Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change-A review

Citation for published version:
Boyd, PW, Collins, S, Dupont, S, Fabricius, K, Gattuso, J-P, Havenhand, J, Hutchins, DA, Riebesell, U, Rintoul, MS, Vichi, M, Biswas, H, Ciotti, A, Gao, K, Gehlen, M, Hurd, CL, Kurihara, H, McGraw, CM, Navarro, JM, Nilsson, GE, Passow, U & Pörtner, H-O 2018, 'Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change-A review' Global Change Biology. DOI: 10.1111/gcb.14102

Digital Object Identifier (DOI):
10.1111/gcb.14102

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Global Change Biology

Experimental strategies to assess the biological ramifications of multiple drivers of ocean global ocean – a review

Journal:	*Global Change Biology*
Manuscript ID	GCB-17-1506.R1
Wiley - Manuscript type:	Research Review
Date Submitted by the Author:	11-Dec-2017
Complete List of Authors:	Boyd, Philip; University of Tasmania, Institute for Marine and Antarctic Studies; Collins, Sinead; University of Edinburgh, Institute of Evolutionary Biology Dupont, Sam; University of Gothenburg, Department of Biological and Environmental Sciences Fabricius, Katherina; Australian Institute of Marine Science, Gattuso, Jean-Pierre; CNRS-UPMC, Laboratoire d'Océanographie de Villefranche Havenhand, Jonathan; Gothenburg University, Biology and Environmental Sciences-Tjärnö Hutchins, Dave; University of Southern California, Marine and Environmental Biology Riebesell, Ulf; GEOMAR, Marine Biogeochemistry, Biological Oceanography Rintoul, Max; University of Tasmania, Institute for Marine and Antarctic Studies Vichi, Marcello; University of Capetown , Oceanography Biswas, Haimanti; National Institute of Oceanography, Biogeochemistry Gao, Kunshan Gehlen, Marion; Laboratoire des sciences du climat et de l'environnement Hurd, Catriona; University of Otago, Botany; University of Tasmania, Institute for Marine and Antarctic Studies Kurihara, Haruko; University of the Ryukyus, Faculty of Science McGraw, Christina; University of Otago, Chemistry Navarro, Jorge; Universidad Austral de Chile, Nilsson, Goran; University of Oslo, Section for Physiology and Cell Biology, Department of Biosciences Passow, Uta; University of California Santa Barbara, Marine Science Institute Poertner, Hans-Otto; Alfred-Wegener-Institute, Biosciences/Integrative Ecophysiology
Keywords:	ocean, multi-drivers, experiments, design, stressors
Abstract:	Marine life is controlled by multiple physical and chemical drivers and by diverse ecological processes. Many of these oceanic properties are being altered by climate change and other anthropogenic pressures. Hence identifying the influences of multi-faceted ocean change, from local to

global scales, is a complex task. To guide policy-making and make projections of the future of the marine biosphere, it is essential to understand biological responses at physiological, evolutionary and ecological levels. Here, we contrast and compare different approaches to multiple driver experiments that aim to elucidate biological responses to a complex matrix of ocean global change. We present the benefits and the challenges of each approach with a focus on marine research, and guidelines to navigate through these different categories to help identify strategies that might best address research questions in fundamental physiology, experimental evolutionary biology, and community ecology. Our Review reveals that the field of multiple driver research is being pulled in complementary directions: the need for reductionist approaches to obtain process-oriented, mechanistic understanding, and a requirement to quantify responses to projected future scenarios of ocean change. We conclude the Review with recommendations on how best to align different experimental approaches to contribute fundamental information needed for science-based policy-formulation.

Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change – a review

Running Head: Ocean Multiple Driver Experimental Strategies

Keywords: Ocean, Multiple-drivers, Experiments, Design, Stressors

Philip.W. Boyd[1,2*], Sinead Collins[3], Sam Dupont[4], Katharina Fabricius[5], Jean-Pierre Gattuso[6], Jonathan Havenhand[7], David A. Hutchins[8], Ulf Riebesell[9], Max S. Rintoul[2], Marcello Vichi[10], Haimanti Biswas[11], Aurea Ciotti[12], Kunshan Gao[13], Marion Gehlen[14], Catriona L. Hurd[1], Haruko Kurihara[15], Christina M. McGraw[16], Jorge Navarro[17], Göran E. Nilsson[18], Uta Passow[19], and Hans-Otto Pörtner[20]

[1]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia

[2]Antarctic Climate and Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Tasmania, Australia

[3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

[4]Department of Biological & Environmental Sciences – Kristineberg, University of Gothenburg, Gothenburg, Sweden

[5]Australian Institute of Marine Science, Townsville, Australia

[6]Observatoire Océanologique, Laboratoire d'Océanographie, CNRS-UPMC, Villefranche-Sur-Mer, France

[7]Department of Marine Sciences – Tjärnö, University of Gothenburg, Gothenburg, Sweden

[8]University of Southern California, Los Angeles, CA 90089, USA

[9]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

[10]Marine Research Institute and Department of Oceanography, University of Cape Town, South Africa

[11]National Institute of Oceanography, Dona Paula, Goa, India

[12]Centro de Biologia Marinha Universidade de São Paulo, Sao Sebastiao, São Paulo, Brazil.

[13]State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, Fujian, China

[14]Laboratoire des Sciences du Climat et de l'Environnement, Gif-Sur-Yvette, France

[15]University of the Ryukyus, Okinawa, Nishihara, Japan

[16]Department of Chemistry, NIWA/University of Otago Research Centre for Oceanography, University of Otago, Dunedin, New Zealand

[17]Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile,Valdivia, Chile

[18]Department of Biosciences, University of Oslo, N-0316 Oslo, Norway

[19]Marine Science Institute, UC Santa Barbara, Santa Barbara, CA, USA

[20]Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Bremerhaven, Germany

*Corresponding Author Philip.boyd@utas.edu.au; 61-3-6226-8554

Abstract

Marine life is controlled by multiple physical and chemical drivers and by diverse ecological processes. Many of these oceanic properties are being altered by climate change and other anthropogenic pressures. Hence identifying the influences of multi-faceted ocean change, from local to global scales, is a complex task. To guide policy-making and make projections of the future of the marine biosphere, it is essential to understand biological responses at physiological, evolutionary and ecological levels. Here, we contrast and compare different approaches to multiple driver experiments that aim to elucidate biological responses to a complex matrix of ocean global change. We present the benefits and the challenges of each approach with a focus on marine research, and guidelines to navigate through these different categories to help identify strategies that might best address research questions in fundamental physiology, experimental evolutionary biology, and community ecology. Our Review reveals that the field of multiple driver research is being pulled in complementary directions: the need for reductionist approaches to obtain process-oriented, mechanistic understanding, and a requirement to quantify responses to projected future scenarios of ocean change. We conclude the Review with recommendations on how best to align different experimental approaches to contribute fundamental information needed for science-based policy-formulation.

Introduction – the challenges of multiple drivers and marine life

The global environment is rapidly being transformed by anthropogenic climate change, altering physical and chemical properties at an accelerating rate and bringing the Earth system into uncharted territory (IPCC, 2013; Gunderson et al., 2016). The imprint of climate change is already evident on multiple ocean properties (Dore et al., 2009; IPCC Summary for Policymakers, 2014) many of which shape the physiology and ecology of marine life. Ocean global change will have detrimental consequences for many organisms and beneficial effects for others, but levels of confidence around the magnitude and direction of these effects are often low, especially when projecting 50 years or more from now (Gattuso et al., 2015). Reducing uncertainty around projections of future change in marine ecosystems, and the goods and services they provide, is thus of paramount importance if we are to better predict responses of marine organisms and ecosystems to ocean global change. However, this represents a formidable challenge since the number of potential permutations of change involved is very large and often requires an interdisciplinary approach.

All approaches to investigate biological responses to environmental changes have benefits and limitations, and there is no single ideal method. Five main strategies have been widely applied to better understand how marine life interacts with environmental change (Fig. 1). Each approach has been employed to provide biological projections in climate change modelling simulations (Ridgwell et al. 2009). Together, they offer diverse insights into the responses of marine biota to multiple drivers. Here we employ the term "driver' in preference to "stressor", because effects of a driver can be either positive or negative, depending on the organism, process, or community being considered (Boyd & Hutchins, 2012).

Proxies for near-future global ocean change have been employed from the geological past, such as the Paleocene Eocene Thermal Maximum (PETM, Gibbs et al., 2016) and from present day marine ecosystems, such as submarine vents that release CO_2 (Hall-Spencer et al. 2008). Such surrogates have the potential to provide a holistic approach to investigating biotic responses to sustained change. During the PETM, and over millennia, the ocean was warmer (~5°C), with more CO_2 (> 1000 µatm ppmv), and more oligotrophic than today. The fossil record provides insights into the influence of long-term change across multiple trophic levels such as species' extinctions and emergences (Gibbs et al., 2016). Submarine CO_2 vents also offer insights into the response of an entire community to altered conditions (particularly acidification) over timescales of months to decades and more (Hall-Spencer et al., 2008). However, proxies do not provide exact analogues for present-day global ocean change. For example: the PETM comprised rates of change that were tenfold slower than those in the modern ocean (Hönisch et al., 2012, Zeebe et al., 2016); submarine vents mainly provide insights into the influence of a single driver (CO_2) rather than multiple drivers (Fig. 1); and CO_2 vent systems reveal responses of a localized benthic community operating in an otherwise un-acidified ocean, rather than the long-term system-wide effects that accrue under ocean global change.

The other approaches presented in Fig. 1 are firstly contemporary observations such as those from long-lived organisms (Thresher et al., 2011), regional or temporal gradients (Cubillos et al., 2007; Beaufort et al. 2011) or ocean time-series (Rivero-Calle et al., 2015). Second, they comprise manipulative experiments (Wernberg et al. 2012) including both small-volume "microcosm" methods often used with single species or strains, and large-volume "mesocosm" techniques that usually incorporate natural assemblages. Observational approaches provide concurrent estimates of long term (decades to centuries) high-resolution changes in environmental properties and responses by marine life, or "space for time" (see

Dunne et al., 2004) substitutes of long-term change (Fig. 1). In contrast, manipulation experiments offer the potential for highly controlled mechanistic insights into the relationship between a driver (or drivers) and the physiological, evolutionary or ecological response of the study organism(s) (Riebesell and Gattuso, 2015).

However, again, there are drawbacks with regard to cost, degree of replication, and ecological relevance to each of these approaches (Fig. 1; Havenhand et al., 2010; Andersson et al., 2015). For example, observational approaches are often confounded by the influence of natural climate variability (Edwards et al., 2013), which may limit their ability to discern global ocean change trends, especially over shorter timespans. Manipulation experiments typically employ highly artificial systems over short periods (weeks (Kroeker et al., 2010), to months, but see Kawecki et al., 2012 or Lenski, 2017), presenting problems with extrapolation to longer timescales (see Hutchins and Boyd, 2016). Microcosm experiments are limited in their ability to predict ecosystem- or food web-level effects, while mesocosm experiments are constrained by their considerable expense and logistical difficulty, and are therefore sometimes difficult to adequately replicate (Fig. 1). Thus, as we move along the continuum from simple, single-species, small-scale experiments through mesocosm studies, to large, open, natural experiments, we increase ecological relevance at the cost of understanding individual mechanisms (Sommer, 2012). Nevertheless, the ability of manipulative experiments to provide mechanistic insights into how multiple drivers will influence marine life in a future ocean makes them powerful and flexible tools, particularly when cross-linked to other approaches presented in Fig. 1. Together, these approaches have the potential to generate the required mechanistic understanding and predictive power to assess the effects of environmental change (Sommer et al., 2012; Dupont & Pörtner, 2013), and thus are particularly suited to providing data for incorporation into models.

In this Review, we commence with a brief historical perspective of ocean global change manipulation studies across a range of disciplines investigating the effects of single drivers. Note, these experimental approaches all rely on well-established conceptual advances in design and analysis that straddle many different disciplines (Table 1). We then chart the development of multiple driver experiments, and how their design and function has evolved. Next, we probe some of the emerging complexities of studying multiple drivers – specifically the increased number of combinations needed to document all the individual and interactive effects of drivers. This imperative leads to a discussion of the design and development of more complex experiments that forge stronger links between physiological, ecological and evolutionary approaches. We advocate the development of scientific questions that are directly relevant for society and therefore focus on solutions, policy formulation, and increased public awareness of these issues. Each of these complex questions can only be answered by its own unique combination of experiments, designs and approaches. We conclude by tackling a central issue that emerges during our synthesis – the need for research strategies that combine testing the effects of holistic ‘IPCC-like’ scenarios, with the development of better mechanistic understanding of specific biological responses to multiple drivers.

Single drivers – physiological, ecological or evolutionary studies

An experimental design which determines the organismal response to a selected range of environmental conditions is termed here the *mechanistic approach*. This strategy, often employed using a gradient of treatments to reveal underlying mechanisms and/or to test theory, has been a cornerstone of organismal physiology for decades. Examples include phytoplankton nutrient uptake studies in which the kinetics were characterised across a wide range of nutrient conditions (Harrison et al., 1989), and physiological research, which has subsequently informed the development of physiological models based on oxygen or irradiance (Pörtner and Grieshaber, 1993; Geider et al., 1996). These models in turn lead to better experimental designs (Table 2). This single driver, gradient approach has also been adopted in an environmental context to study the effects of (e.g.) transient warming or low oxygen concentrations (Baumann, 2016).

In the last two decades, the proliferation of experimental studies into climate change effects on marine life has resulted in a marked divergence from this mechanistic/gradient approach. Multiple climate change scenarios, usually based on model projections for one or more environmental driver for the year 2100 and/or beyond (IPCC WG1, 2013) have been used to create a suite of discrete treatments, relative to a control centred on present day or pre-industrial conditions (termed here the *scenario-based approach*). This scenario-based approach has been widely employed to examine the effects of individual drivers, and combinations of drivers, on biota (see Yang et al., 2016), and is mainly distinguished from the mechanistic approach by the rationale for the choice, and levels, of driver(s) used in experiments to predict biological responses to environmental change.

In marine research, the field of ocean acidification has influenced the refinement of single driver experiments by developing robust recommendations for the replication of treatments, harmonisation of experimental manipulations, and employment of future climate change scenarios (Riebesell et al., 2010). The single driver experimental design has been popular (Yang et al., 2016), not least because of the relatively simple logistics needed to tackle a suite of experiments across a wide range of species or groups, which ultimately permits meta-analysis (Kroeker et al., 2013), and in tandem with modelling accelerates mechanistic understanding (for example, Saito et al., 2008). Furthermore, single driver experiments provide a straightforward conceptual platform to launch more logistically challenging experimental designs such as those that test constant versus fluctuating conditions (see Table 2).

A decade of diversification of the design of single driver manipulation studies enables their categorisation into physiological, ecological and evolutionary studies (Table 2). Physiological scenario-based studies have mainly targeted two to three global change scenarios (for example, CO_2 levels during pre-industrial revolution and the present day, and projected for year 2050 and in particular 2100, Riebesell et al., 2010). These studies have revealed a diverse range of organism-specific responses (ranging from detrimental, to no change, to modal or beneficial effects; Langer et al., 2009; Ries et al., 2009). In contrast, the limited number of treatment levels used (Fig. 2a), and/or inappropriately selected levels (Figure 2b), have often prevented these studies from identifying threshold levels in the relationship between physiological affinity and the environment. For example, differences in the response of planktonic nitrogen-fixers to elevated CO_2 (based on a limited number of treatments) have been reported (Hutchins et al. 2009, Law et al. 2012, Gradoville et al. 2014). Consequently, Hutchins et al. (2013) embarked on an in-depth mechanistic/gradient study of the CO_2 affinities of N-fixers based on a broader range of seven CO_2 concentrations. Their findings revealed distinctive CO_2 functional response curves for these diazotrophs, and provided a compelling explanation for the differences observed in the scenario-based studies.

Single drivers have also been used in more logistically-challenging scenario-based experiments in which the response(s) of entire ecological communities to manipulation have been investigated (Riebesell et al., 2013, Gattuso et al., 2014). Outcomes from such studies reflect the combined influence of direct impacts on individual species, and indirect effects resulting from, for example, shifts in community composition (Schulz et al. 2017; Taucher et al. 2017), prey palatability (Poore et al., 2013) and changes in competition (Hale et al., 2011). Methods for separating direct and indirect effects are available (Alsterberg et al., 2013; see below), but have been applied infrequently in such studies. Inherent in such community/ecosystem-level studies is the need to run the experiment for a longer period (months, often set by the response times of apex predators, such as planktivorous fish; Riebesell et al., 2013) in order to allow the spectrum of ecological interactions to take effect. Consequently, in contrast to the many single-driver physiological studies reviewed by Kroeker et al. (2013), few large-scale, (and hence longer-term) experiments have been performed.

This lack of ecosystem-level and/or longer term (months to years) manipulation studies is an important omission as these spatial and temporal scales are the most relevant for projecting future effects (Riebesell & Gattuso, 2015). For example, a review of 110 marine global change experiments published between 2000 and 2009 reported that ~58% investigated single species and <19% investigated communities (Wernberg et al., 2012). Mesocosms (typically tens to thousands of liters, depending on the ecosystem) provide an important bridge between small, tightly controlled microcosm experiments such as inter-specific competition experiments (Krause et al., 2012), which suffer from limited realism, and the exponentially greater complexity of natural systems in which mechanistic relationships across trophic levels often cannot be identified (Stewart 2013; Table 2). Although mesocosms permit testing hypotheses at the community- and ecosystem-levels, stochastic divergent responses of replicate enclosures, and lack of lateral and/or vertical exchange are considered as potential intrinsic limitations of this approach (see e.g. Chave, 2013; Table 2). To date, the (often logistic) limitations on the number of replicate mesocosms mean that such community/ecosystem approaches have mainly targeted a scenario-based approach (Table 2). Making connections between the results of single species experimental settings and such larger scale mesocosm approaches will be needed to provide a mechanistic understanding at these large scales and will be a challenge for years to come.

The third broad category of single driver experiments has used the principles of experimental evolutionary biology to look at timescales of acclimatisation (plastic responses that involve changes in organismal phenotype without any underlying change in the genetic composition of populations) versus evolution (change in the genetic composition of a population over time) in response to climate-change forcing (Collins, 2014). These experiments have generally been more multi-generational than most other manipulation studies, and have mainly focussed on microbes with short generation times (days), such that micro-evolution could be examined on a timescale of years (i.e., across ~1000 generations, Collins and Bell, 2004). Such evolutionary studies have mainly targeted scenarios (e.g. Lohbeck et al. 2012). More recently, evolutionary studies have begun to focus on interactive effects of multiple drivers (Schlüter et al. 2014; Brennan et al., 2017) and how physiological mechanisms themselves are likely to evolve (Table 2), such as the evolution of thermal reaction norms (e.g. Listmann et al., 2016). For organisms with long generation times, comparative studies of populations in environmental climes offer an indirect option for evolutionary study (see above).

Although single driver studies have been highly versatile and made valuable contributions to our understanding of responses, particularly when coupled with models (Table 2), they also have drawbacks. The complex nature of global ocean change (Fig. 3a) means that investigations of single drivers seldom provide reliable inferences about responses in a multivariate natural environment (but see the example of Hughes et al. (2017) in Table 2): interactive (additive, synergistic or antagonistic) and indirect effects frequently mediate the responses observed in single-driver experiments (Darling and Côté, 2008; Harvey et al., 2013), and can sometimes lead to outcomes that are not readily predictable without a deep understanding of modes of action ("ecological surprises", *sensu* Paine et al., 1998). Hence, estimating the effect(s) of multiple environmental drivers is a major source of uncertainty for projections (Darling and Côté 2008), and so it has been repeatedly recommended that research efforts in this direction should be strengthened (e.g., Crain et al., 2008, Gattuso et al., 2011, Havenhand et al., 2010, Wernberg et al. 2010). Notwithstanding the ongoing valuable contributions made by single-driver ocean-change experiments, it is obvious that a broadening of trajectories is needed in the experimental domain space: from single to multiple drivers, connecting single organism experiments to communities and ecosystems, and linking short (i.e., acclimation) to long (i.e., adaptation) experimental durations (Riebesell & Gattuso, 2015).

From single to multiple drivers – experimental challenges

The transition from an experimental strategy that examines the effect of a single driver to one that has multiple drivers has to deal with three main challenges (Fig. 3). First, is cataloguing the various combinations of drivers (global, regional and local; Boyd and Hutchins 2012), and levels of each driver, that are appropriate for a specific manipulation study. Second, is rationalising the need for a conceptual holistic approach that considers all of these combinations with the need for experimental (mechanistic) reductionism, taking into account the limitations imposed by logistics and resources (Sommer, 2012; Boyd et al., 2010). Third, is designing tractable experiments which address the second challenge and that can be successfully conducted, interpreted, and compared with other manipulation studies to construct a broader picture of responses to ocean global change by biota across trophic levels (Boyd, 2014).

One common approach is adding more variables (drivers) in a fully-factorial matrix experimental system (Fig. 4a). This can quickly become impractical both logistically, and in terms of our ability to interpret the whole range of outcomes (Fig. 3b). This issue is amplified as the number of levels of each driver increases. Such experimental designs are also challenging to present in a clearly organized and intelligible fashion in a typical scientific publication format. In practice, without sacrificing replication, the maximum practical limit in a factorial matrix design is often three variables. However, robust replication (minimum triplicates, and preferably *many* more; see Cumming, 2008) is the foundation of experimental design, and in many cases compromising on replication can result in variable, unrepeatable, and occasionally uninterpretable outcomes. However, it is important to accept that low – or no – replication is sometimes inevitable, for example for community-scale manipulations in the field, behavioural studies where ethics or other concerns may limit sample sizes, monitoring data, observations at CO_2 vents, and the analysis of natural experiments where chance events occur at a single site. Despite low levels of replication, such data can be highly valuable and still amenable to statistical analyses (Davies & Gray, 2015). The dual issues of optimising experimental design and the preferential selection of which drivers to include in experiments are detailed in Sections 4 and 5, respectively.

Multiple driver experiments – design, logistics and analysis

Multiple driver experiments generally involve considerable resources (time, effort, materials) necessitating clarity around experimental hypotheses and aims. Thus, an important consideration is to ensure that the selected design unambiguously addresses these goals, and that resources are well-used. In this context, it is just as important to identify – and accept – what the planned experiment will not address. Incorporation of these principles at the planning and design stage helps to define a more valuable experiment.

An important distinction when moving to studies investigating three or more drivers is that it may necessarily involve a shift from a gradient or mechanistic approach that includes all possible interactions, to an empirical or scenario-testing approach (defined in Section 2). Designs for these approaches are fundamentally different. For relatively simple experiments involving 1 to 3 drivers and designed to provide mechanistic understanding (Fig. 3b) the relevant principles and techniques are well-established (see e.g., Quinn & Keough, 2002 and other references in Table 1). More complex designs call for alternate approaches such as those outlined below. In either case, recent developments in statistical methods have added novel, powerful, and informative techniques that permit analyses to be run that were previously difficult or impossible. These include: analysis of univariate and multivariate data with unknown and heterogeneous variance structures, Bayesian techniques for estimating posterior probability distributions (rather than single P-values), and Structural Equation Modelling that can identify the relative strength – and statistical significance – of direct and indirect effects in networks of many variables (e.g. Alsterberg et al., 2013)

Despite the availability of these powerful new tools, designing and running even relatively "simple" gradient experiments can be logistically challenging, since the aim is often to use multiple levels of each driver to construct response (tolerance) curves. This challenge arises because these designs become unwieldy as the number of drivers and levels increases: the total number of treatment combinations is equal to the product of the number of treatments and the number of treatment levels. Thus, the commendable aim of increasing mechanistic understanding by adding more levels of each driver, causes the experiment to grow exponentially. For example, six levels for each of three drivers results in 216 combinations – without replication (Fig. 3b).

Reducing the number of independent drivers permits greater replication (and, hence, greater statistical power), and/or allows for more levels of each driver (and, hence, better description of response curves). This can be done in one of two, related, ways: by collapsing several variables into one (e.g. Boyd et al., 2015); or by reducing the number of interactions between drivers in the design ("reduced design", Table 2 and Fig. 3). Briefly, the "collapsed design" approach (Fig. 3c left) involves identifying the primary driver of interest, and testing the effects of this driver as one factor with all other drivers (the number of which will be organism-specific) simultaneously "collapsed" into a second combined driver. This creates a two-way design with relatively few treatment combinations (in comparison to the full-factorial alternative), and therefore permits the use of more levels of the factor of interest, and/or greater replication (Boyd et al., 2015). The alternative "reduced design" (see Table 1; Fig 3C centre right) tests the (single) effects of each driver independently and the (combined) interactive effects of all the drivers together, but excludes lower-order (e.g. 2-way) interactions. Like the "collapsed design", this approach permits mechanistic understanding of effects of individual drivers (only), but provides a more holistic understanding of responses to their combined effects. In this case, detailed mechanisms of lower-order interactions among the drivers are sacrificed in order to provide more levels of each driver, and/or greater replication and hence statistical power (see Gunst & Mason, 2009, for alternatives). For both designs, standard statistical analysis techniques such as generalised linear modelling can be used to analyse the results.

Which of these designs is most useful will depend on the question(s) to be addressed and requires a degree of knowledge about the drivers of a particular system. For example, Boyd et al. (2015) used prior information from a literature survey and pilot experiment to determine that one driver (temperature) had an overriding effect on the response variable of interest, and therefore they collapsed all the other drivers into a second combined factor. In the absence of such preliminary information, when it is unclear that one factor has overriding influence or importance, and/or when it is clear that responses to combined scenarios are required, reduced designs, or the fractional factorials of Gunst and Mason (2009), may be more informative. It should be noted, that hybrids between collapsed and reduced designs can provide valuable mechanistic understanding while also testing responses to scenarios (e.g. Xu et al., 2015). The theoretical interaction between two drivers, across all possible treatment levels can be visualised readily using a driver landscape (Fig. 5), a concept borrowed from evolutionary biology in which such visualisations are employed to explore fitness or adaptive landscapes such as between genotypes and reproductive fitness (Mustonen and Lässig, 2009).

It should be noted that even when full-factorial designs using 3 or more drivers are logistically possible, this might not be the most informative approach. Interpreting and understanding the biological significance of statistically significant 3-, 4- and 5-way interactions within a meaningful conceptual framework can be challenging if not impossible.

At larger spatial scales that include multiple drivers, multivariate techniques such as ordination and Structural Equation Modelling can be more informative, especially for large mesocosms, or for observational designs that compare CO_2 seep and vent systems with neighbouring control areas (e.g. Smith et al., 2016). Many of these designs manipulate one (or a few) key driver(s) in the field while measuring additional drivers and responses (e.g.

Albright et al. 2016). These approaches at larger spatial scales epitomise a central issue in experimental design: the lack of statistical independence among drivers can constrain interpretation and inference. Nonetheless, such designs benefit from having strong ecological relevance. In the search for experimental rigour, ecological relevance should not be overlooked, as it is central to understanding how climate change will influence key ecosystem services (Pörtner et al., 2014).

With a few notable exceptions (such as FOCE, see Gattuso et al., 2014), the number of drivers that can be tested in an experimental system is inversely dependent on the size of the study organism – or, more accurately, the experimental unit. For very small experimental units, such as protists in culture, testing many different levels of multiple drivers with a high degree of replication may be possible within the available resources (e.g. Brennan & Collins 2015). Such designs provide vital context in which to interpret the results of single-driver experiments, as well as begin to build a generalizable understanding of the nature and distributions of organismal responses to multiple drivers that is not based mainly on driver identity (Brennan et al., 2017). However, as the size of the experimental unit increases, the capacity to design, conduct, and analyse full-factorial experiments declines because the resources needed to conduct the experiment become limiting. Provision of more resources can remove this limitation, permitting the construction of larger and/or more complex experiments.

At some point, however, the size of the experimental unit becomes severely limiting, allowing few – or perhaps only one – unit for each treatment. As for the multiple-driver examples earlier, reduced or collapsed factorial designs, and multivariate analysis techniques become increasingly important in this situation. It is important to recognise that the "limitation" of large experimental units is a logistical, and not a statistical issue. As noted above, because large mesocosms or FOCE designs encompass more ecological processes, the reduced statistical power that accrues from fewer treatments is offset by ecological relevance (see e.g. Barley and Meeuwig, 2017).

Multiple drivers - rationale for selection of drivers

The wide range of constraints addressed in Section 4 have important ramifications for the selection of drivers used in manipulation studies. The first aspect of selection is to identify the relevant components of the matrix of global ocean change and their projected magnitude in the coming decades. These drivers include pH, temperature, irradiance, nutrients and oxygen (Fig. 2a) and sea-level rise. Superimposed on these global shifts are regional and local anthropogenic changes in marine properties that include underwater penetration of UV radiation (Gao et al. 2012), eutrophication, freshening, point-source pollution, and harvesting pressures (Boyd and Hutchins, 2012). These drivers, individually and interactively, can result in detrimental, beneficial, or no effect on a specific organism. This leads to the second component of driver selection: the assembly of an inventory of biologically-influential drivers that are specific to the study region and/or organism(s)/system of interest (Fig. 2b). Selection of these drivers also depends on the organism(s) of interest. For instance, autotrophs can be strongly influenced by pCO_2 and irradiance, heterotrophs including microbial heterotrophs are more likely to be affected directly by pH than by pCO_2 (Bunse et al., 2016), and the responses of grazers to these drivers are often highly influenced by food availability (Montagnes et al., 2008).

Thus, three of the main considerations for choosing drivers for experiments are: i) that they are relevant in terms of projected change, i.e. they mimic change, test extreme cases, and/or examine known interactions among drivers; ii) that experiments attempt to capture the range of effects of drivers, i.e. the design contains treatments or treatment-levels that could detect both detrimental and beneficial effects; and iii) to keep all other drivers at environmentally relevant levels (if pertinent to the particular experiment). The rationale for selecting drivers will differ depending on where the experimental design falls on the mechanistic versus scenario-testing continuum (see Section 9). In many cases, preliminary experiments may be required to better understand the relationship between the individual and interactive effects of multiple drivers (see Boyd et al., 2015). Such pilot data are also highly valuable for *a priori* Power Analysis to estimate levels of replication needed in the experiment (Havenhand et al., 2010). Both of these practices greatly aid the identification of experimental designs which are both tractable and interpretable (Fig. 3c).

Distinguishing individual and interactive effects of drivers

Distinguishing – and quantifying – the individual and interactive effects of drivers requires statistical analysis of multi-driver designs. Interpreting the results of such analyses can be challenging: not only are designs with 3 or more drivers logistically difficult (Fig. 3), but responses to the hierarchies of multiple drivers may be absent, additive, or multiplicative (see Table 3). Moreover, multiplicative effects of drivers (i.e. statistical interactions, or indirect effects) may often be non-linear, the detection of which requires multiple levels of each driver – which brings the accordant combinatorial problems discussed in Section 4. Interpretation of multiplicative effects of climate drivers has also been complicated by inconsistent terminology – in particular the interpretation of "synergistic" and "antagonistic" effects (Table 3). Therefore, as a first step, we suggest responses to multiple drivers be characterised as 'additive' or 'multiplicative' to specify the absence or presence of an interaction, and ‘aggravating’ or ‘mitigating’ to specify the direction of responses. This should be supplemented by quantification of the effect sizes for various exposure levels through the use of, for example, interaction plots.

Cumulative effects of multiple drivers over time are an even more complex problem field. The successive exposure to varying levels of one driver, and the combined effects of several drivers may lead to cumulative effects on performance. The term ‘cumulative impacts’ has been defined as “the effects of one or more drivers, and their interactions, added to other past, present, and reasonably foreseeable future effects of drivers” (Hegmann et al. 1999). This terminology is often used by environmental protection agencies, and forms the background to multiple driver experiments designed to support environmental impact assessments.

Several approaches are available to tease apart these differing effects on the biota. One approach involves developing suitable experimental designs with powerful statistical modelling to explore the relative influence of individual versus interactive effects (and of increasing the numbers of drivers, without an explicit focus on their identity) in improving our ability to interpret experimental outcomes by characterising averages or distributions of effects over many drivers (e.g., Brennan and Collins, 2015; Brennan et al., 2017). Another approach requires learning from conceptual and modelling approaches to multiple drivers' research from other disciplines such as ecotoxicology (Goussen et al., 2016), and food safety microbiology (Mejlholm and Dalgaard 2009). Ultimately, the goal is to construct broader conceptual frameworks based on unifying principles e.g., metabolic flux theory (Sajitz-Hermstein and Nikoloski, 2013; Kazamia et al., 2016) that are common across taxa.

Findings from multiple driver experiments illustrate that the effects (individual versus interactive) of drivers depend both on driver identity and driver intensity (e.g. Gao et al., 2012; Sett et al., 2014). However, there is growing evidence that the influence of multiple drivers rapidly becomes very complex, is not necessarily additive, and that both individual and interactive driver effects can be species- or process-specific (Boyd et al., 2015; Darling & Côté 2008). In addition, the interaction between any given pair of drivers depends on which other drivers are present, and which scenarios of each driver are being considered in the manipulation study. The underlying forcing across this rapidly expanding number of combinations is both difficult to interpret, and soon becomes logistically impossible to investigate. Such problems can be minimized, or even avoided, by use of the reduced and collapsed designs described earlier (Section 4). Despite the underlying complexity of interpreting such experiments, progress is being made on both discerning emergent patterns between drivers and how it scales with the number of drivers (Brennan et al., 2017) and in identifying physico-chemical interactive mechanisms evident among drivers (Boyd et al., 2015; Brennan and Collins, 2015).

Experiments with two or three drivers based on IPCC climate change scenarios (e.g. projected pH and temperature for the present day, 2050 and 2100), can readily identify interactions among drivers. The interaction can then be categorised as synergistic or antagonistic (Folt et al., 1999; Darling & Coté 2008, but see Table 3), however, there is the wider issue of whether the interactive effect is linear: does it hold across the entire range of the interaction between two drivers or just for a portion of the range being examined i.e. to what extent is the observed interaction a valid description of the relationship between drivers? Consequently, it is important to determine where each of the scenario-based treatments for multiple drivers (such as pCO_2 of 750 μatm and 2° C warming, year 2100) lie on a physiological performance curve (such as CO_2 affinity, see Hutchins et al., 2013) or a toxicant dose response curve (see Goussen et al., 2016). Such performance-based assessments again require a step-function increase in experimental logistics, for example marine photo-autotrophs often have ~6 physiologically-influential drivers (see Bach et al., 2013). This requires assessment of a response curve for each driver, and subsequently the need for curves across a range of conditions of interacting drivers (e.g., CO_2 affinity across a range of environmentally-relevant temperatures, Sett et al., 2014; Fig. 5). Clearly, fundamental underpinning concepts (physiological, ecological, evolutionary) and underlying principles that are common across functional groups, such as primary producers and grazers, are needed to overcome such a Gordian Knot of combinations (Boyd, 2014).

Better understanding of the multiple modes of interaction seen in the marine environment might be obtained by adapting modelling approaches from other fields. For example, the microbial spoilage of foods is also characterised by combinations of many environmental drivers. Modelling, using ~10 relevant drivers, has revealed that sufficiently complex models can accurately predict microbial growth responses, whereas simpler models with fewer drivers do not (Mejlholm et al., 2010). Other fields such as ecotoxicology have focussed on the energetics of organisms as a means to integrate the organismal responses to a wide range of environmental drivers. For example, environmental risk assessments now integrate chemical and ecological drivers, using energy based models (Goussen et al., 2016). Metabolic flux theory (see Kazamia et al., 2016) or other energy flux modelling approach is another promising integrative approach to multiple drivers that might be applied to marine environments to deconvolve individual and interactive effects, and to generalize from experiments on model organisms and systems. For example, planktonic foodwebs are characterised by hundreds of species, strains and ecotypes and their trophodynamics (Worden et al., 2015), yet despite this taxonomic and functional diversity, there are a finite number of cellular processes that occur, and these can be mapped at some level of resolution (Muller and Nisbet, 2014; Lorena et al., 2010). This need not focus solely on shared traits, and indeed could be employed for model species across different (specialized) planktonic functional groups such as calcifiers or nitrogen-fixers.

Bridging between physiological responses and ecosystem impacts

Understanding metabolic and physiological responses provides a baseline for untangling species and population sensitivities to environmental alterations, and hence is highly desirable in the ongoing development of ocean global change research (Fig. 4). However, upscaling physiological responses to community and ecosystem impacts is challenging and remains a major aspiration in ecology (Sutherland et al., 2013). There are many confounding issues associated with such upscaling, including our lack of understanding of the role of intra- and inter-species diversity in defining ecosystem function, which limits the translation of physiological response curves to responses at the ecosystem level (Hillebrand and Matthiessen, 2009). Species deemed tolerant to a driver based on physiological responses derived from lab experiments may display high sensitivities in the natural environment through indirect effects of the same driver, such as modifications of their habitat or other vital resources. For example, the deterioration of habitat complexity in a coral reef exposed to CO_2 venting resulted in the loss of many macroinvertebrate groups, such as crustaceans, in spite of their assumed high physiological tolerance to ocean acidification (Fabricius et al. 2014).

There is also the pressing issue of the context under which experiments are conducted. For example, the response of filter-feeding bivalves and barnacles to ocean acidification depends on the nutritional status of the animals (Thomsen et al., 2013; Pansch et al., 2014). The confounding influences of concurrent direct (e.g. temperature on grazer physiology) and indirect (e.g. food quality and/or quantity) effects on other trophic levels can further complicate the interpretation of community- and ecosystem-level observations (Boyd and Hutchins, 2012). Examples of such indirect effects are alterations of prey quality impacting consumers (Montagnes et al., 2008, Rossoll et al. 2012) or *vice versa* - consumers mediating the effects of experimental ocean acidification and warming on primary producers (Alsterberg et al. 2013).

There is ample evidence now that community and ecosystem interactions (including competition, symbiotic/parasitic relationships, and trophic interactions) can both dampen and amplify physiological sensitivities. Bottom-up and top-down processes may thereby act simultaneously. For instance, elevated CO_2 has the potential to increase primary production by marine algae and plants (Kroeker et al. 2010), thereby increasing food availability, but also to alter food quality and palatability (Arnold et al. 2012, Rossoll et al. 2012). At the same time, ocean acidification raises energetic costs in many consumers, especially calcifying species. These interacting responses generate a complex interplay among the physiological susceptibility of organisms to ocean acidification, the provisioning of resources, and the level of competition (Gaylord et al. 2015).

Compensatory effects may emerge from the diversity among functionally similar taxa, which widens the spectrum of responses to environmental perturbations, with population increases of tolerant taxa counteracting declines of sensitive taxa (Yachi and Loreau 1999). Within a given population, phenotypic diversity will likely buffer population sensitivity to environmental drivers through the portfolio effect or functional redundancy, (see Roger et al., 2012), but testing this with natural communities is not trivial. For example, the increased phenotypic diversity of natural populations, such as obtained in mesocosms, broadens the variance in 'dose-response' relationships determined from laboratory experiments on isolated strains or species (Zhang et al., 2014). Likewise, small or cryptic shifts in physiological responses may be reflected more strongly at the community to ecosystem level. For instance, a 5-10% decline in the specific growth rate of the coccolithophore *E. huxleyi* under ocean acidification can scale up to the failure of bloom formation at the ecosystem level (Riebesell et al., 2017). An assemblage shift may thereby have a greater impact on the integrated community performance and its impact on biogeochemical processes than species-specific responses, highlighting the importance of whole community manipulation experiments for unravelling community level impacts.

A way forward in bridging between physiological responses and community/ecosystem impacts could be in the co-design of up-scaling and down-scaling approaches. Insights gained at the community level could help identify those responses that prevail in the complex texture of natural ecosystems, and which require a more in-depth mechanistic understanding. In turn, improved understanding of physiological sensitivities can help to guide the design and implementation of community-level experiments. A hybrid experimental design in which subsamples from natural community experiments are interrogated physiologically (Sosik and Olsen, 2007), or for their acclimatory (discrete incubators within mesocosms), or evolutionary (Tatters et al. 2013a,b, Scheinin et al. 2015), responses could be a first step in this direction. Research on ocean global change would also greatly benefit from more detailed consideration of ecological theory, which to date has been included only peripherally (Gaylord et al., 2015). Well-founded ecological concepts, when applied in the context of ocean global change, can generate predictions and facilitate the interpretation of a range of community- and ecosystem-level impacts, such as loss in biodiversity and resilience to shifts in species assemblages and geographical ranges.

Evolution under multiple drivers

The majority of the experimental approaches presented in Table 2 can provide insights and information into plastic (i.e., acclimatory, days to months) responses to multiple drivers. However, over longer time scales (dozens or hundreds of generations) marine organisms can evolve in response to multiple drivers due to their high standing genetic variation (Rynearson and Armbrust, 2000; Biller et al., 2015) and rates of mutation. Much has already been learnt from looking at evolution through the lens of an individual (dominant) driver, and only taking other drivers into account when necessary (Fig. 4). A key strength of evolution experiments is that they are usually designed with high statistical power, and are intended to be generalized, since they frame questions in terms of fitness and patterns of environmental change (Shaum and Collins, 2014, Brennan et al., 2017). Evolution experiments can also be used to investigate organism- and driver-specific questions (Lohbeck et al., 2012a, 2102b; 2014; Hutchins et al., 2015). As with all experiments, there is a tradeoff between generality and realism (see Sommer, 2012). For example, experiments may be done in non-marine organisms in order to overcome logistical limitations and achieve the level of replication needed to take a "first pass" at high-level general questions (Collins and Bell, 2004, Low-Décarie et al., 2011). Here, we focus on comparing plastic and evolutionary responses under single drivers, and discuss the challenges in scaling up to multiple drivers and to taking into account the community/ecosystem level.

Evolution experiments using a single driver have provided insights into whether or not plastic responses are maintained, surpassed, or reversed by evolution (Fig. 6). This outcome is trait- and organism-specific, and there is little theory that predicts the evolution of specific traits, even in single-driver environments. Some studies show that the initial (reversible) plastic response is maintained in single-driver environments (Müller et al., 2010), or that plastic responses can become irreversibly fixed traits by evolution (Hutchins et al. 2015, Walworth et al. 2016a). In contrast, other studies show loss of function, or even trait reversion. In the marine alga *Ostreococcus*, an initial response to high CO_2 eventually reverses to some degree under constant high CO_2 conditions, and more or less completely under fluctuating CO_2 conditions (Schaum & Collins, 2014; Schaum et al., 2015). Finally, traits may evolve to surpass the plastic response, which is the expected outcome under directional selection in an environment where fitness is initially low (Elena and Lenski, 2003). Lohbeck et al. (2012a) showed that the evolutionary recovery of calcification in *E. huxleyi* could exceed the plastic response (i.e., cells evolved at high CO_2 were less compromised than expected given their initial decreases in calcification). Other experiments have revealed counter-intuitive effects over long timescales. For example, Tatters et al. (2013a,b) found that the observed growth rate responses of diatoms and dinoflagellates to warming/acidification did not readily translate to enhanced competitive abilities in competitive exclusion manipulation studies.

Conceptually, the Tatters et al. (2013a,b) studies are important because while evolution (genetic change within populations) depends on relative fitness (defined here as the relative growth rates of genotypes when they can interact), the long-term persistence of populations depends on absolute fitness (defined here as net population growth rates). The Lohbeck et al. (2012a) study also illustrates this point; even though there was adaptive evolution after a few hundred generations of growth under high CO_2, growth and calcification rates were still lower than at control CO_2 levels, and it is unclear whether the increase in absolute fitness in the high CO_2 environment was sufficient to allow population persistence. Hence, as we scale up to multiple driver evolutionary experiments, it is evident that we need to consider both absolute and relative fitness in future studies assessing the evolutionary potential of populations, and link that to the likelihood of them persisting (Carlson et al., 2014; Bell, 2017).

A few experiments to date have examined evolution to pairs of drivers (Gao et al., 2012; Tatters et al., 2013a; Schlüter et al. 2014). They suggest that plastic and evolutionary responses differ in both single and multiple driver environments, and that evolution to pairs of drivers differs from evolution to either of the single drivers (Brennan et al., 2017). The single short-term study to investigate the general effect of having different numbers of multiple drivers suggests that when there are many drivers in the environment, a few key drivers determine the strength of selection on average (Brennan at al., 2017). However, there are few data on how and why trait evolution varies between different multi-driver environments. How evolutionary responses to key drivers depend on the multi-driver context in which they occur is another research topic that requires urgent attention to progress this field. Studies that reveal the interactions between specific drivers and driver intensities in key model species provide mechanistic insight, but generalizing from these studies will be difficult without advances in fundamental evolutionary theory; developing such theory will require sustained collaborations between oceanographers and evolutionary biologists. As with physiology studies, a combination of metabolic flux theory, and comparative studies showing how natural populations have adapted to different multi-driver environments (Biller et al., 2015) are two potential ways forward. Empirically-informed theory on the link between plastic and evolutionary responses (Ghalambour et al., 2015, Chevin et al., 2010, Lande, 2014) also has the potential to leverage the results of physiology studies to make predictions about trait evolution.

The challenges of studying evolutionary responses mirror those for plasticity studies in terms of experimental design or logistics. Hence, collapsed or reduced designs (Section 4) in microbial evolution experiments are one way to leverage existing evolutionary theory to address responses to multiple drivers in marine systems. One approach that has been taken to simplify the logistics of evolution experiments is to first evolve populations under a single driver such as high CO_2 (Hutchins et al. 2015), and then subsequently evolve these CO_2-adapted populations in new environments such as nutrient limitation (Walworth et al. 2016b) or warming (Schlüter et al. 2014). This strategy avoids maintaining organisms over long periods of time in full factorial selection regimes. A second challenge is that population genetic theory typically frames organismal responses to environmental change in terms of changes in fitness (Chevin et al. 2010, Lande 2014), while ocean acidification and global change studies are usually concerned with the functional traits of key taxa (Lohbeck et al. 2012a). Reconciling these two approaches – eventually via a functional trait-fitness mapping approach – will help ocean global change research to leverage the body of population genetic theory available.

Finally, the way in which drivers change, in addition to intensity and combinations of drivers involved, has the potential to impact evolutionary responses. Rates of environmental change (Collins & de Meaux, 2009; Lachapelle, et al. 2015), or the presence of environmental fluctuations (Schaum and Collins 2014) impact adaptive responses. This is an area where there is a large body of evolutionary theory (Botero et al. 2015, Collins et al., 2007; Lande 2014), which should be exploited to better guide the design of future experiments.

Multiple driver science that informs society

There is an urgent need to develop multiple-driver science that can directly inform society through improved communication (e.g. stakeholder awareness and acceptance), development of solutions (e.g. adaptation strategies), and policies (e.g. mitigation). Each requires a deep understanding of stakeholder culture, what type of information is needed to drive the changes (for example, the social dimension, Folke et al., 2005), and how to efficiently deliver the message (Dupont et al. 2015; Dupont 2017a,b). This will lead to a wide range of research questions and very different requirements for experimental strategies. A more efficient approach to influence individual behaviour is to develop scientific information directly targeting societal values. However, development of technological or policy solutions often requires more complex information such as models or experiments allowing the prediction of biological impacts at different time scales for a range of scenarios.

Different societal goals will naturally lead to specific research questions that can be better addressed by strategies that combine the different complementary experimental designs described above. Many of these questions have a global context, and yet most researchers work at regional scales. Local mitigation of non-global stressors is also one of the few tools available to management to deal with the near-term effects of global climate change (Magnan et al., 2016). Regional policy-focussed research requires regional projections or forecasts of the changing ocean, which are often not available (but see Meier et al. 2012; Bopp et al. 2013; Capone and Hutchins 2013; Hutchins and Boyd 2016). The drivers selected, and the levels of those drivers used in experiments, will typically be defined by the biological question and organism(s) of interest, and may or may not be cross-referenced to climate-change scenarios (Fig. 2).

The benefits of scenario testing include the development of practical methods to test for multi-driver effects that integrate the modulating effects of interacting drivers, and which can be applied beyond the species-level (i.e. in community-level experimentation). Importantly, for maximum impact the findings should be directly applicable for IPCC-type integrated assessment, in particular for making specific regional mitigation and adaptation recommendations in the coastal ocean (Schmidt and Boyd, 2016). There will inevitably be drawbacks, in particular the risk of design ambiguity with respect to representative scenarios. For example, deciding what combination and range of environmental change parameters to choose can be problematic, as there is a wide range of climate change scenarios across the IPCC (see Magnan et al., 2016). Design issues may also arise if the selection of representative parameter ranges is species- and strain-specific; this form of selection is used primarily to design experiments seeking a mechanistic understanding and often requires *a priori* knowledge of the specific physiological responses of the test species. Another trade-off from such a dedicated scenario-based design includes fewer insights into the additive, antagonistic, or synergistic effects of interacting drivers, which may remain hidden.

The alternative approach that targets mechanistic understanding using scenario-approaches (Table 2) will elucidate the mechanisms underlying individual and interactive physiological responses. While mechanistic studies are essential for developing modelling frameworks, detailed investigation of the many component processes, drivers, and their interactions is likely to create rates of progress too slow to meet societal needs. Experimental designs that comprise a 'hybrid' approach, that span aspects of pure scenario-based information and mechanistic understanding, are possible (see above), however these are largely untried (but see Xu et al. 2015). Such approaches could exploit the harmonisation of experimental design across parts of the scientific community (for example, Boyd et al., 2013). Regardless of the approach employed, the parallel development and application of different approaches will maximise opportunities that scenario-based approaches are timely enough to inform policy, while ensuring that mechanistic approaches continue to contribute to the development of more robust models that then refine existing policy frameworks for ocean global change over longer (decadal) timescales.

Conclusions and recommendations

1) Five main strategies – paleo-proxies, modern proxies, modern observations, manipulative microcosm experiments, and large volume mesoscosm experiment enclosures – have been widely applied to better understand how marine life interacts with environmental change. All approaches to investigate biological responses to change have benefits and limitations, and there is no single ideal method. A combination of approaches targeting a specific question at different levels, often allows for additional insights.
2) Although there is no clear two-way dichotomy in the multi-dimensional space of multiple-driver research, *mechanistic*- and *scenario-based* approaches capture the two main philosophies used to develop mechanistic understanding and to identify the consequences of a projected future state (or series of states), respectively.
3) Which experimental design is most useful will depend on the question(s) to be addressed, and will require a degree of knowledge about the relevant drivers in a particular system.
4) A way forward in bridging between physiological responses and community / ecosystem impacts is to co-design up-scaling and down-scaling approaches.
5) There is a growing body of evolution experiments and theory that can be used to understand biotic responses to multiple driver environmental change. However, these experiments and theory are framed in terms of the action of natural selection and fitness, and are often generic at the cost of being realistic. Understanding how these dynamics will play out in natural populations requires careful interpretation of the evolutionary literature, as well as bridging studies in natural populations or recent isolates.

6) We advocate the development of scientific questions that are directly relevant for society and therefore focus on solutions, policy formulation, and increased public awareness of these issues. Each of these complex questions can only be answered by a unique combination of experiments, designs, and approaches.
7) In addition to selecting the most pertinent experimental designs, the large number of permutations of global, regional, and local drivers raises issues about both the rationale for selecting drivers to be used in experiments, and the subsequent inter-comparability of experimental findings for a wide range of species, communities, locales and provinces.
8) A major challenge for the ocean global change field will be to balance this need for harmonization of multi-driver methodology with the scope and flexibility needed to encourage the continued development of novel approaches. This dynamic balance between intercomparability and creativity in experimental design will not be easy to achieve, but is vital to promote rapid progress in understanding biological responses to ocean global change.
9) This review is part of the platform of SCOR WG149 activities to develop a web-based Best Practice Guide to aid researchers new to the discipline to: navigate through the many permutations of multiple drivers; to optimise the most suitable experimental design for the questions(s) they wish to resolve; and to continue upskilling to further enhance their research into multiple drivers.

Acknowledgments

We wish to thanks SCOR for funding WG149, and the support of the IAEA for a 2017 workshop that helped to evolve many of the ideas in this Review. We acknowledge the valuable insights provided by three anonymous reviewers that helped to improve this manuscript. We are grateful to Peter Dillingham (University of Otago) for providing valuable feedback on Table 3, and to the ACE-CRC for funding an internship for Max Rintoul to carry out the literature search used in Figure 4c.

References

Albright R, Caldeira L, Hosfelt J et al. (2016) Reversal of ocean acidification enhances net coral reef calcification. Nature, 531, 362-365.

Alsterberg C, Eklof JS, Gamfeldt L, Havenhand JN, Sundback K (2013) Consumers mediate the effects of experimental ocean acidification and warming on primary producers. Proceedings of the National Academy of Sciences of the United States of America, 110, 8603-8608.

Andersson, A.J., D.I. Kline, P.J. Edmunds, S.D. Archer, N. Bednaršek, R.C. Carpenter, M. Chadsey, P. Goldstein, A.G. Grottoli, T.P. Hurst, A.L. King, J.E. Kübler, I.B. Kuffner, K.R.M. Mackey, A. Paytan, B.A. Menge, U. Riebesell, A. Schnetzer, M.E. Warner, R.C. Zimmerman. 2015. Understanding ocean acidification impacts on organismal to ecological scales. Oceanography 28, http://dx.doi.org/10.5670/ oceanog.2015.xx.

Arnold, T., C. Mealey, H. Leahey, A. W. Miller, J. M. Hall-Spencer, M. Milazzo, and K. Maers (2012) Ocean acidification and the loss of phenolic substances in marine plants. PLoS ONE 7:e35107

Bach, L. T., Mackinder, L., Schulz, K. G., Wheeler, G., Schroeder, D. C., Brownlee, C. und Riebesell, U. (2013) Dissecting the impact of CO2 and pH on the mechanisms of photosynthesis and calcification in the coccolithophore *Emiliania huxleyi*. New Phytologist, 199, 121–134

Bach, L.T., Riebesell, U., Gutowska, M.A., Federwisch, L., Schulz; K.G. (2015) A unifying concept of coccolithophore sensitivity to changing carbonate chemistry embedded in an ecological framework. Progress in Oceanography 135, 125–138.

Baretta-Bekker, J., B. Riemann, J. Baretta, and E. Rasmussen (1994), Testing the microbial loop concept by comparing mesocosm data with results from a dynamical simulation model, Mar. Ecol. Prog. Ser., 106, 187–198.

Barley SC, Meeuwig JJ (2016) The Power and the Pitfalls of Large-scale, Unreplicated Natural Experiments. Ecosystems*,* **20**, 331-339.

Barry JP, Lovera C, Buck KR *et al.* (2014) Use of a Free Ocean CO2 Enrichment (FOCE) system to evaluate the effects of ocean acidification on the foraging behavior of a deep-sea urchin. Environmental science & technology*,* 48, 9890-9897.

Baumann H., 2016. Combined effects of ocean acidification, warming, and hypoxia on marine organisms. Limnology and Oceanography e-Lectures 6:1-4.

Bell, G., (2017). Evolutionary rescue. Annual Review of Ecology, Evolution, and Systematics 48.1.

Biller, Steven J, Paul M Berube, Debbie Lindell, and Sallie W Chisholm. 2015. “Prochlorococcus: the Structure and Function of Collective Diversity.” Nat Rev Micro 13 (1). Nature Research: 13–27. doi:10.1038/nrmicro3378.

Bopp L, Resplandy L, Orr JC et al. (2013) Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. Biogeosciences, 10, 6225-6245.

Botero, C. A., Weissing, F. J., Wright, J., & Rubenstein, D. R. (2015). Evolutionary tipping points in the capacity to adapt to environmental change. Proceedings of the National Academy of Sciences, 112, 184–189. http://doi.org/10.1073/pnas.1408589111

Boyd, P.W. (2013) Framing biological responses to a changing ocean. Nature Climate Change, 3, 530-533.

Boyd, P.W., Dillingham, C.M. McGraw, E. Armstrong, C.E. Cornwall, Y.-y. Feng, C.L. Hurd, M. Gault-Ringold, M.Y. Roleda, E. Timmins-Schiffman and B.L. Nunn (2015) Physiological responses of a Southern Ocean diatom to complex future ocean conditions. Nat. Clim. Change, DOI: 10.1038/NCLIMATE2811.

Brennan, G., and S. Collins. 2015. “Growth Responses of a Green Alga to Multiple Environmental Drivers.” Nature Climate Change 5: 892–97. doi:10.1038/nclimate2682.

Brennan G.L., Colegrave N., Collins S. (2017) Evolutionary consequences of multidriver environmental change in an aquatic primary producer. Proc Natl Acad Sci U S A, 114, 9930-9935.

Bunse C, Lundin D, Karlsson CMG et al. (2016). Response of marine bacterioplankton pH homeostasis gene expression to elevated CO_2. Nature Climate Change 6: DOI:10.1038/NCLIMATE2914

Byrne M, Przeslawski R (2013) Multistressor Impacts of Warming and Acidification of the Ocean on Marine Invertebrates’ Life Histories. Integrative and Comparative Biology doi:10.1093/icb/ict049

Capone D.G. and Hutchins D.A. (2013). The microbial biogeochemistry of coastal upwelling regimes in a changing ocean. Nature Geoscience 6(9): 711-715. doi:10.1038/ngeo1916

Carlson, Stephanie M, Curry J Cunningham, and Peter A H Westley. 2014. “Evolutionary Rescue in a Changing World.” Trends in Ecology & Evolution 29 (9): 521–30. doi:10.1016/j.tree.2014.06.005.

Chevin, L.M., R. Lande, and G.M. Mace. 2010. “Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory.” PLoS Biology 8. Public Library of Science: e1000357.

Collins, S., & de Meaux, J. (2009). Adaptation to different rates of environmental change in Chlamydomonas. Evolution; International Journal of Organic Evolution, *63*(11), 2952–2965. http://doi.org/10.1111/j.1558-5646.2009.00770.x

Collins S. and A. Gardner (2009) Integrating physiological, ecological and evolutionary timescales: a Price equation approach. Ecol. Lett. 12:744-757 doi: 10.1111/j.1461-0248.2009.01340.x

Collins, S., Rost, B., & Rynearson, T. A. (2013). Evolutionary potential of marine phytoplankton under ocean acidification. Evolutionary Applications, 7, 140–155. http://doi.org/10.1111/eva.12120

Collins, S., Sültemeyer, D., & Bell, G. (2006). Changes in C uptake in populations of *Chlamydomonas reinhardtii* selected at high CO2. Plant, Cell & Environment, 29, 1812–1819. http://doi.org/10.1111/j.1365-3040.2006.01559.x

Collins, S., J. de Meaux, and C. Acquisti. 2007. Adaptive Walks Toward a Moving Optimum. Genetics 176 (2): 1089–99. doi:10.1534/genetics.107.072926.

Cornwall CE, Hepburn CD, McGraw CM et al. (2013) Diurnal fluctuations in seawater pH influence the response of a calcifying macroalga to ocean acidification. Proceedings of the Royal Society B-Biological Sciences, 280.

Crain C. M., Kroeker K. & Halpern B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11:1304-1315.

Cumming G. (2008) Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. Perspectives on Psychological Science, 3, 286-300.

Darling E. S. & Côté I. M., 2008. Quantifying the evidence for ecological synergies. Ecology Letters 11:1278-1286.

Davies G. M. & Gray A., 2015. Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). Ecology and Evolution 5:5295-5304.

Dore, J.E., R. Lukas, D. W. Sadler, M.J. Church, and D. M. Karl (2009) Physical and biogeochemical modulation of ocean acidification in the central North Pacific. PNAS, 106, 12235–12240, doi: 10.1073/pnas.0906044106.

Dunne, J.A., S.R. Saleska, M.L. Fischer, and J. Harte (2004) Integrating Experimental and Gradient Methods in Ecological Climate Change Research. Ecology, 85, 904–916.

Dupont S. & Pörtner H., 2013. Get ready for ocean acidification. Nature 498:429.

Dupont S, Havenhand J, Thorndyke W, Peck L, Thorndyke M (2008) Near-future level of CO_2-driven ocean acidification radically affects larval survival and development in the brittlestar *Ophiothrix fragilis*. Marine Ecology-Progress Series, **373**, 285-294.

Dupont S, Puncher G & Calosi P (2015) Bird is the word – on the importance of ethical and effective scientific communication. Journal of the Marine Biological Association of the United Kingdom. 95: 863-864.

Dupont S, Hall E, Calosi P & Lundve B (2014). First evidence of altered sensory quality in a shellfish exposed to decreased pH relevant for ocean acidification. Journal of Shellfish Research. 33:857-861.

Dupont S (2017a) The OA-Africa network – Putting Africa on the ocean acidification map. Solas event report. 2: 4-5.

Dupont S (2017b) I am the Ocean – arts and sciences to move from ocean literacy to passion for the ocean Journal of the Marine Biological Association of the United Kingdom. DOI: https://doi.org/10.1017/S0025315417000376.

Edwards, M., Beaugrand, G., Helaouet, P., Coombs, S. 2013. Marine ecosystem response to the Atlantic Multidecadal Oscillation. PLoS ONE, 8: e57212.

Elena, S.F, and R.E Lenski. 2003. “Evolution Experiments with Microorganisms: the Dynamics and Genetic Bases of Adaptation.” Nature Reviews Genetics 4 (6). Nature Publishing Group: 457–69.

Edwards, K. F., Thomas, M. K., Klausmeier, C. A., & Litchman, E. (2016). Phytoplankton growth and the interaction of light and temperature: A synthesis at the species and community level. Limnology and Oceanography, 61(4), 1232–1244. http://doi.org/10.1002/lno.10282

Eriander L, Wrange A-L, Havenhand J (2015) Simulated diurnal pH fluctuations radically increase variance in—but not the mean of—growth in the barnacle Balanus improvisus. ICES Journal of Marine Science: Journal du Conseil, fsv214.

Eppley RW (1972) Temperature and phytoplankton growth in the sea. Fish. Bull*,* **70**, 1063-1085.
Fabricius KE, Langdon C, Uthicke S et al. (2011) Losers and winners in coral reefs acclimatized to elevated carbon dioxide concentrations. Nature Clim. Change, 1, 165-169.

Fabricius KE, De’ath G, Noonan S, Uthicke S. (2014) Ecological effects of ocean acidification and habitat complexity on reef-associated macroinvertebrate communities. Proc. R. Soc. B 281: 20132479.

Folke, C., T. Hahn, P. Olsson, J. Norberg (2005) Adaptive governance of social-ecological systems. Annu. Rev. Environ. Resour., 30, 441-473.

Gao, K., Xu, J., Gao, G., Li, Y., Hutchins, D. A., Huang, B., et al. (2012). Rising CO_2 and increased light exposure synergistically reduce marine primary productivity. *Nature Climate Change*. http://doi.org/10.1038/nclimate1507

Gao K., Helbling E.W., Häder D-P., and Hutchins D.A. (2012). Responses of marine primary producers to interactions between ocean acidification, solar radiation, and warming. Marine Ecology Progress Series 470: 167-189. doi: 10.3354/meps10043.

Garland, T., Jr., and M. R. Rose, eds. 2009. Experimental evolution: concepts, methods, and applications of selection experiments. University of California Press, Berkeley, California.

Gattuso J.-P., Bijma J., Gehlen M., Riebesell U. & Turley C., 2011. Ocean acidification: knowns, unknowns and perspectives. In: Gattuso J.-P. & Hansson L. (Eds.), Ocean acidification, pp. 291-311. Oxford: Oxford University Press.

Gattuso J.-P., Kirkwood W., Barry J. P., Cox E., Gazeau F., Hansson L., Hendriks I. E., Kline D. I., Mahacek P., Marker M., Martin S., McElhany P., Peltzer E. T., Reeve J., Roberts D., Saderne V., Tait K., Widdicombe S. & Brewer P., 2014. Free-ocean CO_2 enrichment (FOCE) systems: present status and future developments. Biogeosciences 11:4057-4075.

Gattuso J.-P., Magnan A., Billé R., Cheung W. W. L., Howes E. L., Joos F., Allemand D., Bopp L., Cooley S., Eakin C. M., Hoegh-Guldberg O., Kelly R. P., Pörtner H., Rogers A. D., Baxter J. M., Laffoley D., Osborn D., Rankovic A., Rochette J., Sumaila U. R., Treyer S. & Turley C., 2015. Contrasting futures for ocean and society from different anthropogenic CO_2 emissions scenarios. Science 349:aac4722.

Gaylord B, Kroeker KJ, Sunday JM et al. (2015) Ocean acidification through the lens of ecological theory. Ecology, 96, 3-15.

Geider, R.J., MacIntyre, H.L. & Kana, T.M. (1996). A dynamic model of photoadaptation in phytoplankton. Limnol. Oceanogr., 41: 1-15.

Ghalambor, C. K, K.L Hoke, E.W. Ruell, E. K Fischer, D.N. Reznick, and K.A. Hughes. 2015. "Non-Adaptive Plasticity Potentiates Rapid Adaptive Evolution of Gene Expression in Nature." Nature 525: 372–75. doi:10.1038/nature15256.

Gibbs, S.J., P.R. Bown, A. Ridgwell, J.R. Young, A.J. Poulton and S.A. O'Dea (2016) Ocean warming, not acidification, controlled coccolithophore response during past greenhouse climate change. Geology, 44, 59-62. doi: 10.1130/G37273.1

Godhe, R.F. and A.L Gamfeldt (2012) Genetic Diversity and Ecosystem Functioning in the Face of Multiple Stressors. PLoS one, 7, e45007.

Goussen, B., O.R. Price, C. Rendal & R. Ashauer (2016) Integrated presentation of ecological risk from multiple stressors. Scientific Reports, 6:36004. DOI: 10.1038/srep36004

Gradoville, M.R., White, A.E., Böttjer, D., Church, M.J., Letelier, R.M. (2014). Diversity trumps acidification: Lack of evidence for carbon dioxide enhancement of *Trichodesmium* community nitrogen or carbon fixation at Station ALOHA. Limnol. Oceanogr 59, 645–659.

Gunderson, A.R., E.J. Armstrong, and J.H. Stillman (2016) Multiple Stressors in a Changing World: The Need for an Improved Perspective on Physiological Responses to the Dynamic Marine Environment. Annu. Rev. Mar. Sci. 8:357–78.
doi: 10.1146/annurev-marine-122414-033953

Gunst R.F., and Mason R.L. (2009) Fractional factorial design. Wiley Interdisciplinary Reviews: Computational Statistics, 1, 234-244.

Hairston, N. G., SR. (1989) Ecological experiments: Purpose, design, and execution. Cambridge.

Harrison. P. J., Parslow, J. S., Conway, H. L. (1989). Determination of nutrient uptake kinetic parameters: a comparison of methods. Mar Ecol. Prog. Ser. 52: 301-312.

Harvey BP, Gwynn-Jones D, Moore PJ (2013) Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. Ecology and Evolution*,* **3**, 1016-1030.

Havenhand JN, Dupont S, Quinn GP (2010) Designing ocean acidification experiments to maximise inference. In: Guide to best practices for ocean acidification research and data reporting. (eds Riebesell U, Fabry VJ, Hansson L, J.-P. G) p 67-80. Luxembourg, Publications Office of the European Union.

Hegmann, G., C. Cocklin, R. Creasey, S. Dupuis, A. Kennedy, L. Kingsley, W. Ross, H., Spaling, and D. Stalker. 1999. "Cumulative Effects Assessment Practitioners Guide.", Prepared for the Canadian Environmental Assessment Agency by the Cumulative, Effects Assessment Working Group and AXYS Environmental Consulting. Ltd.http://www.ceaa-acee.gc.ca/default.asp?lang=En&n=43952694-1

Hillebrand H, Matthiessen B (2009) Biodiversity in a complex world: consolidation and progress in functional biodiversity research. Ecology Letters, 12, 1405-1419.

Hughes, T.P. et al. (2017) Global warming and recurrent mass bleaching of corals. Nature 543, 373–377. doi:10.1038/nature21707

Hutchins, D.A., Mulholland, M.R. and Fu, F.-X. (2009). Nutrient cycles and marine microbes in a CO_2-enriched ocean. Oceanography **22**: 128-145

Hutchins, D.A., Fu F.X., Webb E.A., Walworth N., and Tagliabue, A. (2013). Taxon-specific response of marine nitrogen fixers to elevated carbon dioxide concentrations. Nature Geoscience 6(9): 790-795. doi: 10.1038/ngeo1858

Hutchins, D. A., Walworth, N., Webb, E.A., Saito, M. A., Moran, D., McIlvin, M. R., Gale, J., Johnson, C., and Fu, F.-X. (2015). Irreversibly increased N_2 fixation in *Trichodesmium* experimentally adapted to high CO_2. *Nature Communications* 6: 8155. doi:10.1038/ncomms9155

Hutchins, D.A. and Boyd, P.W. (2016). Marine phytoplankton and the changing ocean iron cycle. *Nature Climate Change* 6: 1071-1079.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2014: Summary for policymakers. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1-32.

Kazamia, E., K.E. Helliwell, S. Purton and A.G. Smith (2016) How mutualisms arise in phytoplankton communities: building eco-evolutionary principles for aquatic microbes. Ecology Letters, (2016) 19: 810–822.

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. (2012) Experimental evolution. Trends in ecology & evolution, 27:547-60.

Krause E, Wichels A, Giménez L, Lunau M, Schilhabel MB, Gerdts G (2012) Small changes in pH have direct effects on marine bacterial community composition: a microcosm approach. PloS one, 7, e47035.

Kroeker, K.J., R.L. Kordas, R.N. Crim, G.G. Singh (2010) Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. Ecology Letters (2010) 13: 1419–1434. DOI: 10.1111/j.1461-0248.2010.01518.

Lachapelle, J., Bell, G., & Colegrave, N. (2015). Experimental adaptation to marine conditions by a freshwater alga. Evolution; International Journal of Organic Evolution. http://doi.org/10.1111/evo.12760

Lande, R. 2014. Evolution of Phenotypic Plasticity and Environmental Tolerance of a Labile Quantitative Character in a Fluctuating Environment. Journal of Evolutionary Biology 27 (5): 866–75. doi:10.1111/jeb.12360.

Law, C. S. et al. No stimulation of nitrogen fixation by non-filamentous diazotrophs under elevated CO2 in the South Pacific. Glob. Change Biol. 18, 3004-3014 (2012).

Lenski RE (2017) What is adaptation by natural selection? Perspectives of an experimental microbiologist. PLoS Genet 13(4): e1006668.

Listmann, L., M. LeRoch, L. Schluter, M. K. Thomas and T. B. H. Reusch (2016) Swift thermal reaction norm evolution in a key marine phytoplankton species. Evolutionary Applications ISSN 1752-4571, doi:10.1111/eva.12362

Lohbeck, K.T., Riebesell, U., Reusch, T.B.H. (2012a) Adaptive evolution of a key phytoplankton species to ocean acidification. Nature Geoscience 5, 346–351, DOI: 10.1038/NGEO1441

Lohbeck, K. T., Riebesell, U., Collins, S. and Reusch, T. B. H. (2012b) Functional Genetic Divergence in High CO2 adapted Emiliania huxleyi populations. Evolution 67, 1892-1900

Lohbeck, K.T., Riebesell, U., Reusch, T.B.H. (2014) Gene expression changes in the coccolithophore *Emiliania huxleyi* after 500 generations of selection to ocean acidification. Proc. Royal Soc. B.; doi:10.1098/rspb.2014.0003

Lorena, A., G. M. Marques, S. A. L. M. Kooijman, and T. Sousa. (2010). Stylized facts in microalgal growth: interpretation in a dynamic energy budget context. Philosophical Transactions of the Royal Society B-Biological Sciences 365:3509-3521.

Low-Décarie, E, G F Fussmann, and G Bell. 2011. "The Effect of Elevated CO2 On Growth and Competition in Experimental Phytoplankton Communities." Global Change Biology, 2525–35. doi:10.1111/j.1365-2486.2011.02402.x.

Magnan, A.K., M. Colombier, J.-P. Gattuso (2015) Implications of the Paris agreement for the ocean. Nature Climate Change 6, 732–735

Meier HEM, Andersson HC, Arheimer B et al. (2012) Comparing reconstructed past variations and future projections of the Baltic Sea ecosystem-first results from multi-model ensemble simulations. Environmental Research Letters, 7.

Mejlholm, O., Dalgaard, P., 2009. Development and validation of an extensive growth and growth boundary model for *Listeria monocytogenes* in lightly preserved and ready-to-eat shrimp. Journal of Food Protection 72, 2132–2143.

Miller GM, Watson SA, Donelson JM, McCormick MI, Munday PL (2012) Parental environment mediates impacts of increased carbon dioxide on a coral reef fish. Nature Climate Change*,* **2**, 858-861.

Moustaka-Gouni, M., Kormas, K. A., Scotti, M., Vardaka, E. and Sommer, U. (2016) Warming and Acidification Effects on Planktonic Heterotrophic Pico- and Nanoflagellates in a Mesocosm Experiment Protist, 167, 389-410. DOI 10.1016/j.protis.2016.06.004.

Müller, M. N., Schulz, K. G., & Riebesell, U. (2010). Effects of long-term high CO_2 exposure on two species of coccolithophores. *Biogeosciences*, *7*(3), 1109–1116. http://doi.org/10.5194/bg-7-1109-2010

Muller, E. B., and R. M. Nisbet. (2014). Dynamic energy budget modeling reveals the potential of future growth and calcification for the coccolithophore *Emiliania huxleyi* in an acidified ocean. Global Change Biology 20:2031-2038.

Mustonen, V. and Lässig, M. (2009). "From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation". Trends in genetics : TIG. 25 (3): 111–9.

Oesterwind, D., A. Rau, A. Zaiko (2016) Drivers and pressures: Untangling the terms commonly used in marine science and policy. J. Env. Manag. 181: 8-15.

Paine R. T., Tegner M. J. & Johnson E. A., 1998. Compounded perturbations yield ecological surprises. Ecosystems 1:535-54.

Pansch C, Schaub I, Havenhand J, Wahl M (2014) Habitat traits and food availability determine the response of marine invertebrates to ocean acidification. Global Change Biology, 20, 765-777.5.

Parker LM, O'Connor WA, Raftos DA, Pörtner H-O, Ross PM (2015) Persistence of positive carryover effects in the oyster, Saccostrea glomerata, following transgenerational exposure to ocean acidification. PloS one, 10, e0132276.

Phillimore, A.B., J.D. Hadfield, O.R. Jones, and R.J. Smithers (2010) Differences in spawning date between populations of common frog reveal local adaptation. PNAS, 107, 8292-8297. doi:10.1073/pnas.0913792107

Poore A. G. B., Graba-Landry A., Favret M., Sheppard Brennand H. B., M & Dworjanyn S. A., 2013. Direct and indirect effects of ocean acidification and warming on a marine plant–herbivore interaction. Oecologia 173:1113-1124.

Portner, H.-O. and M.K. Grieshaber (1993) Critical PO2 (s) in oxyconforming and oxyregulating animals: gas exchange, metabolic rate and the mode of energy production. In: The Vertebrate Gas Transort Cascade: Adaptations to Environment and Mode of Life [Bicudo, J.E.P.W (ed)]. CRC Press Inc, Boca Raton, FL, USA, pp. 330-357.

Pörtner H.-O. (2012) Integrating climate-related stressor effects on marine organisms: unifying principles linking molecule to ecosystem-level changes. Mar. Ecol. Prog. Ser. 2012;470:273–290.

Quinn G. and Keough M. (2002) Experimental design and data analysis for biologists, Cambridge University Press.

Ridgwell, A., Schmidt, D. N., Turley, C., Brownlee, C., Maldonado, M. T., Tortell, P., and Young, J. R., From laboratory manipulations to Earth system models: scaling calcification impacts of ocean acidification, Biogeosciences 6, 2611-2623 (2009).

Riebesell, U., Gattuso, J.-P., Thingstad, T. F., and Middelburg, J.J. (2013) Arctic ocean acidification: pelagic ecosystem and biogeochemical responses during a mesocosm study. Biogeosciences, 10, 5619–5626

Riebesell U. and J-P. Gattuso (2014) Lessons learned from ocean acidification research. Nature Climate Change 5, 12–14 (2015) doi:10.1038/nclimate2456

Riebesell, U., Czerny, J., von Bröckel, K., Boxhammer, T., Büdenbender, J., Deckelnick, M., Fischer, M., Hoffmann, D., Krug, S. A., Lentz, U., Ludwig, A., Muche, R., and Schulz, K. G. (2013) Technical Note: A mobile sea-going mesocosm system – new opportunities for ocean change research. Biogeosciences 10, 1835-1847.

Riebesell, U., Bach, L. T., Bellerby, R. G. J., Bermudez, R., Boxhammer, T., Czerny, J., Larsen, A., Ludwig, A., Schulz,, K. G. (2017) Ocean acidification impairs competitive fitness of a predominant pelagic calcifier. Nature Geoscience 10, 19-24, doi:10.1038/ngeo2854

Ries, J.B., A.L. Cohen and D.C. McCorkle (2009) Marine calcifiers exhibit mixed responses to CO2-induced ocean acidification. Geology, 37, 1131-1134. doi: 10.1130/G30210A.1

Rivero-Calle, S., A. Gnanadesikan, C.E. Del Castillo, W.M. Balch, S.D. Guikema (2015) Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO_2. Science 350, 1533-1537.

Roger F, Godhe A, Gamfeldt L (2012) Genetic Diversity and Ecosystem Functioning in the Face of Multiple Stressors. PLoS one, **7**, e45007.

Rossoll, D., R. Bermudez, H. Hauss, K. G. Schulz, U. Riebesell, U. Sommer, and M. Winder. (2012) Ocean acidification-induced food quality deterioration constrains trophic transfer. PLoS ONE 7:e34737.

Rynearson, T A, and E V Armbrust. 2000. "DNA Fingerprinting Reveals Extensive Genetic Diversity in a Field Population of the Centric Diatom Ditylum Brightwellii." Limnology and Oceanography 45: 1329–40.

Saito, M., T.J. Goepfert and J.T. Ritt (2008) Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. Limnol. Oceanogr., 53(1), 2008, 276–290.

Sajitz-Hermstein M, Nikoloski Z (2013) Structural Control of Metabolic Flux. PLoS Comput Biol 9(12): e1003368. doi:10.1371/journal.pcbi.1003368

Schaum, C.E., and S.Collins. 2014. Plasticity Predicts Evolution in a Marine Alga. Proceedings of the Royal Society B: Biological Sciences 281 (October): 20141486. doi:10.1098/rspb.2014.1486.

Schaum, E.C., B. Rost, and S. Collins. 2015. Environmental Stability Affects Phenotypic Evolution in a Globally Distributed Marine Picoplankton..The ISME Journal 10: 75–84. doi:10.1038/ismej.2015.102.

Scheinin, M., Riebesell, U., Rynearson, T. A., Lohbeck, K. T., & Collins, S. (2015). Experimental evolution gone wild. Journal of the Royal Society Interface, 12(106), 20150056–20150056. http://doi.org/10.1098/rstb.2012.0437

Scheiner, S. M., and J. Gurevitch [Eds.]. (1993) The design and analysis of ecological experiments. Chapman & Hall, New York. 445 p. ISBN o-412-0356 1-8.

Schulze, E.D., Scholes, R.J., Ehleringer, J., Hunt, L.A., Canadell, J., Chapin, F.S. & Steffen, W.L. (1999) The study of ecosystems in the context of global change. The Terrestrial Biosphere and Global Change – Implications for Natural and Managed Ecosystems (eds B.Walker, W.Steffen, J.Canadell & J.Ingram), pp. 19–44. Cambridge University Press, Cambridge.

Schlüter, L., K.T Lohbeck, M.A Gutowska, J.P. Gröger, U. Riebesell, and T.B.H. Reusch. 2014. "Adaptation of a Globally Important Coccolithophore to Ocean Warming and Acidification." Nature Climate Change 4 (11): 1024–30. doi:10.1038/nclimate2379.

Schulz K.G., Bach L.T., Bellerby R.G.J., Bermúdez R., Büdenbender J., Boxhammer T., Czerny J., Engel A., Ludwig A., Meyerhöfer M., Larsen A., Paul A.J., Sswat M., Riebesell U. (2017) Phytoplankton blooms at increasing levels of atmospheric carbon dioxide: Experimental evidence for negative effects on prymnesiophytes and positive on small picoeukaryotes. Front. Mar. Sci. 4:64. doi: 10.3389/fmars.2017.00064

Schmidt D. and P.W. Boyd (2016) Forecast ocean variability. Nature, 539, 162-163.

Sett, S., Bach. L.T., Koch-Klavsen, S., Lebrato, M., Oschlies, A., Riebesell, U., Schulz, K.G. (2014) Temperature modulates coccolithophorid sensitivity of growth, photosynthesis and calcification to increasing seawater pCO_2. PLoS ONE 9(2). e88308. DOI 10.1371/ journal.pone.0088308

Smith, J.N., De'ath, G., Richter, C., Cornils, A., Hall-Spencer, J.M., Fabricius, K.E. (2016) Ocean acidification reduces demersal zooplankton that reside in tropical coral reefs. Nature Climate Change. doi: 10.1038/NCLIMATE3122

Sommer, U. (2012) Experimental Systems in Aquatic Ecology In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0003180.pub2]

Sommer U., Adrian R., Bauer B. & Winder M., 2012. The response of temperate aquatic ecosystems to global warming: novel insights from a multidisciplinary project. Marine Biology 159:2367-2377.

Sosik, H. M., and R. J. Olson. 2007. Automated taxonomic classification of phytoplankton sampled with imaging-inflow cytometry. Limnol. Oceanogr. Methods 5:204-216

Stewart R. I. A., Dossena M., Bohan D. A., Jeppesen E., Kordas R. L., Ledger M. E., Meerhoff M., Moss B., Mulder C., Shurin J. B., Suttle B., Thompson R., Trimmer M. & Woodward G., 2013. Mesocosm experiments as a tool for ecological climate-change research. Advances in Ecological Research 48:71-181.

Sutherland WJ, Freckleton RP, Godfray HCJ et al. (2013) Identification of 100 fundamental ecological questions. Journal of Ecology, 101, 58-67.

Taucher, J., Jones, J., James, A., Brzezinski, M. A., Carlson, C. A., Riebesell, U., & Passow, U. (2015). Combined effects of CO_2 and temperature on carbon uptake and partitioning by the marine diatoms *Thalassiosira weissflogii* and *Dactyliosolen fragilissimus*. Limnology and Oceanography, http://doi.org/10.1002/lno.10063

Taucher, J., Haunost, M., Boxhammer, T., Bach, L.T., Alguero-Muniz, M., Riebesell, U. (2017) Influence of ocean acidification on plankton community structure during a winter-tosummer succession: An imaging approach indicates that copepods can benefit from elevated CO_2 via indirect food web effects. PLoS ONE 12(2), e0169737. doi:10.1371/journal.pone.0169737

Tatters, A. O., Roleda, M. Y., Schnetzer, A., Fu, F., Hurd, C. L., Boyd, P. W., et al. (2013a). Short- and long-term conditioning of a temperate marine diatom community to acidification and warming. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1627), 20120437–20120437. http://doi.org/10.1098/rspb.2012.2598

Tatters, A.O., Schnetzer, A., Fu, F.-X., Lie, A.Y.A., Caron, D.A. and Hutchins, D.A. (2013b) Short- versus long-term responses to changing CO_2 in a coastal dinoflagellate bloom: Implications for interspecific competitive interactions and community structure. *Evolution* 67-7: 1879–1891: doi:10.1111/evo.12029

Thomsen J, Casties I, Pansch C, Kortzinger A, Melzner F (2013) Food availability outweighs ocean acidification effects in juvenile Mytilus edulis: laboratory and field experiments. Global Change Biology, 19, 1017-1027.

Thor P., and Dupont S (2015) Transgenerational effects alleviate severe fecundity loss during ocean acidification in a ubiquitous planktonic copepod. Global change biology, 21, 2261-2271.

Thresher RE, Tilbrook BD, Fallon S, Wilson NC, Adkins J. (2011) Effects of chronic low carbonate saturation levels on the distribution growth and skeletal chemistry of deep-sea corals and other seamount megabenthos. Marine Ecology Progress Series. 2011;442:87–99. doi: 10.3354/meps09400.

Vargas C, Lagos N, Lardies M, Duarte C, Manríquez P, Aguilera V, Broiman B, Widdicombe S & Dupont S (2017) Species-specific responses to ocean acidification should account for local adaptation and adaptive plasticity. Nature Ecology and Evolution. 1:84.

Walworth, N.G., Lee, M.D., Fu, F.-X., Hutchins, D.A. and Webb, E.A. (2016a). Molecular and physiological evidence of genetic assimilation to high CO_2 in the marine nitrogen fixer *Trichodesmium.* Proceedings of National Academy of Sciences USA*:* E7367–E7374 doi/10.1073/pnas.1605202113

Walworth, N.G., Fu, F.-X., Webb, E.A, Saito, M.A., Moran, D., Mcllvin, M.R., Lee, M.D., and Hutchins, D.A. (2016b) Mechanisms of increased *Trichodesmium* fitness under iron and phosphorus co-limitation in the present and future ocean. Nature Communications doi10.1038/NCOMMS12081

Wernberg T., Smale D. & Thomsen M., (2012). A decade of climate change experiments on marine organisms: procedures, patterns and problems. Global Change Biology 18:1491-1498.

Worden, A.Z., M.J. Follows, S.J. Giovannoni, S. Wilken, A.E. Zimmerman, P.J. Keeling (2015) Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science, 347, DOI: 10.1126/science.1257594

Xu, K., Fu, F.-X. and Hutchins, D.A. (2014). Comparative responses of two dominant Antarctic phytoplankton taxa to interactions between ocean acidification, warming, irradiance, and iron availability. Limnology and Oceanography 59: 919- 931.

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

Yang Y., Hansson L. & Gattuso J.-P., 2016. Data compilation on the biological response to ocean acidification: an update. Earth System Science Data 8:79-87.

Zhang, Y., Klapper, R., Lohbeck, K.T., Bach, L.T., Schulz, K.G., Reusch, T.B.H., Riebesell, U. (2014) Between- and within-population variations in thermal reaction norms of the coccolithophore *Emiliania huxleyi*. Limnol Oceanogr. 59, 1570–1580.

Tables

Table 1 A selection of seminal reviews, syntheses, and overview papers mainly from the terrestrial literature that present the underlying precepts for the design of physiological, ecological and evolutionary experiments that are discussed here in the context of ocean global change biology and ecology. Note Sommer (2012) is an online electronic version of his 2003 publication.

Discipline	Principles	Reference
Physiology and Ecology	Experiments – design and analysis	Quinn and Keough (2002)
Ecology	Experimental design and analysis	Scheiner and Gurevitch (1993)
Terrestrial Ecology	Experimental methods and their integration	Dunne et al. (2004)
Ecology	Ecosystem studies and global change	Schulze et al. (1999)
Evolution	Experimental design (microbes)	Elena and Lenski (2003)
Evolution	Experiments: theories, approaches, functions	Garland and Rose (2009)
Terrestrial Evolution	Population genetic: space for time substitutions	Phillmore et al. (2010)
Aquatic Ecology	Scale of experimentation; realism versus control	Sommer (2012)
Physiology / Marine Biology	Physiology across scales	Pörtner (2012)
Physiology / Marine Sciences	Multiple drivers and their interplay	Saito et al. (2008)

Table 2 Summary of the main experimental approaches used in multiple driver research, their advantages, disadvantages, and which research themes or fora they have mainly been used in. Note, many of the research questions posed throughout this review cannot be solved by one single experiment or experimental approach. Scenario-based experiments not only permit more replication (because of fewer treatments and treatment combinations), and hence greater statistical power, within the available resources, but also enable tests of more drivers, in different combinations, and/or at more levels. This is essential for identifying emerging patterns of how drivers interact (e.g. Brennan & Collins 2015). The benefits of such scenario testing include the development of practical methods to test for multi-driver effects that integrate the modulating effects of interacting drivers, and which can be applied beyond the species-level (i.e. in community-level experimentation).

Experimental Approach	**Examples**	**Benefits**	**Disadvantages**	**Main uses**
Single driver /mechanistic	Warming (Eppley, 1972)	Intrinsic physiological status; Ability to build models (mathematical or conceptual) from studies of single driver and modes of action, and to iterate this 'loop' (Baretta-Bekker,et al. 1994).	No information on relative influence of other drivers	Reaction norm and Reciprocal interface with models
Single driver /constant conditions	Acidification (Dupont et al., 2008)	Specific response to projected future conditions which can be invaluable if a sole driver is dominant (temperature/coral bleaching, Hughes et al. 2017).	No information on relative influence of other drivers, no information on ecological relevance (lack of realism)	Response to IPCC projections
Single driver/ fluctuations	Acidification (Cornwall et al., 2014, Eriander et al., 2015)	Specific response to projected future conditions and to the influence of natural environmental variability	No information on relative influence of other drivers, no information on ecological relevance (lack of realism)	Response to IPCC projections
Single driver / competition experiment	Acidification (Krause et al., 2012)	Competition as opposed to single species	No information on relative influence of other drivers, limited information on ecological, relevance, (lack of realism)	Comparative physiology, Community ecology
Single driver / community	FOCE, in situ pelagic mesocosms (Riebesell et al., 2013; Barry et al., 2014) seeps (Fabricius et al., 2014)	*In situ* removes many laboratory artifacts Community as opposed to species response. Pre-adapted communities (seeps)	Logistically challenging, no information on relative influence of other drivers	Comparative physiology, Community ecology

Single driver / evolution	Acidification/adaptation Schaum and Collins (2014)	Connects plastic and evolutionary responses, specific responses to projected future conditions	No information on relative influence of other drivers; size of experiments limits use to model species (but see Scheinin et al., 2015).	Microevolution
2 or 3 way multiple driver /one species	Warming and acidification (Parker et al., 2009)	Individual versus interactive effects	no information on ecological relevance (lack of realism)	Comparative physiology
4 way multiple driver / one species	Warming, acidification, light and trace metals Xu et al. (2015)	Individual versus interactive effects	Difficult to conduct and also interpret, no information on ecological relevance (lack of realism)	Comparative physiology
Multiple driver/ competition experiment	Warming/Acidification Moustaka-Gouni et al. (2016) (2 drivers)	Competition as opposed to single species	limited information on ecological relevance (lack of realism)	Comparative physiology
Multiple driver / community	Alsterberg et al. (2013)	Direct and indirect effects, synergies and antagonisms	Logistically difficult and resource intensive,	Response to IPCC projections Community ecology
Multiple driver / evolution	Brennan et al. (2017)	General evolutionary mechanism and limits; connects plastic and evolutionary responses	Logistically challenging and time-intensive, no information on ecological relevance (lack of realism)	Microevolution
Multiple driver / 'collapsed design'	Boyd et al. (2015)	Cumulative effects and influence of individual versus interactive effects	no information on ecological relevance (lack of realism)	Reaction norm Response to IPCC projections
Multiple driver / 'fractional design'	Gunst and Mason (2009)	Efficient testing of main effects in large multi-driver designs	no intermediate driver levels; frequently lack interaction terms	Identify key drivers in multi-driver factorial designs
Multiple driver / 'reduced design'		Cumulative combined effects; Increased power to test hypothesis of interest	no information on ecological relevance (lack of realism)	Reaction norm Response to IPCC projections

Table 3 Definition of terminology relevant for multiple driver research.

Term	**Approximate equivalents/ synonyms and proxies**	**Definitions**	**Comments**
Driver	Stressor, agent, predictor	An environmental factor that is tested for its effect on biological performance/biological systems.	Attempts to harmonize use of "driver" and "pressure"recommend the DPSIRS context (see Oesterwind *et al* 2016 *J. Env. Manag.* 181: 8-15)
Response	Effect, impact	A measure of biological performance following an event/perturbation	Responses may be at the level of genetics, biochemistry, energetics, physiology, population and community ecology, etc.
Response norm, response curve	Reaction norm	The response of a phenotype, or population ("species") to different environments	Typically applied to clones, individuals, or (occasionally) groups of individuals to describe responses to multiple levels of a driver. Rarely applied to multiple drivers although this is possible (e.g. Fig. 5)
Effect size		Magnitude of response, compared to control or reference conditions.	Typically measured by differences in mean, or by slope of regression line, or other statistical model.
Additive effect	Aggravating or mitigating	In a statistical sense – models without interactions.	In a general sense – a term used to describe the response of an organism or ecosystem to multiple drivers, where the presence of one driver does not alter the effect size of another driver.
Multiplicative effect	Aggravating or mitigating	In a statistical sense – models containing a term where one or more variables are multiplied together, and are thus not additive.	This is the most common form of interactive effect (see below) used in statistical models.
Interactive effect		Two or more independent drivers interact if the effect of one of the drivers differs depending on the presence/intensity of another driver (on the modelled scale). Interactions are non-additive (i.e. they are multiplicative).	The presence of an interaction can only be reliably assessed *on the observed scale*, i.e. for unbounded data. In other settings, terms are assessed *on the modelled scale*. For example, a response that shows a multiplicative effect on the observed scale in a linear model in response to two environmental factors, may show no interaction in a generalised linear model on the log scale (here, effects are additive). Hence, the model type and scales need to be specified when assessing the presence of interactions.
Synergistic effect	Aggravating	Several drivers act in the same direction, and their combined effect on a response is greater than the sum of the effects of the individual drivers. Opposite: antagonistic.	Commonly used in multiple driver studies to refer to aggravating interactions, indicating that the presence of one driver amplifies the response to another driver. There has been some confusion about usage and therefore we suggest emphasizing the direction and intensity of the joint effects at any one level of drivers.
Antagonistic effect	Mitigating	Several drivers act in opposition, i.e., the combined effect of several drivers is smaller than the sum of the individual effects. Commonly used in multiple driver studies to refer to	Commonly used in multiple driver studies to refer to mitigating interactions, indicating that the presence of one driver ameliorates the response to another driver. The same caveats apply as for synergism

	mitigating interactions, indicating that the presence of one driver ameliorates the response to another driver.	(above).
Cumulative effect	'Cumulative effects are changes to the environment that are caused by an action in combination with other past, present and future human actions' (Hegman et al. 1999).	This may be caused by either a single driver acting repeatedly or over prolonged periods of time, and/or multiple drivers that coincide or act successively.

Figure 1 Strengths (left column) and limitations (right column) of the five main approaches (center, rectangles) used to understand the effect of environmental drivers on marine biota. Major approaches include: Paleoceanographic studies of past natural climate shifts (Paleo-Proxies) such as the PETM event ~56 million years ago; Modern natural environments that can serve as proxies of particular anthropogenic change processes (Modern Proxies), such as acidification resulting from seafloor CO_2 vents or regions where naturally low-pH seawater is upwelled; Modern observations that capture extended temporal or spatial aspects of global change, including decadal-scale ocean monitoring sites such as the Bermuda Atlantic Time Series (BATS); Manipulative microcosm experiments often used to carry out controlled experimentation on single species or small communities; and Large volume mesoscosm experiment enclosures and Free Ocean CO_2 Enrichment (FOCE) experiments that are used to manipulate entire marine communities.

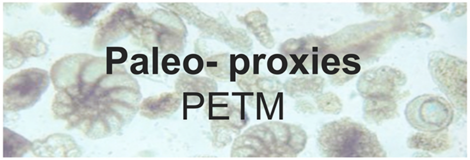
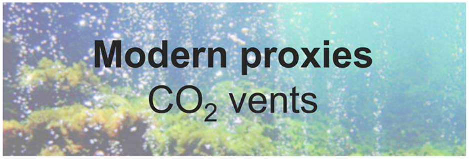
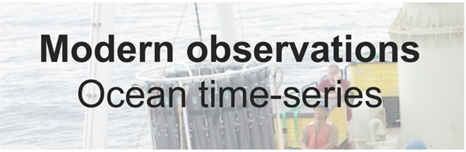
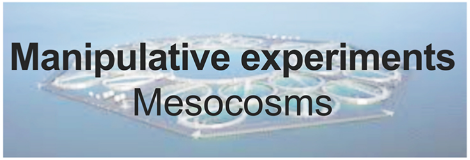
Figure 2 (a) An illustration of the differing degrees of success with which a simple three-level experimental design (using pre-industrial, present day and a year 2100 projection) may capture physiological thresholds. Inspection of the raw data (points) suggests largely similar responses among “species”, however underlying response norms (lines) are very different. (b) Reveals the pitfalls of how small differences among selected driver levels can lead to very different interpretations of underlying physiological response curves when other drivers also change: n=3 (Hoppe et al, orange) captures the response norm reasonably well at 15˚C, whereas n=5 (Iglesias-Rodriguez et al, green) at 19˚C does not; intermediate designs (n=4) perform more, or less, well depending on the overall range of driver levels and location of the optimum (from Bach et al., 2015). Note: scenario approaches, that may lack underpinning mechanistic functions for response norms, may require more driver levels to resolve curvilinear responses.

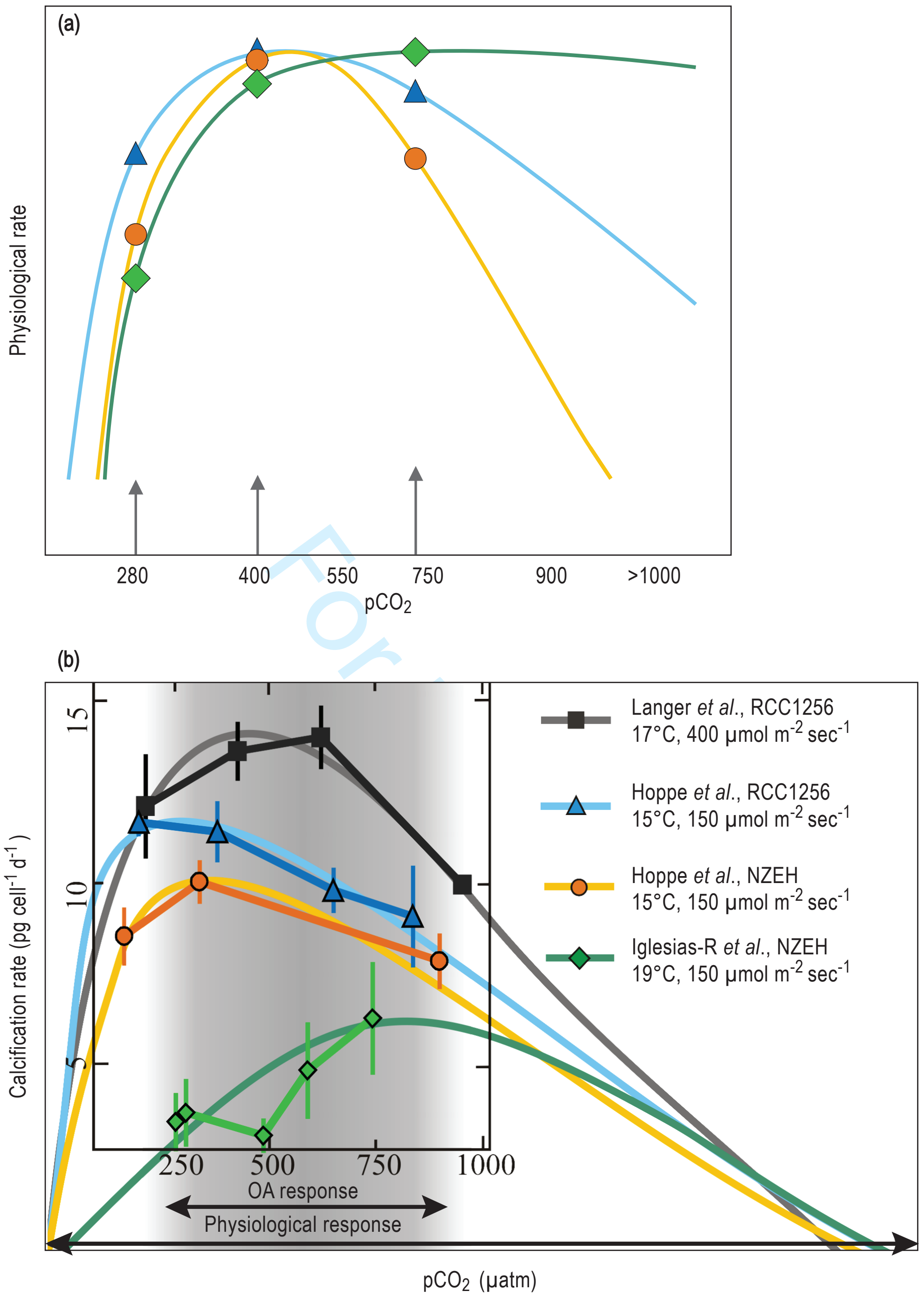
Figure 3 Visual depiction of the steps from formulating a multiple-driver research question to identifying a tractable experimental design that addresses that question within the available resources: (a), identify and quantify all of the key drivers that define the research question; (b), identify an idealised full-factorial design defining all of the drivers (experimental treatments, here illustrated for three factors) and the range of interest for each one; (c), identify the most relevant subset and levels of drivers, and combinations thereof, to create a reduced or collapsed design that best addresses the question(s) of interest (Gunst and Mason 2009; Boyd et al. 2015).

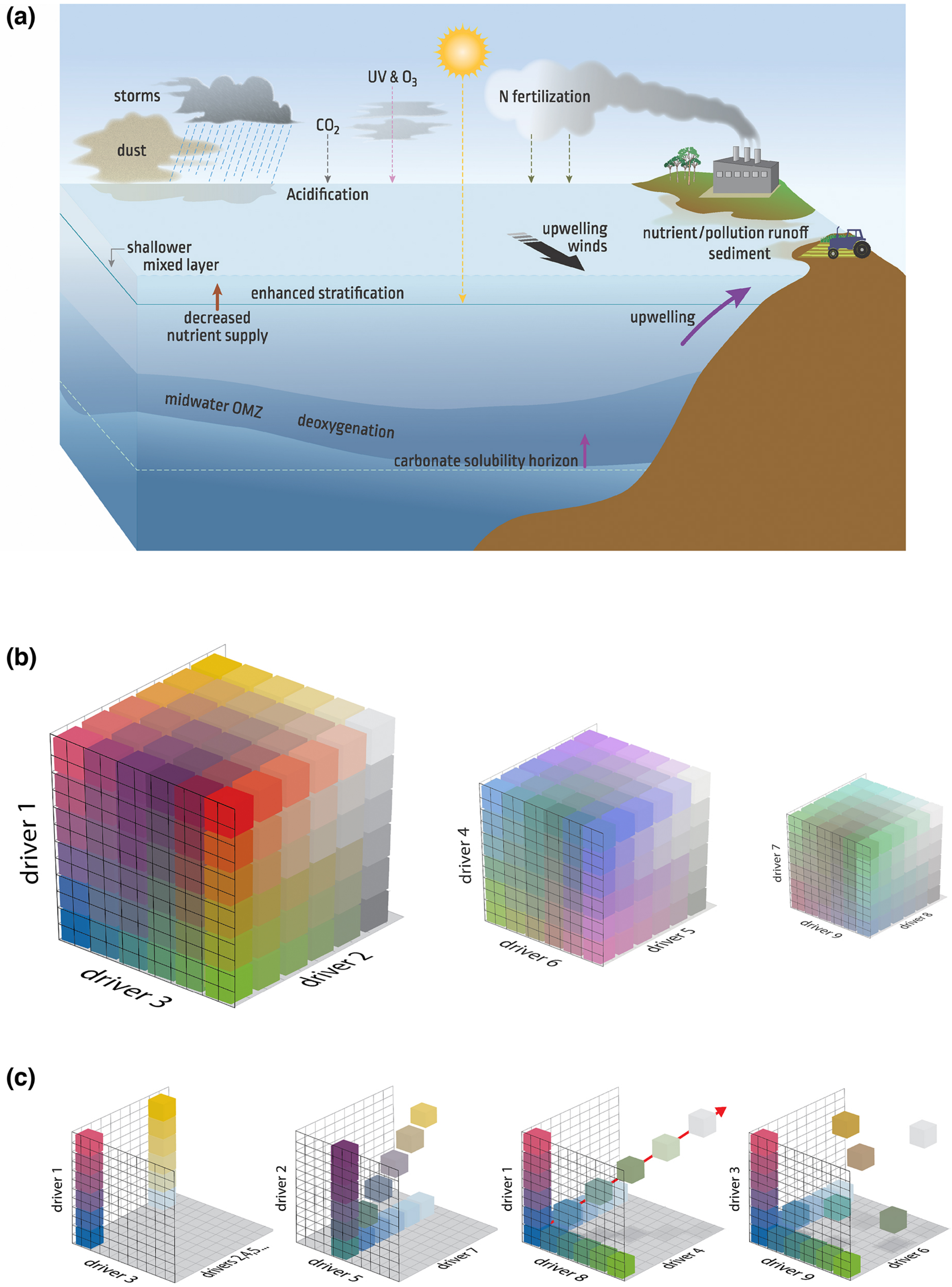
Figure 4 Progress in studies of ocean global change overlaid on the property-property space (termed the ‘RG cube’) developed by Riebesell and Gattuso (2015). (a) denotes the location of Sections in this Review within the ‘ RG cube’; (b) represents different experimental strategies: 1 denotes mesocosms, including FOCE experiments (e.g., Riebesell et al. 2013; Gattuso et al., 2014); 2 are competition experiments, (e.g., Moustaka-Gouni et al., 2016); 3 is a typical acclimated species under acidification (e.g., Hutchins et al., 2013); 4 are long-term (> 400 generations) micro-evolution studies (Lohbeck et al., 2012; Listmann et al., 2016); 5 denote multiple driver studies (e.g., Brennan and Collins, 2015); 6 sites of CO_2 natural enrichment such as CO_2 seeps (e.g., Fabricius et al., 2013) . (c) Progress in populating the 'RG cube' between 2000 and 2016 based on a survey of 171 studies (searched for using the terms 'multiple', 'stressor' and 'marine' between 5 December 2016 and 7 February 2017, see S-materials for bibliography and classifications).

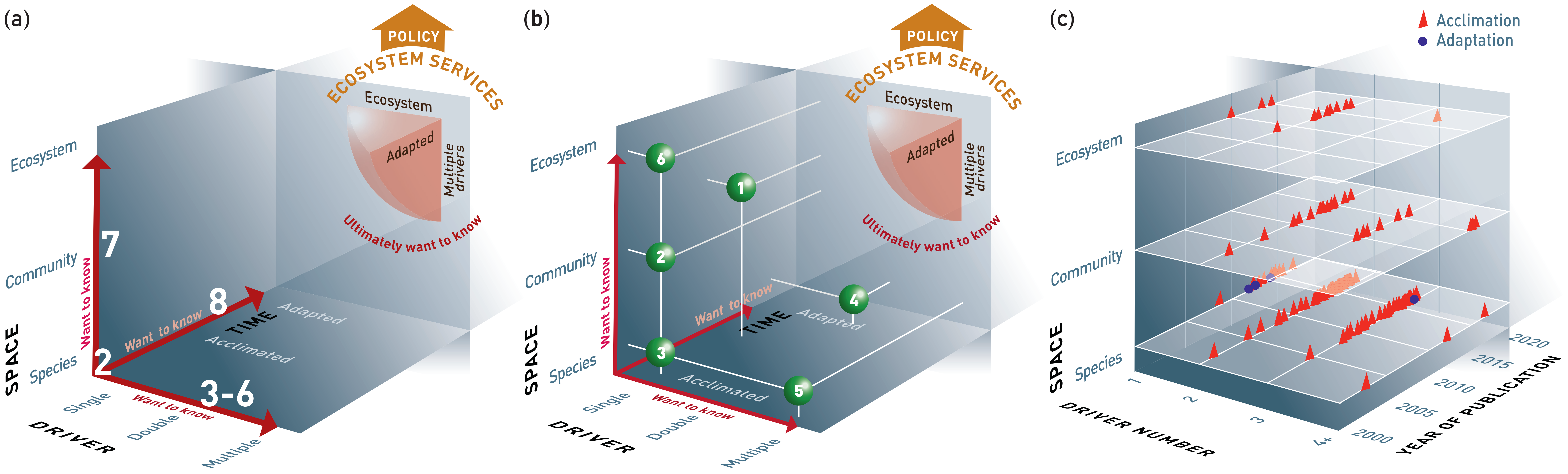
Figure 5 Graphical representation of multiple drivers as a landscape (the number of drivers is reduced for graphical purposes to show the landscape as a three-dimensional surface object). (a) For two environmental drivers mapping out their interplay (as might be captured by a detailed full factorial matrix). (b) as for panel (a) but overlaid with a scenario experimental design (circles) – based on a diagonal sampling strategy (the most efficient approach); and (c) in contrast a random experimental design (circles) which poorly represents the driver landscape. The scale bar denotes the hypothetical intensity of their interactions, based on their mathematical representation. Note that the theoretical entity of such a landscape is likely to have a large stochastic component that is not considered here.

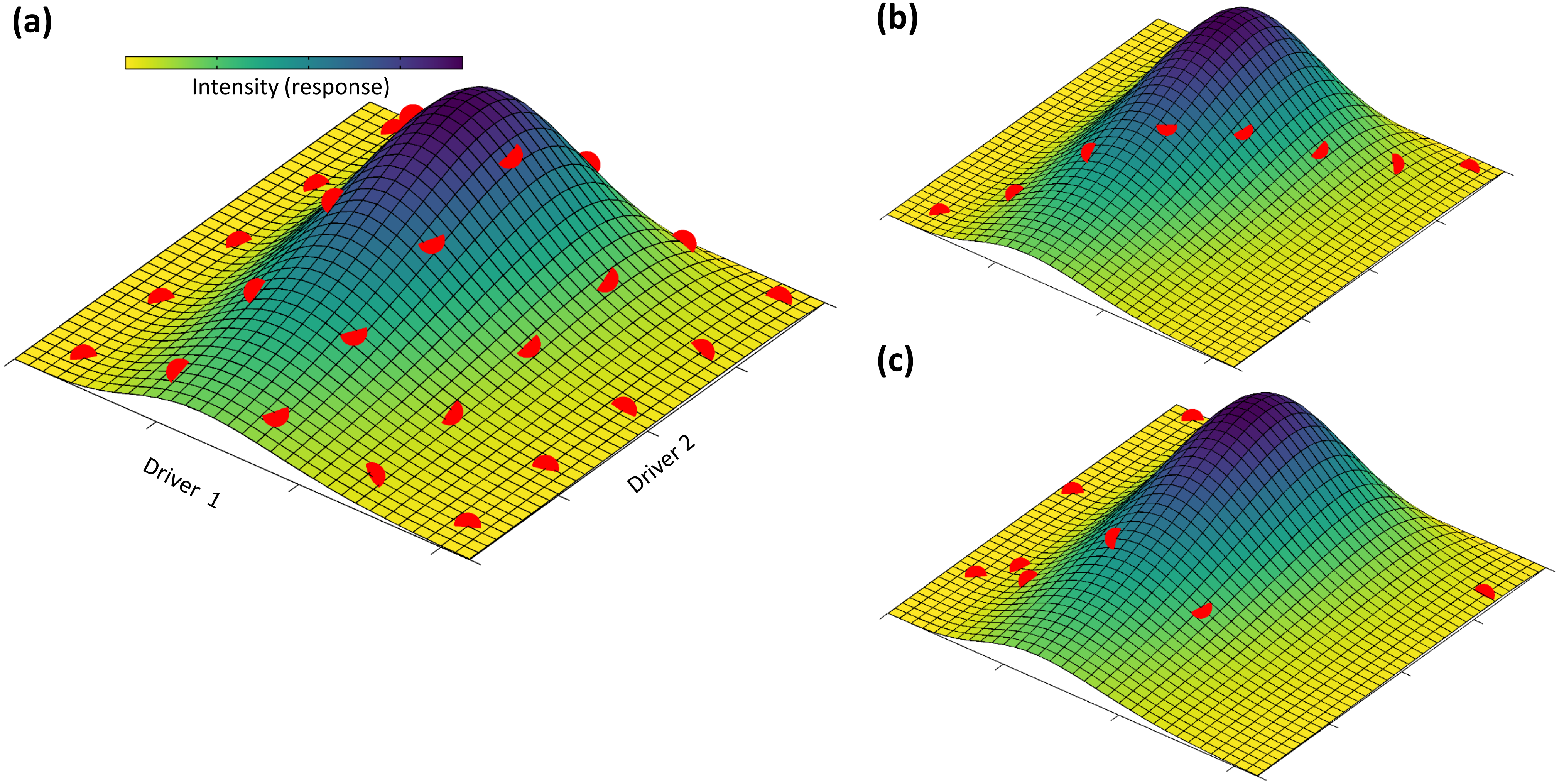
Figure 6: Interaction between physiological and evolutionary processes during trait and community composition changes due to environmental change. Circles delineated by dashed lines enclose entire populations, circles delineated by solid lines show genotypes within populations. (a) Physiological change in a focal lineage. Here, changes in trait values do not require any genetic change within the lineage. (b) Evolutionary change within a population, in the absence of a physiological response. This consists of change in the genetic composition of the population over time, seen as changes in the frequencies of lineages within the population, but the phenotypes of individual lineages do not change over time. Novel genotypes can appear in the population through migration or mutation (c) Physiological and evolutionary change within a population. The phenotypes of individual lineages change, as indicated by a shade shift. The genetic composition of the population also changes such that the frequencies of lineages within the population changes over time. Here, the plastic response of lineages affects their fitness relative to each other, so that the outcome of evolution differs in the presence and absence of a plastic response. Panel (a) corresponds to processes measured during physiology studies; panel (b) corresponds to a subset of current ecosystem models and panel (c) corresponds to processes measured during long-term ecology or evolution studies. Fig. modified from Collins and Gardner (2009).

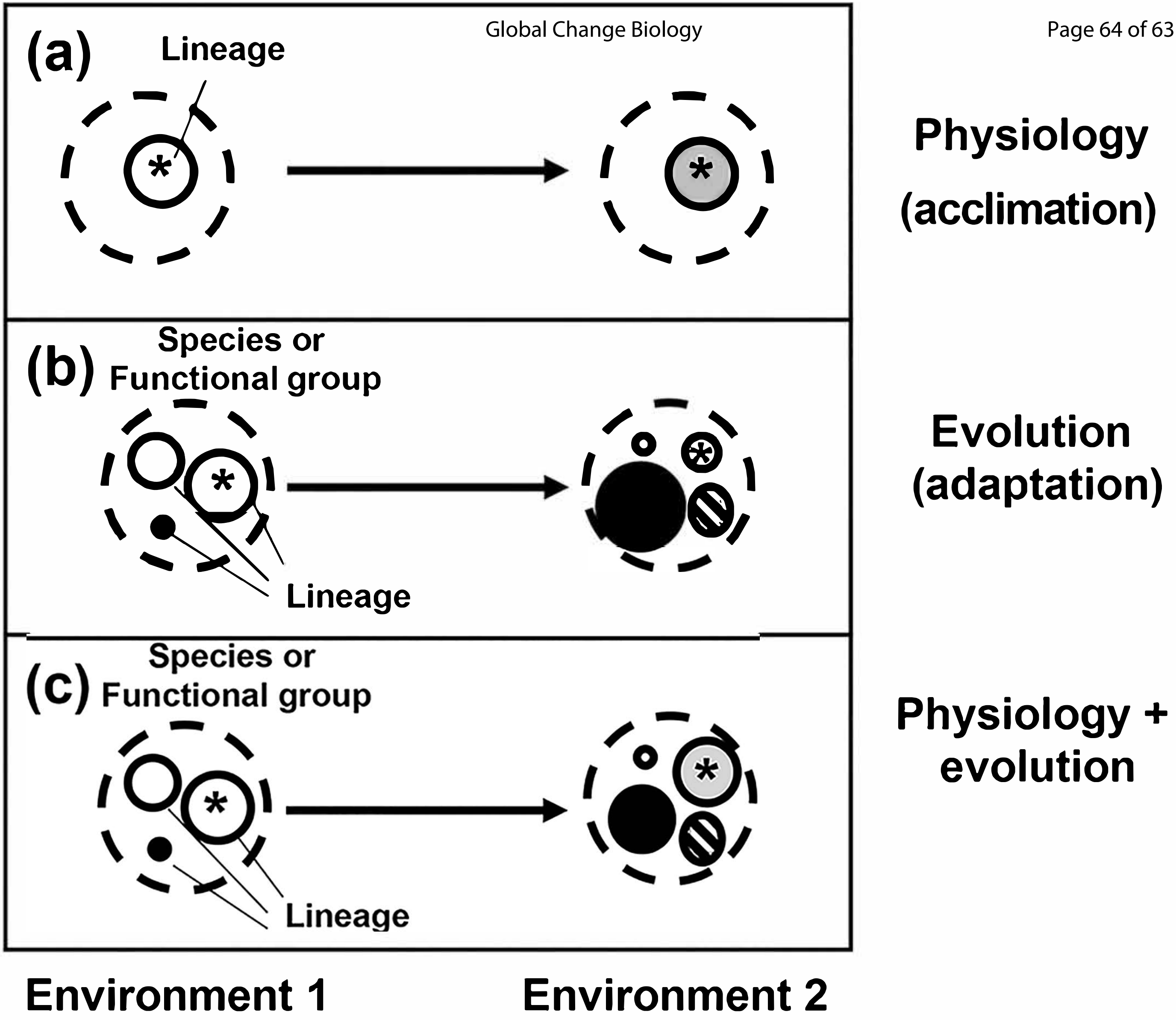
Strengths	Approaches and Examples	Limitations
- *Natural analogs for anthropogenic change* - *Examine globally or regionally integrated ecosystem impacts*		- *Emergence & extinction slower than anthropogenic change scenarios* - *Low temporal & taxonomic resolution*
- *Natural analogs for anthropogenic change* - *Large, observable signals & ecosystem responses*		- *Driver combinations differ from future scenarios* - *Recruitment from outside vent systems*
- *Detailed records over relevant timescales of change* - *Extensive biological, chemical, & physical supporting data sets*		- *Limited spatial resolution* - *Climate variability can obscure long-term trends (low signal:noise ratio)*
- *Many highly controlled and targeted treatments* - *Extensive replication and statistical power possible*		- *Few, or small, species* - *Limited ecological realism*
- *Many species interactions capture indirect effects* - *Strong environmental / ecological relevance*		- *Expensive & logistically difficult (especially for multiple drivers, long-term)* - *Few replicates possible, low statistical power*


(a)
Physiological rate
280
400
550
750
900
>1000
pCO_2
(b)
Calcification rate (pg $cell^{-1}$ d^{-1})
15
10
5
250
500
750
1000
OA response
Physiological response
pCO_2 (µatm)
Langer *et al.*, RCC1256
17°C, 400 µmol m^{-2} sec^{-1}
Hoppe *et al.*, RCC1256
15°C, 150 µmol m^{-2} sec^{-1}
Hoppe *et al.*, NZEH
15°C, 150 µmol m^{-2} sec^{-1}
Iglesias-R *et al.*, NZEH
19°C, 150 µmol m^{-2} sec^{-1}


(a)
storms
dust
UV & O_3
CO_2
Acidification
N fertilization
shallower mixed layer
upwelling winds
nutrient/pollution runoff
sediment
enhanced stratification
decreased nutrient supply
upwelling
midwater OMZ
deoxygenation
carbonate solubility horizon
(b)
driver 1
driver 3
driver 2
driver 4
driver 6
driver 5
driver 7
driver 9
driver 8
(c)
driver 1
driver 3
drivers 2,4,5...
driver 2
driver 5
driver 7
driver 1
driver 8
driver 4
driver 3
driver 9
driver 6


(a)
POLICY
ECOSYSTEM SERVICES
Ecosystem
Adapted
Multiple drivers
Ultimately want to know
Ecosystem
Community
Species
SPACE
Want to know
7
8
2
Want to know
TIME
Adapted
Acclimated
3-6
Single
Double
Multiple
Want to know
DRIVER
(b)
POLICY
ECOSYSTEM SERVICES
Ecosystem
Adapted
Multiple drivers
Ultimately want to know
Ecosystem
Community
Species
SPACE
Want to know
6
1
2
4
3
5
Want to know
TIME
Adapted
Acclimated
Single
Double
Multiple
Want to know
DRIVER
(c)
Acclimation
Adaptation
Ecosystem
Community
Species
SPACE
1
2
3
4+
DRIVER NUMBER
2000
2005
2010
2015
2020
YEAR OF PUBLICATION


(a)
Intensity (response)
Driver 1
Driver 2
(b)
(c)


(a)
Lineage
*
*
Physiology
(acclimation)
(b)
Species or
Functional group
*
Lineage
*
Evolution
(adaptation)
(c)
Species or
Functional group
*
Lineage
*
Physiology +
evolution
Environment 1
Environment 2