Kent Academic Repository

Full text document (pdf)

Citation for published version

Bull, Joseph and Smith, Robert J. (2020) Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems. Conservation Biology . ISSN 0888-8892. (In press)

DOI

Link to record in KAR

https://kar.kent.ac.uk/81783/

Document Version

Author's Accepted Manuscript

Conservation Biology

Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems

Journal:	*Conservation Biology*
Manuscript ID	20-199.R1
Wiley - Manuscript type:	Essay
Keywords:	Baseline, Conservation impact, Impact evaluation, Reference frame, Sum of perceived differences

Abstract:

When evaluating the impact of a biodiversity conservation intervention, a 'counterfactual' is needed, as true experimental controls are typically unavailable. Counterfactuals are possible alternative system trajectories in the absence of an intervention and comparing observed outcomes against the chosen counterfactual allows the impact (change attributable to the intervention) to be determined. Since counterfactuals are hypothetical scenarios, and by definition never occur, they must be estimated. Sometimes there may be many plausible counterfactuals, given that they can include multiple drivers of biodiversity change, and be defined on a range of spatial or temporal scales. Here we posit that, by definition, conservation interventions always take place in social-ecological systems (SES; ecological systems integrated with human actors). Evaluating the impact of an intervention within an SES therefore means taking into account the counterfactuals assumed by different human actors. Use of different counterfactuals by different actors will give rise to perceived differences in the impacts of interventions, which may lead to disagreement about its success or the effectiveness of the underlying approach. Despite that there are biophysical biodiversity trends, it is often true that no single counterfactual is definitively the 'right one' for conservation assessment, so multiple evaluations of intervention efficacy could be considered justifiable. Therefore, we propose the need to calculate a quantity termed the sum of perceived differences, which captures the range of impact estimates associated with different actors within a given SES. The sum of perceived differences gives some indication how closely actors within an SES agree on the impacts of an intervention. We illustrate the concept of perceived differences using a set of global, national and regional case studies. We discuss options for minimising the sum, drawing upon literatures from conservation science, psychology, behavioural economics, management and finance.

Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems

Abstract

When evaluating the impact of a biodiversity conservation intervention, a 'counterfactual' is needed, as true experimental controls are typically unavailable. Counterfactuals are possible alternative system trajectories in the absence of an intervention and comparing observed outcomes against the chosen counterfactual allows the impact (change attributable to the intervention) to be determined. Since counterfactuals are hypothetical scenarios, and by definition never occur, they must be estimated. Sometimes there may be many plausible counterfactuals, given that they can include multiple drivers of biodiversity change, and be defined on a range of spatial or temporal scales. Here we posit that, by definition, conservation interventions always take place in social-ecological systems (SES; ecological systems integrated with human actors). Evaluating the impact of an intervention within an SES therefore means taking into account the counterfactuals assumed by different human actors. Use of different counterfactuals by different actors will give rise to perceived differences in the impacts of interventions, which may lead to disagreement about its success or the effectiveness of the underlying approach. Despite that there are biophysical biodiversity trends, it is often true that no single counterfactual is definitively the 'right one' for conservation assessment, so multiple evaluations of intervention efficacy could be considered justifiable. Therefore, we propose the need to calculate a quantity termed the *sum of perceived differences*, which captures the range of impact estimates associated with different actors within a given SES. The *sum of perceived differences* gives some indication how closely actors within an SES agree on the impacts of an intervention. We illustrate the concept of perceived differences using a set of global, national and regional case studies. We discuss options for minimising the sum, drawing upon literatures from conservation science, psychology, behavioural economics, management and finance.

Introduction

The effectiveness of attempts to conserve the biodiversity of ecosystems – and associated contributions towards human wellbeing – has become an increasingly pressing topic over recent decades. A 'counterfactual' is necessary when quantitatively evaluating the ecological impact of a biodiversity conservation intervention (Ferraro & Hanauer, 2014). Counterfactuals are a type of 'reference scenario', capturing an alternative possible trajectory of a dynamic system in the absence of a given intervention. The 'impact' of the intervention is the change attributable to the intervention, measured as the difference between the actual observed trajectory (the 'outcome') and the predicted counterfactual (Ferraro & Pattanayak, 2006). Numerous counterfactuals can be reasonably specified for most systems relevant to conservation, as it is possible to select from a range of drivers of system change for potential inclusion within the counterfactual (Maron et al., 2018), notwithstanding that counterfactuals can also be specified over various spatiotemporal scales (Bull et al., 2014). Crucially counterfactuals are, by definition, a scenario that does not occur, so they can never be directly observed and monitored – and there is often no single 'correct' counterfactual, but rather various counterfactuals of differing plausibility. This is even true to an extent for 'control' sites used in quasi-experimental methods, as subjective decisions have to be made when choosing such sites to reduce the impacts of confounding factors (e.g. Wiik et al., 2019). Thus, though such control sites may give a good approximation to what would have happened at the 'treatment' sites without the intervention, they are still open to some interpretation.

Typically, conservationists are interested in a counterfactual representing the alternative ecological trajectory of a system, which is often influenced by multiple anthropogenic activities beyond the intervention in question (Ferraro & Pattanayak, 2006; Maron et al., 2018). But – while the literature on impact evaluation is accumulating rapidly, along with tools for implementation – impact evaluation is still rarely carried out in practice (Wiik et al., 2019). Moreover, conservation interventions do not take place in purely ecological systems – they take place in social-ecological systems (SES; Berkes & Folke, 1998), i.e. ecological systems integrated with social systems consisting of human actors. Thus, it is critical to go beyond purely *ecological* counterfactuals when evaluating the impact of a conservation intervention, and consider interlinked social systems (Alagona et al., 2012; Maron et al., 2018). This adds considerable challenges, as different actors may subjectively assume different counterfactuals are most relevant when judging impact (e.g. whether to use a counterfactual at the spatial scale of the project or the landscape; Bull et al., 2014), due to factors such as unconscious biases (Tversky & Kahneman, 1974), temporal starting point (Pauly, 1995), or assumptions made about processes driving change. This is important as the choice of counterfactual not only alters perceptions about intervention success, but also potentially the actions of stakeholders (Bull et al., 2014). Therefore, it is insufficient to consider counterfactuals on a purely ecological basis when judging conservation impact; also deserving attention are the ecological counterfactuals associated with differing interpretations from relevant stakeholders within the SES. We term this set of possible social-ecological counterfactuals – i.e. ecological counterfactuals derived from varying stakeholder perceptions – a *family* of counterfactuals.

Here, our objective is to further formalise the use of counterfactuals for evaluating biodiversity outcomes within an SES. Biodiversity is not the only property of an SES which might require conservation interventions, but it is our focus. We develop an exploratory conceptual framework, illustrated (although not formally tested) with case studies. In particular, we focus on differences of interpretation during quantitative evaluation of the ecological impact of conservation interventions. Approaches for qualitative evaluation exist (Sutherland et al., 2018); however, as the latter do not require quantitative impact estimates, we do not explore them further here. We do, though, explore how an actor's precise choice of counterfactual arises from their personal 'reference frame'. We go on to explore what approaches exist for minimising divergence in personal reference frames and thus the choice of counterfactual, to avoid conflicting perceptions of conservation impact in SES. Ultimately, we aim to enable better impact evaluation by considering not only biophysical outcomes against counterfactuals, but also how multiple stakeholders view the plausibility of different counterfactuals.

Towards families of counterfactuals

We start with a conservation intervention in a hypothetical ecosystem. We do consider 'conservation interventions' to be targeted at biodiversity, but framed more broadly around 'people and nature' (emphasizing "the importance of cultural structures and institutions for developing sustainable and resilient interactions between human societies and the natural environment"; Mace, 2014). By definition, an intervention being implemented means that: (a) at least one human actor has the potential to influence that biodiversity; (b) at least one human actor must be affected by that influence on biodiversity; and, (c) at least one human actor must be responsible for the intervention. In the simplest case, (a), (b) and (c) describe the same actor; for instance, if the SES is a farm, the single actor might be a farmer creating habitat for declining bird species. Note the logic that, for any system within which a conservation intervention is taking place, there must be at least one relevant human actor – that is, *every* system involving a conservation intervention must be an SES.

We focus on social perceptions of the ecological impact of an intervention, a necessary precursor to evaluating its social *impacts*; but though social impacts require similar treatment (e.g. Davidson, 2013), they are beyond scope here. So, to evaluate the impact of the intervention we then specify a counterfactual ("a causal effect of a program is only defined with respect to a well-defined alternative"; Ferraro & Hanauer, 2014). The actor anticipates an alternative biodiversity trend that would have taken place without the intervention. But biodiversity measurement is open to interpretation (Purvis & Hector, 2000), and sub-components of biodiversity are often ascribed wildly different weightings dependent upon the actor (Baylis et al., 2016; Bull & Maron, 2016; Pearson, 2016). Compounding this is the difficulty in determining which components of biodiversity are relevant when going beyond static considerations – which depends upon the chosen spatiotemporal scale (Bull et al., 2014). Finally, the 'shape' of the counterfactual trend will be heavily influenced by actor's expectations (Ferraro & Hanauer, 2014). Yet in the simplest case of our hypothetical SES, the single actor is free to select whatever counterfactual they choose – and the perceived impact of the intervention is the difference between the observed biophysical outcome and that counterfactual (Ferraro & Hanauer, 2014). Continuing our farmer example, the farmer might compare outcomes to the counterfactual scenario in which they had not created new habitat (acknowledging multiple plausible futures).

Consider a more complex hypothetical scenario with *two* actors: one carrying out activities that reduce biodiversity, and the other implementing conservation interventions. Assuming it is relevant for both actors to monitor impact, each specifies a counterfactual and measures the outcome. There is a wealth of reasons why their choice of counterfactual might differ – but even if their choice is the same in theory, expectation or uncertainty might mean that the precise trajectory of their chosen counterfactual diverges. The perceived impact for Actor 1 is the difference between the observed biophysical outcome and Actor 1's chosen counterfactual, and for Actor 2 it is the difference between the same biophysical outcome and Actor 2's chosen counterfactual. If the counterfactuals are different, there is consequently a difference between perceived impact for Actor 1 and Actor 2, which – whatever the actual trajectory – equals the difference in their chosen counterfactuals (assuming they use the same measure of outcome; Fig. 1, Appendix 1). For a real example see Maron et al. (2015), who find Australian policymakers (Actor 1) using different counterfactual rates to those calculated by researchers (Actor 2) for 'biodiversity offset' conservation interventions (Fig. 1, and below); or analogously, multiple reference levels in climate change mitigation (Griscom et al., 2009).

Finally, consider the general case of three or more actors within the hypothetical SES, who all evaluate the impact of an intervention, but at least some specify different counterfactuals. This resultant set of counterfactuals is the aforementioned *family of counterfactuals*. For the SES as a whole, consider the total difference in perceived impact across the family of counterfactuals for the full diversity of stakeholders: the larger it is, the greater the difference in perceived impact for multiple actors, which may imply that some actors are not satisfied with the intervention. We propose calculating a *sum of perceived differences* for the family of counterfactuals, which is the sum of the magnitude of the difference between the counterfactuals assumed by every actor and every other actor within the system (see Appendix 1 for our mathematical formulation, and justification). Note: the sum could conceivably weight the counterfactuals assumed by various individuals differently, e.g. to account for different uncertainties, or for uneven power dynamics – we return to this later in the article. Since no counterfactual is definitively 'correct', but rather is chosen on the basis of actors' value judgments, the sum of perceived differences is necessary to capture the impact of a conservation intervention in an SES. Importantly, the sum incorporates the counterfactuals actually used by actors without differentiating between those counterfactuals that are more or less plausible; but applying this framework *does* make the counterfactuals used more transparent, facilitating discussion around plausibility.

Actual biodiversity trends are, in principle, objective and biophysical. Yet, because the result of conservation interventions will be scrutinised by multiple different actors who each have their own personal counterfactual, which will often diverge, even robustly monitored interventions could lead to wide-ranging and at times conflicting interpretations of efficacy (Pearson, 2016). Equally, since robust evaluation of interventions must inextricably be linked to the initial design of the interventions themselves (Bull et al., 2014), the challenge of divergence in counterfactuals is important when setting conservation objectives. Concerning the sustainability of interventions, it would be preferable to minimise the sum of perceived differences for an SES – implying that most actors use approximately the same counterfactual. Is it possible to reduce the sum of perceived differences? To answer, we first consider the 'reference frame' within which the counterfactual is specified.

Counterfactuals and reference frames

Again, counterfactuals are a type of *reference scenario* (Maron et al., 2018). Reference scenarios are specified within a *reference frame* (or 'frame of reference'), and any number of reference scenarios can be specified within one reference frame (Bull et al., 2014). In Appendix 2 we formally describe a reference frame, and in Appendix 3 discuss usage of the term in different scientific disciplines. To a greater or lesser extent across different disciplines, a reference frame is typically composed of: (i) the relevant 'parameter space' capturing all possible variables of interest, particularly the subset of interest to the relevant 'observer'; and, (ii) the observer themselves, including the social values through which that observer views the chosen parameters. Unlike other disciplines in the natural sciences, which do so *alongside* rather than as *part* of the reference frame, we also include: (iii) the specific *coordinate system* used to make measurements within the parameter space, chosen by the observer, which determines the spatiotemporal scale of the reference frame (see Fig. A2.1). A coordinate system is a set of numbers used to specify quantitative information (e.g. an object's location in space; Appendix 3). In contrast to measurement of physical change in disciplines where standardised units exist, a coordinate system requires specification when assessing quantitative change in biodiversity as it is an imprecise term, and so there are many different ways in which 'biodiversity' could be measured (Purvis & Hector, 2000).

A counterfactual (or indeed any change) cannot be comprehensively specified in the absence of a reference frame: the observer, the parameters they are interested in, and the scale and coordinate system they are using must all be defined first (Fig. A2.1) even if in practice these are typically implicit or assumed (Bull et al., 2014). The reason for discussing reference frames here is that it is differences in the actors' underlying reference frames that give rise to the family of counterfactuals. To expand, reference frames used for evaluating conservation interventions vary depending upon the actor carrying out the evaluation – because different actors have different personal values built into their reference frame when constructing a counterfactual (e.g. Bull & Maron, 2016) which additionally may change over time, or because actors make different assumptions about the appropriate parameter space and scale for evaluation (e.g. Pauly, 1995). Minimising the sum of perceived differences in an SES therefore requires influencing underlying reference frames.

Reference frames can be *divergent* (e.g. actors using different scales) and even *conflicting* (e.g. observers making directly contradictory assumptions about value ascribed to biodiversity components) (Hahn et al., 2014). The idea of conflicting reference frames is not new – Schön & Rein (1994) discuss divergent and conflicting reference frames in policy design respectively as 'disagreements' (resolvable by examining facts) versus 'controversies' (which tend to be intractable). In conservation, it is well established that focusing upon different parameters or coordinate systems can cause conflicting assessments on efficacy (e.g. Naidoo et al., 2008).

Illustration

We now illustrate the preceding concepts using examples across different spatial scales. Multi-scalar illustration is important as counterfactual evaluation is relevant and necessary for conservation interventions on any spatial scale (Ferraro & Pattanayak, 2006). Note that in testing our proposals more thoroughly – and certainly in implementing them – the counterfactuals used would be based on extensive empirical data collected from different actors.

International: influence of Aichi Target 11

We start at the largest scale, treating the global biosphere as our SES. Aichi Target 11 is associated with the Convention on Biological Diversity (CBD) (https://www.cbd.int/sp/targets/): these targets, set in 2010, were intended to be met by 2020, and there are numerous interested stakeholders seeking to understand the impact of setting them (see Butchart et al., 2010). Aichi Target 11 states that, "By 2020, at least 17 per cent of terrestrial and inland water, and 10 per cent of coastal and marine areas, especially areas of particular importance for biodiversity and ecosystem services, are conserved through effectively and equitably managed, ecologically representative and well connected systems of protected areas and other effective area-based conservation measures, and integrated into the wider landscapes and seascapes". Here, for conceptual clarity, we focus on a subset of protected areas (marine areas), tracked via the World Database on Protected Areas (WDPA).

Observations seemingly indicate progress towards Target 11, evidenced by the ongoing trend in marine protected area (MPA) coverage (Fig. 2). But the process of setting the Aichi Targets can be viewed as an attempt to stimulate specific actions across the 196 countries Party to the CBD; as such, it is reasonable to ask whether there has been any associated change in conservation trends since 2010. So, imagine one actor characterises the immediately pre-2010 trend in MPA coverage as linear, with the post-2010 projection of that linear trend taken as their counterfactual. Comparing the latter against the observed trend for MPA coverage post-2010, this first actor treats the difference between the two as resulting from action stimulated by setting Target 11 (Fig. 2). Though not a particularly sound statistical analysis, it represents a plausible interpretation of the data.

A second actor might instead argue that a non-linear trend-line fits the 18-year dataset better, suggesting the MPA network is on a longer-term exponential growth trajectory. This would imply that the counterfactual scenario in the absence of the CBD Strategic Plan was also the observed outcome, insinuating (correctly or otherwise) that setting Target 11 had no influence on net MPA outcomes. The perceived difference between actors one and two, regarding the degree to which Target 11 has stimulated additional growth in MPA coverage, is clear – despite their using the same biophysical outcome, metric and dataset (Fig. 2).

Regional: Australian state biodiversity offset policy

A more pertinent application is in exploring the actions taken towards meeting global conservation policy targets at regional scales; not least because it is on such scales that conservation interventions typically act, and which ecosystem responses can be monitored over reasonable timescales. So, we turn to the aforementioned example of Australian state-level 'biodiversity offset' policies. These policies require that biodiversity losses via clearance of certain native habitats, as a result of economic development activities, are fully compensated through biodiversity gains on sites with comparable habitat elsewhere. These gains can take the form of 'averted losses' (i.e. protection of a habitat that prevents otherwise near-certain degradation), and as such the counterfactual scenario used to evaluate impacts is critical (Maron et al., 2015).

For six offset policies across five Australian states, Maron et al. compared the counterfactual rate of habitat loss assumed by the regulator (the 'crediting baseline' in Fig. 1) against a counterfactual calculated by researchers based on the proxy of 'recent observed deforestation'. These can be considered the counterfactuals assumed by Actors A1 and A2 respectively, and clearly the difference is substantial for most states. Using our mathematical formulation (Appendix 1), we can calculate the sum of perceived differences between the two actors for each state separately (Table 1). In doing so, we show that even the normalised sum varies by 1–2 orders of magnitude between states, and South Australia is a potential outlier (although this may be an artefact of offset data availability for the state; Maron et al., 2015). In isolation, these figures do not substantially advance our understanding of the SES, but if compared to similar statistics calculated for a wide range of regional policy interventions elsewhere (an empirical application of our framework), they could indicate the relative degree of disagreement over policy outcomes.

Local: species protection in Uzbekistan

The previous examples consider two actors, but the framework can be extended to any number, and for different biodiversity components. Consider protection efforts on a subnational scale for an IUCN Red List species, the Critically Endangered saiga antelope *Saiga tatarica*. In remote northwest Uzbekistan, the 'Ustyurt' saiga population has declined in recent years, and consequently attracted intensive conservation efforts; but the region is also experiencing increasing economic development activity (Bull et al., 2013). New development projects typically seek to mitigate any impacts that might exacerbate extinction risk for saigas, and so a counterfactual scenario is needed for evaluating the success of mitigation measures relative to the observed trend in saiga population numbers.

We compare five possible counterfactual population trends for Uzbek saigas over 15 years against the observed population trend. Three counterfactuals relate to known population trends for distinct saiga populations of historically comparable sizes (Association for the Conservation of Biodiversity of

Kazakhstan, unpublished data), and two are hypothetical but realistic counterfactuals (extirpation, and population expansion without die-offs) (Fig. 3). From initially close alignment, the counterfactuals (which we ascribe to five different actors, A1 – A5) diverge substantially over time. This can be tracked using the annual normalised sum of perceived differences across the family of counterfactuals (Fig. 3); showing the dramatic variation possible in terms of interpreting the impact of protection efforts.

If divergent families of counterfactuals in SES do undermine interventions, it is a problem which needs resolution. One solution would be to make the divergence explicit, requiring actors to 'agree to disagree' (Biggs et al., 2017). But next, alternatively, we consider opportunities to minimize the summed difference across a family of counterfactuals via resolving conflicting frames.

Minimising perceived differences

A starting point is to *understand* the relevant actors' reference frames. Extensive literature exists on stakeholder analysis in relationship to natural resource management (Reed et al., 2009; Cummings et al., 2018). Determining reference frames through stakeholder analysis would involve uncovering the perceptions of different actors, associated social discourses, and relationships between actors (Reed et al., 2009; Baynham-Heard et al., 2018). One aspect that needs consideration when implementing our approach is the *power* dynamics between actors (i.e. differentiation between actors' ability or capacity to influence outcomes). This involves assessing who is represented and who is left out of the decision-making, and whether the reference frames of some actors be given greater weighting (e.g. Smith et al., 2010). This is not the same as *weighting* for power dynamics when performing our proposed calculations: that is something we avoid, since our framework relates specifically to differences in counterfactuals *chosen* by various actors, rather than actors' ability to *act* upon those differences. Power dynamics are certainly important to our approach if they result in certain stakeholders being completely excluded from consideration, such that their perspective is not incorporated into the sum of perceived differences; a risk that requires careful treatment. Conversely, consideration of the specific nuance of power dynamics – albeit likely an important next step for research on this topic – is more crucial to discussions around steps taken to *resolve* such differences.

Though we begin to explore the latter in this article, the issue of power dynamics deserves further attention in its own right.

Different authors have explored options for using understanding of actors' perspectives to design effective conservation (Battista et al., 2018; Cummings et al., 2018; Cinner, 2018); here we do so based upon our formalised structure for reference frames (Appendix 2). So – having isolated different actors' reference frames, we can structure possible approaches to conflict resolution in terms of the three key components (Fig. A2.1; Table 2), each of which we discuss below.

Conflict in the physical component

The physical component of the reference frame (i.e. measurable biophysical quantities, whether biotic or abiotic; Appendix 2) is inevitably associated with development of conservation objectives. Consequently, resolution of conflict in the physical component of the reference frame relates strongly to conservation objectives and targets.

Structured decision-making (SDM; an approach designed to "systematically incorporate participant values, objectives and knowledge in decision-making"; Addison et al., 2013) can help develop objectives for natural resource use, given competing actor values (e.g. Robinson et al., 2016). Further, despite involving explicit expression of conflicting reference frames, SDM can strengthen consensus amongst groups during decision-making (Priem et al., 1995) such as design and evaluation of interventions. Indeed, experts typically perform better at making decisions based on well-structured discussions with peers rather than alone (Burgman et al., 2011). Incorporating monitoring and evaluation into SDM is well established (Lyons et al., 2008). Considering these factors, SDM provides a practical platform for using group consensus to reduce conflicts between stakeholders with different reference frames.

When there are intractable differences between stakeholders, consensus may be impossible. International policy decisions can be seen as negotiated decisions constructed through multi-actor interactions (Daniels et al., 2012). In such cases, e.g. the development of quantifiable environmental targets, Maxwell et al. (2015) propose allowing room for manoeuvre, promoting ambiguity in targets with the goal of building trust, cooperation and consensus. This is akin to allowing ambiguity in specification of the parameter space for the physical component of conflicting frames, to promote convergence in social components of those reference frames. An obvious disadvantage is that ambiguity in the definition of any component of the reference frame makes rigorous evaluation impossible – so the approach sacrifices short-term measurability to enable long-term resolution.

Conflict in the social component

More diverse groups often frame likely outcomes more accurately (Sutherland & Burgman, 2015). But this is not only relevant to experts and decision-makers: Addison et al. (2013) suggest using participatory SDM to increase broader social acceptance of conservation objectives. That is, through participatory exercises featuring diverse non-experts, interventions can be designed that integrate multiple social reference frames. Such reference frames typically involve several layers of complexity: deeply held values, important interests, local knowledge, and socially embedded conflicts among them (Daniels et al., 2012). Similarly, participatory modelling is designed to integrate a diversity of perspectives (i.e. social reference frames) through e.g. using role playing games (Jones et al., 2007) – again, for both specialist and non-specialist participants. In seeking consensus across heterogeneous groups of actors, there are many formal processes available e.g. stakeholder engagement, scenario planning (Cummings et al., 2018), and workshops designed to facilitate alignment between individual social values (Kenter et al., 2015).

Beyond participatory mechanisms, extensive literature concerns the degree to which personal reference frames can be modified, including for conservation (Iftekahr & Pannell, 2015). Consider ‘anchoring’ – when estimating numerical outcomes, individuals generally make minor adjustments to some initial value to yield their answer. The initial (anchoring) value is linked to the individual’s personal cognitive frame, and influenced by the formulation of the question itself (Tversky & Kahneman, 1974). Consequently, the extensive theoretical basis underlying the construction of personal reference frames could be leveraged to support resolution of conflicting reference frames – for instance, by purposefully creating common, appropriate anchoring points for evaluating conservation interventions by a range of actors (e.g. the quantitative extent of remaining habitat evaluated as ‘acceptable’; see Cinner, 2018).

The power of setting expectations for assessment is also relevant in terms of qualitative presentation of interventions; policies are likely deemed more acceptable when presented without loaded terms. Again, the presentation of interventions can thus be used proactively to help set personal reference frames amongst actors and reduce potential conflicts; albeit with associated ethical considerations (e.g. Rothschild, 2000). Alternatively, personal reference frames might be open to modification through management of appropriate incentives. Whilst, prosaically, these could be the provision of financial incentives, they could also be 'intrinsic' incentives related to personal desires or values. Some stakeholders might be better motivated by e.g. attachment to the land than by financial incentives (Reddy et al., 2017).

Of course, substantial components of personal reference frames are deeply held and cannot be modified. Individuals distinguish between 'sacred' and 'secular' values (Tetlock et al., 2000). This forms part of an actor's personal reference frame, and they may consider it unacceptable for a conservation intervention to exchange one type of value for the other (Daw et al., 2015). In that case, the solutions would be to explicitly identify sacred values held by different actors and ensure that they are not jeopardised by interventions. Such an approach – requiring an iterative process through which trust and cooperation is built between key parties on a seemingly intractable issue – involves negotiating shared belief structures (Cummings et al., 2018), e.g. squaring different 'mental models' held by stakeholders (Biggs et al., 2017). Alternatively, the intervention could be redesigned around secular values, for which examples in the literature go back decades (Cummings et al., 2018).

Conflict in coordinate systems

One solution to conflict caused by using different coordinate systems is accepting the need to track multiple indicators (e.g. Butchart et al., 2010). Approaches that combine multiple indicators are used widely elsewhere (e.g. finance; Engle & Gallo, 2006). Where this is not satisfactory – say, due to the increased resource requirements – an alternative is again participatory approaches, facilitating consensus between actors on coordinate systems rather than parameter spaces.

Conflict arising from differences in the spatial or temporal scale for evaluation might arise because scales are implicitly assumed by different actors (Bull et al., 2014). An explicit statement of the scale, as part of counterfactual construction, is therefore one straightforward approach towards avoiding conflict. A more nuanced approach would be to consider the appropriate counterfactual scenario for different spatial scales at a given moment: Maron et al. (2018) recommend that the scope of the reference frame be explicitly expanded to approach the largest possible 'overarching' frame. Doing so is comparable with other calls in the literature to evaluate conservation interventions using very long-term timescales (e.g. Willis & Birks, 2006). Finally, Pearson (2016) suggests actors construct their personal reference frames for conservation based partly upon the spatial scale in question: setting the scale could itself be a means for influencing personal reference frames.

Conclusions

All biodiversity conservation interventions take place within a social-ecological system (SES), and if a SES contains more than one actor, this may give rise to a family of counterfactuals. The use of a family of different counterfactuals by different actors will cause perceived differences in the impacts of (change attributable to) an intervention, even when all actors agree on outcomes. This may lead to disagreement between actors on the efficacy of interventions, undermining efforts to conserve biodiversity. But, since no counterfactual can be considered definitive, multiple different impact evaluations can be performed or proposed by different actors, more than one of which might be considered valid. Therefore, we have developed the basis for evaluating a *sum of perceived differences* between actors (defined mathematically in Appendix 1), and explored approaches for minimizing perceived differences. An important next step would be to test the conceptual framework we have developed here against extensive empirical data: not only in terms of the value of our proposed sum in different SES and associated implications, but also in terms of finding the limits to which counterfactuals might be considered 'valid'. Our work provides an exploratory theoretical basis for better quantifying, understanding and ultimately managing multiple diverse perspectives on nature conservation in SES.

References

Addison PFE, et al. (2013). Practical solutions for making models indispensable in conservation decision-making. *Diversity and Distributions*, 19, 490–502.

Battista W, et al. (2018). Behaviour change interventions to reduce illegal fishing. *Frontiers in Marine Science*, 5(403).

Baylis K, et al. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9(1), 58–64.

Baynham-Heard Z, Redpath S, Bunnefeld N, Molony T, & Keane A (2018). Conservation conflicts: Behavioural threats, frames, and intervention recommendations. *Biological Conservation*, 222, 180-188.

Berkes F, & Folke C (1998), Linking social and ecological systems: management practices and social mechanisms for building resilience. Cambridge University Press; Cambridge, UK.

Biggs D, et al. (2017). Breaking the deadlock on ivory. *Science*, 358(6369), 1378–1381.

Bull JW, Gordon A, Law E, Suttle KB, & Milner-Gulland EJ (2014). The importance of baseline specification in evaluating conservation interventions and achieving no net loss of biodiversity. *Conservation Biology*, 28(3), 799–809.

Bull JW, Lloyd SP, & Strange N (2017). Implementation gap between the theory and practice of biodiversity offset multipliers. *Conservation Letters*, 10(6), 656–669.

Bull JW, & Maron M (2016). How humans drive speciation as well as extinction. *Proceedings of the Royal Society B: Biological Sciences*, 283(1833).

Bull JW, Suttle KB, Singh NJ, & Milner-Gulland EJ (2013) Conservation when nothing stands still: moving targets and biodiversity offsets. *Frontiers in Ecology and the Environment*, 11(4), 203-210.

Burgman MA, et al. (2011). Expert status and performance. *PLoS 1*, 6(7), e22998.

Butchart SHM, et al. (2010). Global biodiversity: indicators of recent declines. Science, 328(5982), 1164–1168.

Cinner J (2018). How behavioral science can help conservation. *Science*, 362(6417), 889-890.

Corlett RT (2016). Restoration, reintroduction, and rewilding in a changing world. *Trends in Ecology and Evolution*, 31(6), 453–462.

Cummings JW, et al. (2018). Implicit decision framing as an unrecognized source of confusion in endangered species classification. Conservation Biology, DOI: 10.1111/cobi.13185.

Daniels SE, Walker GB, & Emborg J (2012). The unifying negotiation framework: A model of policy discourse. *Conflict Resolution Quarterly*, 30, 3–31.

Davidson MD (2013). On the relation between ecosystem services, intrinsic value, existence value and economic valuation *Ecological Economics*, 95, 171-177.

Daw TM, et al. (2015). Evaluating taboo trade-offs in ecosystems services and human well-being. *PNAS*, 112(22), 6949–6954.

Engle RF, & Gallo GM (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131, 3–27.

Ferraro PJ, & Pattanayak SK (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, 4(4), e105.

Griscom B, Shoch D, Stanley B, Cortez R, & Virgilio N (2009). Sensitivity of amounts and distribution of tropical forest carbon credits depending on baseline rules. *Environmental Science & Policy*, 12(7), 897-911.

Groves CR, & Game ET (2015). Conservation planning: informed decisions for a healthier planet. Roberts and Company Publishers Inc.; Colorado, USA.

Hahn T, Preuss L, Pinkse J, & Figge F (2014). Cognitive frames in corporate sustainability: managerial sensemaking with paradoxical and business case frames. *Academy of Management Review*, 39(4), 463–487.

Iftekhar MS, & Pannell DJ (2015). 'Biases' in adaptive natural resource management. *Conservation Letters*, 8, 388–396.

Jones NA, et al. (2007). Evaluating participatory modelling: developing a framework for cross-case analysis. CSIRO Working Paper 2008-11; Canberra, Australia.

Kenter JO, et al. (2015). What are shared and social values of ecosystems? *Ecological Economics*, 111, 86–99.

Lyons JE, Runge MC, Laskowski HP, & Kendall WL (2008). Monitoring in the context of Structured Decision-Making and adaptive management. Journal of Wildlife Management, 72(8), 1683–1692.

Mace, GM (2014). Whose conservation? *Science*, 345(6204), 1558-1560.

Maron M, Bull JW, Evans MC, & Gordon A (2015). Locking in loss: baselines of decline in Australian biodiversity offset policies. *Biological Conservation*, 192, 504–512.

Maron M, et al. (2018). The many meanings of No Net Loss in environmental policy. *Nature Sustainability*, 1, 19–27.

Marris E (2013). Charismatic mammals can help guide conservation. *Nature*, doi:10.1038/nature.2013.14396.

Maxwell SL, et al. (2015). Being smart about SMART environmental targets. *Science*, 347(6226), 1075–1076.

Naidoo R, et al. (2008). Global mapping of ecosystem services and conservation priorities. *PNAS*, 105(28), 9495–9500.

Pauly D (1995). Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology and Evolution*, 10(10), 430.

Pearson RG (2016). Reasons to conserve nature. *Trends in Ecology and Evolution,* 31(5), 366–371.

Poiani KA, Goldman RL, Hobson J, Hoekstra JM, & Nelson KS (2011). Redesigning biodiversity conservation projects for climate change: examples from the field. *Biodiversity and Conservation*, 20(1), 185–201.

Priem RL, Harrison DA, & Muir NK (1995). Structured conflict and consensus outcomes in group decision making. *Journal of Management*, 21(4), doi:10.1177/014920639502100406.

Purvis A, & Hector A (2000). Getting the measure of biodiversity. *Nature*, 405(6783), 212–219.

Reddy SMW, et al. (2016). Advancing conservation by understanding and influencing human behaviour. *Conservation Letters*, 10(2), 248–256.

Reed MS, et al. (2009). Who's in and why? A typology of stakeholder analysis methods for natural resource management. *Journal of Environmental Management*, 90, 1933–1949.

Robinson KF, et al. (2016). Structured decision making as a framework for large-scale wildlife harvest management decisions. *Ecosphere*, 7(12), e01613.

Rothschild ML (2000). Ethical considerations in support of the marketing of public health issues. *American Journal of Health Behavior*, 24(1), 26-35.

Schön A, & Rein M (1994). Frame reflection: toward the resolution of intractable policy controversies. Taylor & Francis, Ltd.; Oxfordshire, UK.

Smith RJ, Adams WM, & Leader-Williams N (2010). Another entangled bank: making conservation trade-offs more explicit. In, Trade-offs in conservation: deciding what to save (Eds., Leader-Williams N, Adams WM, Smith RJ). Wiley-Blackwell; UK.

Sutherland WJ, & Burgman MA (2015). Use experts wisely. *Nature*, 526, 317–318.

Sutherland WJ, Dicks LV, Everard M, & Geneletti D (2018) Qualitative methods for ecologists and conservation scientists. *Methods in Ecology & Evolution*, 9(1), 7-9.

Tetlock PE, Kristel OV, Elson SB, Lerner JS, & Green MC (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870.

Tversky A, & Kahneman D (1974). Judgments under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.

Wiik E, et al. (2019). Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice*, 1:e8.

Willis KJ, & Birks HJB (2006). What is natural? The need for a long-term perspective in biodiversity conservation. *Science*, 314(5803), 1261–1265.

***Figure 1**: hypothetical plot of different counterfactuals (**$B_{c(A1)}$**, **$B_{c(A2)}$**) used by two actors (**Actor 1**, **Actor 2**) in an SES, compared to a baseline (**B_{base}**) at time = 0 and the observed biodiversity outcomes through time (**B_{obs}**). **Actor 1**'s perceived impact is based on comparing **B_{obs}** against **$B_{c(A1)}$**, while Actor 2 compares **B_{obs}** against **$B_{c(A2)}$**. The difference in perceived impact is shown. Inset: real life example, of two actors judging different counterfactual deforestation rates in Australian states (from Maron et al., 2015), abbreviations for each state given in Table 1.*

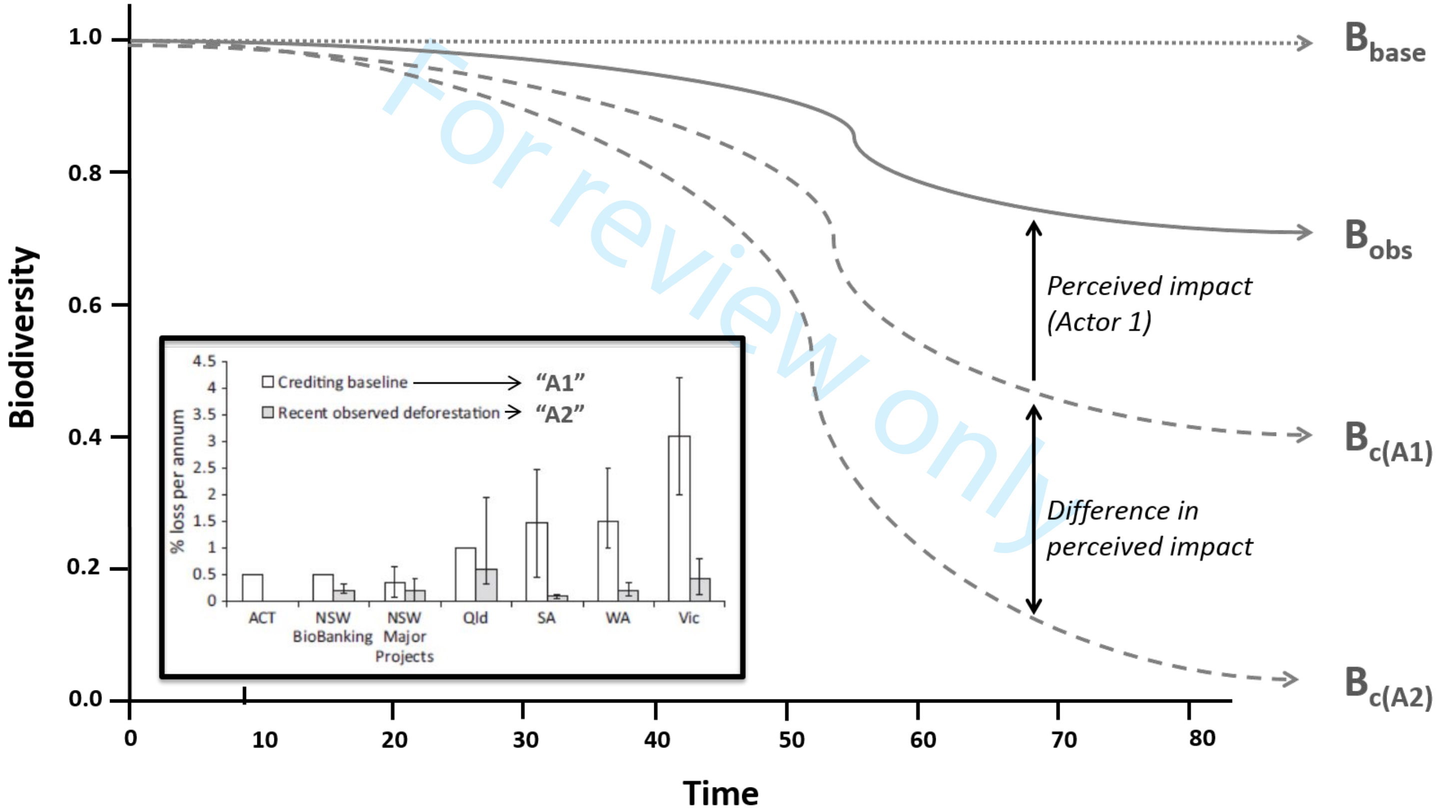

Figure 2: *plot of global marine protected area coverage (%) against time (years). Blue = pre-2010 trend (counterfactual for A1); orange = post-2010 trend; black = exponential trend-line fit to entire time series (counterfactual for A2); grey = Aichi Target 11 (10%). A perceived difference in the impact of setting Aichi Target 11 arises between two actors who use different counterfactuals (the blue and black lines on the graph). Data from the WDPA (UNEP-WCMC).*

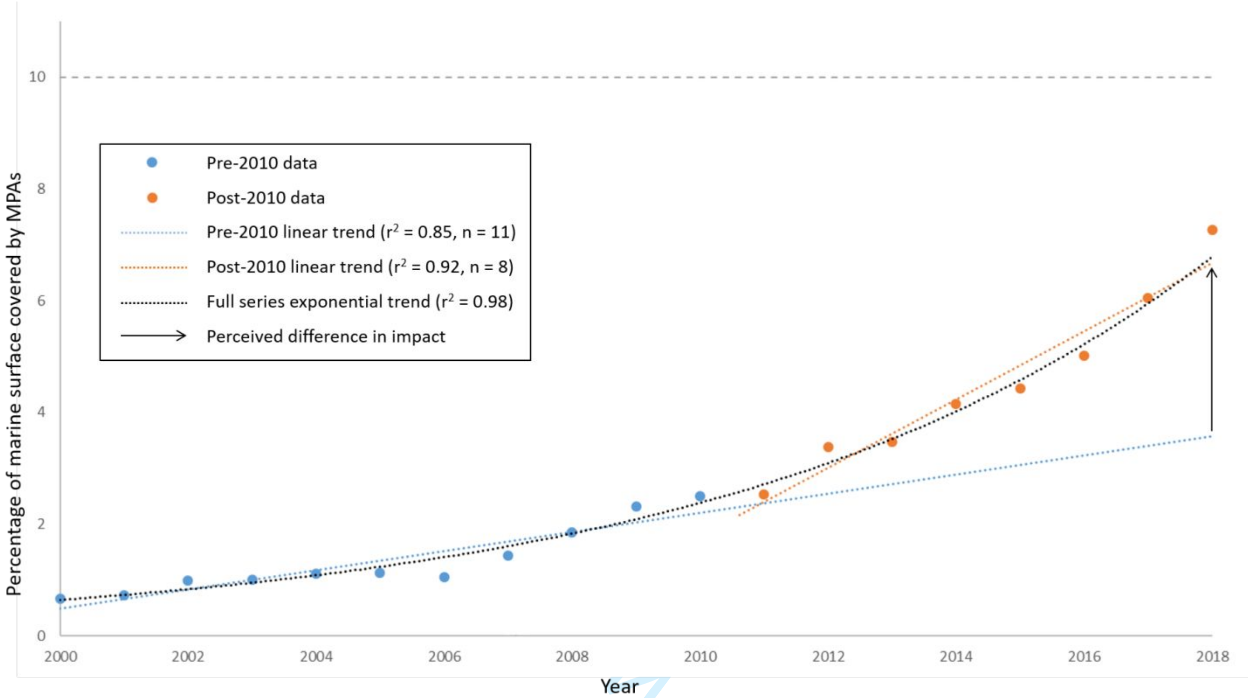

***Figure 3:** evaluating the outcomes of species protection in northwest Uzbekistan, 2006 – 2018. Primary y axis: dashed grey lines = five possible counterfactual trends assumed by different actors 'A1' – 'A5' for the regional saiga antelope* Saiga tatarica *population; solid line = observed population trend 'Ob' (ACBK, unpublished data). All six lines labelled, for clarity, at years 2013 and 2017. Secondary y axis: block dots = annual normalised sum of perceived differences.*

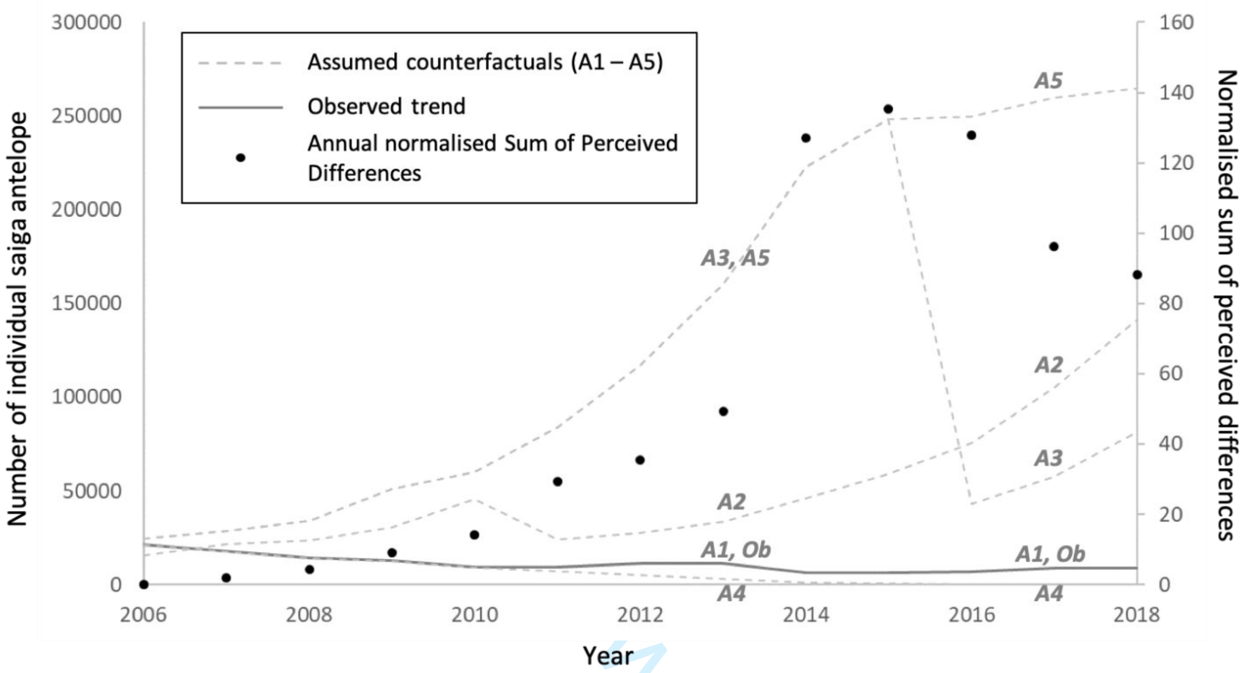

Table 1: *table of the perceived difference in outcomes between the regulator and researchers (A1 and A2, respectively) in the case of six different Australian state biodiversity offset policies, using the formulation of the 'sum of perceived differences' ($\sum \Delta_i$) in Appendix 1. Normalised sum of perceived differences also given. Data from Maron et al. (2015).*

	A1: assumed counterfactual (% habitat loss per annum)	**A2: assumed counterfactual (proxy, % deforestation per annum)**	**Perceived difference ($\sum \Delta_i$)**	**Normalised sum of perceived differences**
New South Wales (NSW) 'BioBanking'	0.55	0.21 (min 0.14, max 0.32)	0.34	1.6
New South Wales (NSW) 'Major Projects'	0.35 (min 0.07, max 0.65)	0.22 (min 0.00, max 0.42)	0.13	0.6
Queensland (Qld)	1.00	0.60 (min 0.32, max 1.96)	0.4	0.7
South Australia (SA)	1.48 (min 0.47, max 2.49)	0.09 (min 0.07, max 0.14)	1.39	15.4
Western Australia (WA)	1.50 (min 1.00, max 1.50)	0.19 (min 0.09, max 0.34)	1.31	6.9
Victoria (Vic)	3.12 (min 2.01, max 4.20)	0.45 (min 0.12, max 0.80)	2.67	5.9

***Table 2**: some possible causes of divergence between different reference frames, for actors observing conservation interventions*

Component of frame (F)	Possible areas of divergence	Examples of divergence potentially leading to conflict	Relevant references
Physical parameter space ($\boldsymbol{E^n}$)	Number and type of parameters	Which parameters to incorporate (e.g. objectives and targets of interventions)	Burgman et al., 2011; Maxwell et al., 2015
		Decision whether to include certain physical parameters (e.g. temperature), or otherwise	Poiani et al, 2011
		Treating the intervention as dynamic (that is, incorporating 'time' as a parameter) or otherwise	Corlett, 2016
Observer's personal reference frame ($\boldsymbol{F_{ob}}$)	Personal values	Degree to which relationship exists between biodiversity and personal well-being	Woodhouse et al., 2015
		Assumption that high biodiversity is 'better' than low biodiversity (e.g. species numbers)	Bull & Maron, 2016
		Different intrinsic incentives to engage in conservation interventions	Reddy et al., 2017
	Social values	Preference for focus on conservation of certain components of biodiversity over others	Marris, 2013
	Cultural values	Attribution of sacred values vs. secular values to components of biodiversity	Daw et al., 2015; Biggs et al., 2017
Coordinate system (*C*)	Choice of indicators	Measurement of outcomes in relation to biological diversity, or to functionality (and consequent ecosystem service provision)	Naidoo et al., 2008
	Choice of scale	Evaluating interventions on the spatial scale of individual projects vs. a landscape scale	Bull et al., 2014; Maron et al., 2018
		Choice of temporal scale over which interventions are to be evaluated for efficacy	Bull et al., 2014
		Decision to include historical context, or otherwise	Willis & Birks, 2006;