Check for updates

Poor hypotheses and research waste in biology: learning from a theory crisis in psychology

Shinichi Nakagawa[1,2,3]* , David W. Armitage[4] , Tom Froese[5] , Yefeng Yang[3] and Malgorzata Lagisz[2,3]

Abstract

While psychologists have extensively discussed the notion of a "theory crisis" arising from vague and incorrect hypotheses, there has been no debate about such a crisis in biology. However, biologists have long discussed communication failures between theoreticians and empiricists. We argue such failure is one aspect of a theory crisis because misapplied and misunderstood theories lead to poor hypotheses and research waste. We review its solutions and compare them with methodology-focused solutions proposed for replication crises. We conclude by discussing how promoting inclusion, diversity, equity, and accessibility (IDEA) in theoretical biology could contribute to ameliorating breakdowns in the theory-empirical cycle.

Keywords Questionable research practice, Replication crisis, Credibility revolution, Team science, SORTEE

Drivers of research waste: is there a theory crisis in biology?

"An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question." John Tukey

*Correspondence:
Shinichi Nakagawa
shinichi.nakagawa@ualberta.ca
[1] Department of Biological Sciences, University of Alberta, CW 405, Biological Sciences Building, Edmonton, AB T6G 2E9, Canada
[2] Theoretical Sciences Visiting Program (TSVP), Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Kunigami District, Okinawa 904-0412, Japan
[3] Evolution & Ecology Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia
[4] Integrative Community Ecology Unit, Okinawa Institute of Science and Technology Graduate University (OIST), 1919-1 Tancha, Onna, Okinawa 904-0495, Japan
[5] Embodied Cognitive Science Unit, Okinawa Institute of Science and Technology Graduate University (OIST), 1919-1 Tancha, Onna, Okinawa 904-0495, Japan

Many scientific fields have faced a *replication crisis* (refer to Glossary in Table 1) due to the low replicability of empirical studies. In biology, this is perhaps most well-known in cancer studies (see, for instance, the "Reproducibility Project: Cancer Biology"; https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology), although it arguably first gained significant attention in the social sciences, particularly in psychology [1–4]. In response, recent reforms have changed how psychologists conduct their research [5, 6] (also see [7]). One example is their adoption of *pre-registration* and *registered reports*, which can involve pre-commitment to a particular set of hypotheses and study design, often involving peer review [8–11]. For example, more than 130 psychology journals now accept registered reports in 2020 (compared to six in 2015) [12] (see also [13]). These innovations can curtail *questionable research practices* (QRPs) such as selective reporting and *HARKing* (hypothesizing after results are known), ubiquitous in all scientific fields, including biology [14–18]. In one example, psychological studies based on pre-registered reports supported the author's hypotheses only 40% of the time,

Table 1 Glossary

Replication crisis: A situation where many scientific studies cannot be replicated or reproduced, casting doubt on the reliability of published findings (see below for the definitions of the narrow-sense and broad-sense replication crisis).

Pre-registration: the practice of registering study aims, hypotheses, methods, and analysis plans before conducting research to increase transparency and reduce bias (usually without peer review).

Registered reports: publishing format where study proposals are peer-reviewed and accepted before data collection begins, emphasizing the importance of the research question and methodology.

Questionable research practices (QRPs): Research behaviors that compromise the integrity and validity of scientific findings, such as selectively reporting results or hypothesizing after results are known.

HARKing: Acronym for "Hypothesizing After Results are Known"; the practice of formulating hypotheses based on data after the fact and presenting them as if they were established before data collection.

Research waste: The inefficient use of resources in research due to issues like unpublished negative results, incomplete reporting, or flawed study design, leading to findings that are unusable or unreliable.

Incomplete reporting: Failure to fully disclose research methods, data, or results, hindering replication and the accumulation of scientific knowledge.

Poor methodological design: Inadequate or flawed research design, data collection, or analysis that undermines the validity and reliability of study findings.

Theory crisis: A situation where a scientific field lacks robust, testable theories or suffers from misinterpretation and misapplication of existing theories, leading to poor hypotheses and research waste.

Narrow-sense replication crisis (method crisis): The aspect of the replication crisis attributable specifically to methodological issues, such as poor study design and statistical practices.

Researcher degrees of freedom: The flexibility researchers have in making decisions during data collection, analysis, and reporting that can influence study outcomes (also referred to as the garden of forking paths).

Broad-sense replication crisis: An expanded view of the replication crisis that encompasses both methodological and epistemological issues (i.e., narrow-sense replication crisis and theory crisis), which acknowledges that difficulties in replicating studies arise not only from flawed methods but also from misapplied or misunderstood theories.

Theoretical model: A formal representation of a theory using mathematical, computational, or conceptual frameworks to make precise predictions although understanding such a model can be difficult.

Null hypothesis significance testing (NHST): A dominant statistical framework where researchers test whether there is enough evidence to reject a null hypothesis in favor of an alternative hypothesis.

P-hacking (P-value hacking): Manipulating data or analyses to achieve statistically significant results, often by trying multiple analyses until a desired p-value is obtained (*p*-hacking is also known as data dredging or fishing expedition).

Cherry-picking (selective reporting): The practice of selectively reporting only favorable results while ignoring or omitting others, leading to biased conclusions.

H-BUTing: Acronym for "Hypothesizing Before Understanding Theory"; formulating hypotheses without fully understanding the underlying theory, which can result in irrelevant or flawed research.

Hypothesis hacking (*h*-hacking): Deliberately manipulating or selectively interpreting theories to generate hypotheses that align with preconceived beliefs or desired outcomes.

Hypothesis cherry-picking: Intentionally ignoring alternative, theoretically valid hypotheses that do not support one's initial beliefs or preferred interpretations.

Strong inference: A scientific approach involving the development of multiple competing hypotheses and designing experiments to systematically eliminate them, strengthening the validity of conclusions.

Construct validity: The extent to which a test or experiment accurately measures the theoretical construct it intends to measure.

Internal validity: The degree to which a study establishes a trustworthy cause-and-effect relationship between variables, free from confounding factors.

External validity: The extent to which study findings can be generalized beyond the specific context of the research.

Non-confirmatory studies: Research conducted without specific hypotheses, aimed at exploring data to generate new hypotheses and insights (also known as exploratory or discovery-oriented studies).

Hackathon: An event where individuals collaborate intensively on projects or problems over a short period, fostering creativity and innovation across various fields, not limited to programming.

Meta-analyses: Statistical analyses that combine results from multiple studies to identify patterns, assess the robustness of findings, and provide a more precise estimate of effects.

Systematic reviews: Comprehensive reviews of literature on a specific topic using systematic and transparent methods to minimize bias and synthesize findings.

Systematic maps: Structured collections of research studies that aim to describe the state of knowledge on a topic, identifying gaps and clusters without necessarily synthesizing results.

Research weaving: An integrative approach that combines systematic mapping and bibliometrics to visualize and understand the structure and development of research fields.

IDEA: Acronym for Inclusion, Diversity, Equity, and Access; principles that promote fair and inclusive practices in academia and research.

Unconference: A participant-driven meeting where attendees create the agenda and lead discussions, fostering open exchange of ideas without predefined topics.

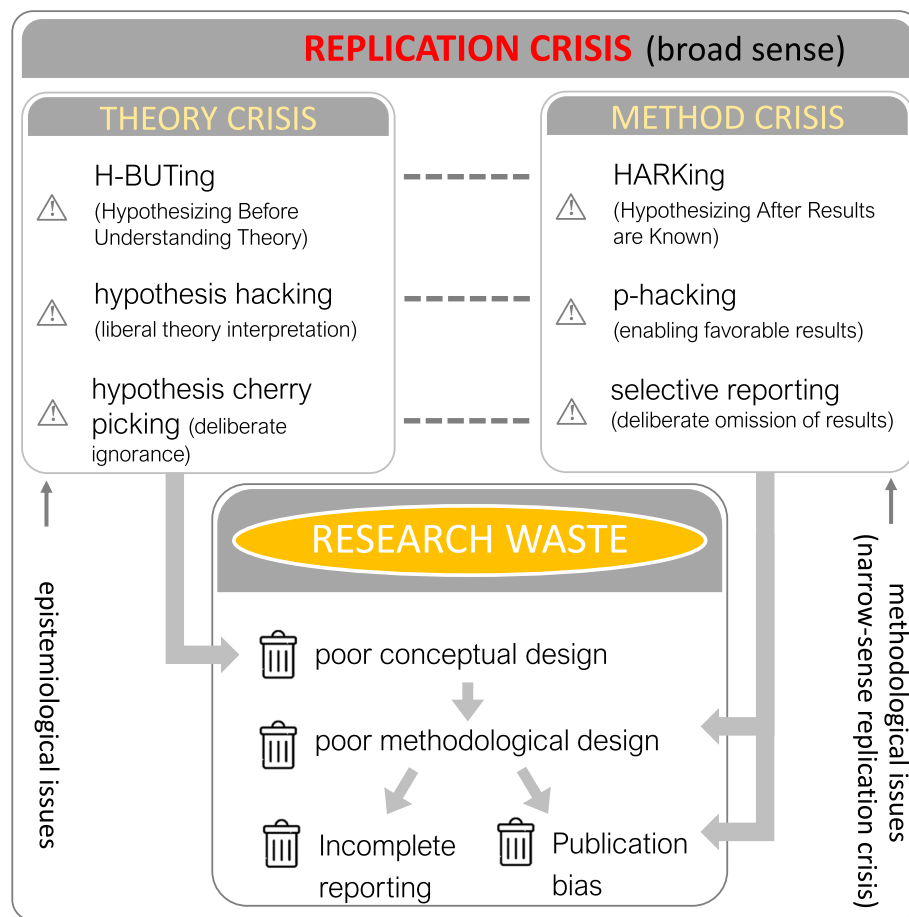Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 3 of 13

Fig. 1 Questionable research practices, QRPs, arising from a theory crisis (via epistemological issues) and method crisis (narrow-sense replication crisis; via methodological issues) and how they relate to research waste. QRPs concerning the theory crisis relate to poor conceptual design, which comes upstream of the other 3 items of research waste. Note that selective reporting and incomplete reporting may sound similar, but the former indicates deliberate selection of positive results while the latter represents the lack of culture in providing all the results and associated outcomes, including associated data and code

whereas traditional non-pre-registered studies supported the authors' hypotheses over 85% [19]. Similar reforms are being implemented in biology, particularly ecology and evolutionary biology, though progress is slower [20–22]. We concentrate on ecology and evolutionary biology for the reasons outlined below, although the issues and solutions are relevant to all branches of biology.

The replication crisis, driven by QRPs, is closely tied to *research waste*. According to Purgar and colleagues [23], research waste occurs when scientific resources are expended without yielding reliable, informative results to researchers and other end users. First, scientists complete projects but never publish or terminate them early, often due to negative results, leading to *publication bias*. Second, scientists publish their results with insufficient reporting of methods, resulting in *incomplete reporting*. Third, the research suffers from poor study design,

inadequate data collection, or flawed analysis, culminating in *poor methodological design*; even if published, results are unreliable and unusable. Alarmingly, Purgar and colleagues estimated that research waste could be as high as 82–89% in ecology [23], mirroring the estimate of 85% for medical sciences by Chalmers and Glasziou [24] (see also [25–29]). Since both the replication crisis and research waste seem to stem from methodological or procedural shortcomings, reforms have appropriately focused on rigorous and transparent methodology and reporting practices, including open data and code [30–33]. Therefore, it appears that a replication crisis should be more appropriately called a *method crisis* (we also refer to it as a *narrow-sense replication crisis*; see Fig. 1).

However, a series of articles in psychology have highlighted a more fundamental problem contributing to both research waste and replication difficulties—the

Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 4 of 13

emergence of a *theory crisis* [34–41]. This theory crisis arises when researchers frequently test vague and incorrect hypotheses because the underlying theories are often verbal (i.e., informal), allowing researchers to interpret them more freely, akin to high *researcher degrees of freedom* [42]. Even when written formally as mathematical models, a theory is often poorly described, leading to misinterpretation by empiricists and a mismatch between theoretical predictions and empirical hypotheses. This situation has been called an interpretation crisis [34] (cf., generalizability crisis [40]). As the opening quote from John Tukey suggests, testing a precise hypothesis originating from a misapplied or misunderstood theory is undoubtedly a form of research waste or, at least, research inefficiency.

Additional signals of a theory crisis include an overall lack of theories (and theoreticians) or the prevalence of poorly reasoned or vague theories. Yet, the lack of formalization and the prevalence of poor interpretations seem to be a substantial part of the discussion on the theory crisis in psychology (cf. [36–38]). Therefore, we use the term a theory crisis to include both (1) a lack of testable, formal theories within a field and (2) misinterpretations of sufficiently developed theories by empirical researchers. These theoretical shortcomings contribute to what we might call the *broad-sense replication crisis*, encompassing both methodological issues (i.e., a narrow-sense replication or method crisis) and fundamental epistemological challenges related to theory development and application (i.e., a theory crisis; Fig. 1).

To our knowledge, theory crises have not been explicitly discussed and linked to research waste in biological research (cf. [21]). Yet there has been much debate about communication gaps and failures between theoreticians and empiricists in biology, especially in ecology and evolution [43–51] (also see [52]; hereafter, we use a *theoretical model* or model as a mathematically formalized version of theory). Such debates indicate a theory crisis, per our definition, may have long existed and remain unresolved in biology. For example, in a 2022 paper [47], Servedio reported survey results from theoreticians in ecology and evolution on how their models are used in empirical work. Models were reported as misinterpreted and used incorrectly 19% of the time, while 36% of the time, they were cited in a non-specific manner. That is, the rest of the time (45%), empiricists cited models correctly, relating their work to specific theoretical predictions. Whether 45% seems low or high, there is much to improve for what both 91% of theoreticians and 80% of empiricists agreed upon—the importance of a tight feedback loop between theoretical and empirical work in biology [51] (see also [43]).

Given these statistics, it is worth asking whether biology currently suffers from a theory crisis and, if so, how it might be remedied. Our aims are four-fold. First, we consider communication breakdowns between theoreticians and empiricists through the lens of a theory crisis to illustrate how such breakdowns could contribute to both research waste and the broad-sense replication crisis. Next, we review proposed solutions for issues associated with the theory crisis, including theory-empirical communication failures from the social and biological sciences. We then highlight a critical gap in a scientific cycle — the lack of systematic mapping of theoretical models — which, when addressed, can facilitate communication between theoreticians and empiricists. Finally, we shift our focus to two issues: the low numbers of pure theoreticians in biology and their diversity with respect to identity, geography, and academic training. We describe two solutions to ameliorate these issues, which could help not only create diverse science teams between theoreticians and empiricists but also turn more people into theoreticians and liaisons between the two groups.

In this essay, our examples will primarily come from ecology and evolutionary biology because they are our primary fields. Also, ecology and evolutionary biology are fields rich in formal theory, making them suitable for illustrating our points, although the literature suggests theory-empiricism communication breakdowns in biology, including neuroscience, molecular biology, and biomedical sciences [53–56]. Yet, the issues we discuss are relevant to biology as a whole, and the proposed solutions can be adapted to other disciplines facing similar challenges.

Theory before method: questionable research practices and research waste in relation to theory

While the replication crisis (narrow-sense) is often attributed to methodological or reporting shortcomings in data collection and analysis, the theory crisis stems from misunderstanding, miscommunication, or misapplication of theory—an epistemic failure rather than a methodological one [37, 38]. In biology, as in psychology, researchers heavily rely on inferential statistics, especially *null hypothesis significance testing (NHST)* [57, 58]. In doing so, researchers often derive their hypotheses or predictions from existing theory, forming the alternative hypothesis within the NHST framework. Then, they statistically assess their hypothesis, using methods such as *t*-tests or generalized linear mixed-effects models to test and potentially reject the null hypothesis.

In psychology, several authors have recently pointed out that this statistical translation of theory is problematic because many theories and hypotheses are verbal, so interpretations of the theories can become unconstrained

Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 5 of 13

or subjective and, therefore, ungeneralizable [34, 36–40]. In biology, while similar issues can occur, many theories are based on precise mathematical logic or computer simulation, providing *directional* or *quantitative predictions* that are then tested in nature (e.g., [45]). However, mathematical formalization alone does not always lead to the development of stronger hypotheses. Even when theoretical predictions are clear, such predictions often depend on parameters that can vary widely and produce different or opposing predictions. These parameters may be difficult or impossible to measure accurately or precisely in empirical systems. As a result, empiricists may struggle to apply the theory appropriately, leading them to use approximations or make subjective choices to fit what they can test into the theoretical framework. This process can lead to misinterpretation or misapplication of formal theories [47] because empiricists might inadvertently oversimplify complex models or ignore critical parameter dependencies.

An example illustrating challenges arising from the oversimplification of complex theories comes from modern coexistence theory (MCT) [59, 60]. As pointed out by Terry and Armitage [61], MCT is less a predictive, testable theory but rather an analytical framework for partitioning growth rates into various coexistence-promoting mechanisms. Nonetheless, MCT has advanced community ecology by concentrating numerous system- and model-dependent theories of species coexistence under a small, shared set of mechanisms. However, this deceptively simple framework, where biological concepts such as "niche differences" are assigned numerical indices, is at risk of being superficially applied by empiricists, since the results from an MCT-based analysis are only as robust as the quality of one's collected data, choice of underlying model, and fitted parameter estimates. This has led to most hypotheses in MCT studies being quite nondirectional and qualitative, with predictions that are difficult to externally validate and generalise [61].

Importantly, theories that produce ambiguous predictions are not inherently poor; they often reflect the complexity of biological systems where multiple processes interact. Using such theories to guide empirical research is valuable but requires careful consideration of all plausible outcomes and understanding the conditions under which different predictions might hold. When researchers formulate hypotheses without fully engaging with the theory's complexities, they may test irrelevant or oversimplified hypotheses, leading to research waste. Such barriers are not unique to community ecology but occur in all fields where training in writing and interpreting formal theory is not a central theme in the scientific curriculum [49].

Turning to evolutionary biology, our following example concerns the "extrinsic mortality hypothesis," proposed in 1957 by G. C. Williams [62]. This nonformal, verbal theory suggested that high extrinsic mortality rates would select for increased growth rates, reproduction, and aging, in turn leading to shorter maximum lifespans (reviewed in [63]; see also [64–68]; note that this example is similar to cases in psychology when nonformal theories motivate empirical studies). However, subsequent formal analyses demonstrated that, depending on the underlying assumptions and trade-offs, extrinsic mortality could either increase or decrease the rate of aging or be unrelated altogether (summarized in [69]); therefore, previous studies were often misdirected because these studies might have studied systems where Williams' original verbal predictions did not apply. This example demonstrates the importance of the formal model. Also, it is critical to understand such formal models, and which set of assumptions (or parameter ranges) is most relevant to one's study system to make more accurate and relevant predictions.

Based on these examples, we suggest that a theory crisis can lead to three types of questionable research practices, QRPs (see Fig. 1). These three "new" extensions of QRPs are counterparts to well-known issues of empirical studies: HARKing (hypothesizing after results are known), *p*-hacking (or data dredging), and cherry-picking (of results to report).

The first is called *H-BUTing* (/hei-tʃ-but-ing/) or "Hypothesizing Before Understanding Theory." H-BUTing occurs when researchers formulate hypotheses without fully understanding the underlying theory, especially when the theory produces ambiguous or context-dependent predictions. H-BUTing may often happen because the theory is complex, involves multiple interacting mechanisms, or depends on parameters that are difficult to measure, as mentioned above. Also, H-BUTing can be equated to the previously identified issue of "premature hypothesis testing" [39]. It is premature because researchers either do not understand a theory or do not understand if their study systems are suitable for testing the theory.

We call the second QRP hypothesis hacking or *h*-hacking; *h*-hacking involves deliberately manipulating or selectively interpreting theories to generate hypotheses that align with preconceived beliefs or are more likely to yield significant results. Researchers might emphasize certain mechanisms within a theory while ignoring others or adjust theoretical assumptions to produce a desired prediction. Also, generating surprising (highly unrealistic) hypotheses can be a part of *h*-hacking because supporting a surprising hypothesis can lead to a high-profile journal publication (cf. [70]; see also [71]).

The third is "hypothesis cherry-picking" (termed by Krämer [34]), where researchers deliberately ignore alternative, theoretically, valid hypotheses that do not support their initial beliefs or the most favorable interpretation of their results. Such selective use of hypotheses theories goes against the concept of strong inference, where multiple plausible hypotheses are considered and tested in a study [72].

All these "new" QRPs are different from the more recognised QPRs, as H-BUTing, *h*-hacking and hypothesis cherry-picking occur before data collection while HARKing *p*-hacking and selective reporting (cherry-picking) occur afterwards. We note that H-BUTing and *h*-hacking are similar and indeed related, yet they are different, as the former is unintentional while the latter is deliberate. Regarding intentionality, *h*-hacking (deliberate misinterpretation) and hypothesis cherry-picking (deliberate negligence of alternatives) are alike.

In addition to these theory-related QRPs, we can add "poor conceptual design" alongside poor methodological design, publication bias, and incomplete reporting as the primary contributors to research waste. Epistemological issues (i.e., conceptualisation issues) lie upstream of methodological issues in the design and conduct of research (Fig. 1). Therefore, solving downstream issues, such as incomplete reporting, does not fix upstream issues, such as improper use of theory. Therefore, we believe that solving upstream poor conceptual design, a part of the theory crisis, is more critical than other downstream methodological issues pertaining to the (narrow-sense) replication crisis.

Notably, if poor conceptualization results in inappropriate experimental design or measurement, it can compromise a study's *construct validity* and *internal validity* (for difficulties in measuring theoretical constraints, see [73]). At the same time, poor conceptualization can also threaten *external validity* (generalizability), for example, through the biases in the selection of study systems, as a certain theory is more easily tested in particular organisms or locations [40] (for reflection on this issue, see "Western, Educated, Industrial, Rich, and Democracies", WEIRD [74] and "Social background, Trappability, Rearing history, Acclimation and habituation, Natural changes in responsiveness, Genetic make-up, and Experience", STRANGE [75]).

Proposed solutions: more development, education, and collaboration

Proposed solutions can be grouped into three initiatives. First, researchers can do more work to understand and operationalize theory before conducting a study with hypothesis testing (also referred to as a confirmatory study). Scheel and colleagues argue that premature hypothesis testing is rampant in psychology, but researchers should conduct more *non-confirmatory studies* (sometimes referred to as "exploratory" or "discovery-oriented" studies; [38], cf. [39]). Non-confirmatory studies can resolve issues of H-BUTing (hypothesizing before understanding theory) because they will identify whether and when a model is relevant to their biological system by understanding model assumptions and parameter space (e.g., a model assuming semelparous organisms without non-overlapping generations) [39].

For example, an ecologist might use a toy theoretical model to investigate how differences in the concentration of a shared resource and the uptake rates by competing species influence their coexistence, before designing experiments to test specific hypotheses. Such exploratory analyses can reveal conditions under which certain hypotheses concerning shifts in coexistence are valid, guiding more effective confirmatory studies and leading to more directional, quantitative hypotheses tailored to the specific system under study. A non-confirmatory exploration is often warranted before embarking on confirmatory work, especially at the start of postgraduate research programs. An alternative approach we find valuable is reproducing a theoretical study's results on our own (typically on a computer). Once this is accomplished, the models can be further modified and played with to more thoroughly understand their scope and assumptions.

However, we acknowledge that current academic incentives often favor confirmatory studies with clear, significant results [76]. Non-confirmatory or exploratory studies may be perceived as less valuable or publishable, which can discourage researchers from undertaking them. To encourage more non-confirmatory research, it is crucial to adjust incentive structures within academia and publishing. Journals and funding agencies can play a role by recognizing the importance of exploratory studies in advancing theoretical understanding and promoting policies that value such contributions more. Also, from our experience, publishing such a non-confirmatory study is often an important precursor, as discussed above, to publishing an important confirmatory work, so this process itself can act as an incentive for conducting and publishing non-confirmatory studies.

Second, universities can offer undergraduate and postgraduate courses on how to understand theoretical and mathematical/computational models and, more broadly, the role of theory in the scientific process, thereby promoting the training of pure theoreticians and theoretically proficient empiricists (i.e., liaisons). Some psychologists have shared their experiences on the effectiveness of such courses [35, 41] (see also [43]). For example, Borsboom, and van der Maas teach a 'theory

Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 7 of 13

construction' course where, in addition to lectures on the methodology, students create computer simulation models that can explain a phenomenon well but also see how a model is limited as it creates non-sensical values under some settings [35]. We also think such a hands-on course would be eye-opening and useful training.

In the biological sciences, however, theoreticians are almost always a minority group who teach courses that a minority of students attend, and often at levels more advanced than the average biosciences student is proficient. Thus, although education is essential and effective, it cannot break the status quo of early separation between theoreticians (who often arrive at biology with math or physics degrees) and empiricists (who tend to avoid most math and physics classes beyond prerequisites). This is also because understanding math takes time, and interventions should start much earlier than university education. Furthermore, we have noticed a concerning trend in ecology and evolution, where advertisements for pure theoreticians seem to be declining. In contrast, appointments in "quantitative biology" are on the rise. We speculate that this may reflect a growing trend of interest in the more lucrative fields of data science and bioinformatics, which compete with the theoretical sciences over a shared pool of students. Nevertheless, online education and technologies can overcome the shortage of theoreticians (more discussed later).

Third, theoreticians and empiricists should collaborate more. Recently, Ou and colleagues have described how theoreticians can write more accessible papers by being mindful of what parts of their analyses are the most difficult for empiricists to understand [44]. Likewise, Grainger and colleagues have produced an excellent guide for empiricists with tips on understanding and testing formal theory [43]. Notably, these two papers largely overlap in their author lists, which comprise a mixture of empiricists, theoreticians, and liaisons in ecology and evolutionary biology. Though the primary means by which theoreticians and empiricists interact is indirectly through citation of one another's publications, direct collaboration might be a more effective means for reducing barriers in understanding, though the relative scarcity of theoreticians in most biological fields can limit such direct interactions.

Interestingly, a group of psychologists have successfully attempted an exercise called the "Many Modelers Hackathon" where a few theoreticians worked with many empiricists to formalize verbal theories during a 3-h workshop [77]. This Many Modelers event was successfully run as part of a 2021 conference for the Society for the Improvement of Psychological Science (SIPS)—a model for the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology (SORTEE). A Many-Modelers Hackathon holds tremendous potential in biology, which has a long history of formalizing theories and designing complex experiments (more on the potential of this type of *hackathon* later).

A gap in the scientific cycle and a proposed solution: systematic mapping of theoretical models

In Fig. 2, we illustrate a scientific cycle where researchers seek to understand a phenomenon and develop a theory via two pathways: empirical research and theoretical development. The last three decades have seen a revolution in empirical research with a rise in research synthesis, especially meta-analysis [78, 79]. *Meta-analyses* (often a part of *systematic reviews*) have gradually been superseding traditional narrative reviews because meta-analyses can bring about unique benefits that are not possible with narrative reviews [80]. For example, meta-analyses can quantitively identify general patterns and knowledge gaps, explain inconsistencies among empirical studies, and generate ideas that can fledge into theories [81].

Unlike meta-analyses, the synthesis of theory is primarily narrative rather than systematic, representing a gap in the scientific cycle (Fig. 2) [82] (cf. [83]). While conducting a quantitative synthesis of theoretical literature may be challenging (cf. [84]), a systematic qualitative synthesis is still possible. Such syntheses might encompass a systematic review or a systematic map (sometimes, known as a scoping review; see [85, 86]). Although often mixed up, systematic reviews and *systematic maps* have different objectives [87, 88]. The former answers a specific question, often relating to "What works?" (e.g., does an intervention have an effect?), while the latter addresses a more general question, such as "What has already been studied and where do gaps remain?" [89]. Therefore, a systematic map is a structured collection of studies, making it suitable for qualitatively summarizing models. Systematic maps can reveal general patterns and knowledge gaps in related models. Additionally, mapping models can identify gaps in empirical observations by examining how parameter values in theoretical models are defined, whether based on empirical results or assumptions requiring further validation. Although their number is limited and often (mis)labeled as a systematic review, systematic maps of models exist in various fields of medicine and biology.

One noteworthy systematic map has been conducted for the field of life history theory by Mathot and Frankenhuis [90]. Many empiricists have tested hypotheses concerning covariation between life history traits within species such as growth rate and age of reproduction with behavioral and physiological traits, theorized
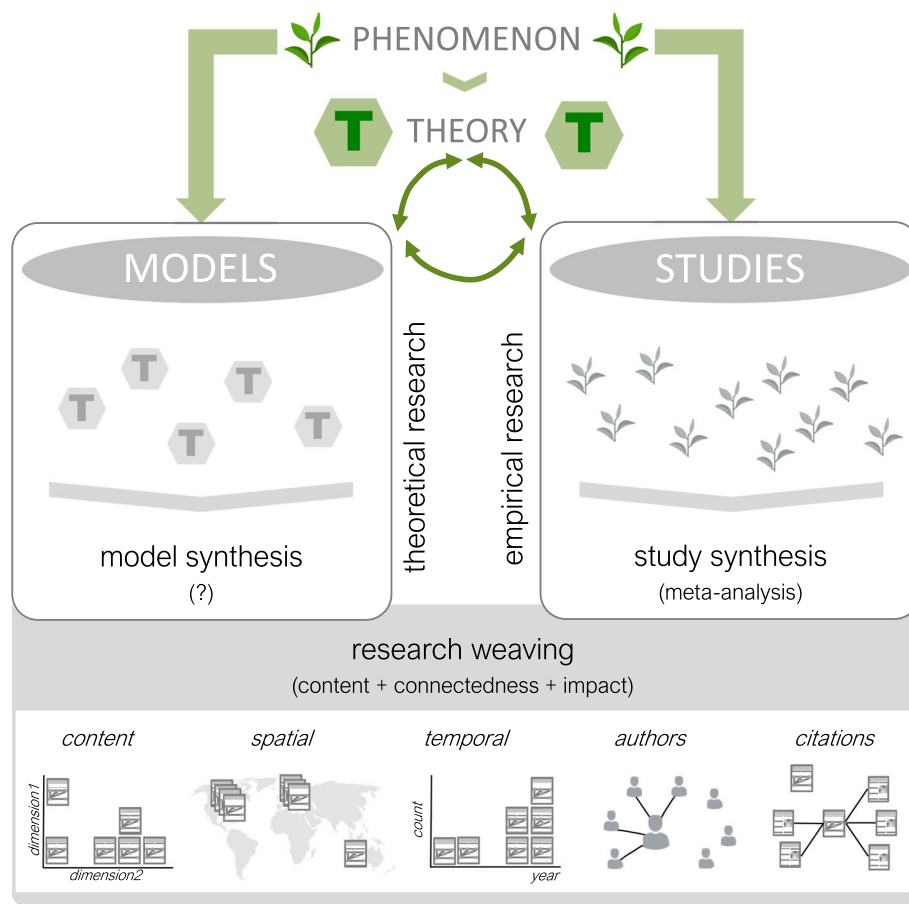
Fig. 2 A depiction of a scientific cycle. Researchers seek to understand a phenomenon and develop a theory while engaging with theoretical research or empirical research. While meta-analysis has revolutionized empirical synthesis, the synthesis of theories (models) is primarily narrative. Research weaving (systematic mapping and bibliometrics) could help not only synthesize theoretical models but also summarize both types of research on a topic

to arise from trade-offs between investment in current and future reproduction, known as "pace-of-life syndromes" [91]. Here, the authors attempted to map theoretical models on this topic, but they only identified two unique models [90]. Importantly, these models provided mechanisms by which covariances could arise between life history and other traits without invoking the current-future reproduction trade-offs. This finding suggested that many empirical studies may have tested hypotheses lacking comprehensive theoretical support, potentially committing H-BUTing and hypothesis cherry-picking (Fig. 1). Here, the systematic mapping revealed a lack of formal theoretical models, and identified ways in which future theoretical development will best allow for their empirical evaluation given their particular assumptions.

This example might give the impression that such mapping activities are futile due to the paucity of theoretical papers in some research areas in biology. However, certain topics enjoy an abundance of theory. A systematic map of theory relating to antimicrobial resistance, for example, identified 273 studies of mathematical models made at the population level [92] (see also [93]); this map elucidated theoretical gaps, such as the scarcity of models on transmission between humans and animals and the consideration of environmental factors. Similarly, another systematic map collated 698 studies of agent-based models of infectious disease transmission published from 2006 to 2015 [94].

Recently, Achter and colleagues have published an opinion article promoting systematic maps (and reviews) of agent-based models so that new models do not "re-invent the wheel" but rather build upon previous models in environmental sciences [82]. They argue that such a systematic map can lead to further refinement and the development of a theory. We concur and join their call for systematic maps of mathematical and computational models. Importantly, systematically mapping models could guard against

hypothesis cherry-picking and *h*-hacking (Fig. 1). This is because a map gives a comprehensive catalog of current theoretical models, which can direct the development of a set of testable hypotheses, hopefully reducing the incidence of cherry-picking hypotheses or liberal interpretations of theories.

Recently, bibliometric analyses in the forms of quantifying articles' citation impacts and connections through citations or collaborations have been conducted for models of antibiotic resistance [95] in addition to the systematic map mentioned above [92]. Some of us (SN and ML) have recently proposed a new way of synthesizing literature, named *research weaving*, which combines systematic mapping and bibliometrics [89]. As outlined above, systematic mapping summarizes the current state of knowledge and evidence, identifying areas with research gaps or an abundance of papers. As a complementary approach, bibliometrics enables researchers to see how pieces of evidence are connected. Such analysis can reveal the structure and development of a field and how influential particular works have been (research weaving examples, see [96, 97]). Therefore, a research weaving of theoretical models holds much potential, for example, for collaboration work between theoreticians and empiricists because both can contribute to such an activity (cf. [82]) and can lead to the synthesis of both theoretical and empirical results (Fig. 2).

Furthermore, some bibliometric analyses (i.e., geographic and collaboration-network analyses) can reveal the lack of inclusion, diversity, equity, and access (IDEA; also referred to as EDI or DEI) in a field of study by examining geographic distribution and collaboration networks. Identifying underrepresented regions or groups can help target efforts to promote diversity and inclusivity in theoretical biology. Enhancing IDEA can help ameliorate the theory crisis by bringing diverse perspectives and expertise, which we outline in the next section.

Ideas to create a tight theory-empirical feedback loop by achieving IDEA: inclusion, diversity, equity, and access

Several articles have pointed out the shortfall of IDEA in the theoretical biology community [43, 98, 99]. Evidence suggests theoreticians are an impactful yet privileged group lacking diversity (e.g., [100]). This lack of diversity may partly stem from the smaller size of the theoretical community or its concentration primarily at top institutes in wealthy nations (c.f., [101]). Increasing the pool of theoreticians with IDEA goals in mind could ameliorate both the theory crisis and the lack of diversity by bringing in new perspectives and fostering innovation in theoretical development. In recent years, many academic societies in biology have initiated efforts to promote IDEA by encouraging historically underrepresented and marginalized groups to join and take on leadership roles, achieving noticeable success (cf. [102]).

Theoretical work is impactful and provides mathematical and computational skills that are highly transferable. It also offers flexibility, as theory development can often be conducted with minimal resources—a pen and paper or a laptop—making it accessible to researchers from institutions with limited funding for large lab or field-based programs. By promoting theoretical training among diverse groups, we can help break down academic inequities and empower researchers globally. Additionally, addressing the scarcity of theoreticians is urgent because researchers now regularly collect and analyze massive, complex datasets with only a small pool of theories to generate a priori hypotheses (e.g., [103]). The overabundance of data highlights the importance of theory and theoreticians in advancing scientific understanding.

To achieve these goals of growing and diversifying the community of theoretical biologists, we propose two strategies. First, academic societies can create free, accessible training resources on understanding and building theoretical models. For example, online video series can democratize knowledge, similar to Richard McElreath's "Statistical Rethinking" lectures, which make Bayesian statistics and causal inference accessible to biologists [104]. Additionally, initiatives like the International Initiative for Theoretical Ecology's online seminar series have successfully attracted a global audience (https://iite.info/seminar/). These resources can help cultivate theoretical skills among researchers from diverse backgrounds and regions.

Second, societies can incorporate collaborative activities like hackathons into their conferences, fostering direct interaction between theoreticians and empiricists. Events such as Many-Modelers hackathons [77] demonstrate how such collaborations can formalize verbal theories and generate testable hypotheses. It is important to note that programming skills are not a prerequisite for participating in a hackathon; the term now widely refers to any event structured around team-based creative problem-solving. For example, assembling ecology and evolution journal editors to develop shared guidelines for accepting registered reports is a form of a hackathon that leads to impactful outcomes [105–107]. Along with hackathons, conferences could include unstructured discussion events where participants are prompted to discuss ideas without a pre-specified task (called *unconference*). These activities promote inclusivity by bringing together researchers from diverse backgrounds and career stages, fostering new collaborations and enhancing theoretical understanding. One example of this type of unstructured conference in ecology is the annual "ANdiNA" meeting
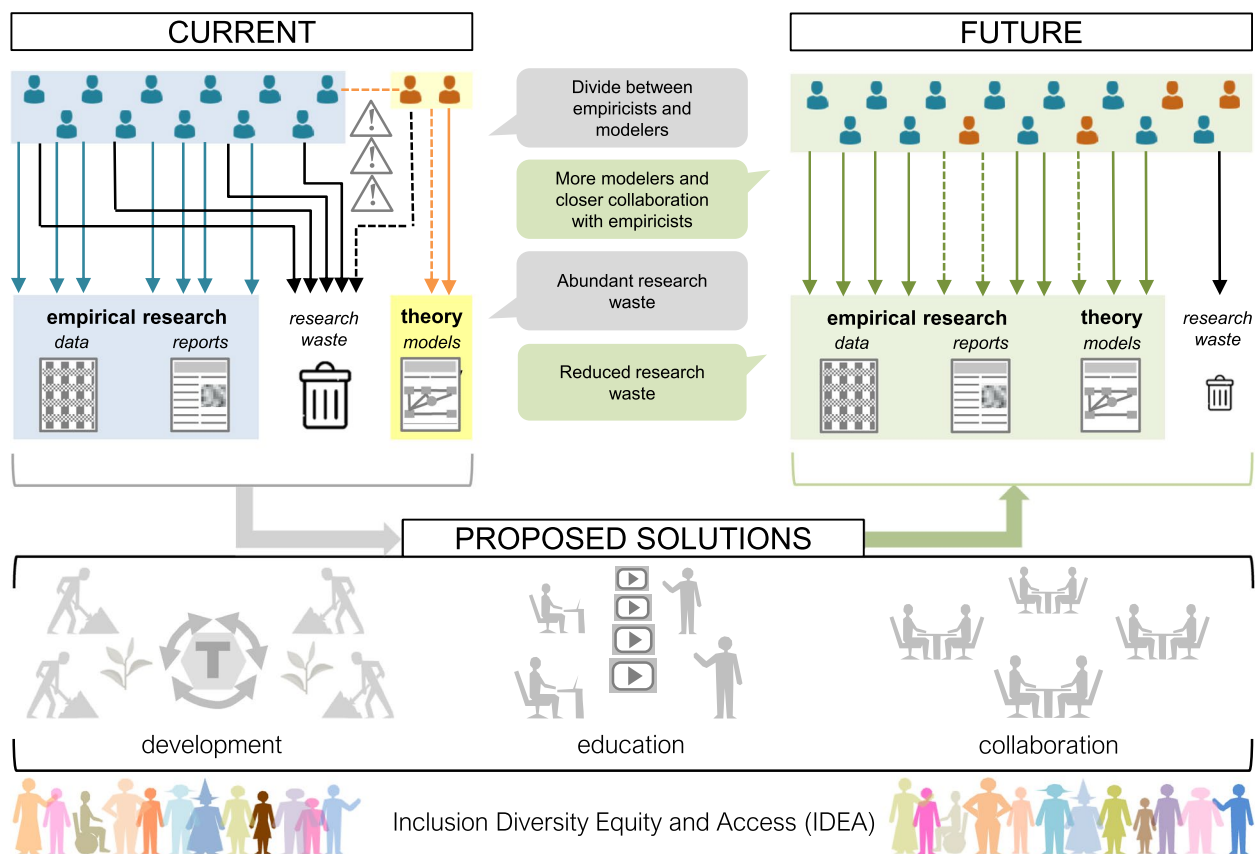
Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 10 of 13

Fig. 3 The current and future of empirical and theoretical research. Currently, due to miscommunications between theoreticians (minority) and empiricists (majority), resulting in research waste. Via the proposed solutions, development, education, and collaboration, the future research community will have more theoreticians working with empiricists, especially if learned societies can embrace the theory crisis and promote the integration of theoretical and empirical work through IDEA. Solid lines represent no direct collaborations, while dotted lines indicate direct collaborations (the upper panels). While research waste may be an unavoidable part of the scientific process, more research efficacy can be attainable

held to "align global priorities and local action in ecological research and practice."

We (SN and ML) have organized several hackathons and unconferences at SORTEE annual online meetings; one of the activities has already resulted in a publication [108] (other examples of SORTEE conference publications [109–111]). These events have facilitated team science by encouraging interdisciplinary work and appropriately acknowledging individual contributions, such as by developing the "Methods Reporting with Initials for Transparency" (MeRIT) system [108]. By embracing IDEA principles in these ways, we can create a more robust and diverse theoretical community. This process, in turn, strengthens the theory-empirical feedback loop, promoting the development of well-founded theories and enhancing the integration of theory and empirical research across biology.

Less waste and more IDEA for the future of biology

Here, we have described how biology may be prone to a theory crisis, which contributes to a replication crisis (broad-sense) and research waste. We have also discussed existing and potential solutions—such as the systematic mapping of theoretical models—to help resolve the theory crisis (Fig. 3). Despite past attempts to fix theory-empirical communication breakdowns, dramatic success has yet to be achieved in creating a tight feedback loop between theory and experimentation (cf. [43]). We contend that pursuing IDEA can, at least partially, ameliorate such breakdowns and presented two concrete ideas—engaging educational videos and Many-Modelers-like-Hackathon events at conferences—that academic societies are encouraged to act upon.

We encourage you and your scientific community to generate more ideas and mobilize these ideas to improve

Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 11 of 13

both theory-empirical communication and IDEA. By fostering a diverse team science approach, we can tackle both big questions in biology and pressing ecological and environmental issues that humanity faces. We believe that embracing these strategies will lead to a future where research waste is minimized, and the integration of theory and empiricism is strengthened, ultimately advancing the field of biology.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests except that SN, YY, and ML are the members of SORTEE.

References
1. Aarts AA, Anderson JE, Anderson CJ, Attridge PR, Attwood A, Axt J, Babel M, Bahník S, Baranski E, Barnett-Cowan M, et al. Estimating the reproducibility of psychological science. Science. 2015;349(6251):aac4716.
2. Camerer CF, Dreber A, Forsell E, Ho TH, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan TZ, et al. Evaluating replicability of laboratory experiments in economics. Science. 2016;351(6280):1433–6.
3. Camerer CF, Dreber A, Holzmeister F, Ho TH, Huber J, Johannesson M, Kirchler M, Nave G, Nosek BA, Pfeiffer T, et al. Evaluating the replicability of social science experiments in and between 2010 and 2015. Nat Hum Behav. 2018;2(9):637–44.
4. Kvarven A, Stromland E, Johannesson M. Comparing meta-analyses and preregistered multiple-laboratory replication projects. Nat Hum Behav. 2020;4(4):423–34.
5. Vazire S. Implications of the Credibility Revolution for Productivity, Creativity, and Progress. Perspect Psychol Sci. 2018;13(4):411–7.
6. Angrist JD, Pischke JS. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J Econ Perspect. 2010;24(2):3–30.
7. Hardwicke TE, Thibault RT, Kosie JE, Wallach JD, Kidwell MC, Ioannidis JPA. Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). Perspect Psychol Sci. 2022;17(1):239–51.
8. Hardwicke TE, Wagenmakers EJ. Reducing bias, increasing transparency and calibrating confidence with preregistration. Nat Hum Behav. 2023;7(1):15–26.
9. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. P Natl Acad Sci USA. 2018;115(11):2600–6.
10. Chambers CD, Tzavella L. The past, present and future of Registered Reports. Nat Hum Behav. 2022;6(1):29–42.
11. Henderson EL, Chambers CD. Ten simple rules for writing a registered report. Plos Comput Biol. 2022;18(10):e1010571.
12. Montoya AK, Krenzer WLD, Fossum JL. Opening the door to registered reports: census of journals publishing registered reports (2013–2020). Collabra-Psychol. 2021;7(1):24404.
13. Ferguson J, Littman R, Christensen G, Paluck EL, Swanson N, Wang ZA, Miguel E, Birke D, Pezzuto JH. Survey of open science practices and attitudes in the social sciences. Nature Communications. 2023;14(1):5401.
14. Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. Questionable research practices in ecology and evolution. Plos One. 2018;13(7):e0200303.
15. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23(5):524–32.
16. Agnoli F, Wicherts JM, Veldkamp CLS, Albiero P, Cubelli R. Questionable research practices among Italian research psychologists. Plos One. 2017;12(3):e0172792.
17. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of P-hacking in science. Plos Biol. 2015;13(3):e1002106.
18. Fanelli D. "Positive" results increase down the hierarchy of the sciences. Plos One. 2010;5(3):e10068.
19. Allen C, Mehler DMA. Open science challenges, benefits and tips in early career and beyond. Plos Biol. 2019;17(5):e3000587.
20. Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, Iorns E, Nosek BA. Investigating the replicability of preclinical cancer biology. Elife. 2021;10:e71601.
21. O'Dea RE, Parker TH, Chee YE, Culina A, Drobniak SM, Duncan DH, Fidler F, Gould E, Ihle M, Kelly CD, et al. Towards open, reliable, and transparent ecology and evolutionary biology. Bmc Biol. 2021;19(1):68.
22. Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. Bmc Biol. 2015;13:88.
23. Purgar M, Klanjscek T, Culina A. Quantifying research waste in ecology. Nat Ecol Evol. 2022;6(9):1390–7.
24. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374(9683):86–9.
25. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, Howells DW, Ioannidis JPA, Oliver S. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383(9912):156–65.
26. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383(9912):166–75.
27. Salman RA, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, Chalmers I. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383(9912):176–85.
28. Chan AW, Song FJ, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, Krumholz HM, Ghersi D, van der Worp HB. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383(9913):257–66.
29. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267–76.
30. Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, Chee YE, Kelly CD, Gurevitch J, Nakagawa S. Transparency in ecology and evolution: real problems. Real Solutions Trends Ecol Evol. 2016;31(9):711–9.
31. Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public data archiving in ecology and evolution: how well are we doing? Plos Biol. 2015;13(11):e1002295.
32. Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, Kokko H, Jennions MD, Kruuk LEB. Troubleshooting public data archiving: suggestions to increase participation. Plos Biol. 2014;12(1):e1001779.
33. Culina A, van den Berg I, Evans S, Sánchez-Tójar A. Low availability of code in ecology: a call for urgent action. Plos Biol. 2020;18(7):e3000763.

34. Krämer B: Why Are Most Published Research Findings Under-Theorized? In: Questions of Communicative Change and Continuity: 2022. Nomos Verlagsgesellschaft mbH & Co. KG: 23–52.

35. Borsboom D, van der Maas HLJ, Dalege J, Kievit RA, Haig BD. Theory construction methodology: a practical framework for building theories in psychology. Perspect Psychol Sci. 2021;16(4):756–66.

36. Eronen MI, Bringmann LF. The theory crisis in psychology: how to move forward. Perspect Psychol Sci. 2021;16(4):779–88.

37. Muthukrishna M, Henrich J. A problem in theory. Nat Hum Behav. 2019;3(3):221–9.

38. Oberauer K, Lewandowsky S. Addressing the theory crisis in psychology. Psychon B Rev. 2019;26(5):1596–618.

39. Scheel AM, Tiokhin L, Isager PM, Lakens D. Why hypothesis testers should spend less time testing hypotheses. Perspect Psychol Sci. 2021;16(4):744–55.

40. Yarkoni T. The generalizability crisis. Behav Brain Sci. 2022;45:e1.

41. van Rooij I, Baggio G. Theory before the test: how to build high-verisimilitude explanatory theories in psychological science. Perspect Psychol Sci. 2021;16(4):682–97.

42. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22(11):1359–66.

43. Grainger TN, Senthilnathan A, Ke PJ, Barbour MA, Jones NT, DeLong JP, Otto SP, O'Connor MI, Coblentz KE, Goel N, et al. An empiricist's guide to using ecological theory. Am Nat. 2021;199(1):1–20.

44. Ou WJA, Henriques GJB, Senthilnathan A, Ke PJ, Grainger TN, Germain RM. writing accessible theory in ecology and evolution: insights from cognitive load theory. Bioscience. 2022;72(3):300–13.

45. Servedio MR, Brandvain Y, Dhole S, Fitzpatrick CL, Goldberg EE, Stern CA, Van Cleve J, Yeh DJ. Not just a theory-the utility of mathematical models in evolutionary biology. Plos Biol. 2014;12(12):e1002017.

46. Rossberg AG, Barabás G, Possingham HP, Pascual M, Marquet PA, Hui C, Evans MR, Meszéna G. Let's train more theoretical ecologists - here is why. Trends Ecol Evol. 2019;34(9):759–62.

47. Servedio MR. An effective mutualism? The role of theoretical studies in ecology and evolution. Am Nat. 2020;195(2):284–9.

48. Otto SP, Rosales A. Theory in service of narratives in evolution and ecology. Am Nat. 2020;195(2):290–9.

49. Fawcett TW, Higginson AD. Heavy use of equations impedes communication among biologists. P Natl Acad Sci USA. 2012;109(29):11735–9.

50. Shou WY, Bergstrom CT, Chakraborty AK, Skinner FK. Theory, models and biology. Elife. 2015;4:e07158.

51. Haller BC. Theoretical and empirical perspectives in ecology and evolution: a survey. Bioscience. 2014;64(10):907–16.

52. Fitzpatrick CL, Hobson EA, Mendelson TC, Rodríguez RL, Safran RJ, Scordato ESC, Servedio MR, Stern CA, Symes LB, Kopp M. Theory meets empiry: a citation network analysis. Bioscience. 2018;68(10):805–12.

53. Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D. Neuroscience needs behavior: correcting a reductionist bias. Neuron. 2017;93(3):480–90.

54. Ioannidis JPA. Why most published research findings are false. Plos Med. 2005;2(8):696–701.

55. O'Malley MA, Soyer OS. The roles of integration in molecular systems biology. Stud Hist Phi Part C. 2012;43(1):58–68.

56. Collins FS. Reengineering translational science: the time is right. Sci Transl Med. 2011;3(90):90cm17.

57. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev. 2007;82(4):591–605.

58. Nakagawa S, Lagisz M, Yang Y, Drobniak SM. Finding the right power balance: Better study design and collaboration can reduce dependence on statistical power. Plos Biol. 2024;22(1):e3002423.

59. Chesson P. Multispecies competition in variable environments. Theor Popul Biol. 1994;45(3):227–76.

60. Chesson P. Mechanisms of maintenance of species diversity. Annu Rev Ecol Syst. 2000;31:343–66.

61. Terry JCD, Armitage DW. Widespread analytical pitfalls in empirical coexistence studies and a checklist for improving their statistical robustness. Methods Ecol Evol. 2024;15(4):594–611.

62. Williams GC. Pleiotropy, natural-selection, and the evolution of senescence. Evolution. 1957;11(4):398–411.

63. Moorad J, Promislow D, Silvertown J. Evolutionary ecology of senescence and a reassessment of williams'"extrinsic mortality" hypothesis. Trends Ecol Evol. 2019;34(6):519–30.

64. Day T, Abrams PA. Density dependence, senescence, and Williams' hypothesis. Trends Ecol Evol. 2020;35(4):300–2.

65. Kozlowski J, Burger O, Danko MJ. Williams' prediction will often be observed in nature. Trends Ecol Evol. 2020;35(4):302–3.

66. Moorad J, Promislow D, Silvertown J: George C. Williams' Problematic Model of Selection and Senescence: Time to Move on. Trends Ecol Evol 2020, 35(4):303-305.

67. da Silva J. Williams' intuition about extrinsic mortality was correct. Trends Ecol Evol. 2020;35(5):378–9.

68. Moorad J, Promislow D, Silvertown J. Williams' intuition about extrinsic mortality is irrelevant. Trends Ecol Evol. 2020;35(5):379–379.

69. de Vries C, Galipaud M, Kokko H: Extrinsic mortality and senescence: a guide for the perplexed. Peer Community Journal 2023, 3.

70. Rubin M. When does harking hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. Rev Gen Psychol. 2017;21(4):308–20.

71. Csete ME, Doyle JC. Reverse engineering biological complexity. Science. 2002;295(5560):1664–9.

72. Platt JR. Strong inference - certain systematic methods of scientific thinking may produce much more rapid progress than others. Science. 1964;146(364):347-+.

73. Flake JK, Fried E. Measurement schmeasurement: questionable measurement practices and how to avoid them. Adv Meth Pract Psych. 2020;3(4):456–65.

74. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci. 2010;33(2–3):61-+.

75. Webster MM, Rutz C. How STRANGE are your study animals? Nature. 2020;582(7812):337–40.

76. Smaldino PE, Mcelreath R. The natural selection of bad science (vol 10, 221306, 2016). Roy Soc Open Sci. 2023;10(9):160384.

77. van Dongen N, Finnemann A, de Ron J, Tiokhin L, Wang SB, Algermissen J, Altmann EC, Bahnik S, Chuang L-C, Dumbrava A et al: Practicing theory building in a many modelers hackathon: A proof of concept. PsyArXiv 2022, https://doi.org/10.31234/osf.io/r5yfz.

78. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature. 2018;555(7695):175–82.

79. Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. Bmc Biol. 2017;15:18.

80. Nakagawa S, Poulin R. Meta-analytic insights into evolutionary ecology: an introduction and synthesis. Evol Ecol. 2012;26(5):1085–99.

81. Cooper H, Hedges LV, Valentine JC, Cooper HM, Hedges LV, Valentine JC, Russell Sage F: The handbook of research synthesis and meta-analysis, 3rd edition. edn. New York: Russell Sage Foundation; 2019.

82. Achter S, Borit M, Cottineau C, Polhill JG, Radchuk V, Meyer M. How to conduct more systematic reviews of agent-based models and foster theory development - Taking stock and looking ahead. Environ Model Software. 2023;173:105867.

83. Heger T, Aguilar-Trigueros CA, Bartram I, Braga RR, Dietl GP, Enders M, Gibson DJ, Gómez-Aparicio L, Gras P, Jax K, et al. The hierarchy-of-hypotheses approach: a synthesis method for enhancing theory development in ecology and evolution. Bioscience. 2021;71(4):337–49.

84. Rossberg AG: Food webs and biodiversity : foundations, models, data. Chichester, West Sussex, UK: Wiley Blackwell; 2013.

85. Campbell F, Tricco AC, Munn Z, Pollock D, Saran A, Sutton A, White H, Khalil H. Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family. Syst Rev-London. 2023;12(1):61.

86. Khalil H, Tricco AC. Differentiating between mapping reviews and scoping reviews in the evidence synthesis ecosystem. J Clin Epidemiol. 2022;149:175–82.

87. Haddaway NR, Callaghan MW, Collins AM, Lamb WF, Minx JC, Thomas J, John D. On the use of computer-assistance to facilitate systematic mapping. Campbell Syst Rev. 2020;16(4):e1129.

88. Haddaway NR, Bernes C, Jonsson BG, Hedlund K. The benefits of systematic mapping to evidence-based environmental management. Ambio. 2016;45(5):613–20.

Nakagawa *et al. BMC Biology*   (2025) 23:33

Page 13 of 13

89. Nakagawa S, Samarasinghe G, Haddaway NR, Westgate MJ, O'Dea RE, Noble DWA, Lagisz M. Research weaving: visualizing the future of research synthesis. Trends Ecol Evol. 2019;34(3):224–38.

90. Mathot KJ, Frankenhuis WE. Models of pace-of-life syndromes (POLS): a systematic review. Behav Ecol Sociobiol. 2018;72(3):1–2.

91. Réale D, Garant D, Humphries MM, Bergeron P, Careau V, Montiglio PO. Personality and the emergence of the pace-of-life syndrome concept at the population level. Philos T R Soc B. 2010;365(1560):4051–63.

92. Niewiadomska AM, Jayabalasingham B, Seidman JC, Willem L, Grenfell B, Spiro D, Viboud C. Population-level mathematical modeling of anti-microbial resistance: a systematic review. Bmc Med. 2019;17:1–20.

93. Birkegård AC, Halasa T, Toft N, Folkesson A, Græsboll K. Send more data: a systematic review of mathematical models of antimicrobial resistance. Antimicrob Resist In. 2018;7:117.

94. Willem L, Verelst F, Bilcke J, Hens N, Beutels P. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006–2015). Bmc Infect Dis. 2017;17:612.

95. Temime L, Hejblum G, Setbon M, Valleron AJ. The rising impact of mathematical modelling in epidemiology: antibiotic resistance research as a case study. Epidemiol Infect. 2008;136(3):289–98.

96. Macartney EL, Drobniak SM, Nakagawa S, Lagisz M. Evidence base for non-genetic inheritance of environmental exposures in non-human animals and plants: a map of evidence syntheses with bibliometric analysis. Environ Evid. 2023;12(1):1.

97. Burke S, Pottier P, Macartney EL, Drobniak SM, Lagisz M, Ainsworth T, Nakagawa S. Mapping literature reviews on coral health: protocol for a review map, critical appraisal and bibliometric analysis. Ecol Solut Evid. 2022;3(4):e12190.

98. Shaw AK. Diverse perspectives from diverse scholars are vital for theoretical biology. Theor Ecol-Neth. 2022;15(2):143–6.

99. Lee SI, Clinedinst L. Mathematical biology: expand, expose, and educate! B Math Biol. 2020;82(9):120.

100. Severson AL, Uricchio LH, Arbisser IM, Glassberg EC, Rosenberg NA. Analysis of author gender in TPB, 1991–2018. Theor Popul Biol. 2019;127:1–6.

101. Smith CJ. Members of the majority need to actively promote diversity, equity, inclusion, and belonging. Plos Biol. 2022;20(12):e3001902.

102. Potvin DA, Burdfield-Steel E, Potvin JM, Heap SM. Diversity begets diversity: a global perspective on gender equality in scientific society leadership. Plos One. 2018;13(5):e0197280.

103. Elliott KC, Cheruvelil KS, Montgomery GM, Soranno PA. Conceptions of good science in our data-rich world. Bioscience. 2016;66(10):880–9.

104. McElreath R: Statistical rethinking : a Bayesian course with examples in R and Stan, Second edition. edn. Boca Raton: Chapman & Hall/CRC; 2020.

105. Goymann W. Ethology now publishes registered reports. Ethology. 2022;128(12):705–6.

106. Nat. Ecol. Evol.: Rolling out Registered Reports. Nat Ecol Evol 2023, 7(5):625–625.

107. Nat. Comm.: Opportunities and challenges for Registered Reports in ecology and evolution. Nature Communications 2022, 13(1):7266.

108. Nakagawa S, Ivimey-Cook ER, Grainger MJ, O'Dea RE, Burke S, Drobniak SM, Gould E, Macartney EL, Martinig AR, Morrison K, et al. Method Reporting with Initials for Transparency (MeRIT) promotes more granularity and accountability for author contributions. Nat Commun. 2023;14(1):1788.

109. Braga PHP, Hébert K, Hudgins EJ, Scott ER, Edwards BPM, Reyes LSL, Grainger MJ, Foroughirad V, Hillemann F, Binley AD, et al. Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. Methods Ecol Evol. 2023;14(6):1364–80.

110. Gomes DGE, Pottier P, Crystal-Ornelas R, Hudgins EJ, Foroughirad V, Sánchez-Reyes LL, Turba R, Martinez PA, Moreau D, Bertram MG, et al. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. P Roy Soc B-Biol Sci. 1987;2022:289.

111. Ivimey-Cook ER, Pick JL, Bairos-Novak KR, Culina A, Gould E, Grainger M, Marshall BM, Moreau D, Paquet M, Royauté R, et al. Implementing code review in the scientific workflow: insights from ecology and evolutionary biology. J Evolution Biol. 2023;36(10):1347–56.

Publisher's Note