Received: 6 April 2023 | Revised: 14 August 2023 | Accepted: 6 September 2023

DOI: 10.1111/csp2.13024

CONTRIBUTED PAPER

Conservation Science and Practice
A journal of the Society for Conservation Biology
Open Access

WILEY

Assessing diverse evidence to improve conservation decision-making

Alec P. Christie[1,2,3] | **William H. Morgan**[1,2] | **Nick Salafsky**[4] | **Thomas B. White**[1] | **Robyn Irvine**[5] | **Nicolas Boenisch**[6] | **Rafael M. Chiaravalloti**[7] | **Kate Kincaid**[1] | **Ali Mohammad Rezaie**[8] | **Hiromi Yamashita**[1,9,10] | **William J. Sutherland**[1,2]

[1]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK

[2]BioRISC (Biosecurity Research Initiative at St Catharine's), St Catharine's College, Cambridge, UK

[3]Downing College, Cambridge, UK

[4]Foundations of Success, Bethesda, Maryland, USA

[5]Ecosystem Conservation Team, Parks Canada Agency, Gatineau, Canada

[6]Foundations of Success Europe, Eefde, The Netherlands

[7]Anthropology Department, University College London, London, UK

[8]Department of Civil Engineering, University of Asia Pacific, Dhaka, Bangladesh

[9]Ritsumeikan Asia Pacific University, Beppu-City, Oita, Japan

[10]Nagoya University Graduate School of Environmental Studies, Aichi, Japan

Correspondence
Alec P. Christie, Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK.
Email: apc58@cam.ac.uk

Abstract

Meeting the urgent need to protect and restore ecosystems requires effective decision-making through wisely considering a range of evidence. However, weighing and assessing evidence to make complex decisions is challenging, particularly when evidence is of diverse types, subjects, and sources, and varies greatly in its quality and relevance. To tackle these challenges, we present the Balance Evidence Assessment Method (BEAM), an intuitive way to weigh and assess the evidence relating to the core assumptions underpinning the planning and implementation of conservation projects, strategies, and actions. Our method directly tackles the question of how to bring together diverse evidence whilst assessing its relevance, reliability, and strength of support for a given assumption, which can be mapped, for example to a Theory of Change. We consider how simple principles and safeguards in applying this method could help to respectfully, and equitably, include more local forms of knowledge when assessing assumptions, such as by ensuring diverse groups of individuals contribute and assess evidence. The method can be flexibly applied within existing decision-making tools, platforms, and frameworks whenever assumptions (i.e., claims and hypotheses) are made. This method could greatly facilitate and improve the weighing of diverse evidence to make decisions in a range of situations, from local projects to global policy platforms.

KEYWORDS

conservation evidence, decision-making, evidence assessment, evidence-based, evidence-informed, local knowledge, project management

1 | INTRODUCTION

There is an urgent need for action to protect and restore natural ecosystems. Finding effective solutions through good decision-making requires the wise use of a diverse range of evidence from different sources (MacLeod et al., 2022; Sutherland et al., 2004, 2022; Walsh et al., 2015). However, those making decisions often find that assessing this diversity of evidence is challenging when trying to account for variability in its quality and usefulness

for decision-making in complicated socioeconomic and cultural contexts (Adams & Sandbrook, 2013; Evans et al., 2017).

Evidence itself is sometimes challenging and controversial to define, but here we adopt the broad, inclusive definition from Salafsky et al. (2022, p. 4) that evidence can be any "relevant data, information, knowledge, and wisdom used to assess an assumption" related to a question of interest. In conservation, there has been a broad movement toward evidence-based or evidence-informed conservation (i.e., promoting more effective and efficient actions based on sound consideration of available evidence; Sutherland et al., 2004). This has led to a growing scientific evidence base on conservation solutions (e.g., Conservation Evidence [www.conservationevidence.com; Sutherland et al., 2019], Evidensia [https://www.evidensia.eco/], Collaboration for Environmental Evidence [https://environmentalevidence.org/]), an increasing availability of global information on biodiversity status and threats (e.g., IUCN Red List; Stephenson & Stengel, 2020), and identification of important knowledge gaps (Christie et al., 2020; Christie, Amano, et al., 2021). Approaches toward gathering, synthesizing, and assessing evidence to inform conservation efforts are often focused on identifying general patterns, recommendations, and answers (Christie et al., 2020; Christie et al., 2022) across many scientific studies in a collated evidence base through meta-analyses, various types of reviews, or subject-wide evidence synthesis (Cook et al., 2017; Sutherland et al., 2004).

Whilst this information may be useful in many contexts, there has been confusion and criticisms over whether such approaches adequately recognize, or value, local forms of evidence (such as local, expert, traditional, and Indigenous knowledge, wisdom, and experience) or properly consider the practicalities of implementing conservation actions in diverse and complex socio-economic settings (Adams & Sandbrook, 2013). These approaches have also (historically and currently) included colonial structures for documenting and sharing evidence in ways that have not honored data sovereignty nor sources and have unequally weighted particular approaches more heavily than others due to the dominant paradigm (Tengö et al., 2017; Wheeler & Root-Bernstein, 2020). Indeed, local information from diverse sources can be highly relevant and reliable in specific conservation contexts, and considerations of local factors influence the effectiveness, costs, acceptability, and feasibility of actions (Adams & Sandbrook, 2013; Christie, Downey, et al., 2022). Meaningfully involving local communities and partners in conservation decisions has also been shown to add value to local projects and strategies, and ultimately underpin their success (Cote et al., 2021), as

BOX 1 System 1 and System 2 thinking, Theory of Change, and formulating assumptions.

There are two recognized types of thinking based on Information Processing Theory: "fast" System 1 or "slow" System 2 thinking (Kahneman et al., 2011; Papworth, 2017). Decisions requiring System 1 thinking are those that are rapid, low-effort, and intuitive, relying on heuristics, instinct, and pattern-matching, whilst decisions requiring System 2 thinking require a more time-intensive, deliberative, and analytical cognitive process, involving assessing and processing all available information (Papworth, 2017). Using System 1 thinking for complex problems instead of System 2 thinking will typically result in poor decisions.

When thinking about conservation projects, an example of a decision requiring System 2 thinking could be how a local NGO team can reduce and reverse the decline in seabird populations on an island on which rats have been introduced (Irvine et al., 2021). During the planning stage, the team might create a Situation Analysis (to identify relevant threats, opportunities, and stakeholders) and a Theory of Change (to outline the logical steps by which a strategy will contribute to achieving set targets). Doing so will highlight that the project is making several key assumptions (e.g., the linkages between different steps of a Theory of Change). In a Situation Analysis, these may include, for example: the seabird population is declining rapidly and problematically due to the introduction of rats. In a Theory of Change, assumptions may include: (1) rat eradication through trapping (without poisoning) is socially acceptable to local partners and communities; and (2) rat eradication through trapping (without poisoning) will lead to the recovery of the seabird population within the project timescale. Others might include that rat eradication using trapping is feasible or cost-effective in terms of the budgets and resources available.

Clearly, the evidence required to check whether these assumptions are valid (i.e., that the Situation Analysis and Theory of Change are valid) should come from a range of sources including, but not limited to, Indigenous and Local Knowledge, experience, and wisdom of practitioners and partners, scientific studies and syntheses, expert

judgments, reports, databases (e.g., citizen science), local records, and observations.

well as fill knowledge gaps in the scientific literature (MacLeod et al., 2022; Rowland et al., 2023). There is therefore growing recognition that evidence that comes in a variety of types and subjects from local and external sources is required to make successful, pragmatic decisions in conservation under challenging time, resource, and cost constraints (Game et al., 2018; Malmer et al., 2020; Schuster et al., 2019; Sutherland et al., 2022; Taper et al., 2021; Tengö et al., 2017).

Weighing such diverse evidence to make better decisions (i.e., ones that contribute to successfully completing a project's goals) is seen as a major, and somewhat controversial, challenge in conservation (Kadykalo, Cooke, & Young, 2021; Sutherland et al., 2017; Wheeler & Root-Bernstein, 2020). In this paper, we want to highlight that we are focusing on conservation decisions that require System 2 thinking (see Box 1)—that is, complex situations where more analytical, deliberative, and cognitive thinking is employed (Papworth, 2017). In such scenarios, decision-makers will often weigh conflicting evidence from different sources where there may be great uncertainty over the quality of evidence and how well external evidence from other contexts may apply to the context of interest (e.g., evidence drawn from other countries, taxa, cultures, sectors, habitats, disciplines, etc.).

Whilst structured Decision Support Frameworks and tools exist to assess evidence and make decisions (e.g., structured decision-making, argument maps, and evidence-to-decision tool; Schwartz et al., 2018), there is often insufficient guidance that addresses the quality and relevance of evidence and how to weigh evidence that can be messy, conflicting, and/or uncertain (Christie, Downey, et al., 2022). We also believe that a particular focus on using assumption-based thinking could be integrated into, and improve, existing decision support frameworks and tools (Salafsky et al., 2022; Specht et al., 2022), particularly by linking to the Conservation Standards through tools such as situation analyses and theories of change (Box 1). The Conservation Standards is a framework to help support conservation decision-making, consisting of a cycle of five steps (Assess, Plan, Implement, Analyze, and Adapt, Share). When planning a project, the Conservation Standards (CMP, 2020; Conservation Standards, 2023) suggests the use of a Situation Analysis—a process to identify the scope, targets, direct and indirect threats, opportunities, rightsholders, and stakeholders relating to the project. This can be followed by developing a Theory of Change, a process to outline the series of assumptions relating to how a particular strategy will contribute to achieving certain targets (often illustrated by a Results Chain; CMP, 2020; Conservation Standards, 2023).

Here, we outline an intuitive approach to consider and visualize weighing evidence of different types and sources to understand the confidence a decision-maker should have in the key assumptions underpinning a conservation project, program, or strategy. We test the inter-rater reliability of our assessment approach and show how it could support and be integrated within different existing Decision Support Frameworks, platforms, and tools to improve decision-making when considering diverse sources of evidence.

2 | A PROCESS TO ASSESS ASSUMPTIONS USING EVIDENCE

2.1 | An overview of the Balance Evidence Assessment Method (BEAM)

Here, we describe the BEAM. We adopt Salafsky et al.'s (2022) framing of the practice of conservation, whereby conservation and natural resource management occur through specific projects and programs, which aim to achieve a set of goals relating to a system of interest. Plans, strategies, and actions are then decided upon to achieve these goals, and help in making these decisions are available from many different planning and Decision Support Frameworks applicable at different temporal and spatial scales (Schwartz et al., 2018).

In these decision-making contexts, we believe that there is great value in laying out a Situation Analysis and Theory of Change for planned projects and programs (CMP, 2020) and identifying the core *assumptions* that they make (see Box 1 for examples). Assumptions can be defined as: "something that we believe to be true about a system; often a more detailed 'assessable' articulation of a question of interest" (Salafsky et al., 2022, p. 4) and can be supported or refuted by available evidence, which may come in many forms (as per the earlier inclusive definition of evidence).

Different types of assumption may exist relating to a claim (an assumption about a system supported by existing evidence) or a hypothesis (an assumption that requires new evidence to be gathered)—but in the interests of simplicity and using practitioner-friendly terms, we follow Salafsky et al.'s (2022) recommendation to use the term "assumption."

The BEAM (Figure 1) is designed to provide the user (any decision-maker or decision-making group in conservation) with a graphical indication of the confidence they

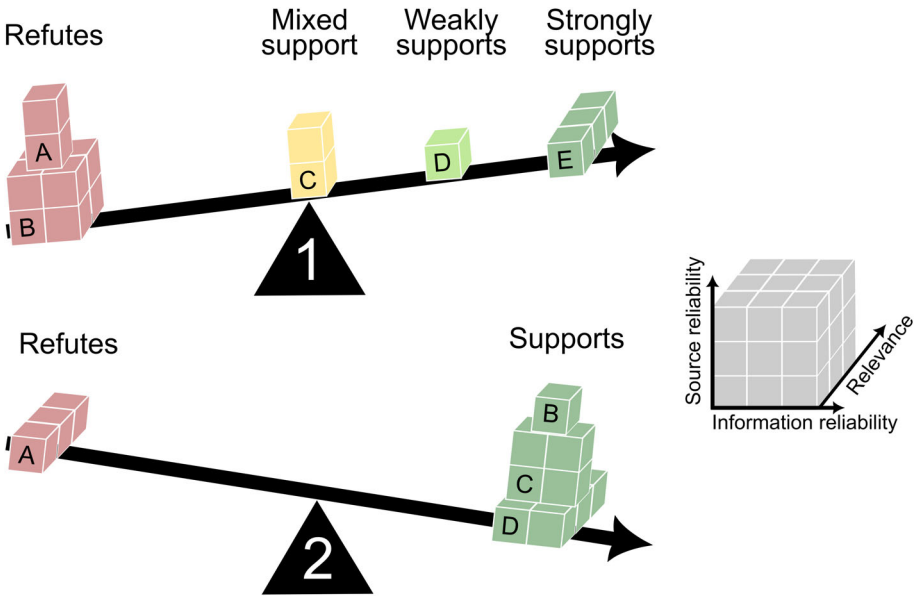

FIGURE 1 A diagram illustrating the Balance Evidence Assessment Method (BEAM), an intuitive way to visualize weighing different pieces of evidence supporting or refuting an assumption. Note that if the relevance or reliability of a piece of evidence is zero, then the block of evidence has no weight and disappears. Balance 1 shows an assumption that can be assessed by five different pieces of evidence (A)–(E) of varying weights (shown by their size) that can support or refute an assumption on an ordinal scale. Balance 2 shows a situation where an assumption can be assessed by four different pieces of evidence (A)–(D) that can only either support or refute an assumption (in a binary manner). In many situations, Balance 1 (using an ordinal scale for support) is most likely to be appropriate.

have in a stated assumption, given the available evidence, visualizing this as an intuitive balance (a.k.a. seesaw). Pieces of evidence relevant to an assumption are placed as blocks onto a balance dependent on their *strength of support* (i.e., the degree to which they support or refute the assumption) and vary in their *weight* (i.e., the quality and relevance of the evidence). Visualizing the overall balance of evidence and its strength of support and weight provides us with an indication of our confidence in an assumption. The BEAM can be visualized in an ordinal way if evidence can refute or support an assumption to varying degrees (i.e., refutes, mixed, weak, or strong support; Figure 1 Balance 1). In some cases, evidence might only be able to either refute or support an assumption (i.e., true or false) in a binary manner (Figure 2 Balance 2). A suggested process for using the BEAM and further details on each of its components will be explained in the following sections.

2.2 | Defining an assumption

Defining the assumption to be assessed is a crucial part of the process and needs to be carefully considered. We recommend following the guidance provided by the Conservation Measures Partnership's Conservation Standards 4.0 in formulating core assumptions underpinning a project (CMP, 2020) or the SIFT framework (Specht et al., 2022). These assumptions may be best expressed within useful planning tools such as a Theory of Change for action-related assumptions and a Situation Analysis for threat or status-related assumptions (these can be presented in text form or a diagram such as a Results Chain, which visualizes a Theory of Change; CMP, 2020).

These core assumptions are important to define clearly and carefully because they will influence the degree to which different pieces of evidence support or refute the assumption, and how relevant the evidence is. Ideally, the assumption should be specific, with intentional directionality (i.e., so it is clear whether evidence supports or refutes it), and include a reference to the decision context. An assumption such as "Protecting bird nests from predation will be beneficial" will typically be too general and will make assessing the relevance of evidence difficult later. An assumption such as "Protecting bird nests using fencing from mammalian predators on Island X will increase breeding success" can be more readily assessed with evidence, as the assumption specifies a particular action (fencing nests to protect nests from mammalian predators), an outcome of interest (breeding success) and a specific local context (Island X). Specifying assumptions in this way, alongside supplementing them with a clear background statement on the context in which the project and decision are taking

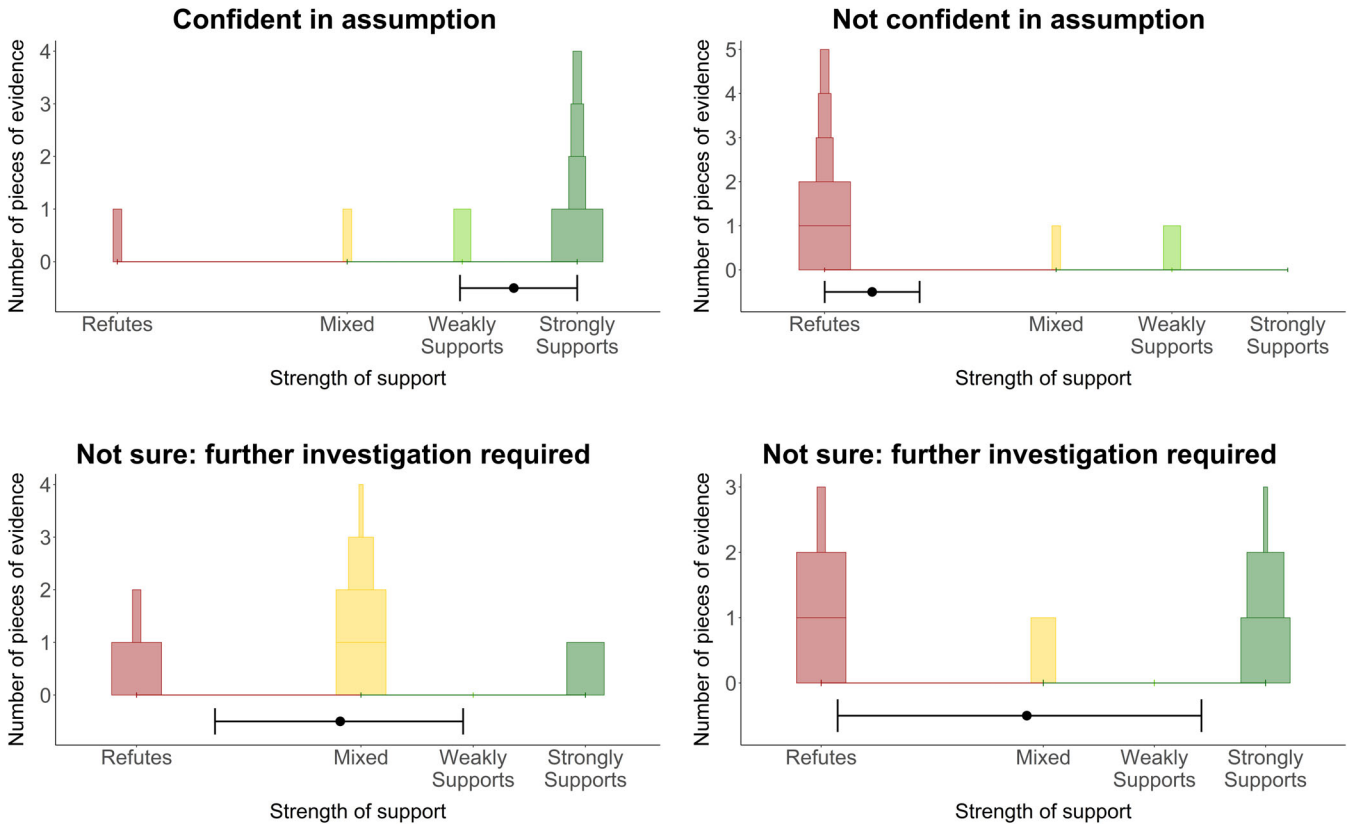

FIGURE 2 Four Ziggurat plots demonstrating how the weight of evidence refuting and supporting an assumption can be visualized using the Balance Evidence Assessment Method (BEAM). The top-left panel shows a situation where the balance of the weight of evidence is clearly in support of the assumption, whilst the top-right panel shows the opposite situation where the available evidence clearly refutes the assumption. The bottom two panels show situations where the available evidence is equivocal: the bottom left shows where the evidence is mainly mixed, whereas the bottom right shows conflicting pieces of evidence and is mixed overall. The weighted average strength of support effectively shows where the central point of balance lies in the overall support of evidence for the assumption (black point with bootstrapped 95% Confidence Intervals). Shiny application available here: https://alecchristie888.shinyapps.io/ziggurat-plot-app/.

place, will help to ensure the strength of support, information reliability, source reliability, and relevance of evidence can be properly considered.

2.3 | Assessing the weight of different pieces of evidence

Once one or multiple assumptions have been carefully defined, evidence can be gathered. Each piece of evidence is represented as a block, which holds a certain weight directly related to the *information reliability*, *source reliability*, and *relevance* of the piece of evidence (Figure 1).

Building on Salafsky et al.'s definition (2022) of *the reliability* of evidence (used in a broad sense, rather than the strict statistical definition), we specify two dimensions that reflect the quality and validity of the information provided (*information reliability*) and the source of the evidence (*source reliability*). Evidence with higher *information reliability* would reflect that the methodology or process of acquiring that information has high validity and builds our trust in the evidence. Therefore, we have greater confidence that the information provided is sound and valid, giving it greater weight (e.g., observations corroborated by several people, or collected over many years, observational or experimental data collected using a well-designed study; Tables 1; S1).

Evidence with greater *source reliability* will come from sources that have greater credibility and trustworthiness (i.e., a widely respected and experienced knowledge holder with a proven track record, or a scientific study from a peer-reviewed, high-quality journal authored by well-respected, experienced researchers). An observation, study, report, or anecdote with little information about the source or from an entity with a clear conflict of interest or questionable record would hold far less *source reliability* (Tables 1; S2).

The *relevance* of evidence reflects that certain evidence will provide more context-specific, transferable, and useful information than others based on what the

TABLE 1 Criteria for providing assessments for different aspects of evidence.

Evidence criteria	Categories			
Strength of support	Refutes Evidence refutes the assumption.	Mixed support Evidence does not support or refute the assumption, OR both support AND refute (to roughly an equal extent)	Weakly supports Evidence weakly supports the assumption	Strongly supports Evidence strongly supports the assumption
Weight of evidence (ISR: Information reliability × Source reliability × Relevance)				
Information reliability (I)	None The information provided has been derived or collected in a way that is misleading, unreliable, and invalid, and cannot be relied upon	Low The information provided has been derived or collected in a way that is not very reliable or valid—there are several areas of concern	Moderate The information provided has been derived or collected in a way that is generally reliable and valid, but there are some aspects that reduce our confidence in it	High The information provided has been derived or collected in a way that is highly reliable and valid
Source reliability (S)	None The evidence source is unreliable and cannot be trusted. The information they provide may be misleading, false, or untrue	Low The evidence source is either unknown or there are several concerns over their credibility. The information provided should be treated with caution	Moderate The evidence source is probably trustworthy, but there are some concerns that reduce our confidence that they are a reliable source	High The evidence source is highly trustworthy. There are no concerns over their credibility or reliability as a source of evidence
Relevance (R)	None Evidence has no relevance whatsoever to the assumption or specific context of interest—the information has no use	Low Evidence has limited relevance to the assumption and specific context of interest, and several small additional or a few large assumptions are required to relate the evidence to the assumption and context of interest	Moderate Evidence has some relevance to the assumption and specific context of interest, but a few small additional assumptions are required to relate the evidence to the assumption and context of interest	High Evidence is highly relevant to the assumption and specific context of interest. No additional assumptions are required to relate the evidence to the assumption and context of interest

Note: Strength of support assesses whether a piece of evidence refutes and/or supports the assumption of interest. Determining the weight of evidence for each piece of evidence involves considering three different criteria denoted by ISR: Information reliability, Source reliability, and Relevance.

evidence pertains to and from which system or context it was derived (Tables 1; S3). More relevant evidence comes from *specific evidence* (Salafsky et al., 2022) that directly applies to the assumption being assessed and the project or decision context (i.e., a local pilot study, report, or observations from a local knowledge holder), as well as external *generic evidence* that is likely to be highly transferable and applicable (i.e., a scientific study from a similar location, culture, or taxon being considered; a synthesis demonstrating transferability and generality; or experience and wisdom from a local ranger working in a similar nature reserve; Table 1). Less relevant evidence might come from an evidence synthesis with many studies that are out-of-date or from another region (or maybe relating to a broader group of taxa) where transferability is uncertain, or from a case study drawn from a very different habitat or location. A guiding principle for judging relevance is to consider how many additional assumptions are required (and how large those assumptions are) to relate the findings in the evidence to the assumption of interest; the more additional assumptions that are required, the lower the relevance of the evidence.

BOX 2 Testing the inter-rater reliability of evidence assessment category scores.

We conducted an inter-rater reliability test to assess the consistency with which different individuals scored the information reliability, source reliability, relevance, and strength of support of different pieces of evidence for different assumptions (Files S1–S4). We used an online questionnaire to collect participants' individual ratings of evidence (using numerical scores from 0 to 3 based on the categories in Table 1) for three different assumptions relating to a real conservation project adapted from (Irvine et al., 2021). Ethical approval was obtained through the Department of Geography, University of Cambridge. The full survey, collected data, and information on ethical approval are provided in the Supplementary Information (Files S1–S4). Participants were able to opt-out at any time and could choose to assess one, two, or all three assumptions.

We tested for consistency between participants' individual ratings using the Finn coefficient as this accounted for the limited range in scores that could be provided for each assessment category (i.e., 0 to 3; Finn, 1970; Gamer et al., 2012). The Finn coefficient ranges from 0 (no agreement) to 1 (perfect agreement), with scores greater than 0.7 considered to be indicative of strong agreement, and values between 0.5 and 0.7 considered as satisfactory (Finn, 1970; James et al., 1984; Lindstädt et al., 2020). We obtained ratings from 15 to 20 participants for each assumption, finding strong agreement for 10 out of 12 evidence ratings (Finn coefficients of 0.71–0.84), and satisfactory agreement for the other two ratings (Finn coefficient of 0.58 and 0.67)—detailed results are presented in Table S4. Finn's coefficient for strength of support, information reliability, source reliability, and relevance ranged between 0.75–0.83 for assumption 1, whilst assumptions 2 and 3 had smaller sample sizes ($n = 17$ and 15, respectively) and ranged between 0.578–0.749 and 0.738–0.841 for all categories (see Table S4). This indicates there was mostly strong agreement (coefficients of >0.7), and at least satisfactory agreement (coefficients between 0.5 and 0.7), between individuals when assessing the different evidence criteria.

The weight of evidence for a piece of evidence is therefore determined by the combination of information reliability, source reliability, and relevance (Figure 1). The higher the overall weight of evidence, the more likely that the balance will tilt. Key questions and guidance to help assess these different aspects of evidence are detailed in Tables S1–S3.

To visualize the overall weight of evidence using the BEAM, we suggest assigning simple numerical scores (ISR scores) to each block of evidence based on the previous assessments (Table 1) of information reliability, source reliability, and relevance (e.g., 0 = no reliability/relevance, 1 = low, 2 = medium, 3 = high; Sutherland et al., 2022). Each piece of evidence would receive a triplicate ISR score denoting the information reliability, source reliability, and relevance (e.g., ISR = 2|2|3; Sutherland et al., 2022). We have chosen a scale of 0–3 because this is the simplest scale possible with enough resolution to separate evidence of different qualities. However, different scales could be used depending on the user's preferences—for example, Sutherland et al. (2022) suggest a scale of 0–5 for ISR scores to allow greater distinction between levels. A standardized weight of a piece of evidence is determined by multiplying the scores together and dividing by the total possible score (e.g., using a 0–3 scale, for a maximum score of 27 (3 x 3 x 3), a piece scoring 2|2|3 would have an overall weight of evidence of $12/27 = 0.44$) and the cumulative weight of evidence would be obtained by summing these weights across all pieces of evidence. This gives an overall indication of the weight of the evidence across the balance (given that an individual weight of evidence of 1 represents a "perfect" block of evidence—that is, a total weight of 2 would be equivalent to two "perfect" blocks of evidence).

As scoring evidence could be subjective, we conducted an inter-rater reliability test (Box 2) to assess the consistency with which different individuals scored the different aspects of the weight and strength of support of a range of pieces of evidence for different assumptions (rating using numerical scores from 0 to 3 based on the Table 1 categories). We found mostly "strong," or at least "satisfactory," agreement between individuals in how they applied this scoring system (Finn, 1970; Gamer et al., 2012). As will be discussed later, it is important that the composition of the decision-making body or group assessing the evidence is as diverse and inclusive as possible, with a range of expertise and experience so that the collation and assessment of evidence (and ultimately decision-making; Hemming, Burgman, et al., 2018; Hemming, Walshe, et al., 2018) is high quality and not systematically biased against any particular source of evidence.

2.4 | Strength of support

Having assessed the weight of each piece of evidence, these evidence "blocks" can be placed on the balance by the degree to which the evidence supports the assumption (its *strength of support*). If there is gradation in support, evidence can be placed according to whether it refutes the assumption (left), provides mixed support (middle), weakly supports (partially to the right), or strongly supports (right) the assumption (Figure 1 Balance 1). We decided to have only one "refutes" category as refuting evidence is typically rare overall and is difficult to justify differentiating into categories of weakly and strongly refutes. If evidence can only support or refute an assumption (i.e., in a binary way), then evidence is either placed on the left (refuting evidence) or right (supporting evidence) of the balance (Figure 1 Balance 2).

2.5 | Visualizing the weight of evidence and evaluating confidence in an assumption

An intuitive way to visualize the collective weight and strength of support of evidence relating to an assumption is a "Ziggurat plot" (Figure 2), for which we have created an interactive Shiny App platform to demonstrate its use and provide guidance (https://alecchristie888.shinyapps.io/ziggurat-plot-app/). This type of plot has been piloted in the Conservation Learning Initiative (2022) and is customisable depending upon the categories and scale being used for the evidence criteria.

A Ziggurat plot stacks blocks of evidence in order of their weight (the size of blocks is proportional to their ISR scores) along the y-axis, separated out into each strength of support category along the x-axis. After assigning numerical values to each support category (e.g., Refutes $= -2$, Mixed $= 0$, Weakly supports $= 1$, Strongly supports $= 2$), a weighted average strength of support can be calculated (weighted by the standardized ISR scores for each piece of evidence) to show where the balance of evidence lies. A bootstrapped 95% confidence interval is plotted alongside this to show the uncertainty (i.e., the variability in the strength of support and weight of evidence). This 95% confidence interval is sensitive to the number of evidence pieces (i.e., narrowing when more evidence is available).

It may be desirable to convert the *strength of support* into a judgment of *confidence in an assumption*. There are two options to do this. The first option is to use the Ziggurat plot as an advisory tool to assess our confidence in an assumption by considering three key aspects: 1. the average strength of support; 2. the variability in this support; and 3. the overall weight of evidence. A practical rule of thumb may be that if the error bar overlaps several categories, then there is mixed support in assumption, whilst if the error bar overlaps two closely related categories, the category with the greatest weight of evidence should be selected. Based on this assessment, it can then be concluded as to whether the available evidence suggests we are: (1). Not confident in the assumption; (2). Not sure and require further investigation; or (3). Confident in the assumption (Figure 2; Table 2).

The second option requires the incorporation of a measure called the required level of proof. The *required level of proof* is related to the *burden of proof* (usually carried by those responsible for making a given decision) and can be defined as "the weight of evidence required to believe the assumption is valid" (Salafsky & Redford, 2013). This is dependent on the nature of the assumption (its *implausibility*—i.e., *how implausible is the assumption?*) following Laplace's Principle that: "the weight of evidence for an extraordinary claim must be proportioned to its strangeness" (Gillispie et al., 1999), or put another way, "extraordinary claims require extraordinary evidence." The required level of proof also grows as the wider decision's consequences and the relative risks of action versus inaction increase—that is, if we wrongly act based on the assumption, what are the risks and consequences? And vice versa—that is, what happens if we wrongly fail to act? (Salafsky & Redford, 2013).

Determining the required level of proof before assessing an assumption is useful as this can ensure that the investment in evidence collation and assessment is proportional to the risks and consequences involved and therefore as efficient as possible (Sutherland et al., 2021). It is also important when considering the sufficiency of the weight of evidence to conclude that an assumption is valid. One way to incorporate this into the BEAM is by giving greater influence to evidence that refutes the assumption. This can be done by mentally enlarging the length of the refutation side of the balance (Figure 3) for more implausible assumptions or those that entail greater levels of risk. For example, adjusting the numeric value assigned to refuting evidence from -2 to -4 effectively leads to a doubling of the influence of refuting evidence when calculating the average strength of support (this is built into the Ziggurat plot Shiny App, as shown in Figure 3). Caution should be exercised, however, when determining the required level of proof given this is a subjective judgment. A useful exercise would be to conduct a sensitivity analysis to see the effect that altering the required level of proof has on the Ziggurat plot and the average strength of support. This in turn can help to judge our confidence in the validity of an assumption

TABLE 2 Three potential levels of confidence in an assumption that could be concluded from weighing the evidence and visualizing it using a Ziggurat plot, including what the implications would be for decision-making.

Confidence in assumption	Consulting Ziggurat plot	Implications for decision-making
Not confident	1. Average strength of support firmly refutes the assumption 2. AND variation in strength of support is acceptably small or negligible 3. AND there is a large enough weight of evidence	The available evidence clearly refutes your assumption. It may be useful to reconsider or modify the strategy or action(s) you were intending to implement and assumptions being made, or defer action completely. It may not be useful to invest more effort in collecting evidence for this assumption, at least until this is revisited in the future, or the available evidence changes substantially.
Not sure: further investigation required (lack of evidence or highly variable evidence)	1. Average strength of support is around mixed support 2. AND/OR there is highly variable support amongst the evidence 3. AND/OR there may also be a small weight of evidence	There is not currently sufficient evidence to confidently refute or support the assumption. There are several possible implications: 1. If there is a lack of evidence, consider investing in acquiring and assessing more evidence, including additional research or synthesis of evidence; 2. If there is a lot of variability in the strength of support, this may suggest some moderating variable or context-dependency needs to be addressed. For example, a reintroduction may only be effective if the original cause of decline has been addressed. You may decide to move forward with your decision (despite the uncertainty) but to periodically revisit this assumption alongside collecting monitoring data and generating evidence to feed into adaptive management plans and future assessments of evidence. If the assumption does not represent a critical part of your strategy or project plans, you may also decide to revisit this later and/or accept the uncertainty around this assumption. Alternatively, you may want to reconsider your strategy or action(s) if you feel the available evidence for this assumption is too mixed and equivocal to make a decision.
Confident	1. Average strength of support firmly supports the assumption 2. AND variation in strength of support is acceptably small or negligible 3. AND large enough weight of evidence	The available evidence has given you a satisfactory level of confidence in your assumption and you can proceed to the next one to be assessed. You should monitor the effects of any actions taken carefully and be ready to undertake the assessment again if new evidence arises (either through conducting monitoring or from new or existing evidence sources).

based on the available evidence—if there is substantial sensitivity, this would lower our confidence.

2.6 | Applying the BEAM to a set of assumptions

The process of assessing an assumption can be repeated if several separate assumptions are being assessed as part of a Theory of Change or Situation Analysis. To work out an overall level of confidence in a set of assumptions, we suggest following the *critical weakest link* approach set out by Salafsky et al. (2022). A set of assumptions, for example, could relate to different impacts of the conservation action (e.g., its acceptability, feasibility, costs, and effectiveness) or different stages in a strategy (e.g., status, threats, actions, and alternatives; Box 1). The critical weakest link approach simply means that the overall confidence in a set of assumptions is determined by the lowest level of confidence amongst the assumptions that are rated as critical or essential (i.e., if we are not confident about these assumptions, then we would not be confident about the collective set of assumptions given their connectedness in a Theory of Change; Salafsky et al., 2022). This approach also helps to prioritize effort in assessing different assumptions—clearly, targeting the

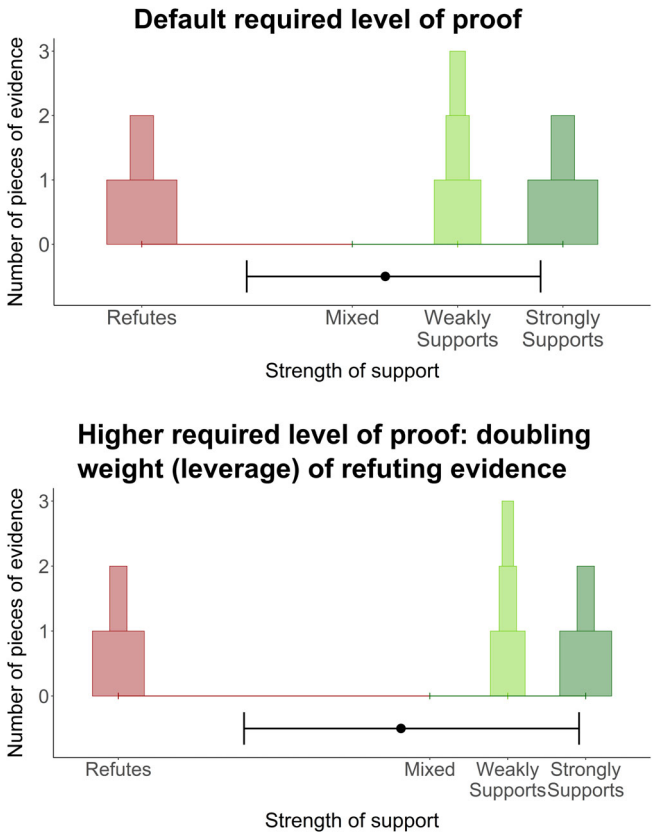

FIGURE 3 Top panel with a numeric value for refutes set to −2 and bottom panel with numeric value for refutes set to −4. This effectively doubles the influence of refuting evidence when determining the weighted average strength of support and drags this further to the left (toward the refuting side of the balance), lowering the confidence we have in the assumption. Shiny application available here: https://alecchristie888.shinyapps.io/ziggurat-plot-app/.

most critical or essential assumptions in a Theory of Change, for example, would be the most efficient use of time and resources.

For complex interventions, it is important to appropriately assess the different components of an intervention using multiple sets of assumptions—for example, not just assessing the ecological effects of a strategy, but also its feasibility, acceptability, and costs (both social and economic; Christie, Downey, et al. 2022). This often requires cross-sectoral and/or cross-disciplinary thinking to ensure diagrams representing Theories of Change, such as Results Chains, are complete (Tallis et al., 2019).

3 | DISCUSSION

Our approach, the BEAM, presented here represents a novel and useful way to overcome the challenge of assessing the diverse sources of evidence that typify many conservation challenges. Here we discuss the strengths and limitations of our approach, as well as how it can be applied to existing tools and frameworks for making decisions in conservation.

3.1 | Tackling the challenges of assessing evidence of diverse types and sources

One of the core strengths of the BEAM is its flexibility, enabling us to assess a diverse range of evidence from a wide variety of sources, which using our broad definition of evidence could include local expert and practical knowledge, Indigenous and Local Knowledge (ILK), studies and syntheses from the scientific literature, and the gray literature across the social and natural science spectrum (Bennett et al., 2017; Salafsky et al., 2022).

Suggestions on how to consider evidence from a diverse range of sources in an equitable and fair manner have been proposed by several authors, particularly regarding the use of ILK (sometimes called Traditional knowledge) from knowledge systems distinct from Western science. For example, Two-Eyed Seeing (Etuaptmumk in Mi'kmaw) is a conceptual framework developed by Mi'kmaw Elder Albert Marshall that encourages using the strengths of different knowledge systems to better understand complex systems (Bartlett et al., 2012; No'kmaq et al., 2021; Reid et al., 2021). Another framework set out by Tengö et al. (2014, 2017), and further detailed by Malmer et al. (2020), is the Multiple Evidence Base framework developed collaboratively with the Intergovernmental Science Policy Platform on Biodiversity and Ecosystem Services (IPBES). This framework is composed of three stages: (1) joint problem formulation; (2) generating an enriched picture with contribution from multiple sources of evidence; and (3) joint analysis and evaluation of knowledge. The Multiple Evidence Base framework places great emphasis on triangulation across knowledge systems, recognizing that different forms of evidence can be equally valid and that the relevance of evidence is a critical part of making decisions. Such frameworks and many others (e.g., Extreme Citizen Science; Chiaravalloti et al., 2022; Cross-disciplinary evidence principles; Game et al., 2018) were developed partly due to concerns over the underutilisation or even exclusion (explicitly or implicitly) of certain sources of evidence to make decisions—in particular, local forms of knowledge (e.g., expert, practical, traditional, and/or ILK; Malmer et al., 2020; Smith et al., 2009).

Whilst these conceptual frameworks exist, few attempts have been documented to develop practical and pragmatic processes to assess diverse forms of evidence

(Malmer et al., 2020). This was a major motivation for developing the BEAM since better decisions will typically come from a considered and thoughtful usage of a wide, diverse range of sources and types of evidence assessed by a diverse group of individuals (Herzog & Hertwig, 2009; Lord et al., 1984). It is also likely that in many cases, decision-makers are already (formally or informally) weighing diverse sources of evidence to make decisions and there is therefore a pressing need to formally define and discuss transparent, pragmatic, and structured guidance and processes, like the BEAM, that can help decision-makers assess this information.

In its present form, our method can be used to assess a diverse range of evidence, but important considerations and safeguards need to be put in place to ensure fair assessment of evidence, particularly in the case of Indigenous knowledge. Like Two-Eyed Seeing, we designed the BEAM to avoid any domination by one worldview or another, or assimilation by one worldview of the knowledge of another (No'kmaq et al., 2021; Reid et al., 2021). Pieces of evidence are each treated by the user as individual and distinct pieces of knowledge or information in their own right, much like the lines or streams of knowledge in the diagram of the Multiple Evidence Base framework presented by Tengö et al. (2017). Although Two-Eyed Seeing generally aims to avoid having different sources of evidence conflicting with one another, we respectfully believe here that this is not a pragmatic approach to take in the context of conservation project decision support. In the BEAM, different pieces of evidence could conflict whether they are from the same or different sources of evidence (i.e., one refuting and one supporting an assumption). This is not to be feared, but to be embraced because sometimes conflict between pieces of evidence can contribute to learning and a better understanding of the problem or system (i.e., what are the reasons behind these differences?).

Another key feature of the BEAM is that no external or a priori hierarchy or system is enforced on pieces of evidence. Instead, all pieces of evidence start out on a level playing field and it is up to the user to assess how much they trust each piece of evidence based upon three suggested aspects of its weight: (1) the reliability of the information; (2) the reliability of the source; and (3) the relevance of the information. This means that regardless of the type or source of evidence, any piece of evidence can be given a high weight of evidence—that is, evidence from a scientific experiment and evidence from detailed observations of changes over time by a community could both score highly. We supply guidelines of possible important questions that those assessing evidence could ask (Tables S1–S3), which aim to encompass a wide range of possible questions that could be asked of a variety of types and sources of evidence. Our method also places great emphasis on the relevance of evidence, which enables the recognition that often the expertise or wisdom of a local reserve manager, for example, may represent more relevant evidence than the general recommendations of a systematic review, for example.

However, whilst the freedom given to those assessing evidence using the BEAM can be regarded as a strength (in terms of enabling inclusive use and weighing of diverse sources of evidence), it could also be regarded as a limitation. Without proper safeguards and considerations in place when implementing the BEAM, the method could be misused and systematically exclude or disadvantage evidence from particular sources when assessing assumptions. For this reason, we want to highlight that it is of critical importance that anyone applying or adapting the BEAM to assess evidence applies the following principles, particularly when including ILK.

3.2 | Principles and considerations for applying the BEAM

First, as a general rule, the entire process of defining assumptions, assessing evidence, and making decisions will benefit from involving a diverse set of individuals and perspectives (Hemming, Burgman, et al., 2018; Hemming, Walshe, et al., 2018). However, a relatively low-risk, simple decision where available evidence sources come from a single source or sources may only require one or two appropriately qualified individuals (e.g., a private landowner) to assess the evidence.

Nevertheless, for higher-risk, complex decisions where a diverse range of evidence is available, it is particularly important to involve a diverse, inclusive group of individuals. This will help mitigate the risks that different types of decision-makers and project partners could classify different pieces of evidence with different weights and strengths of support for an assumption. For example, inconsistencies in rating evidence could come from different levels of understanding about how the information provided by the evidence was derived or collected, different past experiences, viewpoints, or perceptions of different types of evidence from different sources (e.g., preconceived views on the relative quality of scientific literature, gray literature, or ILK). Ensuring that a diverse group of evidence assessors participate actively and fairly in the process will help to ensure that imbalances in the power dynamics at play during such decision-making processes do not side-line any one form of evidence or knowledge (scientific, local, or Indigenous), thus making the best use of all the available evidence.

Our pilot study found that there was satisfactory to strong consistency between different individuals in how they applied scores to different pieces of evidence, which were deliberately designed to range across a wide continuum of sources (Table S4; Files S1, S2). Whilst this is encouraging, our sample was biased towards academics, scientists, and practitioners from NGOs in Western countries. Further testing of how the backgrounds of those assessing evidence influence the scores given to evidence from different sources would be an important step to quantify and mitigate such assessment bias.

Nevertheless, clearly, a central principle to applying the BEAM should be that *those assessing evidence should be as representative of the sources of evidence as possible*. For example, if Indigenous knowledge is part of the available evidence, Indigenous partners need to be fully involved in assessing it (and ideally throughout the project) so that there is a shared understanding and appreciation for the methods and context associated with all the evidence. A good example of this comes from IPBES when the 2017 IPBES Plenary in decision IPBES-5/1 approved an approach to "recognising and working with ILK within IPBES assessments," which included the establishment of "Indigenous and local knowledge liaison groups" (IPBES, 2017). Whilst this has its challenges (McElwee et al., 2020) and many smaller organizations and projects may not have the resources to replicate this process in full, the general principles and approach could be applied and adapted into other decision-making workflows at a variety of scales.

A similar principle can be applied to assessing evidence from different disciplines (Game et al., 2018)—for example, when evidence from the social sciences is being assessed, there should be appropriate representation from social scientists in the team assessing the evidence. Furthermore, appropriate representation is also important to ensure that a complete set of assumptions are assessed for projects with complex Theories of Change—particularly those involving complex interventions that cross-sectoral and disciplinary boundaries (Tallis et al., 2019).

The BEAM also encourages users to make any judgments on scoring different pieces of evidence clear and transparent, such as by providing a means to document these scores (e.g., using the Ziggurat plot Shiny App: https://alecchristie888.shinyapps.io/ziggurat-plot-app/) to enable scrutiny and critique of how evidence was assessed and whether decisions should be revisited.

Second, in decision-making situations where Indigenous participation is important, the application of our method should be led by, and its design co-adapted to the needs and viewpoints of, local and Indigenous scientists and partners. This will help to ensure that the questions (e.g., Tables S1–S3) and modes of assessing evidence and assumptions are equitable, respectful, inclusive, and above all useful to those making the decisions and designing conservation projects and strategies. In addition, applying this principle to the use of the BEAM provides the opportunity to establish more "evidence bridges" and "knowledge brokers" in conservation (Kadykalo, Buxton, et al., 2021). This will help to achieve more inclusive decision-making groups and help to enhance the information on the attributes and context behind the evidence (e.g., metadata, details and explanation of methodology, its validity, and context which the decision-makers may not be aware of or appreciate). This is an important consideration because we risk ignoring or excluding certain forms of evidence from decision-making processes if such evidence is not respectfully and equitably brought to the decision-making table or "the balance."

3.3 | Value of assumption-based thinking and applying the BEAM to existing tools

We believe that the BEAM provides an intuitive process that can help decision-makers identify and critically examine the assumptions they are making in conservation projects, as well as the important attributes of evidence to consider (e.g., information reliability, source reliability, and relevance). Explicitly defining key assumptions of conservation projects, critically evaluating the strength of support and weight of the available evidence, and our confidence in an assumption's validity, would be a crucial step forward toward ensuring more effective, efficient, and equitable conservation (Specht et al., 2022; Sutherland et al., 2022; White et al., 2022). This is because the more confidence we have in the validity of assumptions underpinning conservation projects, the more confident we can be that conservation projects will succeed at achieving their goals. In addition, the BEAM can be used to assess the evidence that a conservation project has achieved its goals (formulated as an assumption) by evaluating evidence generated by the project itself. In this way, the BEAM can be used both before, during, and after the beginning of a project.

The use of the BEAM should also encourage transparency in decision-making, as the process will typically involve the documentation of evidence used to guide decisions and their assessment (e.g., using the Ziggurat plot Shiny App). This means that resultant decisions can be updated if new evidence becomes available, or if others disagree with how that evidence has been assessed. Doing so could help push organizations towards greater information sharing, due diligence, and professional practice in conservation, particularly through generating "decision libraries" (Christie, Downey, et al., 2022) that can help

TABLE 3 Examples of how the Balance Evidence Assessment Method (BEAM) can be applied to different existing decision support tools, platforms, and frameworks.

Decision support tool, platform, or framework	Application of the BEAM
Structured Decision Making	Tools deployed through a Structured Decision Making framework seek to find optimal actions to achieve desired outcomes, whilst also balancing a range of stakeholder objectives under uncertainty (Gregory et al., 2012). A tool such as a consequence table to compare different alternatives could be formulated in terms of assumptions (Table S5), where each assumption may differ in its importance based on different stakeholder objectives (e.g., a given alternative will increase the population size by at least 20%, acceptable to local partners, cost-effective, or feasible). The consequence table could be color-coded by the level of confidence in each assumption reflecting the weight of evidence and strength of support (e.g., for effects on population change, costs, acceptability, and feasibility) alongside the data usually included in consequence tables such as effect sizes, costs, and values (Table S5). This would add an explicit consideration of the weight of evidence for different assumptions to the consequence table approach.
Bayesian Belief Networks/Bayesian Models	A more quantitative tool, Bayesian Belief Networks (BBNs) or Bayesian models of which BBNs are a subset (Marcot et al., 2006), could also be used to apply the principles of the BEAM by bringing in the information gained in the Ziggurat plots to define the prior probability distribution for variables of interest described by the assumption. With the primary goal of the reduction of uncertainty, the models benefit from informed priors that reflect the state of knowledge, or the evidence base going into the model to inform the decision-making.
Argument Maps	Argument maps are less well-known as a decision-making tool in the environmental sciences but offer a useful way to lay out the logical steps of reasoning behind a decision (Keith et al., 2017). These maps start with a claim (a form of *assumption*) backed up by different reasons (*sub-assumptions*) that are supported or rebutted by different pieces of evidence. Therefore, the BEAM could be used to work out the confidence in each of the reasons based on different pieces of evidence (Figure S2). Reasons in the argument map could be labeled with levels of confidence (and pieces of evidence could also be labeled with their weight of evidence and strength of support). These could then be combined (using the *critical weakest link* approach) to work out the confidence in the overall claim that the argument map is considering (Figure S1).
Situation Analyses and Theories of Change	The BEAM can also be applied to Situation Analysis and Theory of Change diagrams that can be built using *Miradi Software* to support the Conservation Measures Partnership's Conservation Standards 4.0 (CMP, 2020). The standards help explicitly lay out the assumptions made during the design of a conservation project. The Conservation Learning Initiative (2022) has already begun implementing some of the concepts illustrated here, including Ziggurat plots.
Evidence-to-Decision Tool	The BEAM could also be directly integrated into the Evidence-to-Decision tool (Christie, Downey, et al., 2022). The different steps within the tool can be reframed as sub-assumptions (when assessing effectiveness, costs, acceptability, feasibility, and modifications of potential alternative actions) and the confidence in each of these can be weighed up using the available evidence. This tool's main aim is to document the reasoning and evidence behind decisions and this would therefore provide a useful way to document and visualize the BEAM for different decisions.
Intergovernmental Science Policy Platform on Biodiversity and Ecosystem Services (IPBES) Assessments	Claims made in IPBES assessments could be assessed using the BEAM. For example, 18 categories of Nature's Contributions to People (NCP) were assessed in the Global Assessment (IPBES, 2019), for which evidence was gathered from a variety of sources, including Indigenous and Local Knowledge (McElwee et al., 2020). One way the BEAM could help is by providing a structured approach to assess the weight and support of a diverse range of evidence for, and overall levels of confidence in, these NCP-related claims.

to show how and why decisions were made, as well as to monitor how well organizations are using evidence to make decisions using particular metrics (e.g., criteria for "Evidence Champions"; Conservation Evidence, 2023; or an evidence-use index being developed by the US Fish and Wildlife Service for funding decisions; Sutherland et al. 2022).

A key consideration in the development of the BEAM was that it should be able to be applied to any decision-making tool, platform, or framework that requires making assumptions. Although assumptions may not be explicitly referred to, many units and parts of these processes can often be reframed as an assumption (they may be claims or hypotheses, for example). As discussed in

the previous section, the application of the BEAM could help enable these existing tools, platforms, and frameworks to better consider a diverse range of evidence (Table 3).

4 | CONCLUSION

Here we have presented the BEAM, an approach for weighing and assessing a diverse range of evidence for different assumptions that are made during conservation decision-making and the planning of conservation projects and strategies. This represents a key step forward in recognizing that diverse forms of evidence can and should feed into conservation decisions and that this can be done in a way that usefully assesses the weight of evidence, placing the power of assessing evidence in the hands of project teams and partners. Whilst this power could be misused, intentionally or not, following some simple principles to applying the BEAM (particularly if Indigenous knowledge forms part of the available evidence base) can help to safeguard and avoid the exclusion or underweighting of any particular type of evidence. Our method has been designed to be able to be adapted and used by anyone within any existing decision-making tool, platform, or framework that makes an assumption (or a claim or hypothesis that could be reframed as an assumption), potentially ranging from small conservation projects up to international policy platforms such as IPBES. We explicitly talk about the method in a conservation setting, but a similar method and framing could be useful in other disciplines where evidence from diverse sources needs to be assessed together to judge confidence in certain assumptions being made. We hope that the BEAM can be applied or adapted to help enable and improve the weighing of diverse evidence to make complex decisions in a range of situations.

AUTHOR CONTRIBUTIONS

Conceptualisation: Alec P. Christie, Thomas B. White, William H. Morgan, Nick Salafsky, Robyn Irvine, William J. Sutherland. *Methodology*: Alec P. Christie. *Statistical analysis*: Alec P. Christie. *Writing—original draft preparation*: Alec P. Christie. *Writing—review and editing*: All authors. *Visualization*: Alec P. Christie. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

Thanks to everybody involved in implementing the Conservation Learning Initiative (https://conservation-learning.org/), which provided the space to test and develop approaches presented in this paper, as well as 20 anonymous individuals who pilot-tested the scoring systems described in this paper. Thank you also to Hannah Wauchope for help in writing code for the Ziggurat plot shiny app, Howard Nelson for helping us through the ethics review process for the pilot survey, and to Mark Burgman for advice and comments on the manuscript. Comments provided by two anonymous reviewers greatly helped to improve the manuscript. For the purpose of open access, the lead author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data analyzed in this article are provided in the appendix.

ETHICS STATEMENT

Ethical approval for data collected in this study was given by the Department of Geography's ethics committee, University of Cambridge: application ID: 2346. Free, prior, and informed consent was obtained for all participants, and participants were informed their responses would be anonymous.

ORCID

Alec P. Christie https://orcid.org/0000-0002-8465-8410
Nick Salafsky https://orcid.org/0000-0001-9665-7477
Thomas B. White https://orcid.org/0000-0002-0536-6162
Rafael M. Chiaravalloti https://orcid.org/0000-0003-4271-3749
William J. Sutherland https://orcid.org/0000-0002-6498-0437

REFERENCES

Adams, W. M., & Sandbrook, C. (2013). Conservation, evidence and policy. *Oryx*, *47*, 329–335. https://doi.org/10.1017/S0030605312001470

Bartlett, C., Marshall, M., & Marshall, A. (2012). Two-eyed seeing and other lessons learned within a co-learning journey of bringing together indigenous and mainstream knowledges and ways of knowing. *Journal of Environmental Studies and Sciences*, *2*, 331–340. https://doi.org/10.1007/s13412-012-0086-8

Bennett, N. J., Roth, R., Klain, S. C., Chan, K., Christie, P., Clark, D. A., Cullman, G., Curran, D., Durbin, T. J., Epstein, G., Greenberg, A., Nelson, M. P., Sandlos, J., Stedman, R., Teel, T. L., Thomas, R., Veríssimo, D., & Wyborn, C. (2017). Conservation social science: Understanding and integrating human dimensions to improve conservation. *Biological Conservation*, *205*, 93–108. https://doi.org/10.1016/j.biocon.2016.10.006

Chiaravalloti, R. M., Skarlatidou, A., Hoyte, S., Badia, M. M., Haklay, M., & Lewis, J. (2022). Extreme citizen science: Lessons

learned from initiatives around the globe. *Conservation Science and Practice*, *4*, e577. https://doi.org/10.1111/csp2.577

Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. R., & Sutherland, W. J. (2020). Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation*, *248*, 108666. https://doi.org/10.1016/j.biocon.2020.108666

Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. R., & Sutherland, W. J. (2021). The challenge of biased evidence in conservation. *Conservation Biology*, *35*, 249–262. https://doi.org/10.1111/cobi.13577

Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., & Sutherland, W. J. (2022). Innovation and forward-thinking are needed to improve traditional synthesis methods: A response to Pescott and Stewart. *Journal of Applied Ecology*, *59*, 1191–1197. https://doi.org/10.1111/1365-2664.14154

Christie, A. P., Downey, H., Frick, W. F., Grainger, M., O'Brien, D., Tinsley-Marshall, P., White, T. B., Winter, M., & Sutherland, W. J. (2022). A practical conservation tool to combine diverse types of evidence for transparent evidence-based decision-making. *Conservation Science and Practice*, *4*, e579. https://doi.org/10.1111/csp2.579

CMP. (2020). The Open Standards for the Practice of Conservation, Version 4.0 [WWW Document]. Conservation Measures Partnership. www.conservationmeasures.org

Conservation Evidence. (2023). The Evidence Champions [WWW Document]. https://www.conservationevidence.com/content/page/112

Conservation Learning Initiative. (2022). Conservation Learning Initiative.

Conservation Standards. (2023). Assess the Conservation Situation: Step 1. Assess. [WWW Document]. URL express.adobe.com/page/5LOUbotBsqHCF/

Cook, C. N., Nichols, S. J., Webb, J. A., Fuller, R. A., & Richards, R. M. (2017). Simplifying the selection of evidence synthesis methods to inform environmental decisions: A guide for decision makers and scientists. *Biological Conservation*, *213*, 135–145. https://doi.org/10.1016/j.biocon.2017.07.004

Cote, D., van Leeuwen, T. E., Bath, A. J., Gonzales, E. K., & Cote, A. L. (2021). Social-ecological management results in sustained recovery of an imperiled salmon population. *Restoration Ecology*, *29*, e13401. https://doi.org/10.1111/rec.13401

Evans, M. C., Davila, F., Toomey, A., & Wyborn, C. (2017). Embrace complexity to improve conservation decision making. *Nature Ecology & Evolution*, *1*, 1588. https://doi.org/10.1038/s41559-017-0345-x

Finn, R. H. (1970). A note on estimating the reliability of categorical data. *Educational and Psychological Measurement*, *30*, 71–76. https://doi.org/10.1177/001316447003000106

Game, E. T., Tallis, H., Olander, L., Alexander, S. M., Busch, J., Cartwright, N., Kalies, E. L., Masuda, Y. J., Mupepele, A.-C., Qiu, J., Rooney, A., Sills, E., & Sutherland, W. J. (2018). Cross-discipline evidence principles for sustainability policy. *Nature Sustainability*, *1*, 452–454. https://doi.org/10.1038/s41893-018-0141-x

Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package 'irr.' Var. coefficients interrater Reliab. Agreem. 22, 1–32.

Gillispie, C. C., Gratton-Guinness, I., & Fox, R. (1999). *Pierre Simon Laplace, a life in exact science*. Princeton University Press.

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2012). *Structured decision making: A practical guide to environmental management choices*. John Wiley & Sons.

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, *9*, 169–180. https://doi.org/10.1111/2041-210X.12857

Hemming, V., Walshe, T. V., Hanea, A. M., Fidler, F., & Burgman, M. A. (2018). Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. *PLoS One*, *13*, e0198468. https://doi.org/10.1371/journal.pone.0198468

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*, 231–237. https://doi.org/10.1111/j.1467-9280.2009.02271.x

IPBES. (2017). Report of the Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on the work of its fifth session: Annex II to decision IPBES-5/1. https://www.ipbes.net/sites/default/files/inline/files/ipbes_ilkapproach_ipbes-5-15.pdf

IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Version 1). Zenodo. https://doi.org/10.5281/zenodo.6417333

Irvine, R., Graham, B., Harper, P., Laurin, A., Lavers, A., Persohn, C., Prior, K., Bisaillon, J.-F., Neufeld, L., Boshoven, J., & Salafsky, N. (2021). A practical approach to developing the evidence base for conservation and restoration (CoRe) projects in Canadian national parks. Gatineau, QC, Canada: Parks Canada. https://docs.google.com/presentation/d/1CnwEBUb-nJtC1xXtkxo9_LOe9RtJwH5_GaPRe-Il_Ng/edit#slide=id.gb52ad4f6d8_0_11.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *The Journal of Applied Psychology*, *69*, 85–98. doi:10.1037/0021-9010.69.1.85

Kadykalo, A. N., Buxton, R. T., Morrison, P., Anderson, C. M., Bickerton, H., Francis, C. M., Smith, A. C., & Fahrig, L. (2021). Bridging research and practice in conservation. *Conservation Biology*, *35*, 1725–1737. https://doi.org/10.1111/cobi.13732

Kadykalo, A. N., Cooke, S. J., & Young, N. (2021). The role of western-based scientific, indigenous and local knowledge in wildlife management and conservation. *People and Nature*, *3*, 610–626. https://doi.org/10.1002/pan3.10194

Kahneman, D. A., Butchart, S. H. M., Regan, H. M., Harrison, I., Akçakaya, H. R., Solow, A. R., & Burgman, M. A. (2011). Inferring extinctions I: A structured method using information on threats. *Biological Conservation*, *214*, 320–327.Thinking, Fast and Slow (New York: Macmillan). https://doi.org/10.1016/j.biocon.2017.07.026

Keith, D. A., Butchart, S. H., Regan, H. M., Harrison, I., Akçakaya, H. R., Solow, A. R., & Burgman, M. A. (2017). Inferring extinctions I: a structured method using information on threats. *Biological Conservation*, *214*, 320–327.

Lindstädt, R., Proksch, S.-O., & Slapin, J. B. (2020). When experts disagree: Response aggregation and its consequences in expert

surveys. *Political Science Research and Methods*, *8*, 580–588. https://doi.org/10.1017/psrm.2018.52

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*, 1231–1243. https://doi.org/10.1037//0022-3514.47.6.1231

MacLeod, C. J., Brandt, A. J., & Dicks, L. V. (2022). Facilitating the wise use of experts and evidence to inform local environmental decisions. *People and Nature*, *4*, 904–917. https://doi.org/10.1002/pan3.10328

Malmer, P., Vanessa, M., Austin, B., & Tengö, M. (2020). Mobilisation of indigenous and local knowledge as a source of useable evidence for conservation partnerships. In W. Sutherland, P. Brotherton, Z. Davies, N. Ockendon, N. Pettorelli, & J. Vickery (Eds.), *Conservation research, policy and practice (ecological reviews)* (pp. 82–113). Cambridge University Press. https://doi.org/10.1017/9781108638210.006

Marcot, B. G., Steventon, J. D., Sutherland, G. D., & McCann, R. K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, *36*, 3063–3074. https://doi.org/10.1139/x06-135

McElwee, P., Fernández-Llamazares, Á., Aumeeruddy-Thomas, Y., Babai, D., Bates, P., Galvin, K., Guèze, M., Liu, J., Molnár, Z., Ngo, H. T., Reyes-García, V., Roy Chowdhury, R., Samakov, A., Shrestha, U. B., Díaz, S., & Brondízio, E. S. (2020). Working with indigenous and local knowledge (ILK) in large-scale ecological assessments: Reviewing the experience of the IPBES global assessment. *Journal of Applied Ecology*, *57*, 1666–1676. https://doi.org/10.1111/1365-2664.13705

No'kmaq, M's., Marshall, A., Beazley, K. F., Hum, J., Joudry, S., Papadopoulos, A., Pictou, S., Rabesca, J., Young, L., & Zurba, M. (2021). "Awakening the sleeping giant": re-indigenization principles for transforming biodiversity conservation in Canada and beyond. *Facets*, *6*, 839–869. https://doi.org/10.1139/facets-2020-0083

Papworth, S. (2017). Decision-making psychology can bolster conservation. *Nature Ecology & Evolution*, *1*, 1217–1218. https://doi.org/10.1038/s41559-017-0281-9

Reid, A. J., Eckert, L. E., Lane, J.-F., Young, N., Hinch, S. G., Darimont, C. T., Cooke, S. J., Ban, N. C., & Marshall, A. (2021). "Two-eyed seeing": An indigenous framework to transform fisheries research and management. *Fish and Fisheries*, *22*, 243–261. https://doi.org/10.1111/faf.12516

Rowland, J. A., Moore, J. L., & Walsh, J. C. (2023). How expert insight into alpine peatland conservation complements global scientific evidence. *Conservation Biology*, *00*, e14081. https://doi.org/10.1111/cobi.14081

Salafsky, N., & Redford, K. H. (2013). Defining the burden of proof in conservation. *Biological Conservation*, *166*, 247–253.

Salafsky, N., Irvine, R., Boshoven, J., Lucas, J., Prior, K., Bisaillon, J.-F., Graham, B., Harper, P., Laurin, A. Y., Lavers, A., Neufeld, L., & Margoluis, R. (2022). A practical approach to assessing existing evidence for specific conservation strategies. *Conservation Science and Practice*, *4*, e12654. https://doi.org/10.1111/csp2.12654

Schuster, R., Germain, R. R., Bennett, J. R., Reo, N. J., & Arcese, P. (2019). Vertebrate biodiversity on indigenous-managed lands in Australia, Brazil, and Canada equals that in protected areas. *Environmental Science & Policy*, *101*, 1–6. https://doi.org/10.1016/j.envsci.2019.07.002

Schwartz, M. W., Cook, C. N., Pressey, R. L., Pullin, A. S., Runge, M. C., Salafsky, N., Sutherland, W. J., & Williamson, M. A. (2018). Decision support frameworks and tools for conservation. *Conservation Letters*, *11*, e12385. https://doi.org/10.1111/conl.12385

Smith, R. J., Veríssimo, D., Leader-Williams, N., Cowling, R. M., & Knight, A. T. (2009). Let the locals lead. *Nature*, *462*, 280–281. https://doi.org/10.1038/462280a

Specht, H., Golding, J. D., Pero, E. M., Crane, M. T., Ortiz-Calo, W., McDevitt, M. C., Karlen, J. G., Branch, J. V., Hansen, C. P., & Millspaugh, J. J. (2022). A framework for evaluating the implications of assumptions in broad conservation strategies. *Conservation Science and Practice*, *4*, e12758. https://doi.org/10.1111/csp2.12758

Stephenson, P. J., & Stengel, C. (2020). An inventory of biodiversity data sources for conservation monitoring. *PLoS One*, *15*, e0242923. https://doi.org/10.1371/journal.pone.0242923

Sutherland, W. J., Cheng, S. H., Christie, A. P., Cooke, S. J., Koricheva, J., Randall, N., Smith, R. K., & Stewart, R. (2022). Assessing collated and synthesised evidence. In W. Sutherland (Ed.), *Transforming conservation: A practical guide to evidence and decision making*. Open Book Publishers. https://doi.org/10.11647/OBP.0321

Sutherland, W. J., Downey, H., Frick, W. F., Tinsley-Marshall, P., & McPherson, T. (2021). Planning practical evidence-based decision making within time constraints: The strategic evidence assessment framework. *Journal for Nature Conservation*, *60*, 125975. https://doi.org/10.1016/j.jnc.2021.125975

Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution*, *19*, 305–308. https://doi.org/10.1016/j.tree.2004.03.018

Sutherland, W. J., Shackelford, G., & Rose, D. C. (2017). Collaborating with communities: co-production or co-assessment? *Oryx*, *51*, 569–570. https://doi.org/10.1017/S0030605317001296

Sutherland, W. J., Taylor, N. G., MacFarlane, D., Amano, T., Christie, A. P., Dicks, L. V., Lemasson, A. J., Littlewood, N. A., Martin, P. A., Ockendon, N., Petrovan, S. O., Robertson, R. J., Rocha, R., Shackelford, G. E., Smith, R. K., Tyler, E. H. M., & Wordley, C. F. R. (2019). Building a tool to overcome barriers in research-implementation spaces: The conservation evidence database. *Biological Conservation*, *238*, 108199. https://doi.org/10.1016/j.biocon.2019.108199

Tallis, H., Kreis, K., Olander, L., Ringler, C., Ameyaw, D., Borsuk, M. E., Fletschner, D., Game, E., Gilligan, D. O., Jeuland, M., Kennedy, G., Masuda, Y. J., Mehta, S., Miller, N., Parker, M., Pollino, C., Rajaratnam, J., Wilkie, D., Zhang, W., … Zobrist, S. (2019). Aligning evidence generation and use across health, development, and environment. *Current Opinion in Environment Sustainability*, *39*, 81–93. https://doi.org/10.1016/j.cosust.2019.09.004

Taper, M. L., Lele, S. R., Ponciano, J. M., Dennis, B., & Jerde, C. L. (2021). Assessing the global and local uncertainty of scientific evidence in the presence of model misspecification. *Frontiers in Ecology and Evolution*, *9*, 679155. https://doi.org/10.3389/fevo.2021.679155

Tengö, M., Brondizio, E. S., Elmqvist, T., Malmer, P., & Spierenburg, M. (2014). Connecting diverse knowledge Systems for Enhanced Ecosystem Governance: The multiple evidence

base approach. *Ambio*, *43*, 579–591. https://doi.org/10.1007/s13280-014-0501-3

Tengö, M., Hill, R., Malmer, P., Raymond, C. M., Spierenburg, M., Danielsen, F., Elmqvist, T., & Folke, C. (2017). Weaving knowledge systems in IPBES, CBD and beyond—Lessons learned for sustainability. *Current Opinion in Environment Sustainability*, *26–27*, 17–25. https://doi.org/10.1016/j.cosust.2016.12.005

Walsh, J. C., Dicks, L. V., & Sutherland, W. J. (2015). The effect of scientific evidence on conservation practitioners' management decisions. *Conservation Biology*, *29*, 88–98. https://doi.org/10.1111/cobi.12370

Wheeler, H. C., & Root-Bernstein, M. (2020). Informing decision-making with indigenous and local knowledge and science. *Journal of Applied Ecology*, *57*, 1634–1643. https://doi.org/10.1111/1365-2664.13734

White, T. B., Amano, T., Boersch-Supan, P., Christie, A. P., Freckleton, R. C., Quinzin, M. C., Rezaie, A. M., Sutherland, W. J., & Yamashita, H. (2022). Gathering and assessing pieces of evidence. In W. Sutherland (Ed.), *Transforming conservation: A practical guide to evidence and decision making*. Open Book Publishers. https://doi.org/10.11647/OBP.0321

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Christie, A. P., Morgan, W. H., Salafsky, N., White, T. B., Irvine, R., Boenisch, N., Chiaravalloti, R. M., Kincaid, K., Rezaie, A. M., Yamashita, H., & Sutherland, W. J. (2023). Assessing diverse evidence to improve conservation decision-making. *Conservation Science and Practice*, e13024. https://doi.org/10.1111/csp2.13024