011). However, at small spatial scales with limited environmental heterogeneity, the potential for niche partitioning is limited (Hart *et al.* 2017) and theory predicts that the likelihood of transgressive overyielding should increase if niche partitioning is stronger (Cardinale *et al.* 2004; Chesson *et al.* 2013).
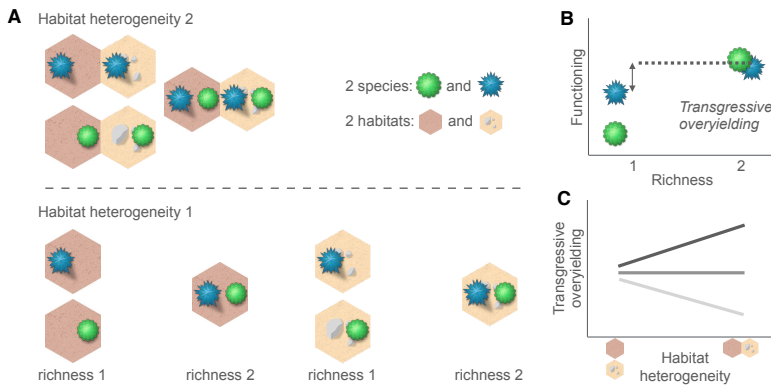

Fig. 13: (a) Conceptual illustration of my approach to manipulate habitat heterogeneity. I increase habitat heterogeneity by creating landscapes via the aggregation of different habitat types. At each level of habitat heterogeneity, all possible landscape configurations are created (e.g. two landscapes at habitat heterogeneity=1 and one landscape at habitat heterogeneity=2). For each landscape, all possible species combinations are assembled as in a BEF experiment. **(b)** For each landscape, I calculate transgressive overyielding. **(c)** I then examine if and how transgressive overyielding changes with habitat heterogeneity. The figure only shows two levels of richness and habitat heterogeneity, but the approach is the same for higher dimensions.

To address this, I asked whether transgressive overyielding changes with environmental heterogeneity. Instead of trying to experimentally vary spatial scale, we instead used BEF experiments replicated in different environmental conditions. I then aggregated BEF experiments and re-calculated transgressive overyielding at different scales with different levels of environmental heterogeneity (overviewed in Fig. 13). First, I did this using a simple theoretical simulation to generate my expectations. Second, I used a novel experimental system to create BEF experiments using different *Escherichia coli* strains in five different environmental conditions defined by sub-inhibitory concentrations of antibiotics. This experimental system allows significant replication, and I generated 10800 individual bacterial communities. Finally, I compiled a database of 26 published BEF experiments that had been replicated in at least two environmental conditions and used a type of meta-analysis to test whether the results applied to other ecosystems.

The theoretical simulations showed that, on average, transgressive overyielding increased with habitat heterogeneity because monoculture functioning decreased with habitat heterogeneity. On average, the experimental results concurred with the simulations. Moreover, our meta-analysis confirmed that transgressive overyielding tended to increase with habitat heterogeneity but only when species were specialised to different habitats and were not inhibited in mixtures by negative species

interactions. This was not the case in several experiments used in our meta-analysis where one species maximised functioning across all habitats, contrary to the assumptions of many ecological models.

3.3: Paper III

As discussed, in **Paper II** I compiled a set of 26 BEF published experiments that were replicated in at least two environmental conditions. For example, a study might do a BEF experiment under ambient nutrient levels and under nutrient addition. I was surprised to find that in many of these experiments, the highest functioning monoculture was the same in all the different environmental conditions. This contradicted several theoretical models (Chesson *et al.* 2013; Loreau *et al.* 2003) that assume that different species are highest functioning on different parts of an environmental gradient. This finding also raised questions about a whole host of observational studies (e.g. Genung *et al.* 2023; Schiettekatte *et al.* 2022; Winfree *et al.* 2018) that had examined whether different species drive ecosystem functioning in different environmental conditions or "places". These studies generally conclude that, at larger scales that encompass large environmental gradients, more species are required to sustain ecosystem function. However, none of these studies show that one or a few species *could not* maintain functioning across the environmental gradient under a scenario where species loss occurred. Doing this requires knowing if some species can compensate for the loss of others and, therefore, maintain functioning (e.g. Chaves & Smith 2021; Diaz & Ernest 2022; Pan *et al.* 2016).

To try and address this, I wanted to study species along an environmental gradient and observe how they respond to being moved to different parts of that environmental gradient. I reasoned that, if species reduce their growth rates (i.e. a proxy of functions like biomass productivity) when moved to a different part of the environmental gradient, this would suggest that, if one species were lost, others may not be able to compensate for the lost functioning. In contrast, if species' growth rates remained relatively stable when moved, this would suggest that they could compensate for species loss. If species can compensate for the loss of others along an environmental gradient this would show that studies like Schiettekatte *et al.* (2022) and Winfree *et al.* (2018) may be unreliable indicators of how ecosystem functioning would respond to species loss on an environmental gradient.

I approached these ideas using a model intertidal marine macroalgal system from sheltered rocky shores on the Swedish West Coast. The intertidal on sheltered shores in this region is usually dominated by four fucoid brown algae species that tend to occupy four relatively distinct zonal bands from high to low shore (Cervin *et al.* 2004). The upper limits of intertidal algal species are often assigned to their tolerance to abiotic factors such as desiccation, whereas biotic factors such as competition and predation often explain the lower limits (Connell 1972; Hawkins & Hartnoll 1985). I transplanted monocultures of adults of the four fucoid species into each of the four different zones and monitored relative growth rate as a proxy of functioning.

Despite these species occupying distinct depth zones under natural conditions, the growth rates of three of the four species showed limited responses to being transplanted to different depth zones (Fig. 14). This suggests that species loss in this system could be compensated by other species expanding their distributions in contrast to many ecological models (Loreau *et al.* 2003) and observational studies (Schiettekatte *et al.* 2022; Winfree *et al.* 2018).
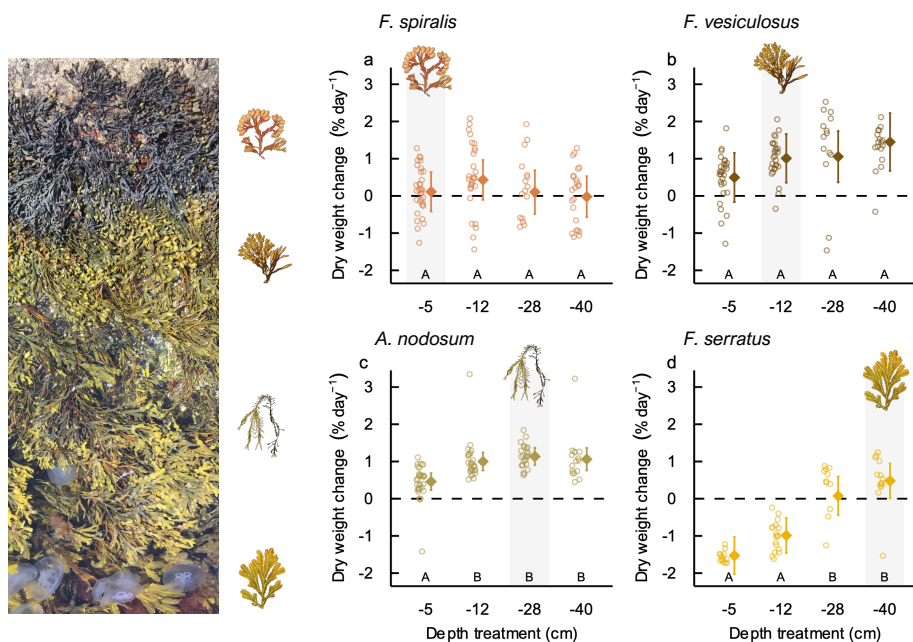
Fig. 14: (a-d) The relative growth rates measured as change in dry weight per day (g g^{-1} % day^{-1}) for the four species that occupy distinct zones on sheltered shores in the intertidal zone of the Swedish West Coast: *Fucus spiralis*, *Fucus vesiculosus*, *Ascophyllum nodosum* and *Fucus serratus*. Only *F. serratus* responded strongly to a change in depth and the three other species maintained their growth rates in all zones. Points and error bars are the mean and 95% confidence intervals. Photograph on the left shows the species' natural depth distribution.

3.4: Paper IV

In **Paper III**, I was able to show that transplanting four species of fucoid macroalgae out of the depth zones that they typically occur in had limited effects on their growth rates. To me, this suggested that species loss in this system could potentially be compensated by other species expanding their distributions. However, in **Paper III**, the evidence was relatively indirect. I simply observed a lack of growth rate response to transplantation. Although an important first step, I wanted to examine in more detail whether some species could compensate for the loss of others and to what extent. To do this, I supplemented the growth rate data for the four species from **Paper III** with additional transect data on the standing stock biomass distribution of the four species. Using data on the distribution of standing stock biomass and the growth rate data, I estimated dry biomass productivity of the four species in the four depth zones.

Using the estimated biomass productivity data, I calculated the number of species required to maintain biomass productivity across the four depth zones using two commonly used observational approaches (Schiettekatte *et al.* 2022; Winfree *et al.* 2018). I then explored how the number of species required to maintain biomass productivity across the different depth zones changed under counterfactual scenarios where the loss of species can be compensated by other species along the gradient. To do this, I assumed that a species went extinct along the whole gradient (i.e. its standing stock biomass went to zero). I then assumed that another species could partially compensate for the

extinct species by increasing its standing stock biomass. Combined with the growth rate for all species across all depth zones (**Paper III**), I was then able to estimate dry biomass productivity given an extinction of a certain species and assuming partial compensation by others (see Fig. 15 for an overview).
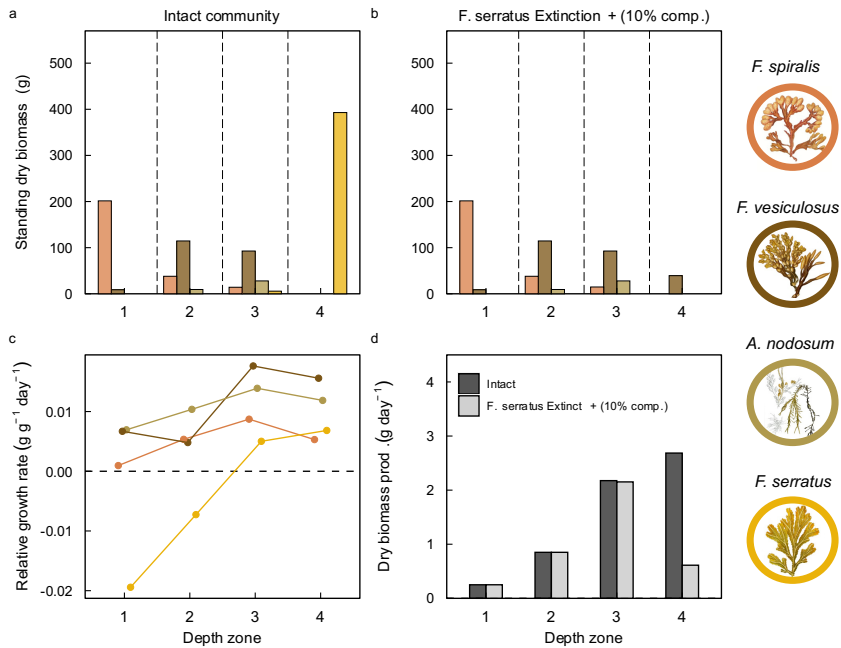

Fig 15: Example of a simulated extinction-compensation scenario using one of the 3000 samples of relative growth rates. The standing stock dry biomass (g) of **(a)** the intact community containing all four species and **(b)** a hypothetical community where *Fucus serratus* is extinct but other species (in this case *Fucus vesiculosus*) compensate 10% of the standing stock dry biomass of *F. serratus* in each depth zone. **(c)** Using one sample from the posterior distribution and obtaining relative growth rates (g g^{-1} day^{-1}) for each species across the four depth zones, I calculated **(d)** the dry biomass productivity (g day^{-1}) of each depth zone for the intact community and the hypothetical community where *F. serratus* is extinct. Under this extinction compensation scenario, dry biomass productivity declined in the two deepest zones but was unaffected in the two shallowest zones.

The results of the observational analyses suggested that between three and four species were required to maintain productivity across the depth zones. In contrast, the simulated counterfactual extinction scenarios gave a more nuanced picture. Specifically, the simulations showed that decreases in biomass productivity due to the loss of some species (e.g. *Fucus spiralis*, *Ascophyllum nodosum*) were easily compensated by other species (e.g. *Fucus vesiculosus*). However, for the simulated extinction of some species like *F. vesiculosus*, the simulations suggested that compensation would be unlikely. The conclusion I took from this analysis is that commonly used observational approaches may overestimate the number of species required to maintain ecosystem functioning as spatial scale increases. In addition, I demonstrated that using counterfactual extinction scenarios parameterised with observational and experimental data may be a useful tool to understand the potential consequences of ecosystem changes such as species extinctions or other kinds of community change.

3.5: Paper V

Comparing the functioning of mixtures of species to their constituent species in monoculture is one of the clearest and most direct ways to quantify how biodiversity effects ecosystem functioning. For example, calculating transgressive overyielding or applying Loreau and Hector's (2001) statistical partition to calculate net biodiversity effects along with the contribution of complementarity and selection effects both rely on monoculture data. However, in large-scale natural ecosystems, monoculture data is difficult to obtain. Therefore, BEF research in natural ecosystems has generally relied on correlating measures of biodiversity with ecosystem functioning, a much less robust approach. The consequence is that it is unclear how strong biodiversity effects on ecosystem function are in natural systems at large scales of space and time.

Recently, Isbell *et al.* (2018) developed an extension to Loreau and Hector's (2001) approach to calculate the net biodiversity effect and partition it into complementarity and selection effects (see Fig. 16 for an overview). Isbell *et al.*'s (2018) extension calculates a net biodiversity effect across a set of communities distributed across space and measured at multiple time points (i.e. multiples places and times). This net biodiversity effect is then partitioned into a set of effects which describe how biodiversity can affect ecosystem functioning at local and larger scales of space and time. The effects operating at larger scales of space and time are termed the *insurance effects* of biodiversity (Fig. 16a). For example, the *spatial insurance effect* quantifies whether species that are high functioning at a given place in monoculture tend to dominate mixtures in that place (Fig. 16b). The spatial insurance effect therefore quantifies the extent to which spatial niche partitioning among species can increase ecosystem functioning at large spatial scales across many places. But, despite the potential of this approach to calculate biodiversity effects on ecosystem functioning at larger scales of space and time, it has only been applied in Isbell *et al.*'s (2018) original paper.

The main problem with Isbell *et al.*'s (2018) approach is that the data required to apply it are rarely available, especially in field data from natural systems. First, the approach requires monoculture data for all species at all places and times for which mixture data are available. Second, it requires initial relative abundance data for the mixtures (i.e. expected relative yields, RY_{exp}). Such extensive monoculture data are only available in some of the largest BEF experiments (e.g. Cedar Creek: Big Biodiversity Experiment, BioCON experiment at Cedar Creek, Jena experiment). To solve this, I developed a methodological pipeline that uses Bayesian data imputation techniques to deal with the incomplete monoculture data along with uncertainty assumptions for the expected relative yields (RY_{exp}). I then used this pipeline to apply Isbell *et al.*'s (2018) partition to two marine ecosystems: intertidal rockpool macroalgae from Plymouth, United Kingdom and marine fouling communities from the Swedish West Coast.

I found that a combination of local-scale species interactions, dominance across space and time by a high functioning species and spatial niche partitioning contributed to a positive effect of biodiversity on ecosystem functioning in both ecosystems. Moreover, I found that the effect of spatial niche partitioning on ecosystem function increased with greater spatial environmental heterogeneity in the marine fouling communities. This latter point indicates that the BEF studies done on small spatial scales may have failed to quantify an important feature of biodiversity that contributes to functioning. Moreover, the approach highlights the importance of taking a multiscale perspective in BEF research

and provides a methodological pipeline that may help future researchers quantify the many ways that biodiversity can contribute to ecosystem functioning across scales in natural ecosystems.
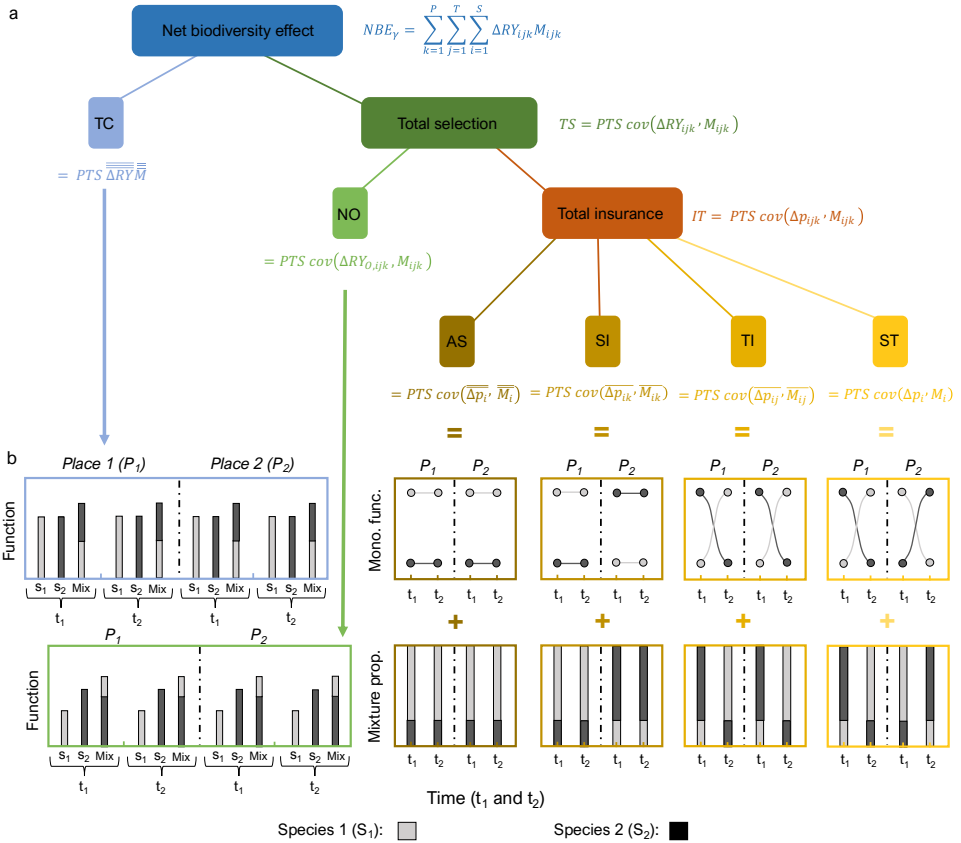

Fig. 16: (a) How the net biodiversity effect (*NBE*) is calculated and partitioned into total complementarity (*TC*) and total selection (*TS*). Total selection is then further partitioned into five terms that describe how biodiversity can affect function across times and places (i.e. insurance effects): average selection (*AS*), temporal insurance (*TI*), spatial insurance (*SI*), spatio-temporal insurance (*ST*) and a residual term called non-random overyielding (*NO*). Total insurance (*IT*) is the sum of *AS*, *IT*, *SI* and *ST* (see **Paper V** for details). **(b)** Hypothetical examples with two species, two times and two places to illustrate what these biodiversity effects measure. For example, total complementarity quantifies the part of the net biodiversity effect that is due to local-scale interactions among species whilst spatial insurance quantifies the extent to which species that are high functioning in monoculture at a particular place tend to dominate the mixtures in those places.

Section 4: BEF partitioning methods

In *Section 1*, I described Loreau and Hector's (2001) statistical partition to compare selection and complementarity in BEF experiments. I wrote that the *complementarity effect* quantifies the extent to which the increase in mixture functioning relative to the monoculture expectation is above a zero-sum expectation. And I wrote that the *selection effect* quantifies the extent to which the increase in mixture yield relative to the monoculture expectation is due to species increasing their relative yield at the expense of others and those increases being correlated with ecosystem functioning in monoculture. This interpretation, which is commonly repeated, has been shown several times to be inaccurate (Fox 2005; Petchey 2003). Here, I take this opportunity to resolve these inaccuracies and propose a workflow that can more accurately detect the different types of biodiversity effects that we can and cannot quantify. In addition, I provide R-code in a Github repository (https://github.com/haganjam/BEF_partitions) that synthesises the different partitions and provides code to implement my proposed workflow.

Fox's (2005) tripartite partition

I start with the interpretation of the selection effect as quantifying the extent to which species with high monoculture yields dominate the mixture at the expense of other species (Loreau & Hector 2001). In a seminal paper, Fox (2005) showed that this interpretation of the selection effect was not correct except in certain marginal circumstances (e.g. see the example in Fig. 5a). Rather, as Fox (2005) showed, the selection effect as calculated by Loreau and Hector (2001) only partially reflected this dominance effect. To see this, we can start with the original Loreau and Hector (2001) partition equation where Y_i is the observed yield in mixture of species i, $M_{obs,i}$ is the observed monoculture yield of species i and $RY_{obs,i}$ and $RY_{exp,i}$ are the observed and expected relative yields of species:

Equation 20-21

$$NBE = \sum_{i=1}^{S} \Delta RY_i M_i = S\,\overline{\Delta RY}\,\overline{M} + S\,cov(\Delta RY_i, M_i)$$

$$\Delta RY_i = RY_{obs,i} - RY_{exp,i} = Y_i/M_{obs,i} - RY_{exp,i}$$

Loreau and Hector (2001) interpret the second term in this equation (i.e. the covariance term) as indicating the increase in some species at the expense of others. As pointed out by Fox (2005), the problem with this interpretation is that the $RY_{obs,i}$ term is not a true frequency (i.e. the sum across species can be greater than one). Rather, the sum of the following expression can (and does frequently) exceed one:

Equation 22

$$\sum_{i=1}^{S} Y_i/M_{obs,i} = \sum_{i=1}^{S} RY_{obs,i}$$

Given that the sum of the $RY_{obs,i}$ terms across species can exceed one, high $RY_{obs,i}$ of a species does not necessarily come at the expense of other species even though this is the common interpretation of the selection effect (Fox 2005). Fox (2005) then developed a new partition with three terms. In this updated partition, one term is the dominance effect where species increase at the expense of others and the other two represent different kinds of complementarity effect. For this, Fox (2005) defined the observed frequency of species as:

Equation 23-24

$$RY_{obs,i}/RY_{obs,T}$$
$$RY_{obs,T} = \sum_{i=1}^{S} RY_{obs,i}$$

$RY_{obs,T}$ is the total observed relative yield. Using this observed species frequency term ($RY_{obs,i}/RY_{obs,T}$), Fox (2005) extended Loreau and Hector's (2001) original partition to three terms (see original paper for details):

Equation 25

$$NBE = S\,\overline{\Delta RY}\,\overline{Mono} + S\,cov(\frac{RY_{obs,i}}{RY_{obs,T}} - RY_{exp,i}, Mono_i)$$
$$+ S\,cov(RY_{obs,i} - \frac{RY_{obs,i}}{RY_{obs,T}}, Mono_i)$$

Summarising from Fox (2005): The first term is the same as the complementarity effect term in Loreau and Hector's (2001) partition and is referred to as *trait-independent complementarity*. This is positive if all species have higher functioning in mixture than expected based on their monoculture functioning and this positive effect is similar across species. The second term is defined as the *dominance effect*. When this effect is positive, species that are high functioning in monoculture dominate the mixture at the expense of other species. The third term is *trait-dependent complementarity,* and this term is positive when species that are high functioning in monoculture attain high observed relative yields but this does not come at the expense of other species. Thus, Fox's (2005) partition can directly quantify the part of the net biodiversity effect that is due to high functioning species in monoculture outcompeting other species.

To illustrate the Fox (2005) partition, we can revisit the hypothetical examples presented in Fig. 5. In the first example, we had a 50:50 mixture of two species where species 2 was higher functioning in monoculture than species 1 (500 *versus* 200 ecosystem function units). We then assumed that species 2 completely outcompeted species 1 leading to a positive net biodiversity effect driven completely by the selection effect (Fig. 5a). However, if we use Fox's (2005) partition, we get the same net biodiversity effect, but it is driven completely by the dominance effect as species 2 outcompetes species 1 (Fig. 17a, d). In the second example, species 1 and 2 coexisted in mixture and both increased their functioning 1.4-fold above their monoculture expectations. This led to a net biodiversity effect driven completely by the complementarity effect (Fig. 17b, d). With Fox's (2005) partition, we get the same net biodiversity effect but it is driven by trait-independent complementarity because both species increase their functioning 1.4-fold relative to their monocultures. In the final

example, species 1 and 2 coexist and both increase their functioning beyond their monoculture expectations but now species 2 increases its mixture functioning more than species 1 (1.7-fold versus 1.3-fold), (Fig. 17c, d). In this case, applying Fox's (2005) partition, we have all three of the different biodiversity effects operating at some level: dominance, trait-independent complementarity and trait-dependent complementarity.

In summary, Fox (2005) improved on Loreau and Hector's (2001) original complementarity-selection partition and provided researchers with three statistical terms that had clearer interpretations than its predecessor. In particular, the dominance term provided a way to quantify the part of the net biodiversity effect that comes from species that are high functioning in monocultures outcompeting species that are lower functioning in monoculture in mixtures which was a major advance. Despite this improvement, Fox's (2005) partition never gained the same popularity as Loreau and Hector's (2001). The reasons for this are not clear to me but even very recent papers (e.g. Hong *et al.* 2022) still use Loreau and Hector's (2001) partition.
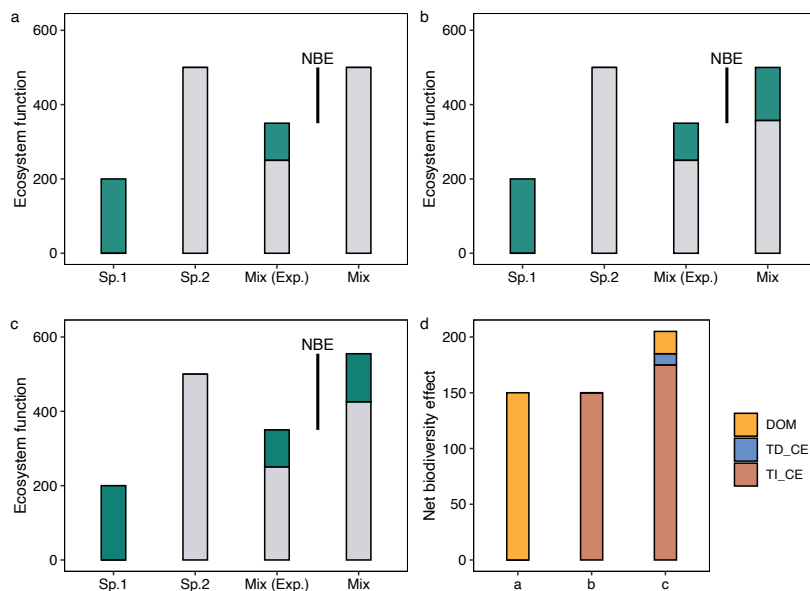

Fig. 17: (a-c) The effect on ecosystem functioning of combining two species (1 and 2, different colours) in a 50:50 mixture (Mix) compared to their respective functioning in monoculture (Sp.1 and Sp.2) and the expected mixture functioning (Mix (Exp.)) under the null hypothesis of no species interactions. The difference between Mix (Exp.) and Mix is the net biodiversity effect (*NBE*, vertical black line) and we partition this *NBE* using Fox's (2005) method. In **(a)** the *NBE* is fully accounted for by the dominance effect (*DOM*) as species two dominates the mixture. This leads to a positive *NBE* but no transgressive overyielding (i.e. the mixture functioning does not exceed that of the highest functioning monoculture). However, in **(b)**, the magnitude of the *NBE* is the same as in **(a)** but it is fully accounted for by trait-independent complementarity (*TI_CE*) as both species are higher functioning in mixture than expected based on their monoculture functioning. In **(c)** a combination of dominance, trait-dependent complementarity (*TD_CE*) and trait-independent complementarity is operating and there is transgressive overyielding. **(d)** Summary of the *NBE*, *DOM*, *TD_CE* and *TI_CE* for the three examples **(a-c)**. Ecosystem function is an arbitrary quantity but could represent, for example, primary productivity (g m^{-2} day^{-1}).

Ecosystem functions can be measured as fluxes or stocks (Hooper *et al.* 2005). As mentioned in the introduction, some workers only consider fluxes as valid ecosystem functions (Cardinale *et al.* 2012; Manning *et al.* 2018) but this is not universal. Either way, explicitly thinking about the difference is crucial to correctly interpreting the result of the statistical partitions of biodiversity effects. This point was made by Petchey (2003) but, like Fox's (2005) partition, its messages have not been widely taken up.

Petchey (2003) observed that different BEF experiments were measuring different types of response variables. Some measured short-term fluxes, others long-term fluxes. Some measured changes in biomass stocks of different species and some measured both biomass stocks and fluxes. What Petchey (2003) pointed out was that these choices had consequences for the interpretation of biodiversity experiments. For example, some workers conducted short-term experiments with mixtures of species and measured a flux (e.g. grazing, nutrient uptake etc.). In these experiments (e.g. Griffin et al. 2008, Bracken et al. 2008), mixtures and monocultures are created whilst controlling for overall abundance and some flux measurement is taken. In such an experiment, there is no time for species to alter their overall biomass stocks and, as a result, any increase in the flux of the mixture relative to the monoculture expectation (i.e. net biodiversity effect) is due to a short-term species interaction.

This contrasts with experiments that allow changes in overall biomass stocks of species to occur and which measure a flux. In such a case, species in monoculture and mixture can change their abundances over time which contributes to changes in the fluxes observed in the mixtures compared to the monocultures. Therefore, in a long-term experiment where species' biomass stocks can change, any net biodiversity effect in the flux can be due to changes in species' biomass stocks and/or species interactions that do not result in any biomass stock changes.

How can we separate out the influence of changes in biomass stocks on mixture fluxes versus the influence of other kinds of species interactions that do not rely on changes in biomass stocks on mixture fluxes? Here, I will illustrate a method for how this can be done which builds upon and uses Fox's (2005), Loreau and Hector's (2001) and Petchey's (2003) work on these statistical partitions of biodiversity effects.

Let us consider a hypothetical BEF experiment with two species of macroalgae. For this experiment, we use outdoor mesocosms to create monocultures of each species and a two-species mixture. The total weight of algae added to each mesocosm is 30 g which means that 15 g of each species was added to the two-species mixture. We fill the mesocosms with seawater and supply ammonia initially at a concentration of 60 μmol l^{-1}. We then allow the algae to grow for two weeks and after the two weeks, we measure the biomass of the algae and the concentration of ammonia. Thus, we are left with i. biomass data (i.e. a stock) and ii. the uptake of ammonia (i.e. a flux).

How should we analyse these data? If our main goal is to quantify biodiversity effects on nutrient uptake, then we could calculate a net biodiversity effect for the flux which quantifies the difference in uptake of ammonia in the mixture relative to what we expected based on the monocultures. We would do this as follows:

Equation 26-27

$$NBE_{total} = F_Y - F_{Y,exp}$$

$$F_{Y,exp} = \sum_{i=1}^{S} RY_{exp,i} \times F_{M,i}$$

Where F_Y is the observed mixture flux (i.e. in this case, ammonia uptake), $F_{Y,exp}$ is the expected mixture flux, $RY_{exp,i}$ is the expected relative yield of species i which is simply 0.5 for both species because the experiment uses a substitutive design and $F_{M,i}$ which is the flux of species i in monoculture. This NBE_{total} term quantifies the total increase in ammonia uptake in the mixtures relative to the expectation based on the monocultures due to both changes in biomass stocks and other species interactions.

Now, if we want to quantify the increase in the flux that is due to changes in biomass stocks, we can do the following. First, for each species, we quantify the biomass specific levels of the flux ($R_{M,i}$) as follows:

Equation 28

$$R_{M,i} = {F_{M,i}}\big/{S_{M,i}}$$

Where $F_{M,i}$ which is the flux of species i in monoculture and $S_{M,i}$ is the biomass of species i in monoculture. Therefore, $R_{M,i}$ quantifies the biomass-specific flux level of each species in the absence of influence from other species. Next, we simply multiply the $R_{M,i}$ for each species by their biomass in mixture ($S_{Y,i}$). Summing across species we get the expected mixture flux under the assumption that all changes in the flux in mixture were due to changes in the biomass stocks of the species (F_{Y,EXP_ABUN}):

Equation 29

$$F_{Y,EXP\_ABUN} = \sum_{i=1}^{S} S_{Y,i} \times R_{M,i}$$

It follows that we can then calculate the part of the total net biodiversity effect (NBE_{total}) that is not due to changes in abundance (NBE_{no_abun}) as:

Equation 30

$$NBE_{no\_abun} = F_Y - F_{Y,EXP\_ABUN}$$

Finally, we can calculate the part of the total net biodiversity effect (NBE_{total}) that is due to changes in abundance (NBE_{abun}) as:

Equation 31

$$NBE_{abun} = F_{Y,EXP\_ABUN} - F_{Y,EXP}$$

And, of course, these two different net biodiversity effects are additive and sum to give the total net biodiversity effect:

Equation 32
$$NBE_{total} = NBE_{abun} + NBE_{no\_abun}$$

In addition to dividing the total net biodiversity effect into contributions from changes in abundance from other types of species interactions, we can simply apply the Fox (2005) partition to the biomass data and we can divide the NBE_{abun} term further:

Equation 33
$$NBE_{abun} = TI\_CE_{abun} + DOM_{abun} + TD\_CE_{abun}$$

Which gives us the following where TI_CE is trait independent complementarity, DOM is dominance and TD_CE is trait-dependent complementarity (Fox 2005, see previous section).

Equation 34
$$NBE_{total} = TI\_CE_{abun} + DOM_{abun} + TD\_CE_{abun} + NBE_{no\_abun}$$

Why is this useful? Well, if we want to understand why mixtures of species have higher levels of fluxes than monocultures, this method provides a way to separate out two distinct effects that occur through changes in biomass from other types of interactions. Moreover, by coupling this with Fox's (2005) partition, we can go further and figure out exactly what types of biomass changes are responsible.

The BEF field is weak when it comes to understanding mechanisms (Cardinale *et al.* 2012). This type of partition may help because it at least gives hints into the types of mechanisms that we should be thinking about. For example, if an increase in a flux in a mixture relative to the monoculture expectation is largely not due to changes in biomass, this demands a very different explanation than if it were due to changes in biomass. Future experimental work may find this method useful in dissecting underlying biodiversity effects.

Case study

To illustrate the utility of this method, I applied it to a biodiversity experiment performed on freshwater phytoplankton (Gamfeldt & Hillebrand 2011). The experiment consisted of monocultures of five freshwater phytoplankton species: *Ankistrodesmus* sp. (Chlorophyta), *Chlamydomonas terricola* (Chlorophyta), *Cylindrospermum* sp. (Cyanobacteria), *Fragilaria sapucina* (Bacillariophyta), and *Gymnodinium* sp. (Dinophyta) and a mixture with all five species. These experimental communities were set-up in 50 ml of medium in 60 ml Nunc flasks. The experiment used a substitutive design so that all mixtures and monocultures started with a biovolume of 231×10^3 μm^3 mL^{-1}. For all monocultures and mixtures, there were three replicates and three nutrient treatments. Specifically, the N:P ratio was varied between 2, 16, and 128 whilst holding the initial P concentration at 5.02 µmol L^{-1}. At the end of the experiment, which lasted 31 days in a 15°C climate chamber (12 h : 12 h light-dark cycle), biovolume was measured (cells were counted and measured before estimating biovolume as per Hillebrand *et al.* 1999). In addition, P-uptake was measured (%) by subtracting the end P concentration from the initial P concentration and dividing by the initial P concentration. Thus, biovolume is the measure of stock and P-uptake is the flux measurement.

Before applying the partition, I averaged both biovolume and P-uptake across the three replicates for the monocultures and mixtures within the three treatments. This is because there is no clear way to assign a given replicate monoculture to a given replicate mixture. This left me with monoculture and mixture data for all five species for three different N:P ratios. Using these data, I applied the partition as described in the previous section.
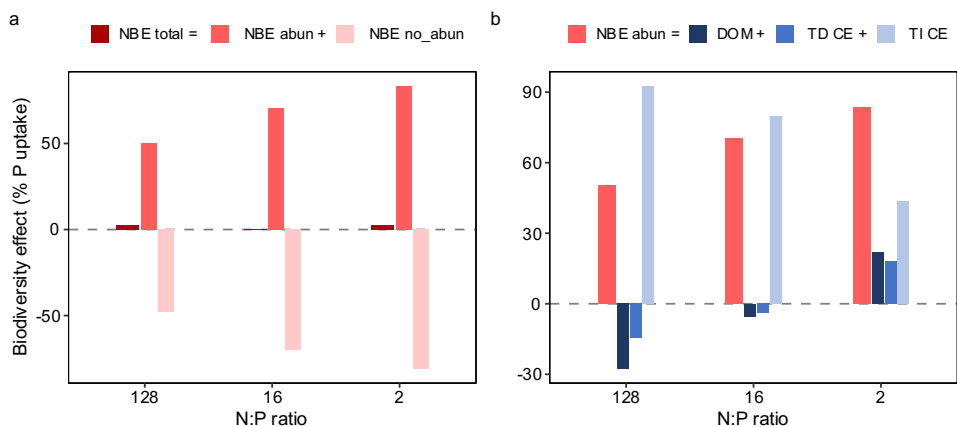

Fig. 18: (a) The net biodiversity effect on P-uptake (%, i.e. a flux) partitioned into the total net biodiversity effect (NBE_{total}), the part of the net biodiversity effect that is due to species changing abundance (NBE_{abun}, in this case biovolume) and the part of the net biodiversity effect that is not due to species changing their abundance (NBE_{no_abun}, in this case biovolume). **(b)** The part of the net biodiversity that is due to changing abundance (NBE_{abun}) can be further partitioned (as per Fox 2005) into three terms: dominance (DOM), trait-dependent complementarity (TD CE) and trait-independent complementarity (TI CE).

Using this partition, I found that total net biodiversity effect (NBE_{total}) on P-uptake was very weak irrespective of N:P treatment (between 0.5 and 2.5%, Fig. 18a). This means that, in this experiment, going from one to five species only increased P-uptake by less than 3%. However, the low NBE_{total} was due to opposite effects of changing abundance and other interactions not linked to changes in abundance (Fig. 18a). Specifically, the part of the NBE_{total} that was not due to changes in abundance was strongly negative (Fig. 18a). This suggests that interactions among species that are not linked to changes in biovolume have strong negative effects on P-uptake. In contrast, species interactions that are linked to changes in biovolume tended to positively affect P-uptake (Fig. 18a). Dissecting this pattern further, increases in biovolume due to trait-independent complementarity increased P-uptake across the three treatments (Fig. 18b). However, dominance and trait-dependent complementarity decreased biovolume and, therefore, P-uptake in high and middle N:P ratio treatments. These results are still preliminary and the method itself is still under development. But, I think it illustrates how we might think about the effect of biodiversity on fluxes in future experiments.

All code and data used to reproduce this analysis can be found in the following Github repository: https://github.com/haganjam/BEF_partitions. In addition, this repository contains code to implement the Fox (2005), Loreau and Hector (2001), this extended partition and the Isbell *et al.* (2018) biodiversity effect partition across spatial and temporal scales.

Section 5: Conclusions and future directions

I will now provide my conclusions and thoughts on future directions for the BEF field. First, I layout my vision for the next phase of BEF research and how some of the papers in my thesis relate to this vision. Next, I analyse the BEF literature from a causal inference perspective, a topic that I became interested in during my PhD. I think it provides a useful framework for thinking about scientific problems and potentially has some insights to offer the BEF field. Finally, I provide some comments on why, despite considerable evidence for the positive effects of biodiversity on ecosystem functions and services (IPBES 2019), many land managers (e.g. foresters, farmers etc.) still persist with practices that harm biodiversity. Whilst I have tried to be balanced throughout this thesis, I take this opportunity to give my personal views on these topics.

5.1: What's next for the BEF field?

New tests of BEF theory

I attended the 2022 British Ecological Society's Annual Meeting in Edinburgh and I was excited to listen to a talk from a prominent BEF researcher called: "The past and future of diversity-functioning research". I thought it would be the perfect moment to get my finger on the pulse of what the next phase of BEF research might look like. However, I was rather disappointed. I distinctly remember a question from the audience which was something to the effect of: "If you were starting your career now, what kind of BEF questions would you be addressing?". The response was that they might consider doing BEF experiments on organisms that had not been studied in a BEF context. I think the example was corals.

There's nothing inherently wrong with this response of course. Moreover, I am not so pompous to say that we absolutely *could not* learn something interesting from such experiments. Indeed, such BEF experiments with corals have recently been done and revealed that coral neighbours do affect each other's productivity and that this may be due to the release of secondary metabolites (Engelhardt *et al.* 2023). However, I am sceptical about how much more we can really learn from doing more BEF experiments.

As I see it, there are at least two ways to test theoretical models. The first way is to design experiments that match the assumptions of the theoretical model and then evaluate whether the experimental results produce the same patterns as the theoretical model. This essentially tells us whether real organisms behave in a similar way to our theoretical model. We can then use the theoretical model to help us interpret the experimental results. In my view, the experimental BEF approach represents this approach to testing theory. Indeed, as summarised by Tilman *et al.* (2014), BEF experiments are tightly connected to certain theoretical models. And, this feedback between theory and experiments, has led to a number of robust conclusions about how to interpret BEF experiments (Tilman *et al.* 2014). We take a similar approach when we correlate measures of biodiversity with measures of ecosystem functioning in field data from natural systems (i.e. we ask whether the patterns we see in natural systems vaguely match those that the theoretical models produce).

The limitation of this first way to test theory is that, if the experimental system exactly matches the theoretical model, then the results of the experiment are not in question (Bolnick 2019). I think this is important in the BEF field because the theoretical models are only really defined by a few, simple assumptions. I am simplifying slightly but what BEF theory shows is that, if we assume that i. intraspecific competition is stronger than interspecific competition (without specifying exactly how much, Barabás *et al.* 2016), ii. species differ in their competitive abilities (i.e. in the absence of niche differentiation, one species out-competes another, Chesson 2000) and iii. species differ in their levels of ecosystem functioning in the absence of competition (i.e. in monoculture) then, in the context of a BEF experiment, biodiversity should increase ecosystem functioning. The additional assumption is that species compete and are from one trophic level (but there are some multitrophic extensions: e.g. Poisot *et al.* 2013; Wang & Brose 2018).

Ecologists have studied competition for a long time and I think most would agree that these three assumptions (or at least some combination of them) are met in almost all natural ecosystems. Unless you take Hubbell's (2011) neutral theory very literally, I would even go out on a limb and say that these assumptions are probably extremely general for sets of competing species in most natural systems. In my view, this may explain why the results of BEF experiments are so consistent (Cardinale *et al.* 2011; Gamfeldt *et al.* 2015) and even apply to other scales of ecological organisation like genetic diversity (Raffard *et al.* 2019). But, what this also means is that the results of BEF experiments are not particularly surprising because we know a lot about these assumptions in many natural systems. Of course, BEF experiments have revealed many interesting details like what kinds of interactions lead to positive BEF relationships (e.g. habitat partitioning: Cardinale 2011; Williams *et al.* 2017; pathogens: Lambers *et al.* 2004). However, unless we find competitive ecological systems that do not meet these assumptions, traditional BEF experiments are probably not going to fundamentally teach us much more.

The second way to test a theoretical model is to simply test its assumptions and ask how common those assumptions are in different natural systems. In the context of BEF theory, we might test the strength of competition using removal experiments (Sears & Chesson 2007), causal inference methods (Rinella *et al.* 2020), or field-parameterised competition models (Lanuza *et al.* 2018). We might also try to evaluate differences in functioning among species in monoculture (Vile *et al.* 2006; **Paper V**). If we find evidence that the assumptions of the theoretical BEF models are met in a natural system, then it follows logically that, if we were to perform a BEF experiment using the set of species studied, it would likely result in a positive BEF relationship. This approach also has the advantage of studying species and their interactions under natural conditions as opposed to the artificial conditions imposed by most BEF experiments.

An additional advantage of this second approach to testing theoretical models is that it is easily applied to more complex models like those of the insurance hypothesis that I discussed in *Section 3* (Loreau *et al.* 2003; Shanafelt *et al.* 2015). Models of the insurance hypothesis tend to make three key assumptions i. species respond asynchronously to environmental variation in space and time in the absence of interspecific competition (i.e. in monoculture), ii. these asynchronous responses to the environment coupled with dispersal allow different species to dominate in places and times where and when they are highest functioning in monoculture, and iii. each species has some combination of environmental conditions in which they are highest functioning in monoculture (Loreau *et al.* 2021). As a result, these models predict that species loss should reduce ecosystem functioning across places and times because no species can fully compensate for the loss of another species.

Instead of trying to build an experimental system to mimic models of the insurance hypothesis (e.g. Venail *et al.* 2010), the second approach to testing theoretical models says that we must simply test their assumptions. This approach is already common for one of the most highly developed and general bodies of ecological theory: *modern coexistence theory* (Chesson 2000). Over the last 40 years or so, Peter Chesson and his co-workers have developed conditions that are required for stable coexistence of competing species. For example, the spatial storage effect (a coexistence mechanism) promotes coexistence when two assumptions are met. First, species differ in their vital rate (e.g. growth, survival, fecundity) responses to spatial environmental heterogeneity. Second, species experience stronger competition under environmental conditions where they have high vital rates. Towers *et al.* (2020) tested these assumptions by examining annual plant fecundity rates with and without competitors under a range of environmental conditions and found limited support for them. The inference is then that the spatial storage effect is probably a weak determinant of coexistence in this system. Adler *et al.* (2006) took a similar approach but using the temporal environmental heterogeneity equivalent (i.e. the temporal storage effect) and many other studies have taken this kind of assumption-testing approach to evaluating modern coexistence theory (e.g. Angert *et al.* 2009; Staples *et al.* 2016).

What might this look like for the insurance hypothesis? It could take many different forms. For example, in **Paper III**, I predicted that, for a macroalgae system, species should grow fastest (i.e. be most productive) in the environmental conditions where they naturally dominate (assumptions i and ii described above). I then tested this prediction using a transplant experiment and found a rather limited growth rate response to the environmental gradient (see also **Paper IV**). This indicates that this system may not be well-described by models of the insurance hypothesis at least during the adult life-stages that we studied. Another example which myself and some colleagues are currently working on is to use removal experiments combined with data imputation techniques to try and estimate the functioning of different species in monoculture across a variety of different environmental conditions (as I did in **Paper V**). Using this monoculture data and data on the relative abundance of different species in natural mixtures, we can directly evaluate the assumptions of some of these insurance hypothesis models.

I am excited by this kind of assumption-testing because it will also allow us to refine our theoretical models. For example, in **Papers III**, **IV** and **V**, we found evidence from different systems that strongly contradicted with assumption iii (i.e. each species is highest functioning in monoculture under some combination of environmental conditions). For example, in the meta-analysis from **Paper II**, we found that one species was highest functioning in monoculture under all the different experimental conditions (e.g. control *versus* nutrient addition) in around 50% of the experiments analysed. Similar patterns were found in **Papers IV** and **V**. Perhaps then, we need to build insurance hypothesis models that relax this third assumption or at least explore the potential consequences of assuming higher levels of compensation (e.g. **Paper IV**).

Overall, I hope that more BEF studies start to take this approach of directly evaluating the assumptions of the theoretical models in natural systems. Undoubtedly, a lot has been learnt from BEF experiments and from correlations between measures of biodiversity and functioning in field data from natural systems. But, in my view, it might be time to try something a bit new and see if it can teach us something different. The modern coexistence theory literature already has some excellent examples of this kind of assumption testing that may serve as inspiration.

Monocultures, monocultures, monocultures

There is something very satisfying about comparisons between monocultures and species mixtures. Specifically, they allow us to compare a mixture where species interact to a hypothetical mixture without any inter-specific interactions. If the functioning of the mixtures where species are interacting is higher than the hypothetical mixture without interspecific interactions, then interactions between species increase ecosystem functioning (i.e. we have a positive *net biodiversity effect*). Moreover, through the use of statistical partitions (see *Section 4*), we can then isolate different kinds of species interactions that increase (or decrease) functioning. Such species interactions are, in my view, the closest thing we will get to an unambiguous biodiversity effect. And, therefore, a net biodiversity effect can be seen as a direct quantification of the effect of species interactions on ecosystem functioning.

In nature, we very rarely have monocultures. Rather, we have communities of species (i.e. mixtures) that are interacting in different ways and at different scales of space and time (Chesson *et al.* 2013). In this context, if we can calculate a net biodiversity effect then we are asking how different the functioning of this observed community of species would be if species were prevented from interacting. Using the statistical partitions, we can also try and quantify what kinds of species interactions are important in these observed communities of species. This is the approach I took in **Paper V** using Isbell *et al.*'s (2018) statistical partition of the net biodiversity effect and I think it holds a lot of promise. For example, in **Paper V**, we were able to show that local-scale species interactions, dominance of a high functioning species and spatial niche partitioning all positively affected ecosystem functioning in two marine ecosystems. So, how do we solve the problem of insufficient monoculture data?

In **Paper V**, we used removals to obtain some observed monoculture data and then we used Bayesian data imputation techniques to obtain the rest. However, some innovation is required if this approach is going to be used more frequently in the BEF field. In both the systems we worked with in **Paper V**, it was relatively easy to create monocultures. But, this may not always be the case and, as I argued in that paper, we might need other modelling approaches that do not require observed monoculture data. This could be statistical models (e.g. Diversity-interactions models, Connolly *et al.* 2013; Joint species distribution models, Ovaskainen *et al.* 2017) or field-parameterised population models (Chalmandrier *et al.* 2021, 2022). There may also be other ways to quantify something like a net biodiversity effect. I have already been impressed with some methods for quantifying the extent that asynchronous dynamics among species and asynchronous dynamics among local populations and communities contribute to ecosystem stability (Lamy *et al.* 2019).

In proposing this as a future avenue for BEF research, I acknowledge that there has been a lot of debate about these methods (which are all based on the Price equation) to calculate and partition net biodiversity effects (Bourrat *et al.* 2023; Pillai & Gouhier 2020). Some of the criticisms are relatively minor. For example, all these partitions assume that monoculture functioning and initial relative abundances (i.e. relative expected yields, RY_{exp}) are linearly related (Pillai & Gouhier 2020). If this assumption is violated (and it frequently is), then net biodiversity effects may be substantially overestimated because the expected monoculture functioning in mixture will be underestimated (Pillai & Gouhier 2020). This is undoubtedly true but there are available methods that can solve this issue (Baert *et al.* 2017). We could also, for example, test how sensitive our conclusions are to such violations. Some of the other criticisms, however, are more existential.

According to Pillai and Gouhier (in Bourrat *et al.* 2023) for example: "*every 'Price equation' inspired approach published over the last twenty years has been foundationally flawed*". This comment holds for some of the most popular methods (Fox 2005) and for the method I used in **Paper V** (Isbell *et al.* 2018). One of the main problems cited by Pillai and Gouhier (in Bourrat *et al.* 2023) is that all these methods involve dividing observed change (i.e. the net biodiversity effect) into a set of additive but, in their view, arbitrary terms that have no ecological meaning. The division of the observed change into different terms is not necessarily based on any specific theory which, in their view, makes them meaningless. van Veelen (in Bourrat *et al.* 2023) sums this problem up as: "*there are a million tautological ways to rewrite or partition change*".

I actually agree with this characterisation. The net biodiversity effect can indeed be partitioned in many different ways and we, as researchers, decide what those ways are. However, to me, this does not make such partitions meaningless. As an example, the terms in Isbell *et al.*'s (2018) spatio-temporal partition of the net biodiversity were derived based on models of the insurance hypothesis (Loreau *et al.* 2003, 2021). Specifically, the different terms measure *patterns* in the data that would be consistent with a prediction from the models (Isbell *et al.* 2018). For example, the spatial insurance effect measures the extent to which species dominate in certain spatial locations where they are also high functioning in monoculture (Isbell *et al.* 2018). I cannot speak to the mathematical rigour of the partition (see Pillai & Gouhier 2018 for a discussion) but I have performed many simulations using metacommunity models (Thompson *et al.* 2020) while writing **Paper V**. These simulations have shown me, for example, that when species differ in their monoculture functioning in space but not time and there is no local complementarity (i.e. intraspecific competition is equal to interspecific competition), the net biodiversity effect is completely due to the spatial insurance effect (*unpublished data*). In my view, the additive terms in these partitions can, therefore, help identify certain *patterns* in the data and these patterns may correspond (albeit fairly loosely) to theoretical models. They do not, however, tell you why something happened nor do they represent some stable ecological phenomenon that we are trying to objectively measure. But, to me, that does not make them useless.

Could the use of these partitioning schemes be improved? As with anything in science, we can always make improvements. First, I do think the linearity assumption between monoculture functioning and initial relative abundance (i.e. RY_{exp}) is important and we should try to deal with this problem in the future or at least test how sensitive our conclusions are to this assumption. Second, there are question marks over whether the interpretation of these effects should be done relative to some null expectation (Massol and Peres-Neto in Bourrat *et al.* 2023) and this may be true. However, it is not clear to me what such a null model would look like and Massol and Peres-Neto (in Bourrat *et al.* 2023) provide no guidance. Finally, and perhaps most importantly, we need to be clear on what these methods can and cannot do. In my view, the terms in these statistical partitions represent fractions of the net biodiversity effect that correspond to a certain pattern in the data. Thus, in that sense, they are human constructs just as, for example, a species is a human construct along with any number of objects of scientific investigation. However, I believe these terms, when used carefully, have the potential to help us understand patterns and processes occurring in natural systems.

5.2: Causal inference concepts and BEF

During my PhD, I became more and more interested in *causal inference*, a sub-discipline of statistics and computer science that aims to infer causes from data. Having never heard of it during

my education, I was surprised to learn that causal inference methodologies are standard practice in several quantitative fields like economics and epidemiology. Learning about causal inference was extremely enlightening because it gave me a framework for thinking carefully about several causal problems that I had encountered whilst reading and working in the BEF field.

The first causal inference problem that I found myself grappling with during my studies of the BEF literature is the debate about whether it is species richness (i.e. the number of different species) or species composition (i.e. the identity and abundance of different species) that causes ecosystem functioning (Bengtsson 1998). This goes back to some of the earliest controversies in the BEF field because, in a typical BEF experiment (Fig. 4), we usually cannot do a fully factorial manipulation where all possible combinations of species richness and species composition are created. This means that, as species richness increases, certain species or combinations of species are more likely to be included (Aarssen 1997; Huston 1997). To solve this issue, many early BEF studies tested separately for the effect of species richness and species composition by including both factors in their statistical models (e.g. species composition was operationalised as assemblage identity or the identity of the functional groups present) (Hector *et al.* 1999; Tilman *et al.* 2001). Moreover, other experimental designs and analysis methods were developed that mitigate this problem. For example, one popular method mitigated this problem by only including each species once across all replicates in each species richness treatment (Bell *et al.* 2009). Other methods instead used the ability of differences in species richness and differences in species composition (quantified as Bray-Curtis dissimilarity) to predict differences in functioning as a way to try and separate out these effects (Sandau *et al.* 2017).

Similar issues are present in observational field data. Usually, when studies test for an effect of a biodiversity variable (e.g. species richness) on some measure of ecosystem functioning in field data from natural systems, various covariates are included as statistical controls (Duffy *et al.* 2017). However, looking at van der Plas' (2019) systematic review of such studies, it is clear that there is no agreement on the kinds of covariates that should be included. Whilst most studies seem to agree that environmental variables should be controlled for, there is no agreement on whether species composition should be. This means that some studies include a species composition covariate such as the assemblage identity (Ratcliffe *et al.* 2017), an ordination axis (Kahmen *et al.* 2005) or the relative abundance of each species (Jacob *et al.* 2013) whilst other studies do not. The exception to this are studies that analyse functional trait metrics where there seems to be a tacit assumption that both a measure of functional diversity and the community-weighted mean (as a measure of species composition) should be included in the statistical models (van der Plas 2019).

So who is correct? Do we need to statistically control for species composition if we want to properly measure the effect of biodiversity on ecosystem functioning? And, if so, how should this be done? Does this change our interpretation of what constitutes a biodiversity effect? What are the consequences of doing so for the interpretation of BEF experiments? In practical terms, what does a biodiversity effect measure? These are some of the questions I will try to answer using causal inference principles.

Applying causal inference principles to interpret BEF studies

One of the main tools of causal inference is the causal graph and one of the most popular kinds of causal graph is the Directed Acyclic Graph (DAG, Digitale *et al.* 2022). DAGs consist of variables (represented as nodes) and arrows which depict suspected causal relationships between variables. For

example, a DAG with two nodes (X and Y) connected by an arrow from X to Y says that X causes Y and that Y does not cause X. Using DAGs allows one to express assumptions and hypotheses about a system which can help guide both the interpretation and analysis of any given scientific project (Digitale *et al.* 2022).

Before developing a DAG for BEF studies, I will illustrate the use of a DAG with a simpler ecological experiment. In the central grassland regions of the United States of America, many observational studies have shown that precipitation is a major correlate of net primary productivity (NPP) (e.g. Sala *et al.* 1988). However, understanding the causal effect of precipitation is important because of the predicted changes in precipitation in the region under climate change and because the effect may be confounded by other factors (e.g. temperature, growing season length etc). We can represent this situation, where precipitation and NPP are confounded in a DAG (Fig. 19a).
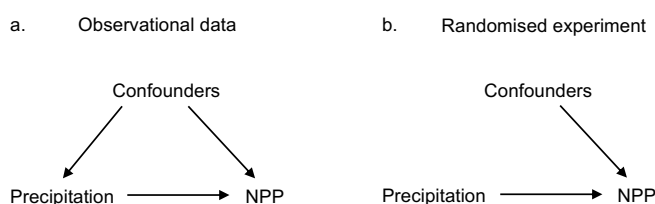

Fig. 19: (a) A Directed Acyclic Graph (DAG) representing causal assumptions about how precipitation affects Net Primary Productivity (NPP) and how other confounding variables (e.g. temperature, humidity, growing season duration etc.) may affect both precipitation and NPP. **(b)** In this DAG, precipitation is experimentally manipulated randomly such that it becomes decoupled from the confounding variables. This allows the causal effect of precipitation on NPP to be estimated.

To experimentally eliminate any confounding between precipitation and NPP, we need to change precipitation and we need to do so randomly (e.g. assign different precipitation treatments randomly to different plots). In a DAG, we represent this randomised experimental treatment by deleting the arrow between the confounding variables and precipitation (Fig. 19b). The logic is that, because we have uncoupled the different precipitation treatments from variables that typically covary with precipitation and NPP in the field through randomisation, the confounding variables no longer affect precipitation (Pearl & Mackenzie 2018). Practically, we may manipulate precipitation by using rain shelters (Byrne *et al.* 2013) or by redirecting some precipitation during rainfall (February *et al.* 2013) but the method of manipulation is not particularly important for this discussion.

Using these principles, I will now construct a DAG for a typical BEF experiment (i.e. Fig. 4). I am not the first one to use causal inference techniques to think about a BEF experiment (Grace *et al.* 2022; Schoolmaster Jr. *et al.* 2020, 2022). But, my views differ slightly from these studies. For a typical BEF experiment, we need to randomise the species richness treatment because we want to know whether biodiversity (which we often measure as species richness) affects ecosystem functioning (Fig. 20). However, to set the level of species richness, we need to manipulate species composition (i.e. the identity of species, Fig. 20). Indeed, it is not possible to change species richness without changing species composition (Sandau *et al.* 2017). Therefore, whilst we can randomise the species richness treatment and we can randomise the species composition of any given replicate, the species richness treatment is not, in my view, independent of species composition if species composition is defined as the identity and abundance of species (Fig. 20). Rather, as has been pointed

out previously, species richness is a treatment that has *multiple versions* (Dee *et al.* 2023; Kimmel *et al.* 2021).
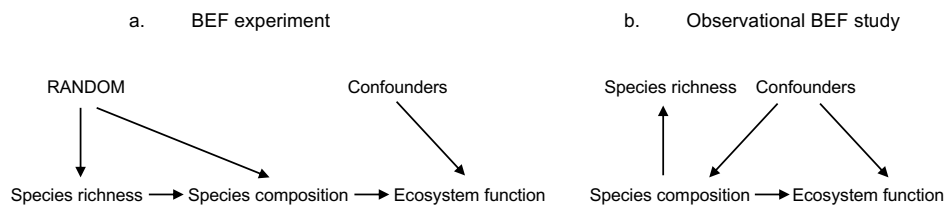
a. BEF experiment

b. Observational BEF study

RANDOM   Confounders   Species richness   Confounders

Species richness → Species composition → Ecosystem function   Species composition → Ecosystem function

Fig. 20: (a) A Directed Acyclic Graph (DAG) that represents my view of a BEF experiment. In this DAG, we randomise both species richness and species composition within any given species richness treatment which is assigned first. The species composition then fully determines ecosystem function (Wanis *et al.* 2019). **(b)** In an observational BEF study, the DAG is slightly different. Here, the version of the treatment (i.e. species composition) precedes and fully determines species richness. We can estimate a causal effect of species richness on ecosystem function but if we statistically adjust for species composition then we are blocking any effect that species richness may have.

What do I mean by multiple versions? Well, most simply, if we have a plot with one species (e.g. species A) and we have a pool of 20 species, there are 3876 ways to increase the plot to a species richness of five (e.g. {A, B, C, D, E}, {A, B, C, D, F} etc.). Unlike some have argued (Schoolmaster Jr. *et al.* 2020, 2022), this does not mean that we cannot calculate a valid causal effect of biodiversity on ecosystem functioning. The causal inference literature makes no such claim. Rather, there are many similar cases in the literature such as: What is the causal effect of heart surgery on survival if patients have a certain heart problem? In this case, surgery is the treatment but the surgery is done by different surgeons who have different competences, experience levels etc. which may considerably alter the effectiveness of the surgery (VanderWeele & Hernan 2013). Moreover, there are cases where these kinds of multiple treatment versions require special statistical attention (VanderWeele & Hernan 2013) although, as far as I can tell, this is generally not the case for a well-designed BEF experiment. However, I do think we need to think carefully about what a causal effect of biodiversity on ecosystem function means.

In my view, it is important for the community of BEF researchers to recognise and communicate that the average effect of biodiversity on ecosystem functioning that has been measured in BEF experiments can only be used to make vague statements about what may happen to ecosystem functioning if biodiversity declines. We can make interventional statements like: if we reduce species richness from five to one, on average, ecosystem functioning will decline. However, the caveat is that, depending on which species go extinct, this effect could be strengthened or reversed (i.e. ecosystem function may increase if a specific high functioning monoculture is left over after species loss) (Dee *et al.* 2023). This is because, depending on how species are lost, a community can follow many different trajectories which can have a considerable impact on the functioning of the system (Fig. 21; Bannar-Martin *et al.* 2018; De Laender *et al.* 2016). But, even these vague statements do not necessarily tell us what may happen if a species goes extinct from some intact community. Rather, as I argued in **Paper I**, they tell us what may happen if local species pools become impoverished. This is much less tangible and more difficult to measure.
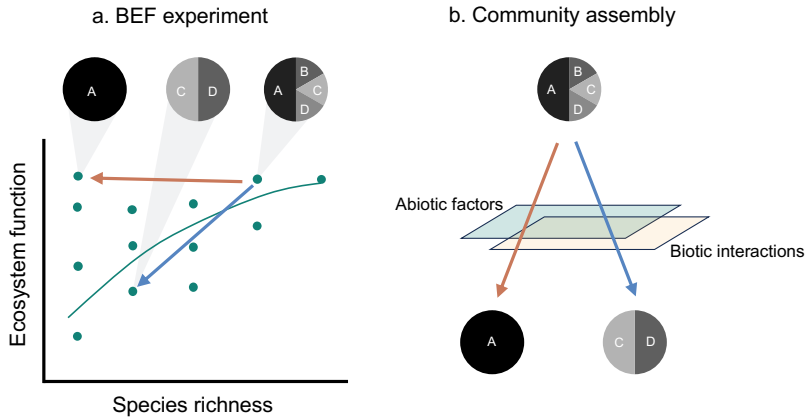
Fig. 21: Changes in species richness can have very different effects on ecosystem functioning depending on how species composition changes. **(a)** Results from a typical BEF experiment showing changes in average ecosystem functioning with species richness (green line). However, three points are highlighted to illustrate how going from four to two species can decrease functioning (blue arrow) whilst going from four to one species can increase functioning (orange arrow) depending on the trajectory of change. **(b)** The trajectory of change will depend on the community assembly processes (e.g. abiotic factors and biotic interactions). Figure redrawn and modified from Bannar-Martin *et al.* (2018)

In observational studies, we also have the problem of multiple versions when we examine the effect of species richness on ecosystem functioning (Dee *et al.* 2023). However, it takes a slightly different form. In the case of a BEF experiment, the treatment precedes the version i.e. we assign an experimental unit a biodiversity treatment (e.g. a level of species richness) and then we decide on a specific species composition (Fig. 20a). However, in my opinion, the same cannot be said of an observational study. Take the example presented previously of studying the effect of surgery on survival for a certain heart problem. If we gather observational data in this example, we first know the version of the treatment (i.e. which surgeon performed the surgery etc.). We then classify the treatment as 'surgery' or 'no surgery'. This is, in my view, what happens in observational studies where a plot's species composition is measured (i.e. counting and identifying species in a plot) and then assigned a species richness value post-hoc (see Grace *et al.* 2022 who make somewhat similar points). This kind of version-precedes-treatment situation can also require specific statistical methods (VanderWeele & Hernan 2013). However, most importantly, the version of the treatment should not be included in statistical models because this will block any effect of the treatment (VanderWeele & Hernan 2013) (Fig. 20b). Thus, the causal inference literature gives us an answer as to whether we need to statistically control for species composition and the answer is no. We do, however, need to control for other confounding variables like aspects of the environment because these variables affect both species composition and ecosystem functioning (see Dee *et al.* 2023 for an excellent example) (Fig. 20b).

A few anecdotal examples from the literature illustrate what happens when species composition is statistically controlled. For example, Fotis *et al.* (2018) found a significant, positive relationship between tree species richness and aboveground biomass but this effect disappeared in a multiple regression model including a species composition variable (in this case, an nMDS axis). Ratcliffe *et*

al. (2017) included an assemblage identity factor in their models of the effect of species richness on various ecosystem functions. In general, they found weak effects of species richness on different functions and species richness explained considerably less variation than assemblage identity. A counterexample is Gamfeldt *et al.* (2013) who included the biomass of six dominant species as statistical controls in their models of how species richness affects different ecosystem functions in Swedish forests. However, this is probably because the plots included up to 10 species and, therefore, species richness provided compositional information about the additional species not included in the models. In general, I would hypothesise that statistically controlling for species composition, in most cases, would cause any effect of species richness on ecosystem functioning to disappear.

This discussion is a long way to say that I do not think there are such things as unambiguous biodiversity effects on ecosystem functioning in the same way that there are unambiguous effects of precipitation on net primary productivity. In my view, this is because we cannot manipulate species richness without changing species composition, neither practically nor theoretically. For example, it does not make sense to ask what the effect of species richness is when holding species composition constant (Schoolmaster Jr. *et al.* 2022). However, that does not mean that we cannot calculate a valid causal effect of species richness on ecosystem functioning as argued by Schoolmaster Jr. *et al.* (2020). But, we must recognise that such a causal effect can only ever represent a vague statement about what we expect to see when biodiversity changes (Grace *et al.* 2022) especially when certain versions of the treatment (i.e. species compositions) may be rarely found in natural systems (Jochum *et al.* 2020; Lepš 2004; VanderWeele & Hernan 2013). In certain contexts (e.g. high-level policy goals) this can be useful information and should not be discounted (Isbell *et al.* 2017). Moreover, BEF experiments have undoubtedly provided fundamental ecological knowledge (Eisenhauer *et al.* 2016; Petermann *et al.* 2010). But, if we want to deepen our understanding of specific ecological systems and answer questions like how ecosystems will respond to climate change, pollution and other global change drivers or predict how an invasive species may alter ecosystems, I do not think the standard BEF approach is necessarily the way to do it.

Back to functional traits

Although I disagree with Schoolmaster Jr. *et al.*'s (2020) main point that any causal effects of species richness on ecosystem functioning cannot be calculated and are fundamentally non-causal, I do find myself agreeing with much of what they have to say about alternatives. Schoolmaster Jr. *et al.* (2020) propose a research program aimed at expressing and quantifying trait-based mechanisms that link changes in species composition to changes in ecosystem functioning. Of course, this is not new. For example, Díaz *et al.* (2007) outlined such a research program more than 15 years ago. And, such a research program is already being actively pursued. Indeed, most observational BEF studies test for relationships between multiple metrics of biodiversity such as functional diversity, the community weighted mean of different traits and taxonomic diversity indices (e.g. species richness) on ecosystem functioning (van der Plas 2019). However, I envision a broader research program where these studies make way for new studies with a stronger focus on the questions that arguably kicked off BEF as a sub-field of ecology.

When I started my PhD, I read Schulze and Mooney's (1993) edited volume: *Biodiversity and Ecosystem Function*. The volume was published around the time when Naeem *et al.* (1994) and Tilman *et al.* (1996) were performing the first BEF experiments. In the preface of this book, Schulze

and Mooney outlined the goal of the volume and, in my view, laid out the kinds of questions that they wanted the BEF research program to address. Specifically, they wrote:

"The biota of the earth is being altered at an unprecedented rate. ... There are many reasons for concern about these trends. One is that we unfortunately do not know in detail the consequences of these massive alterations in terms of how the biosphere as a whole operates or even, for that matter, the functioning of localized ecosystems. We do know that the biosphere interacts strongly with the atmospheric composition, contributing to potential climate change. We also know that changes in vegetative cover greatly influence the hydrology and biochemistry of a site or region. Our knowledge is weak in important details, however. How are the many services that ecosystems provide to humanity altered by modifications of ecosystem composition? Stated in another way, what is the role of individual species in ecosystem function? We are observing the selective as well as wholesale alteration in the composition of ecosystems. Do these alterations matter in respect to how ecosystems operate and provide services? This book represents the initial probing of this central question."

In my opinion, over the last 20 years, many ecologists (myself included although I haven't been around for 20 years) have focused a bit too much on trying to find and quantify a causal effect of biodiversity on ecosystem functioning. Ecologists like biodiversity and, therefore, want to show that it is important (Frank 2022). However, I think this has come at the expense of answering arguably more fundamental questions about how ecosystems work. Given my work in the BEF field over the last few years and with the benefit of hindsight I now think it is less important to prove that, in grassland ecosystems, experimentally removing species causes productivity to, on average, decline than to be able to answer questions like: What will happen to the species composition in this grassland if we add nitrogen or if precipitation increases? How will these changes affect biogeochemical cycles? Can we forecast potential changes in biogeochemical cycles under future expected climate change? What are the consequences of such changes in biogeochemical cycles for the benefits we derive from ecosystems? There is no doubt that BEF studies have contributed to answering questions like this (Cardinale *et al.* 2012; Eisenhauer *et al.* 2016) and provided fundamental ecological insights (Tilman *et al.* 2014). But, in general, I think that answering these questions requires a different kind of research program than that pursued currently in the BEF literature.

I will now lay out my thoughts on such a research program. These thoughts reflect my reading of the general ecological literature and how we might use insights from community assembly theory, functional trait ecology and coexistence theory to improve the BEF research program.

Community assembly has shown us that the species composition (i.e. the identity and abundance of species) at any given site is based on the combination of dispersal, the environmental tolerances of different species and species interactions (Germain *et al.* 2018; Kraft *et al.* 2015a; Mayfield & Levine 2010). More specifically, dispersal determines which species arrive at a given site and species' responses both to the environment and to other species determines their demographic rates (i.e. growth, survival, reproduction) and, ultimately, their abundances (Kraft *et al.* 2015a; Laughlin *et al.* 2018). As I see it, ecosystem functions (or processes as I prefer to call them) are then the emergent properties of that community assembly process (Bengtsson 1998).

A crucial component of the community assembly process are species' traits. However, species' traits do not only determine their response to the environment (Laughlin *et al.* 2018) and to other

species (Kraft *et al.* 2015b), traits are also important determinants of ecosystem processes (Chacón-Labella *et al.* 2023). Traits that determine species responses to the environment and other species are often called *response traits* whilst those that determine ecosystem processes are often called *effect traits* (Suding *et al.* 2008). In practice, these can overlap substantially in certain cases (e.g. Blanco *et al.* 2007). There is considerable evidence showing that response traits can be used to predict species' responses to the environment and other species and can be used to parameterise models of community assembly (Laughlin 2014; Laughlin *et al.* 2012). Moreover, community distributions of effect traits have been shown to be good predictors of ecosystem processes in some cases (Hagan *et al.* 2023; van der Plas *et al.* 2020). Thus, community assembly theory combined with trait-based approaches, in my view, have a lot of potential to answer some of the questions I laid out for a future BEF research program.

What do I mean by this? In my view, the first step when trying to understand connections between biodiversity and ecosystem processes is to understand the community assembly processes in the system. There are many different ways to go about this (reviewed in Keddy & Laughlin 2021) but I think the most useful approach in a BEF context is to build predictive models of species composition based on species' response traits and the environment (e.g. Chalmandrier *et al.* 2021, 2022; Laughlin *et al.* 2012). For example, Chalmandrier *et al.* (2022) used species' response traits to estimate species' responses to the abiotic environment along with the strength of species interactions. This allowed them to calibrate a model of plant species abundances including both of these factors. Not only does such a model provide clear insights into community assembly mechanisms (e.g. environmental filtering acts primarily through the response of certain response traits to the environment and species interactions are based on other response traits), it is fully predictive. This means that, using species' response traits and environmental parameters, the abundance of different species can be directly modelled (see also Chalmandrier *et al.* 2021). Such a model gives us the power to answer questions posed previously like how will this plant community respond to changes in, for example, precipitation or nutrient enrichment?

Secondly, once we understand the community assembly process and can predict species composition, it is easy to derive effect trait distributions for a given community (provided we have at least species-level data on effect traits). This means that we can derive a multidimensional effect trait distribution for any given community which, in theory, should provide considerable information about different ecosystem processes (Chacón-Labella *et al.* 2023). There are many examples where effect trait distributions are able to predict ecosystem processes with high levels of accuracy (Cadotte 2017; Garnier *et al.* 2004; Liu *et al.* 2015; Smart *et al.* 2017) especially when combined with environmental information (Hu *et al.* 2020; Krix & Murray 2022). Such empirically derived relationships between trait distributions and ecosystem processes, combined with community assembly models that predict species composition, allow other questions to be answered such as: How will changes in species composition in response to environmental changes affect ecosystem processes?

Although the first two steps can provide a predictive model of ecosystem processes from species traits and the environment, they do not necessarily answer certain 'why questions'. By 'why questions', I mean specific causal questions such as why species with certain traits are selected for in a given environment or why a certain trait distribution increases the rates of some ecosystem process. The third step is then to try to answer such questions. There is no generalised approach for doing this but I provide a few examples that, in my view, are instructive. In a Mediterranean climate region, Yates *et al.* (2010) showed that small leaves were selected for in hot, dry and nutrient poor

environments (i.e. step 1 in this framework). Using ecophysiological theory, they hypothesised that small leaves increased sensible heat loss in the hot, dry summers and increased transpiration in the cold, wet winter months and thus facilitated nutrient uptake via mass flow (Cramer *et al.* 2009) and used a series of experiments to support their hypothesis. Thus, Yates *et al.* (2010) provided an answer to why small leaves are selected for in dry and nutrient poor environments which facilitated a deeper understanding of community assembly in their system.

A second example is the response of grassland plant communities to nutrient enrichment. It has been shown in many different grassland plant communities that the addition of nutrients (i.e. fertilisation) leads to changes in plant traits (e.g. plant height, leaf nutrients and specific leaf area) (Dwyer *et al.* 2014; Firn *et al.* 2019; Hautier *et al.* 2009) and a reduction in plant species richness (Harpole & Tilman 2007; Silvertown *et al.* 2006). Thus, it is well-understood what happens to traits and species composition in grassland plant communities in response to fertilisation. Hautier *et al.* (2009) then showed why this is the case. Using experimental manipulations of light in the understorey, Hautier *et al.* (2009) showed that fertilisation led to increased canopy cover which increased competition for light and, therefore, caused declines in species richness. As with Yates *et al.*'s (2010) leaf size example, this provided additional mechanistic insights into how changes in traits of species in response to the environment cause changes in species composition.

Similar examples can be found for connections between trait distributions and ecosystem processes. Probably the most famous example is the leaf economics spectrum and its relation to productivity. Plants species are generally positioned along a trait axis with resource acquisition traits (e.g. high specific leaf area) on one side and resource conservation traits (e.g. long leaf lifespan) on the other (Wright *et al.* 2004). Species with traits facilitating resource acquisition are typically found in fertile conditions (i.e. high soil nutrients and water availability) which causes these ecosystems to have higher ecosystem productivity (Li *et al.* 2021; Smart *et al.* 2017). There are many potential causes for these patterns, but one that has been thoroughly investigated is the relationship between leaf traits and net photosynthetic rates (Marino *et al.* 2010). Therefore, with these additional studies, we can conclude that resource acquisition traits are positively correlated with productivity and one of the causes is that resource acquisition traits lead to higher rates of photosynthesis.

Pearl and Mackenzie (2018) argue that the highest level of causal understanding is the ability to construct counterfactual outcomes. The logic is that if we understand how a system works causally, then we can answer questions about how a system would have behaved if, for example, something in the system had been different. Some of the most successful scientific research programs have been based on this kind of counterfactual reasoning. Probably the most famous example is the current climate models (Pearl & Mackenzie 2018). Scientists have constructed incredibly elaborate causal models of the earth's climate which has allowed us to answer questions like: What would the earth's climate look like now if we had not started emitting greenhouse gases during the industrial revolution? Answering such questions is not only key to scientific understanding but is also often how humans think causally (Pearl & Mackenzie 2018) and has been called for in biodiversity science (Gonzalez *et al.* 2023).

The research program I have outlined here allows a range of useful counterfactual outcomes to be constructed. This is possible because such a research program is fully predictive. Specifically, using response traits and the environment to model species composition with predictive community assembly models (step 1) allows counterfactual species compositions to be generated under environmental conditions or species interactions that were not observed but may occur in the future.

Secondly, the predicted species compositions can then be used to derive multidimensional effect trait distributions which can be converted to ecosystem functions using empirically derived effect trait – ecosystem function relationships (step 2). With additional answers to certain 'why questions' (step 3), this general approach can help build a predictive, theory-driven understanding of community assembly and ecosystem functioning. In my view, this kind of predictive, theory-driven approach is important because we now know that the changes we are seeing in biodiversity are considerably more complicated than simple decreases in species richness (Blowes *et al.* 2019; Hillebrand *et al.* 2018; McGill *et al.* 2015).

I truly hope that, in this discussion, I did not make it seem that these ideas are all mine and that they are extremely novel. This is indeed not the case and many authors have proposed and implemented similar research programs (Benkwitt *et al.* 2020; Díaz *et al.* 2007, 2013; Laughlin 2014). But, if I had my PhD over again and if I knew what I know now, this is the kind of research I would be putting my energy into. Predicting species composition and ecosystem functioning from causal principles is arguably one of the major goals in ecology (Currie 2019) and, in my view, we need to move beyond typical BEF approaches if we want to reach such a goal.

Finally, to end this section, I also think that BEF researchers need to start thinking bigger again. As quoted previously, Schulze and Mooney (1993) presented questions related to, for example, feedbacks between the biosphere and earth's climate. Such questions are, for the most part, well-beyond the current BEF literature but why should they be? In my view, there are some truly exciting questions about connections between the biosphere, ecosystem functions/processes and ecosystem services that are being currently being pursued. These include how herbivores alter the distribution of nutrients on continental scales (Doughty 2017), how species interactions affect the movement of resources between land and sea (Lapiedra *et al.* 2023) and how feedbacks between drivers like geology, climate and fire determine the distribution of major biomes (Cramer *et al.* 2019; Staver *et al.* 2011). If we view biodiversity as more than just species richness, functional trait variation or phylogenetic diversity and more as a concept that encompasses the whole biosphere and its interconnections at multiple spatial and temporal scales, I think we will produce a much richer understanding of the natural world and how it mediates ecosystem processes.

5.3: Closing remarks

At the 2023 Swedish Oikos conference in Göteborg, Janne Bengtsson gave an interesting talk on multispecies forest plantations. Bengtsson was part of some of the early debates about BEF research (Bengtsson 1998) and has, since then, done a fair amount of work on forest biodiversity and ecosystem functioning (Gamfeldt *et al.* 2013; Jonsson *et al.* 2019). The crux of Bengtsson's talk was that, despite considerable evidence that mixed species tree plantations generally improve production relative to monoculture plantations (e.g. Jonsson *et al.* (2019) found transgressive overyielding for certain species mixtures in Swedish forests), forest owners still persist with monoculture plantations. This, in his view, was because Swedish forest research was dominated by foresters and that this has filtered to society which has made it difficult to make changes in light of new evidence.

I agree with Bengtsson that the evidence that mixed species plantations often have higher production (and other ecosystem services) than monoculture plantations is undeniable (Feng *et al.* 2022; Jonsson *et al.* 2019). Moreover, there is some evidence that mixed plantations are less susceptible to diseases which can be important over longer timescales (Kelty 2006). It seems so

absurd then. Mixed plantations are often more productive and usually they provide higher levels of many ecosystem services than monoculture plantations yet they are still very uncommon (Liu et al. 2018). Forest plantations are just one example. Similar arguments are made about wild pollinators whereby ecologists promote pollinator-friendly agricultural practices because pollinators can enhance fruit production and quality (e.g. Garratt *et al.* 2014) and I could name several, similar examples.

So, why, despite all this evidence, do foresters, farmers and land managers continue with the status quo (IPBES-Food 2016)? I think we, as ecologists, may forget sometimes that forest owners, farmers and land managers are simply trying to make a living. Indeed, research has shown that economic constraints are a key factor in their decision making (Bartkowski & Bartke 2018). As a result, foresters, farmers and land managers have to think about more than just production or yield benefits. If we take the forestry example, compared to mixture plantations, monoculture plantations result in more uniform final products which are both easier to harvest and to work with post-harvesting (Liu *et al.* 2018). Moreover, mixed plantations often require more maintenance than monocultures (Nichols *et al.* 2006) and more expertise which land managers may not have. Thus, despite higher production, mixed plantations have several direct costs. Given uncertainties in these costs and the need for specific expertise, I think it makes more sense that monocultures still dominate forestry and, more generally, why biodiversity-friendly forestry, farming and land management are not very common (IPBES-Food 2016).

Ecologists have, so far, demonstrated quite thoroughly how certain biodiversity-friendly practices (e.g. mixed tree plantations) can improve both yields (e.g. biomass production, fruit production etc.) and biodiversity (e.g. Albrecht *et al.* 2020; Feng *et al.* 2022). And, the BEF literature shows that there can be benefits of biodiversity for several ecosystem functions (Cardinale *et al.* 2012). However, if we, as ecologists are serious about these kinds of practices, I think we need to take it upon ourselves to demonstrate that it can work operationally at scale whilst doing a thorough accounting of all the different costs involved (e.g. maintenance, expertise, product quality etc.). An excellent recent example of this was Scheper *et al.*'s (2023) analysis of the economic costs and benefits associated with pollinator-friendly grassland management for promoting sunflower yields in adjacent sunflower crops. Scheper *et al.* (2023) showed that reducing grassland harvesting rates improved wild bee biodiversity and, therefore, pollination rates of adjacent sunflower crops. This, in turn, translated to increased sunflower crop revenues of between 10% and 17% (depending on the type of grassland). However, reducing grassland harvesting rates led to a decline in revenue from forage biomass which was considerably greater than the pollination-driven revenue benefits.

Of course, Scheper *et al.*'s (2023) analysis only focused on pollination benefits from reduced grassland harvesting. And, there may have been improvements in several other ecosystem services like pest-control, carbon sequestration and many others which were not quantified. These are public goods that are not easily monetised. If we were to monetise them, there may be net increases in revenue. However, without such monetisation, I think it is unlikely that farmers would adopt such practices at scale. As a result, in my view, there is some serious innovation to be done in ecology to figure out how to support biodiversity and ecosystem functioning in managed landscapes without reducing revenues.

There are already some fascinating examples of this kind of innovation in the literature. Recently, for example, Zemp *et al.* (2023) showed that planting islands of trees in large oil palm plantations drastically improved both the biodiversity of multiple taxonomic groups and several indicators of ecosystem functioning with negligible effects on yield at the plantation scale. They did

not, however, examine the costs that come with setting up and managing such tree islands which is an obvious next step. However, if the management costs are low, tree islands may provide a simple solution to improve biodiversity and ecosystem functioning outcomes in oil palm plantations with minimal economic impact. Another interesting example comes from Alignier *et al.* (2020) who showed that increasing crop heterogeneity at the landscape scale led to considerable increases in wild plant diversity in the agricultural fields. This shows that, without changing the total amount of agricultural land but only modifying the number of different crops in a landscape, there can be substantial benefits for biodiversity. These examples show that it may be possible to support biodiversity and ecosystem functioning with limited impact on revenues for foresters, farmers and land-managers.

In my view, all the work we have done to demonstrate the importance of biodiversity for supporting ecosystem functioning and services will be worthless unless we can come up with more of these innovative solutions for supporting biodiversity and ecosystem functioning in managed landscapes and, I should not forget, seascapes. Biodiversity is remarkably resilient and, in protected areas which make up around 10% of both land and sea (UNEP-WCMC and IUCN 2016), it can largely take care of itself in my view. But, more than 70% of both land and sea has been directly modified in some way by humans (Allan *et al.* 2017; Jones *et al.* 2018). More than 35% of land is agricultural (FAO 2016) and the amount of actively managed ocean for aquaculture is predicted to increase considerably in the future (Galparsoro *et al.* 2020). Being the cynic that I am, I do not see a world where we give significant amounts of land and sea back to nature. As a result, I think we need to focus on innovation that is going to make tangible gains for biodiversity and ecosystem functioning in modified lands and seas. This is where, in my view, we can make the biggest gains for biodiversity and, as a result, maintain the ecosystems functions that the planet depends on.

Reference list

Aarssen, L.W. (1997). High Productivity in Grassland Ecosystems: Effected by Species Diversity or Productive Species? *Oikos*, 80, 183–184.

Adler, P.B., HilleRisLambers, J., Kyriakidis, P.C., Guan, Q. & Levine, J.M. (2006). Climate variability has a stabilizing effect on the coexistence of prairie grasses. *Proceedings of the National Academy of Sciences*, 103, 12793–12798.

Albrecht, J., Peters, M.K., Becker, J.N., Behler, C., Classen, A., Ensslin, A., *et al.* (2021). Species richness is more important for ecosystem functioning than species turnover along an elevational gradient. *Nature Ecology and Evolution*, 5, 1582–1593.

Albrecht, M., Kleijn, D., Williams, N.M., Tschumi, M., Blaauw, B.R., Bommarco, R., *et al.* (2020). The effectiveness of flower strips and hedgerows on pest control, pollination services and crop yield: a quantitative synthesis. *Ecology Letters*, 23, 1488–1498.

Alignier, A., Solé-Senan, X.O., Robleño, I., Baraibar, B., Fahrig, L., Giralt, D., *et al.* (2020). Configurational crop heterogeneity increases within-field plant diversity. *Journal of Applied Ecology*, 57, 654–663.

Allan, J.R., Venter, O. & Watson, J.E.M. (2017). Temporally inter-comparable maps of terrestrial wilderness and the Last of the Wild. *Scientific Data*, 4, 170187.

Andre, M., Brechignac, F. & Thibault, P. (1994). Biodiversity in model ecosystems. *Nature*, 375, 565.

Angert, A.L., Huxman, T.E., Chesson, P. & Venable, D.L. (2009). Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences.*, 106, 11641–11645.

Baert, J.M., Eisenhauer, N., Janssen, C.R. & De Laender, F. (2018). Biodiversity effects on ecosystem functioning respond unimodally to environmental stress. *Ecology Letters*, 21, 1191–1199.

Baert, J.M., Jaspers, S., Janssen, C.R., De Laender, F. & Aerts, M. (2017). Nonlinear partitioning of biodiversity effects on ecosystem functioning. *Methods in Ecology and Evolution*, 8, 1233–1240.

Balvanera, P., Pfisterer, A.B., Buchmann, N., He, J.-S., Nakashizuka, T., Raffaelli, D., *et al.* (2006). Quantifying the evidence for biodiversity effects on ecosystem functioning and services: Biodiversity and ecosystem functioning/services. *Ecology Letters*, 9, 1146–1156.

Bannar-Martin, K.H., Kremer, C.T., Ernest, S.K.M., Leibold, M.A., Auge, H., Chase, J., *et al.* (2018). Integrating community assembly and biodiversity to better understand ecosystem function: the Community Assembly and the Functioning of Ecosystems (CAFE) approach. *Ecology Letters*, 21, 167–180.

Barabás, G., D'Andrea, R. & Stump, S.M. (2018). Chesson's coexistence theory. *Ecol Monogr*, 88, 277–303.

Barabás, G., J. Michalska-Smith, M. & Allesina, S. (2016). The Effect of Intra- and Interspecific Competition on Coexistence in Multispecies Communities. *The American Naturalist*, 188, E1–E12.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., *et al.* (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471, 51–57.

Barry, K.E., Pinter, G.A., Strini, J.W., Yang, K., Lauko, I.G., Schnitzer, S.A., *et al.* (2021). A graphical null model for scaling biodiversity–ecosystem functioning relationships. *Journal of Ecology*, 109, 1549–1560.

Bartkowski, B. & Bartke, S. (2018). Leverage Points for Governing Agricultural Soils: A Review of Empirical Studies of European Farmers' Decision-Making. *Sustainability*, 10, 3179.

Bell, T., Lilley, A.K., Hector, A., Schmid, B., King, L. & Newman, J.A. (2009). A Linear Model Method for Biodiversity–Ecosystem Functioning Experiments. *The American Naturalist*, 174, 836–849.

Bengtsson, J. (1998). Which species? What kind of diversity? Which ecosystem function? Some problems in studies of relations between biodiversity and ecosystem function. *Applied Soil Ecology*, 10, 191–199.

Benkwitt, C.E., Wilson, S.K. & Graham, N.A.J. (2020). Biodiversity increases ecosystem functions despite multiple stressors on coral reefs. *Nature Ecology and Evolution*, 4, 919–926.

Blanco, C.C., Sosinski, E.E., dos Santos, B.R.C., da Silva, M.A. & Pillar, V.D. (2007). On the overlap between effect and response plant functional types linked to grazing. *Community Ecology*, 8, 57–65.

Blowes, S.A., Supp, S.R., Antão, L.H., Bates, A., Bruelheide, H., Chase, J.M., *et al.* (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, 366, 339–345.

Bohner, T. & Diez, J. (2020). Extensive mismatches between species distributions and performance and their relationship to functional traits. *Ecology Letters*, 23, 33–44.

Bolnick, D. (2019). The parable of Bob and Dr. Doom. Or, "When the seminar speaker craps on your idea". *ecoevoevoeco: eco-evolutionary dynamics*. http://ecoevoevoeco.blogspot.com

Bond, E.M. & Chase, J.M. (2002). Biodiversity and ecosystem functioning at local and regional spatial scales. *Ecology Letters*, 5, 467–470.

Bourrat, P., Godsoe, W., Pillai, P., Gouhier, T.C., Ulrich, W., Gotelli, N.J., *et al.* (2023). What is the price of using the Price equation in ecology? *Oikos*, 2023, e10024.

Bracken, M.E.S., Friberg, S.E., Gonzalez-Dorantes, C.A. & Williams, S.L. (2008). Functional consequences of realistic biodiversity changes in a marine ecosystem. *Proceedings of the National Academy of Sciences.*, 105, 924–928.

Bracken, M.E.S. & Williams, S.L. (2013). Realistic changes in seaweed biodiversity affect multiple ecosystem functions on a rocky shore. *Ecology*, 94, 1944–1954.

Buchmann, T., Schumacher, J., Ebeling, A., Eisenhauer, N., Fischer, M., Gleixner, G., *et al.* (2018). Connecting experimental biodiversity research to real-world grasslands. *Perspectives in Plant Ecology, Evolution and Systematics*, 33, 78–88.

Byrne, K.M., Lauenroth, W.K. & Adler, P.B. (2013). Contrasting Effects of Precipitation Manipulations on Production in Two Sites within the Central Grassland Region, USA. *Ecosystems*, 16, 1039–1051.

Byrnes, J.E.K., Gamfeldt, L., Isbell, F., Lefcheck, J.S., Griffin, J.N., Hector, A., *et al.* (2014). Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods in Ecology and Evolution*, 5, 111–124.

Cadotte, M.W. (2013). Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proceedings of the National Academy of Sciences*, 110, 8996–9000.

Cadotte, M.W. (2017). Functional traits explain ecosystem function through opposing mechanisms. *Ecology Letters*, 20, 989–996.

Cardinale, B.J. (2011). Biodiversity improves water quality through niche partitioning. *Nature*, 472, 86–89.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., *et al.* (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.

Cardinale, B.J., Harvey, C.T., Gross, K. & Ives, A.R. (2003). Biodiversity and biocontrol: emergent impacts of a multi-enemy assemblage on pest suppression and crop yield in an agroecosystem. *Ecology Letters*, 6, 857–865.

Cardinale, B.J., Ives, A.R. & Inchausti, P. (2004). Effects of species diversity on the primary productivity of ecosystems: extending our spatial and temporal scales of inference. *Oikos*, 104, 437–450.

Cardinale, B.J., Matulich, K.L., Hooper, D.U., Byrnes, J.E., Duffy, E., Gamfeldt, L., *et al.* (2011). The functional role of producer diversity in ecosystems. *American Journal of Botany*, 98, 572–592.

Cardinale, B.J., Venail, P., Gross, K., Oakley, T.H., Narwani, A., Allan, E., *et al.* (2015). Further re-analyses looking for effects of phylogenetic diversity on community biomass and stability. *Functional Ecology*, 29, 1607–1610.

Carlander, K.D. (1955). The Standing Crop of Fish in Lakes. *Journal of the Fisheries Board of Canada*, 12, 543–570.

Carpenter, S.R., Kitchell, J.F., Hodgson, J.R., Cochran, P.A., Elser, J.J., Elser, M.M., *et al.* (1987). Regulation of Lake Primary Productivity by Food Web Structure. *Ecology*, 68, 1863–1876.

Cervin, G., Lindegarth, M., Viejo, R.M. & Åberg, P. (2004). Effects of small-scale disturbances of canopy and grazing on intertidal assemblages on the Swedish west coast. *Journal of Experimental Marine Biology and Ecology*, 302, 35–49.

Chacón-Labella, J., Hinojo-Hinojo, C., Bohner, T., Castorena, M., Violle, C., Vandvik, V., *et al.* (2023). How to improve scaling from traits to ecosystem processes. *Trends in Ecology & Evolution*, 38, 228–237.

Chalmandrier, L., Hartig, F., Laughlin, D.C., Lischke, H., Pichler, M., Stouffer, D.B., *et al.* (2021). Linking functional traits and demography to model species-rich communities. *Nat Commun*, 12, 2724.

Chalmandrier, L., Stouffer, D.B., Purcell, A.S.T., Lee, W.G., Tanentzap, A.J. & Laughlin, D.C. (2022). Predictions of biodiversity are improved by integrating trait-based competition with abiotic filtering. *Ecology Letters*, 25, 1277–1289.

Chalmers, A. F. (1999). What is this thing called Science? Open University Press.

Chaves, F.A. & Smith, M.D. (2021). Resources do not limit compensatory response of a tallgrass prairie plant community to the loss of a dominant species. *Journal of Ecology*, 109, 3617–3633.

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology, Evolution, and Systematics*, 31, 343–366.

Chesson, P., Pacala, S. & Neuhauser, C. (2013). Environmental Niches and Ecosystem Functioning. In: *The Functional Consequences of Biodiversity* (eds. Tilman, D., Kinzig, A.P. & Pacala, S.). Princeton University Press, pp. 213–245.

Chisholm, R.A. & Dutta Gupta, T. (2023). A critical assessment of the biodiversity–productivity relationship in forests and implications for conservation. *Oecologia*, 201, 887–900.

Chisholm, R.A., Muller-Landau, H.C., Abdul Rahman, K., Bebber, D.P., Bin, Y., Bohlman, S.A., *et al.* (2013). Scale-dependent relationships between tree species richness and ecosystem function in forests. *Journal of Ecology*, 101, 1214–1224.

Connell, J (1972). Community Interactions on Marine Rocky Intertidal Shores. *Annual Reviews of Ecology and Systematics*, 3, 169-192.

Connolly, J., Bell, T., Bolger, T., Brophy, C., Carnus, T., Finn, J.A., *et al.* (2013). An improved model to predict the effects of changing biodiversity levels on ecosystem function. *Journal of Ecology*, 101, 344–355.

Cornell, H.V. & Harrison, S.P. (2014). What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, 45, 45–67.

Cramer, M.D., Hawkins, H.-J. & Verboom, G.A. (2009). The importance of nutritional regulation of plant water flux. *Oecologia*, 161, 15–24.

Cramer, M.D., Power, S.C., Belev, A., Gillson, L., Bond, W.J., Hoffman, M.T., *et al.* (2019). Are forest-shrubland mosaics of the Cape Floristic Region an example of alternate stable states? *Ecography*, 42, 717–729.

Craven, D., van der Sande, M.T., Meyer, C., Gerstner, K., Bennett, J.M., Giling, D.P., *et al.* (2020). A cross-scale assessment of productivity–diversity relationships. *Global Ecology and Biogeography*, 29, 1940–1955.

Currie, D.J. (2019). Where Newton might have taken ecology. *Global Ecology and Biogeography*, 28, 18–27.

De Angelis, D.L. (1975). Stability and Connectance in Food Web Models. *Ecology*, 56, 238–243.

De Laender, F., Rohr, J.R., Ashauer, R., Baird, D.J., Berger, U., Eisenhauer, N., *et al.* (2016). Reintroducing Environmental Change Drivers in Biodiversity–Ecosystem Functioning Research. *Trends in Ecology & Evolution*, 31, 905–915.

Dee, L.E., Ferraro, P.J., Severen, C.N., Kimmel, K.A., Borer, E.T., Byrnes, J.E.K., *et al.* (2023). Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nature Communications*, 14, 2607.

deLaplante, K. & Picasso, V. (2011). The Biodiversity-Ecosystem Function Debate in Ecology. In: *Philosophy of Ecology*, Handbook of the Philosophy of Science (eds. deLaplante, K., Brown, B. & Peacock, K.A.). North-Holland, Amsterdam, pp. 169–200.

Diaz, R.M. & Ernest, S.K.M. (2022). Maintenance of community function through compensation breaks down over time in a desert rodent community. *Ecology*, 103, e3709.

Díaz, S., Lavorel, S., de Bello, F., Quétier, F., Grigulis, K. & Robson, T.M. (2007). Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences*, 104, 20684–20689.

Díaz, S., Purvis, A., Cornelissen, J.H.C., Mace, G.M., Donoghue, M.J., Ewers, R.M., *et al.* (2013). Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and Evolution*, 3, 2958–2975.

Digitale, J.C., Martin, J.N. & Glymour, M.M. (2022). Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142, 264–267.

Díaz, S. & Cabido, M. (2001). Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution*, 16, 646–655.

Díaz, S., Symstad, A.J., Stuart Chapin, F., Wardle, D.A. & Huenneke, L.F. (2003). Functional diversity revealed by removal experiments. *Trends in Ecology & Evolution*, 18, 140–146.

Doak, D.F., Bigger, D., Harding, E. K., Marvier, M. A., O'Malley, R. E. & Thomson, D. (1998). The statistical inevitability of stability-diversity relationships in community ecology. *The American Naturalist*, 151, 264–276.

Donohue, I., Petchey, O.L., Montoya, J.M., Jackson, A.L., McNally, L., Viana, M., *et al.* (2013). On the dimensionality of ecological stability. *Ecology Letters*, 16, 421–429.

Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., *et al.* (2014). Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. *Science*, 344, 296–299.

Doughty, C.E. (2017). Herbivores increase the global availability of nutrients over millions of years. *Nature Ecology and Evolution*, 1, 1820–1827.

Duffy, J.E. (2009). Why biodiversity is important to the functioning of real-world ecosystems. *Frontiers in Ecology and the Environment*, 7, 437–444.

Duffy, J.E., Godwin, C.M. & Cardinale, B.J. (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature*, 549, 261–264.

Dwyer, J.M., Hobbs, R.J. & Mayfield, M.M. (2014). Specific leaf area responses to environmental gradients through space and time. *Ecology*, 95, 399–410.

Eisenhauer, N., Barnes, A.D., Cesarz, S., Craven, D., Ferlian, O., Gottschall, F., *et al.* (2016). Biodiversity-ecosystem function experiments reveal the mechanisms underlying the consequences of biodiversity change in real world ecosystems. *Journal of Vegetation Science*, 27, 1061–1070.

Elahi, R., O'Connor, M.I., Byrnes, J.E.K., Dunic, J., Eriksson, B.K., Hensel, M.J.S., *et al.* (2015). Recent Trends in Local-Scale Marine Biodiversity Reflect Community Structure and Human Impacts. *Current Biology*, 25, 1938–1943.

Elton, C.E. (1957). Stability arguments from Elton's (1957) book on invasion biology.

Engelhardt, K., Vetter, J., Wöhrmann-Zipf, F., Dietzmann, A., Proll, F., Reifert, H., *et al.* (2023). Contact-free modulation of stony coral productivity by sessile reef organisms (preprint). *Research Square*.

February, E.C., Higgins, S.I., Bond, W.J. & Swemmer, L. (2013). Influence of competition and rainfall manipulation on the growth responses of savanna trees and grasses. *Ecology*, 94, 1155–1164.

Feng, Y., Schmid, B., Loreau, M., Forrester, D.I., Fei, S., Zhu, J., *et al.* (2022). Multispecies forest plantations outyield monocultures across a broad range of conditions. *Science*, 376, 865–868.

Firn, J., McGree, J.M., Harvey, E., Flores-Moreno, H., Schütz, M., Buckley, Y.M., *et al.* (2019). Leaf nutrients, not specific leaf area, are consistent indicators of elevated nutrient inputs. *Nature Ecology and Evolution*, 3, 400–406.

Fotis, A.T., Murphy, S.J., Ricart, R.D., Krishnadas, M., Whitacre, J., Wenzel, J.W., *et al.* (2018). Above-ground biomass is driven by mass-ratio effects and stand structural attributes in a temperate deciduous forest. *Journal of Ecology*, 106, 561–570.

Fox, J.W. (2005). Interpreting the 'selection effect' of biodiversity on ecosystem function. *Ecology Letters*, 8, 846–856.

Fox, J.W. (2006). Using the Price Equation to Partition the Effects of Biodiversity Loss on Ecosystem Function. *Ecology*, 87, 2687–2696.

Fox, J.W. (2012). When should we expect microbial phenotypic traits to predict microbial abundances? *Frontiers in Microbiology*, 3.

Frank, D.M. (2022). Science and values in the biodiversity-ecosystem function debate. *Biology & Philosophy*, 37, 7.

Fraser, L.H., Pither, J., Jentsch, A., Sternberg, M., Zobel, M., Askarizadeh, D., *et al.* (2015). Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science*, 349, 302–305.

Freitag, M., Hölzel, N., Neuenkamp, L., van der Plas, F., Manning, P., Abrahão, A., *et al.* (2023). Increasing plant species richness by seeding has marginal effects on ecosystem functioning in agricultural grasslands. *Journal of Ecology*, 111, 1968–1984.

Fukami, T., Naeem, S. & Wardle, D.A. (2001). On Similarity among Local Communities in Biodiversity Experiments. *Oikos*, 95, 340–348.

Galparsoro, I., Murillas, A., Pinarbasi, K., Sequeira, A.M.M., Stelzenmüller, V., Borja, Á., *et al.* (2020). Global stakeholder vision for ecosystem-based marine aquaculture expansion from coastal to offshore areas. *Reviews in Aquaculture*, 12, 2061–2079.

Gamfeldt, L. & Hillebrand, H. (2011). Effects of Total Resources, Resource Ratios, and Species Richness on Algal Productivity and Evenness at Both Metacommunity and Local Scales. *PLoS ONE*, 6, e21972.

Gamfeldt, L., Hillebrand, H. & Jonsson, P.R. (2008). Multiple Functions Increase the Importance of Biodiversity for Overall Ecosystem Functioning. *Ecology*, 89, 1223–1231.

Gamfeldt, L., Lefcheck, J.S., Byrnes, J.E.K., Cardinale, B.J., Duffy, J.E. & Griffin, J.N. (2015). Marine biodiversity and ecosystem functioning: what's known and what's next? *Oikos*, 124, 252–265.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., *et al.* (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, 4, 1340.

Garnier, E., Cortez, J., Billès, G., Navas, M.-L., Roumet, C., Debussche, M., *et al.* (2004). Plant Functional Markers Capture Ecosystem Properties During Secondary Succession. *Ecology*, 85, 2630–2637.

Garratt, M.P.D., Breeze, T.D., Jenner, N., Polce, C., Biesmeijer, J.C. & Potts, S.G. (2014). Avoiding a bad apple: Insect pollination enhances fruit quality and economic value. *Agriculture, Ecosystems & Environment*, 184, 34–40.

Genung, M.A., Fox, J. & Winfree, R. (2020). Species loss drives ecosystem function in experiments, but in nature the importance of species loss depends on dominance. *Global Ecology and Biogeography*, 29, 1531–1541.

Genung, M.A., Reilly, J., Williams, N.M., Buderi, A., Gardner, J. & Winfree, R. (2023). Rare and declining bee species are key to consistent pollination of wildflowers and crops across large spatial scales. *Ecology*, 104, e3899.

Germain, R.M., Mayfield, M.M. & Gilbert, B. (2018). The 'filtering' metaphor revisited: competition and environment jointly structure invasibility and coexistence. *Biology Letters*, 14, 20180460.

Germain, R.M., Strauss, S.Y. & Gilbert, B. (2017). Experimental dispersal reveals characteristic scales of biodiversity in a natural landscape. *Proceedings of the National Academy of Sciences.*, 114, 4447–4452.

Givnish, T.J. (1994). Does diversity beget stability? *Nature*, 471, 113–114.

Gonzalez, A., Cardinale, B.J., Allington, G.R.H., Byrnes, J., Arthur Endsley, K., Brown, D.G., *et al.* (2016). Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology*, 97, 1949–1960.

Gonzalez, A., Chase, J.M. & O'Connor, M.I. (2023). A framework for the detection and attribution of biodiversity change. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 378, 20220182.

Gonzalez, A., Germain, R.M., Srivastava, D.S., Filotas, E., Dee, L.E., Gravel, D., *et al.* (2020). Scaling-up biodiversity-ecosystem functioning research. *Ecology Letters*, 23, 757–776.

Gonzalez, A. & Loreau, M. (2009). The Causes and Consequences of Compensatory Dynamics in Ecological Communities. *Annual Review of Ecology, Evolution, and Systematics*, 40, 393–414.

Grace, J.B., Anderson, T.M., Seabloom, E.W., Borer, E.T., Adler, P.B., Harpole, W.S., *et al.* (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, 529, 390–393.

Grace, J.B., Loreau, M. & Schmid, B. (2022). A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations: comment. *Ecology*, 103, e03378.

Griffin, J.N., Byrnes, J.E.K. & Cardinale, B.J. (2013). Effects of predator richness on prey suppression: a meta-analysis. *Ecology*, 94, 2180–2187.

Griffin, J.N., Jenkins, S.R., Gamfeldt, L., Jones, D., Hawkins, S.J. & Thompson, R.C. (2009). Spatial heterogeneity increases the importance of species richness for an ecosystem process. *Oikos*, 118, 1335–1342.

Grime, J.P. (1973). Competitive Exclusion in Herbaceous Vegetation. *Nature*, 242, 344–347.

Grime, J.P. (1997). Biodiversity and Ecosystem Function: The Debate Deepens. *Science*, 277, 1260–1261.

Hagan, J.G., Henn, J.J. & Osterman, W.H.A. (2023). Plant traits alone are good predictors of ecosystem properties when used carefully. *Nature Ecology and Evolution*, 7, 332–334.

Harpole, W.S., Sullivan, L.L., Lind, E.M., Firn, J., Adler, P.B., Borer, E.T., *et al.* (2016). Addition of multiple limiting resources reduces grassland diversity. *Nature*, 537, 93–96.

Harpole, W.S. & Tilman, D. (2007). Grassland species loss resulting from reduced niche dimension. *Nature*, 446, 791–793.

Hart, S.P., Usinowicz, J. & Levine, J.M. (2017). The spatial scales of species coexistence. *Nature Ecology and Evolution*, 1, 1066–1073.

Hawkins, S. J., and R. J. Hartnoll. 1985. Factors determining the upper limits of intertidal canopy-forming algae. Marine Ecology-Progress Series 20:265-271.1985.

Hautier, Y., Niklaus, P.A. & Hector, A. (2009). Competition for Light Causes Plant Biodiversity Loss After Eutrophication. *Science*, 324, 636–638.

Hector, A. (1998). The Effect of Diversity on Productivity: Detecting the Role of Species Complementarity. *Oikos*, 82, 597–599.

Hector, A. & Bagchi, R. (2007). Biodiversity and ecosystem multifunctionality. *Nature*, 448, 188–190.

Hector, A. & Hooper, R. (2002). Darwin and the First Ecological Experiment. *Science*, 295, 639–640.

Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M.C., Diemer, M., Dimitrakopoulos, P.G., *et al.* (1999). Plant Diversity and Productivity Experiments in European Grasslands. *Science*, 286, 1123–1127.

Hector, A. & Wilby, A. (2009). Biodiversity and ecosystem functioning. In: *Princeton Guide to Ecology*. Princeton University Press, Princeton, New Jersey.

Hillebrand, H., Bennett, D.M. & Cadotte, M.W. (2008). Consequences of Dominance: A Review of Evenness Effects on Local and Regional Ecosystem Processes. *Ecology*, 89, 1510–1520.

Hillebrand, H., Blasius, B., Borer, E.T., Chase, J.M., Downing, J.A., Eriksson, B.K., *et al.* (2018). Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology*, 55, 169–184.

Hillebrand, H., Dürselen, C.-D., Kirschtel, D., Pollingher, U. & Zohary, T. (1999). Biovolume Calculation for Pelagic and Benthic Microalgae. *Journal of Phycology*, 35, 403–424.

Hong, P., Schmid, B., De Laender, F., Eisenhauer, N., Zhang, X., Chen, H., *et al.* (2022). Biodiversity promotes ecosystem functioning despite environmental change. *Ecology Letters*, 25, 555–569.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., *et al.* (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, 75, 3–35.

Houlahan, J.E., Currie, D.J., Cottenie, K., Cumming, G.S., Ernest, S.K.M., Findlay, C.S., *et al.* (2007). Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences*, 104, 3273–3277.

Hu, Z., Chen, H.Y.H., Yue, C., Gong, X.Y., Shao, J., Zhou, G., *et al.* (2020). Traits mediate drought effects on wood carbon fluxes. *Global Change Biology*, 26, 3429–3442.

Hubbell, S.P. (2011). The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). In: *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press.

Huston, M. (1979). A General Hypothesis of Species Diversity. *The American Naturalist*, 113, 81–101.

Huston, M.A. (1997). Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia*, 110, 449–460.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., *et al.* (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, 477, 199–202.

Isbell, F., Cowles, J., Dee, L.E., Loreau, M., Reich, P.B., Gonzalez, A., *et al.* (2018). Quantifying effects of biodiversity on ecosystem functioning across times and places. *Ecology Letters*, 21, 763–778.

Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Díaz, S., Hector, A., *et al.* (2017). Linking the influence and dependence of people on biodiversity across scales. *Nature*, 546, 65–72.

Jacob, A., Hertel, D. & Leuschner, C. (2013). On the significance of belowground overyielding in temperate mixed forests: separating species identity and species diversity effects. *Oikos*, 122, 463–473.

Jochum, M., Fischer, M., Isbell, F., Roscher, C., van der Plas, F., Boch, S., *et al.* (2020). The results of biodiversity–ecosystem functioning experiments are realistic. *Nature Ecology and Evolution*, 4, 1485–1494.

Jones, K.R., Klein, C.J., Halpern, B.S., Venter, O., Grantham, H., Kuempel, C.D., *et al.* (2018). The Location and Protection Status of Earth's Diminishing Marine Wilderness. *Current Biology*, 28, 2506-2512.e3.

Jonsson, M., Bengtsson, J., Gamfeldt, L., Moen, J. & Snäll, T. (2019). Levels of forest ecosystem services depend on specific mixtures of commercial tree species. *Nature Plants*, 5, 141–147.

Kahmen, A., Perner, J., Audorff, V., Weisser, W. & Buchmann, N. (2005). Effects of plant diversity, community composition and environmental parameters on productivity in montane European grasslands. *Oecologia*, 142, 606–615.

Kardol, P., Fanin, N. & Wardle, D.A. (2018). Long-term effects of species loss on community properties across contrasting ecosystems. *Nature*, 557, 710–713.

Keddy, P.A. & Laughlin, D.C. (2021). *A Framework for Community Ecology: Species Pools, Filters and Traits*. Cambridge University Press.

Kelty, M.J. (2006). The role of species mixtures in plantation forestry. *Forest Ecology and Management*, Improving Productivity in Mixed-Species Plantations, 233, 195–204.

Kennedy, T.A., Naeem, S., Howe, K.M., Knops, J.M.H., Tilman, D. & Reich, P. (2002). Biodiversity as a barrier to ecological invasion. *Nature*, 417, 636–638.

Kimmel, K., Dee, L.E., Avolio, M.L. & Ferraro, P.J. (2021). Causal assumptions and causal inference in ecological experiments. *Trends in Ecology & Evolution*, 36, 1141–1152.

Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S. & Levine, J.M. (2015a). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, 29, 592–599.

Kraft, N.J.B., Godoy, O. & Levine, J.M. (2015b). Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, 112, 797–802.

Krix, D.W. & Murray, B.R. (2022). A Predictive Model of Leaf Flammability Using Leaf Traits and Radiant Heat Flux for Plants of Fire-Prone Dry Sclerophyll Forest. *Forests*, 13, 152.

Ladouceur, E., Stanley Harpole, W., Blowes, S.A., Roscher, C., Auge, H., Seabloom, E.W., *et al.* (2020). Reducing dispersal limitation via seed addition increases species richness but not above-ground biomass. *Ecology Letters*, 23, 1442–1450.

Lambers, J.H.R., Harpole, W.S., Tilman, D., Knops, J. & Reich, P.B. (2004). Mechanisms responsible for the positive diversity-productivity relationship in Minnesota grasslands. *Ecology Letters*, 7, 661–668.

Lamy, T., Wang, S., Renard, D., Lafferty, K.D., Reed, D.C. & Miller, R.J. (2019). Species insurance trumps spatial insurance in stabilizing biomass of a marine macroalgal metacommunity. *Ecology*, 100, e02719.

Lanuza, J.B., Bartomeus, I. & Godoy, O. (2018). Opposing effects of floral visitors and soil conditions on the determinants of competitive outcomes maintain species diversity in heterogeneous landscapes. *Ecology Letters*, 21, 865–874.

Lapiedra, O., Morales, N., Yang, L., Fernández-Bellon, D., Michaelides, S., Giery, S., *et al.* (2023). Predator-driven behavioral shifts in a common lizard shape resource-flow dynamics from marine to terrestrial ecosystems (preprint). *Authorea*.

Laughlin, D.C. (2014). Applying trait-based models to achieve functional targets for theory-driven ecological restoration. *Ecology Letters*, 17, 771–784.

Laughlin, D.C., Joshi, C., van Bodegom, P.M., Bastow, Z.A. & Fulé, P.Z. (2012). A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters*, 15, 1291–1299.

Laughlin, D.C., Strahan, R.T., Adler, P.B. & Moore, M.M. (2018). Survival rates indicate that correlations between community-weighted mean traits and environments can be unreliable estimates of the adaptive value of traits. *Ecology Letters*, 21, 411–421.

Le Provost, G., Schenk, N.V., Penone, C., Thiele, J., Westphal, C., Allan, E., *et al.* (2022). The supply of multiple ecosystem services requires biodiversity across spatial scales. *Nature Ecology and Evolution*, 1–14.

Lefcheck, J.S., Byrnes, J.E.K., Isbell, F., Gamfeldt, L., Griffin, J.N., Eisenhauer, N., *et al.* (2015). Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nature Communications*, 6, 6936.

Lefcheck, J.S., Edgar, G.J., Stuart-Smith, R.D., Bates, A.E., Waldock, C., Brandl, S.J., *et al.* (2021). Species richness and identity both determine the biomass of global reef fish communities. *Nature Communications*, 12, 6875.

Lehman, C.L. & Tilman, D. (2000). Biodiversity, Stability, and Productivity in Competitive Communities. *The American Naturalist*, 156, 534–552.

Leibold, M.A., Chase, J.M. & Ernest, S.K.M. (2017). Community assembly and the functioning of ecosystems: how metacommunity processes alter ecosystems attributes. *Ecology*, 98, 909–919.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.

Lepš, J. (2004). What do the biodiversity experiments tell us about consequences of plant species loss in the real world? *Basic and Applied Ecology*, 5, 529–534.

Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73, 1943–1967.

Levine, J.M. (2000). Species Diversity and Biological Invasions: Relating Local Process to Community Pattern. *Science*, 288, 852–854.

Levine, J.M., Adler, P.B. & Yelenik, S.G. (2004). A meta-analysis of biotic resistance to exotic plant invasions. *Ecology Letters*, 7, 975–989.

Li, Y., Li, Q., Xu, L., Li, M., Chen, Z., Song, Z., *et al.* (2021). Plant community traits can explain variation in productivity of selective logging forests after different restoration times. *Ecological Indicators*, 131, 108181.

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., *et al.* (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354, aaf8957.

Lisner, A., Konečná, M., Blažek, P. & Lepš, J. (2023). Community biomass is driven by dominants and their characteristics – The insight from a field biodiversity experiment with realistic species loss scenario. *Journal of Ecology*, 111, 240–250.

Liu, C.L.C., Kuchma, O. & Krutovsky, K.V. (2018). Mixed-species versus monocultures in plantation forestry: Development, benefits, ecosystem services and perspectives for the future. *Global Ecology and Conservation*, 15, e00419.

Liu, J., Zhang, X., Song, F., Zhou, S., Cadotte, M.W. & Bradshaw, C.J.A. (2015). Explaining maximum variation in productivity requires phylogenetic diversity and single functional traits. *Ecology*, 96, 176–183.

Loreau, M. (1998a). Biodiversity and ecosystem functioning: A mechanistic model. *Proceedings of the National Academy of Sciences*, 95, 5632–5636.

Loreau, M. (1998b). Separating Sampling and Other Effects in Biodiversity Experiments. *Oikos*, 82, 600–602.

Loreau, M. (2000). Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos*, 91, 3–17.

Loreau, M. (2004). Does functional redundancy exist? *Oikos*, 104, 606–611.

Loreau, M., Barbier, M., Filotas, E., Gravel, D., Isbell, F., Miller, S.J., *et al.* (2021). Biodiversity as insurance: from concept to measurement and application. *Biological Reviews*, 96, 2333–2354.

Loreau, M. & Hector, A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature*, 412, 72–76.

Loreau, M., Mouquet, N. & Gonzalez, A. (2003). Biodiversity as spatial insurance in heterogeneous landscapes. *Proceedings of the National Academy of Sciences*, 100, 12765–12770.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., *et al.* (2001). Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science*, 294, 804–808.

Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11, 995–1003.

MacArthur, R. (1955). Fluctuations of Animal Populations and a Measure of Community Stability. *Ecology*, 36, 533.

Manning, P., van der Plas, F., Soliveres, S., Allan, E., Maestre, F.T., Mace, G., *et al.* (2018). Redefining ecosystem multifunctionality. *Nature Ecology and Evolution*, 2, 427–436.

Mao, Z., van der Plas, F., Corrales, A., Anderson-Teixeira, K.J., Bourg, N.A., Chu, C., *et al.* (2023). Scale-dependent diversity–biomass relationships can be driven by tree mycorrhizal association and soil fertility. *Ecological Monographs*, 93, e1568.

Marino, G., Aqil, M. & Shipley, B. (2010). The leaf economics spectrum and the prediction of photosynthetic light–response curves. *Functional Ecology*, 24, 263–272.

Maureaud, A., Hodapp, D., van Denderen, P.D., Hillebrand, H., Gislason, H., Spaanheden Dencker, T., *et al.* (2019). Biodiversity–ecosystem functioning relationships in fish communities: biomass is related to evenness and the environment, not to species richness. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20191189.

May, R.M. (1972). Will a large complex system be stable? *Nature*, 238, 413–414.

Mayfield, M.M. & Levine, J.M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093.

McGill, B.J. (2012). Trees are rarely most abundant where they grow best. *Journal of Plant Ecology*, 5, 46–51.

McGill, B.J., Dornelas, M., Gotelli, N.J. & Magurran, A.E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology & Evolution*, 30, 104–113.

McNaughton, S.J. (1977). Diversity and Stability of Ecological Communities: A Comment on the Role of Empiricism in Ecology. *The American Naturalist*, 111, 515–525.

Mellinger, M.V. & McNaughton, S.J. (1975). Structure and Function of Successional Vascular Plant Communities in Central New York. *Ecological Monographs*, 45, 161–182.

Midolo, G., Wellstein, C. & Faurby, S. (2021). Individual fitness is decoupled from coarse-scale probability of occurrence in North American trees. *Ecography*, 44, 789–801.

Morgan Ernest, S.K. & Brown, J.H. (2001). Delayed Compensation for Missing Keystone Species by Colonization. *Science*, 292, 101–104.

Morin, P.J. (2011). *Community Ecology*. Wiley-Blackwell, West Sussex.

Mouquet, N. & Loreau, M. (2002). Coexistence in Metacommunities: The Regional Similarity Hypothesis. *The American Naturalist*, 159, 420–426.

Naeem, S., Iii, F.S.C., Costanza, R., Ehrlich, P.R., Golley, F.B., Hooper, D.U., *et al.* (1999). Biodiversity and Ecosystem Functioning: Maintaining Natural Life Support Processes.

Naeem, S. & Li, S. (1997). Biodiversity enhances ecosystem reliability. *Nature*, 390, 507–509.

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

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., *et al.* (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45–50.

Nichols, J.D., Bristow, M. & Vanclay, J.K. (2006). Mixed-species plantations: Prospects and challenges. *Forest Ecology and Management*, 233, 383–390.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., *et al.* (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.

Pan, Q., Tian, D., Naeem, S., Auerswald, K., Elser, J.J., Bai, Y., *et al.* (2016). Effects of functional diversity loss on ecosystem functions are influenced by compensation. *Ecology*, 97, 2293–2302.

Pearl, J. & Mackenzie, D. (2018). *The Book of Why*. Penguin Books, Great Britain.

Petchey, O.L. (2003). Integrating methods that investigate how complementarity influences ecosystem functioning. *Oikos*, 101, 323–330.

Petchey, O.L. & Gaston, K.J. (2002). Functional diversity (FD), species richness and community composition. *Ecol Letters*, 5, 402–411.

Petchey, O.L. & Gaston, K.J. (2006). Functional diversity: back to basics and looking forward. *Ecology Letters*, 9, 741–758.

Petchey, O.L., Hector, A. & Gaston, K.J. (2004). How Do Different Measures of Functional Diversity Perform? *Ecology*, 85, 847–857.

Petermann, J.S., Fergus, A.J.F., Roscher, C., Turnbull, L.A., Weigelt, A. & Schmid, B. (2010). Biology, chance, or history? The predictable reassembly of temperate grassland communities. *Ecology*, 91, 408–421.

Pillai, P. & Gouhier, T.C. (2018). On the use and abuse of Price equation concepts in ecology (preprint). *ArXiv*.

Pillai, P. & Gouhier, T.C. (2020). On the correct mathematical derivation and ecological application of unbiased estimators in biodiversity research. *Methods in Ecology and Evolution*, 11, 1550–1558.

Pimm, S.L. (1980). Food Web Design and the Effect of Species Deletion. *Oikos*, 35, 139.

Pimm, S.L. (1984). The complexity and stability of ecosystems. *Nature*, 307, 321–326.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., *et al.* (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344, 1246752.

van der Plas, F. (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews*, 94, 1220–1245.

van der Plas, F., Schröder-Georgi, T., Weigelt, A., Barry, K., Meyer, S., Alzate, A., *et al.* (2020). Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. *Nature Ecology and Evolution*, 4, 1602–1611.

Poisot, T., Mouquet, N. & Gravel, D. (2013). Trophic complementarity drives the biodiversity–ecosystem functioning relationship in food webs. *Ecology Letters*, 16, 853–861.

Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., *et al.* (1996). Challenges in the Quest for Keystones. *BioScience*, 46, 609–620.

Pulliam, H.R. (1988). Sources, Sinks, and Population Regulation. *The American Naturalist*, 132, 652–661.

Qiu, J. & Cardinale, B.J. (2020). Scaling up biodiversity–ecosystem function relationships across space and over time. *Ecology*, 101, e03166.

Raffard, A., Santoul, F., Cucherousset, J. & Blanchet, S. (2019). The community and ecosystem consequences of intraspecific diversity: a meta-analysis. *Biological Reviews*, 94, 648–661.

Ratcliffe, S., Wirth, C., Jucker, T., van der Plas, F., Scherer-Lorenzen, M., Verheyen, K., *et al.* (2017). Biodiversity and ecosystem functioning relations in European forests depend on environmental context. *Ecology Letters*, 20, 1414–1426.

Ratzke, C., Barrere, J. & Gore, J. (2020). Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology and Evolution*, 4, 376–383.

Reich, P.B., Tilman, D., Isbell, F., Mueller, K., Hobbie, S.E., Flynn, D.F.B., *et al.* (2012). Impacts of Biodiversity Loss Escalate Through Time as Redundancy Fades. *Science*, 336, 589–592.

Rinella, M.J., Strong, D.J. & Vermeire, L.T. (2020). Omitted variable bias in studies of plant interactions. *Ecology*, 101, e03020.

Roger, F. (2017). *Biodiversity and Ecosystem Functioning - What Diversity? Which Functioning?*. PhD Thesis.

Rosenzweig, M.L. (1981). A Theory of Habitat Selection. *Ecology*, 62, 327–335.

Rosenzweig, M.L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.

Sala, O.E., Parton, W.J., Joyce, L.A. & Lauenroth, W.K. (1988). Primary Production of the Central Grassland Region of the United States. *Ecology*, 69, 40–45.

Sandau, N., Fabian, Y., Bruggisser, O.T., Rohr, R.P., Naisbit, R.E., Kehrli, P., *et al.* (2017). The relative contributions of species richness and species composition to ecosystem functioning. *Oikos*, 126, 782–791.

Scheffer, M. & van Nes, E.H. (2006). Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences.*, 103, 6230–6235.

Scheper, J., Badenhausser, I., Kantelhardt, J., Kirchweger, S., Bartomeus, I., Bretagnolle, V., *et al.* (2023). Biodiversity and pollination benefits trade off against profit in an intensive farming system. *Proceedings of the National Academy of Sciences*, 120, e2212124120.

Schiettekatte, N.M.D., Brandl, S.J., Casey, J.M., Graham, N.A.J., Barneche, D.R., Burkepile, D.E., *et al.* (2022). Biological trade-offs underpin coral reef ecosystem functioning. *Nature Ecology and Evolution*, 6, 701–708.

Schmid, B. (2002). The species richness–productivity controversy. *Trends in Ecology & Evolution*, 17, 113–114.

Schmid, B., Hector, A., Saha, P. & Loreau, M. (2008). Biodiversity effects and transgressive overyielding. *Journal of Plant Ecology*, 1, 95–102.

Schmid, B., Schmitz, M., Rzanny, M., Scherer-Lorenzen, M., Mwangi, P.N., Weisser, W.W., *et al.* (2022). Removing subordinate species in a biodiversity experiment to mimic observational field studies. *Grassland Research*, 1, 53–62.

Schoolmaster Jr., D.R., Zirbel, C.R. & Cronin, J.P. (2020). A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations. *Ecology*, 101, e03070.

Schoolmaster Jr., D.R., Zirbel, C.R. & Cronin, J.P. (2022). A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations: Reply. *Ecology*, 103, e03593.

Schulze, Ernst-Detlef & Mooney, Harold A. (1993). *Biodiversity and Ecosystem Functioning*. Springer Berlin, Heidelberg.

Schwartz, M.W., Brigham, C.A., Hoeksema, J.D., Lyons, K.G., Mills, M.H. & van Mantgem, P.J. (2000). Linking biodiversity to ecosystem function: implications for conservation ecology. *Oecologia*, 122, 297–305.

Sears, A.L.W. & Chesson, P. (2007). New Methods for Quantifying the Spatial Storage Effect: An Illustration with Desert Annuals. *Ecology*, 88, 2240–2247.

Shanafelt, D.W., Dieckmann, U., Jonas, M., Franklin, O., Loreau, M. & Perrings, C. (2015). Biodiversity, productivity, and the spatial insurance hypothesis revisited. *Journal of Theoretical Biology*, 380, 426–435.

Silvertown, J., Poulton, P., Johnston, E., Edwards, G., Heard, M. & Biss, P.M. (2006). The Park Grass Experiment 1856–2006: its contribution to ecology. *Journal of Ecology*, 94, 801–814.

Simpson, D.T., Weinman, L.R., Genung, M.A., Roswell, M., MacLeod, M. & Winfree, R. (2022). Many bee species, including rare species, are important for function of entire plant–pollinator networks. *Proceedings of the Royal Society B: Biological Sciences*, 289, 20212689.

Smart, S.M., Glanville, H.C., Blanes, M. del C., Mercado, L.M., Emmett, B.A., Jones, D.L., *et al.* (2017). Leaf dry matter content is better at predicting above-ground net primary production than specific leaf area. *Functional Ecology*, 31, 1336–1344.

Smith, M.D. & Knapp, A.K. (2003). Dominant species maintain ecosystem function with non-random species loss. *Ecology Letters*, 6, 509–517.

Smith, M.D., Koerner, S.E., Knapp, A.K., Avolio, M.L., Chaves, F.A., Denton, E.M., *et al.* (2020). Mass ratio effects underlie ecosystem responses to environmental change. *Journal of Ecology*, 108, 855–864.

Solan, M., Cardinale, B.J., Downing, A.L., Engelhardt, K.A.M., Ruesink, J.L. & Srivastava, D.S. (2004). Extinction and Ecosystem Function in the Marine Benthos. *Science*, 306, 1177–1180.

Solan, M., Godbold, J.A., Symstad, A.J., Flynn, D.F.B. & Bunker, D.E. (2009). Biodiversity-ecosystem function research and biodiversity futures: early bird catches the worm or a day late and a dollar short? In: *Biodiversity, ecosystem functioning, and human wellbeing: an ecological and economic perspective*. Oxford University Press, Oxford ; New York, pp. 30–45.

Spaak, J.W., Baert, J.M., Baird, D.J., Eisenhauer, N., Maltby, L., Pomati, F., *et al.* (2017). Shifts of community composition and population density substantially affect ecosystem function despite invariant richness. *Ecology Letters*, 20, 1315–1324.

Srivastava, D.S., Cadotte, M.W., MacDonald, A.A.M., Marushia, R.G. & Mirotchnick, N. (2012). Phylogenetic diversity and the functioning of ecosystems. *Ecology Letters*, 15, 637–648.

Srivastava, D.S. & Vellend, M. (2005). Biodiversity-Ecosystem Function Research: Is It Relevant to Conservation? *Annual Reviews of Ecology, Evolution and Systematics*, 36, 267–294.

Staples, T.L., Dwyer, J.M., Loy, X. & Mayfield, M.M. (2016). Potential mechanisms of coexistence in closely related forbs. *Oikos*, 125, 1812–1823.

Staver, A.C., Archibald, S. & Levin, S.A. (2011). The Global Extent and Determinants of Savanna and Forest as Alternative Biome States. *Science*, 334, 230–232.

Suding, K.N., Lavorel, S., Chapin Iii, F.S., Cornelissen, J.H.C., Díaz, S., Garnier, E., *et al.* (2008). Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. *Global Change Biology*, 14, 1125–1140.

Thompson, P.L., Guzman, L.M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., *et al.* (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecology Letters*, 23, 1314–1329.

Thompson, P.L., Isbell, F., Loreau, M., O'Connor, M.I. & Gonzalez, A. (2018). The strength of the biodiversity–ecosystem function relationship depends on spatial scale. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20180038.

Thompson, P.L., Kéfi, S., Zelnik, Y.R., Dee, L.E., Wang, S., de Mazancourt, C., *et al.* (2021). Scaling up biodiversity–ecosystem functioning relationships: the role of environmental heterogeneity in space and time. *Proceedings of the Royal Society B: Biological Sciences*, 288, 20202779.

Thomsen, M.S., Garcia, C., Bolam, S.G., Parker, R., Godbold, J.A. & Solan, M. (2017). Consequences of biodiversity loss diverge from expectation due to post-extinction compensatory responses. *Scientific Reports*, 7, 43695.

Thomsen, M.S., Godbold, J.A., Garcia, C., Bolam, S.G., Parker, R. & Solan, M. (2019). Compensatory responses can alter the form of the biodiversity–function relation curve. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20190287.

Tilman, D. (1997). Distinguishing between the Effects of Species Diversity and Species Composition. *Oikos*, 80, 185–185.

Tilman, D. (1999). The ecological consequences of changes in biodiversity: A search for general principles. *Ecology*, 80, 1455–1474.

Tilman, D. & Downing, J.A. (1994). Biodiversity and stability in grasslands. *Nature*, 367, 363–365.

Tilman, D., Isbell, F. & Cowles, J.M. (2014). Biodiversity and Ecosystem Functioning. *Annual Reviews of Ecology, Evolution and Systematics*, 45, 471–493.

Tilman, D., Lehman, C.L. & Bristow, C.E. (1998). Diversity-Stability Relationships: Statistical Inevitability or Ecological Consequence? *The American Naturalist*, 151, 277–282.

Tilman, D., Lehman, C.L. & Thomson, K.T. (1997). Plant diversity and ecosystem productivity: Theoretical considerations. *Proceedings of the National Academy of Sciences,* 94, 1857–1861.

Tilman, D., Reich, P.B., Knops, J., Wedin, D., Mielke, T. & Lehman, C. (2001). Diversity and Productivity in a Long-Term Grassland Experiment. *Science*, 294, 843–845.

Tilman, D., Wedin, D. & Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature*, 379, 718–720.

Towers, I.R., Bowler, C.H., Mayfield, M.M. & Dwyer, J.M. (2020). Requirements for the spatial storage effect are weakly evident for common species in natural annual plant assemblages. *Ecology*, 101, e03185.

Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A., *et al.* (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92, 698–715.

Turnbull, L.A., Levine, J.M., Loreau, M. & Hector, A. (2013). Coexistence, niches and biodiversity effects on ecosystem functioning. *Ecology Letters*, 16, 116–127.

VanderWeele, T.J. & Hernan, M.A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference*, 1, 1–20.

Vasseur, D.A., Fox, J.W., Gonzalez, A., Adrian, R., Beisner, B.E., Helmus, M.R., *et al.* (2014). Synchronous dynamics of zooplankton competitors prevail in temperate lake ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20140633.

Vellend, M. (2016). *The Theory of Ecological Communities (MPB-57)*. Princeton University Press.

Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D., *et al.* (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences.*, 110, 19456–19459.

Venail, P., Gross, K., Oakley, T.H., Narwani, A., Allan, E., Flombaum, P., *et al.* (2015). Species richness, but not phylogenetic diversity, influences community biomass production and temporal stability in a re-examination of 16 grassland biodiversity studies. *Functional Ecology*, 29, 615–626.

Venail, P.A., Maclean, R.C., Meynard, C.N. & Mouquet, N. (2010). Dispersal scales up the biodiversity–productivity relationship in an experimental source-sink metacommunity. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2339–2345.

Vile, D., Shipley, B. & Garnier, E. (2006). Ecosystem productivity can be predicted from potential relative growth rate and species abundance. *Ecology Letters*, 9, 1061–1067.

Vitousek, P.M., Walker, L.R., Whiteaker, L.D., Mueller-Dombois, D. & Matson, P.A. (1987). Biological Invasion by *Myrica faya* Alters Ecosystem Development in Hawaii. *Science*, 238, 802–804.

Wacker, L., Baudois, O., Eichenberger-Glinz, S. & Schmid, B. (2008). Environmental heterogeneity increases complementarity in experimental grassland communities. *Basic and Applied Ecology*, 9, 467–474.

Wagg, C., Roscher, C., Weigelt, A., Vogel, A., Ebeling, A., de Luca, E., *et al.* (2022). Biodiversity–stability relationships strengthen over time in a long-term grassland experiment. *Nat Commun*, 13, 7752.

Wang, S. & Brose, U. (2018). Biodiversity and ecosystem functioning in food webs: the vertical diversity hypothesis. *Ecology Letters*, 21, 9–20.

Wanis, K.N., Madenci, A.L., Dokus, M.K., Orloff, M.S., Levstik, M.A., Hernandez-Alejandro, R., *et al.* (2019). The meaning of confounding adjustment in the presence of multiple versions of treatment: an application to organ transplantation. *European Journal of Epidemiology*, 34, 225–233.

Wardle, D.A. (1999). Is "Sampling Effect" a Problem for Experiments Investigating Biodiversity-Ecosystem Function Relationships? *Oikos*, 87, 403.

Wardle, D.A. (2016). Do experiments exploring plant diversity-ecosystem functioning relationships inform how biodiversity loss impacts natural ecosystems? *Journal of Vegetation Science*, 27, 646–653.

Wardle, D.A., Bonner, K.I. & Nicholson, K.S. (1997). Biodiversity and Plant Litter: Experimental Evidence Which Does Not Support the View That Enhanced Species Richness Improves Ecosystem Function. *Oikos*, 79, 247–258.

Wardle, D.A., Huston, M.A., Grime, J.P., Berendse, F., Garnier, E., Lauenroth, W.K., *et al.* (2000). Biodiversity and Ecosystem Function: An Issue in Ecology. *Bulletin of the Ecological Society of America*, 81, 235–239.

Williams, L.J., Paquette, A., Cavender-Bares, J., Messier, C. & Reich, P.B. (2017). Spatial complementarity in tree crowns explains overyielding in species mixtures. *Nature Ecology and Evolution*, 1, 0063.

Winfree, R. & Kremen, C. (2008). Are ecosystem services stabilized by differences among species? A test using crop pollination. *Proceedings of the Royal Society B: Biological Sciences*, 276, 229–237.

Winfree, R., Reilly, J.R., Bartomeus, I., Cariveau, D.P., Williams, N.M. & Gibbs, J. (2018). Species turnover promotes the importance of bee diversity for crop pollination at regional scales. *Science*, 359, 791–793.

Winfree, R., W. Fox, J., Williams, N.M., Reilly, J.R. & Cariveau, D.P. (2015). Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecology Letters*, 18, 626–635.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., *et al.* (2004). The worldwide leaf economics spectrum. *Nature*, 428, 821–827.

Yachi, S. & Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Sciences,* 96, 1463–1468.

Yates, M.J., Anthony Verboom, G., Rebelo, A.G. & Cramer, M.D. (2010). Ecophysiological significance of leaf size variation in Proteaceae from the Cape Floristic Region. *Functional Ecology*, 24, 485–492.

Zemp, D.C., Guerrero-Ramirez, N., Brambach, F., Darras, K., Grass, I., Potapov, A., *et al.* (2023). Tree islands enhance biodiversity and functioning in oil palm landscapes. *Nature*, 618, 316–321.

Zhou, Z., Wang, C. & Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nature Communications*, 11, 3072.