Are hypotheses necessary in ecology and evolution?

Matthew G. Betts*, Adam S. Hadley*, David Frey, Sarah Frey, Dusty Gannon, Scott Harris, Hankyu Kim, Urs Kormann, Kara Leimberger, Katie Moriarty, Joseph Northrup, Ben Phalan, Josée S. Rousseau, Thomas S. Stokely, Jonathon Valente, Diego Zarrate Charry

Forest Biodiversity Research Network, Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA.

*Corresponding authors (matt.betts@oregonstate.edu, adam.hadley@oregonstate.edu). These authors contributed equally to this manuscript.

Abstract

Research hypotheses have been a cornerstone of science since before Galileo. Many have argued that inclusion of multiple hypotheses (1) encourage discovery of mechanisms, and (2) reduce bias – both features that should increase transferability and reproducibility. However, we are entering a new era of big data and highly predictive models where some argue the hypothesis is outmoded. Indeed, using a detailed literature analysis, we found prevalence of hypotheses in eco-evo research is very low (6.7-26%) and static from 1990-2015, a pattern mirrored in an extensive literature search (N=302,558 articles). Our literature review also indicates that neither grant success or citation rates were related to the inclusion of hypotheses, which may provide disincentive for hypothesis formulation. Here we confront common justifications for avoiding hypotheses and present new arguments based on benefits to the individual. Although hypotheses are not always necessary, we expect their continued and increased use will help our fields move toward greater understanding, reproducibility, prediction, and effective conservation of nature.

Key words: hypothesis, prediction, mechanisms, multiple-working hypotheses, scientific method

Introduction

Why should ecologists have hypotheses? At the beginning of most science careers there comes a time of "hypothesis angst" where students question the need for the hypothetico-deductive approach their elders have deemed essential for good science. Why is it not sufficient to just have a research objective or question? Why can't we just collect observations and describe those in our research papers?

Hypotheses are explanations for an observed phenomenon (Loehle 1987, Wolff and Krebs 2008) (See Box 1) and have been proposed as a central tool of science since Galileo and Francis Bacon in the mid-1500s. Over the past century, there have been repeated calls for rigorous application of hypotheses in science, and arguments that hypothesis use is the cornerstone of the scientific method (Chamberlin 1965, Popper 1959, Romesburg 1981). In a seminal paper in *Science,* Platt (1964) challenged all scientific fields to adopt and rigorously test multiple hypotheses (sensu Chamberlin 1965); he argued that without such hypothesis tests, disciplines would be prone to "stamp collecting" (Landy 1986). To constitute "strong inference" Platt required the scientific method to be a three-step process including (1) developing alternative hypotheses, (2) devising a set of "crucial" experiments to eliminate all but one hypothesis, and (3) performing the experiments (Elliott and Brook 2007).

The commonly touted strengths of hypotheses are two-fold. First, by adopting multiple plausible explanations for a phenomenon (hereafter "*multiple alternative hypotheses*"; Box 1), a researcher reduces the chance that he or she will become attached to a single possibility, thereby biasing research in favor of this outcome (Chamberlin 1965); this “confirmation bias” is a well-known human trait (Loehle 1987, Rosen 2016) and likely decreases reproducibility. Second, various authors have argued that the hypothesis framework forces one to think in advance about, and then test various *causes* for patterns in nature (Wolff and Krebs 2008), rather than simply examining the patterns themselves and coming up with explanations after the fact (so called ‘inductive research’; Romesburg 1981). By understanding and testing mechanisms, science becomes more reliable and transferable (Ayres and Lombardero 2017, Houlahan et al. 2017, Sutherland et al. 2013) (Fig. 1). Importantly, both of these strengths should have strong, positive impacts on reproducibility of ecological and evolutionary studies (see Discussion).

However, we are entering a new era of ecological and evolutionary science that is characterized by massive datasets on genomes, species distributions, climate, land cover, and other remotely sensed information (e.g., bioacoustics, camera traps; Pettorelli et al. 2017). Exceptional computing power and new statistical and machine-learning algorithms now enable thousands of statistical models to be run in minutes. Such datasets and methods allow for recognizing patterns at unprecedented spatial scales and for huge numbers of taxa and processes. Indeed, there have been recent arguments in both the scientific literature and popular press to do away with the traditional scientific method and ditch the notion of a priori hypotheses (Glass and Hall 2008, Golub 2010). These arguments go something along the lines of “if we can get predictions right most of the time, why do we need to know the cause”?

Here we sought to understand if hypothesis use in ecology and evolution has shifted in response to these pressures on the discipline. Is the pattern in the direction of *more* frequent hypothesis use – reflecting a response to various high-profile papers that have called for them? Or is hypothesis use declining – reflecting an increased move toward big data-pattern recognition and away from mechanistic studies?

We also present some common justifications for absences of hypotheses and suggest potential counterpoints researchers should consider prior to dismissing hypothesis use. We also evaluate the potential for benefits to the individual researcher from hypothesis use. Our hope for this communication is that we provide practical recommendations for improving hypothesis use in ecology and evolution – particularly for new practitioners in the field (Box 2).

Results

Trends in hypothesis use in ecology and evolution

In the ecology and evolution journals we examined in detail (see Methods) the prevalence of multiple alternative hypotheses (6.7%) and mechanistic hypotheses (26%) is very low and showed no temporal trend (Fig. 2A-D; GLMM, multiple alternative: $\hat{\square}$ = 0.098±0.247 SE, z=0.40, p=0.69, mechanistic: $\hat{\square}$ = 0.131±0.143 SE, z=0.92, p=0.36; see Supporting Methods for definitions). This pattern is consistent with a Web of Science search (N= 302,558 articles) for the term ‘hypothesis’ in titles or abstracts that shows essentially no trend over the same time period (Fig. 2E & F; see Supporting Methods).

Interestingly, applied and basic journals did not show a stastistically significant difference in the prevalence of either mechanistic (GLMM: z=0.15, p=0.875, $\hat{\square}$ = 0.054±0.344 SE) or multiple alternative hypotheses (GLMM: z=0.88, p=0.375, $\hat{\square}$ = 0.517±0.583 SE). However, there was substantial variation across both basic and applied journals in the prevalence of hypotheses (Fig. 3).

Do hypotheses 'pay'?

Why are hypotheses so uncommon? Could it be that use of hypotheses conveys few individual benefits? Hypotheses might be "useful" for overall progress in science (i.e., discovery of mechanism, reduced bias, increased reproducibility (Platt 1964)); but for their use to be propagated in the population of scientists, one would also expect them to confer benefits to the individuals conducting the science. Perhaps hypotheses are relatively rare because in this competitive era, scientists are not perceiving rewards of adopting this approach in terms that promote individual progress: grants, publications and citations (Weinberg 2010)? If hypothesis use conveys individual-level advantages in the quantities typically measured for academic success, then having hypotheses should result in getting articles published in top-ranked journals, higher citation rates, and higher funding rates.

We found little evidence that presence of hypotheses increased paper citation rates. Papers with mechanistic (LME: t = 0.042, p=0.97) or multiple alternative hypotheses (LME: t= 0.971, p=0.33) did not have higher average annual citation rates (Fig. 4A); nor did papers with at least one of these hypothesis types show a signal of being cited slightly more frequently (LME: $t=0.218$, $p = 0.83$], top panel of Fig. 4A)

On the other hand, journal articles containing mechanistic hypotheses did tend to be published in higher impact journals (GLM: $t=2.74$, $p=0.006$) but only slightly so (Fig. 4B); including multiple-alternative hypotheses in papers did not have statistically significant effect (GLM: $t=1.80$, $p=0.072$, Fig. 4B, bottom panel).

Finally, we found no association between obtaining a competitive national or international grant and the presence of a hypothesis (Fig. 4C; logistic regression: mechanistic: $z=0.36$, $p=0.75$, multiple alternative: $z=0.49$, $p=0.87$).

Discussion

Overall, the prevalence of hypothesis use in the ecological and evolution literature is strikingly low, and has been so for the past 25 years despite repeated calls to reverse this pattern (Elliott and Brook 2007, Peters 1991, Rosen 2016, Sells et al. 2018). Why is this the case?

Clearly hypotheses are not always necessary and a portion of the sampled articles may represent situations where hypotheses are truly not useful (See Box 3: “When Are Hypotheses Not Useful?”). Some authors (Wolff and Krebs 2008) overlook knowledge gathering and descriptive research as a crucial first step for making observations about natural phenomenon – from which hypotheses can be formulated. This descriptive work is an important part of ecological science

(Tewksbury et al. 2014), but may not benefit from strict use of hypotheses. Similarly, some efforts are simply designed to be predictive, such as machine-learning-driven auto-recognition of species (Briggs et al. 2012) or for prioritizing conservation efforts (Wilson et al. 2006), where the primary concern is correct identification and prediction rather than the biological or computational reasons for correct predictions (Box 3). However, it would be surprising if 75% of ecology since 1990 has been purely descriptive work from little-known systems or purely predictive in nature. Indeed, the majority of the articles we observed did not fall into these categories.

Alternatively, researchers may not include hypotheses because they see little individual-level incentive for their inclusion. Our results suggest that currently there are relatively few measurable benefits to individuals. Articles with mechanistic hypotheses do tend to be published in higher impact factor journals, which, for better or worse, is one of the key predictors in obtaining an academic job (van Dijk et al. 2014). However, few of the other typical academic metrics appear to reward this behavior. Although hypotheses might be 'useful' for overall progress in science (Platt 1964), for their use to be propagated in the population of scientists, one would also expect them to provide benefits to the individuals conducting the science. Interestingly, the few existing papers on hypotheses (Loehle 1987, Romesburg 1981, Sells et al. 2018) tended to explain the advantages in terms of benefits to the group by offering arguments such as "because hypotheses help the field move forward more rapidly".

Here we confront some of the common justifications for hypotheses not being necessary and show how one's first instinct to avoid hypotheses may be mistaken. We also present four reasons that use of hypotheses may be of individual self interest.

Common Justifications for the Absence of Hypotheses Dispelled

During our collective mentoring at graduate and undergraduate levels, as well as examination of the literature, we have heard a number of common justifications for why hypotheses are not included. We must admit that many of us have, on occasion, rationalized absence of hypotheses in our own work using the same logic! We understand that clearly formulating and testing hypotheses can often be challenging, but propose that the justifications for avoiding hypotheses should be carefully considered.

*1. "**But I do have hypotheses**"*. Simply using the word "hypothesis" does not a hypothesis make. A common pattern in the literature we reviewed was for researchers to state their guess about the results they expect and call this the "hypothesis" (e.g., "I hypothesize trees at higher elevation will grow slowly"). But these are usually predictions derived from an implicit theoretical model (Symes et al. 2015) or are simply descriptive statements with the word 'hypothesis' in front of them (see Box 1). A research hypothesis must contain explanations for an observed phenomenon (Loehle 1987, Wolff and Krebs 2008). Such explanations are derived from existing or new theory (Symes et al. 2015). Making the link between the expected mechanism (hypothesis) and logical outcome if that mechanism were true (the prediction), is a key element of strong inference. Similarly, using "statistical hypotheses" and "null hypothesis testing" is not the same as developing mechanistic research hypotheses (Romesburg 1981, Sells et al. 2018).

2. "***Not enough is known about my system to formulate hypotheses*". ** This is perhaps the most common defense against needing hypotheses (Golub 2010). The argument goes that due to lack of previous research no mature theory has developed, so formal tests are impossible. Such arguments may have basis in some truly novel contexts (e.g., exploratory research on genomes) (Golub 2010). But on close inspection, similar work has often been conducted in other geographic regions, systems or with different taxa. If the response by a researcher is "but we really need to know if X pattern also applies in this region as well" (e.g., does succession influence bird diversity in forests of Western North America the same way as it does in Eastern forests), this is fine and it is certainly useful to accumulate descriptive studies globally for future synthetic work. However, continued efforts at description alone constitute missed opportunities for understanding the mechanisms behind a pattern (e.g., why does bird diversity decline when the forest canopy closes?). The key is for students to fight the inertia that often results in purely descriptive studies and to exhaust avenues for attempting to formulate mechanistic hypotheses. Often with a little planning, both the initial descriptive local interest question (e.g., "is it?") and the broader interest question (i.e., "why?") can both be tackled with little additional effort.

3. "***What about Darwin? Many important discoveries have been made without hypotheses***". Several authors (and many students) have argued that many important and reliable patterns in nature have emerged outside of the hypothetico-deductive (H-D) method (Brush 1974). For instance, Darwin's discovery of natural selection as a key force for evolution has been put forward as an example of how reliable ideas can emerge without the H-D method (May 1981, Milner 2018). Examination of Darwin's notebooks has suggested that he did not propose explicit hypotheses and test them (Brush 1974). However, Darwin himself wrote "all observation must be for or against some view if it is to be of any service!" (Ayala 2009) In fact, Darwin actually put forward and empirically tested hypotheses in multiple fields, including geology, plant morphology and physiology, psychology, and evolution (Ayala 2009). This debate suggests that, like Darwin, we should continue to value systematic observation and descriptive scince (Tewksbury et al. 2014), but whenever possible it should be with a view toward developing theory and testing hypotheses.

The statement that "many important discoveries have been made without hypotheses" stems from a common misconception that somehow hypotheses spring fully formed into the mind, and that speculation, chance and induction play no role in the H-D method. As noted by Loehle (1987; p 402) "The H-D method and strong inference, however, are valid no matter how theories are obtained. Dreams, crystal balls, or scribbled notebooks are all allowed. In fact, induction may be used to create empirical relations which then become candidates for hypothesis testing even though induction cannot be used to prove anything". So, although induction has frequently been used to develop theory, it is an unreliable means to test theory (Popper 1959). As is well known, Darwin's theory of natural selection was heavily debated in scientific circles at the time, and it is only through countless hypothesis tests that it remains the best explanation for evolution even today (Mayr 2002).

4. "***Ecology is too complex for hypotheses***". In one of the most forcefully presented arguments for the H-D method, Karl Popper (1959) argued that science should be done through a process of falsification; that is, multiple hypotheses should be constructed and the researcher's role is to successively eliminate these one at a time via experimentation, until a single plausible hypothesis remains. This approach has caused some consternation among ecologists because the idea of single causes to phenomena doesn't match most of our experiences (Quinn and Dunham 1983); rather, multiple interacting processes often overlap to drive observed patterns; for example Robert Paine's finding that the distribution of a common seaweed was best explained by competition, physical disturbance *and* dispersal ability.

It would be interesting if Popperian logic has inoculated ecology and evolution against the frequent application of hypotheses in research. Perhaps because the bar of falsification and testable mutually exclusive hypotheses is so high, many have opted to ignore the need for hypotheses altogether? If this is the case, our response is that in ecology and evolution we must not let Popperian perfection be the enemy of strong inference. With sufficient knowledge of a system, formal a priori hypotheses can be formulated that directly address the possibility of non-linear relationships and interactions among variables. An example from conservation biology is the well-explored hypothesis that the effects of habitat fragmentation should be greatest when habitat amount is low due to dispersal limitation (i.e., there should be a statistical interaction between fragmentation and habitat loss (Andrén 1994)).

5. "***But I am not a physiologist***". A common misconception has to do with the hierarchical aspect of mechanisms (Fig. 5). Many think that they are not testing the mechanism for a pattern because they have not managed to get to the bottom of a causal hierarchy (which reflects a sort of physics envy that commonly occurs in ecology and evolution (Egler 1986)). However, hierarchy theory (O'Neill et al. 1989), states that the cause of a given phenomenon usually occurs at the level of organization just below the observed phenomenon. So, for example, species distributions might be best understood by examining hypotheses about the spatial composition and configuration of landscapes (Fahrig 2003); explanations for population regulation might be best explored through observing the reproductive success and survival of individual organisms (Lack 1954); to understand variation among individuals in fecundity one might test hypotheses relating to individual behavior or physiology. Hypothesis generation is possible at all levels of organization (Fig. 5). Support for a hypothesis at one level often generates a subsequent question and hypotheses at the next (e.g., Observation: variation in animal densities can best be explained by forest patch size; Question: why are densities lower in small patches? H_1: small patches have more edge, and predation rates are higher at the edge). However, in a single research project it is not necessary to develop hypotheses that address mechanisms at all scales.

6. "***But my model predicts patterns well***". An increasingly common justification for not presenting and testing research hypotheses seems to be the notion that if large datasets and complex modeling methods can predict outcomes effectively, what is the need for hypothesizing a mechanism (Glass and Hall 2008, Golub 2010)? Indeed, some have argued that prediction is a gold standard in ecology and evolution (Houlahan et al. 2017). However, underlying such arguments is the critical assumption that the relationship between predictors (i.e., independent variables, 'x's) and responses ('y's) exhibit *stationarity* in time and space. Although this appears to be the case in cosmology (e.g., relativity is thought to apply wherever you are in the universe

(Einstein 1920)), the assumption of stationarity has repeatedly been shown to be violated in ecological and evolutionary studies (Betts et al. 2006, Osborne Patrick et al. 2007, Thompson 2005). Hence the well-known maxim "correlation does not equal causation"; correlates of a phenomenon often shift, even if the underlying cause remains the same.

The advantage of understanding mechanism is that the relationship between cause and effect is less likely to shift in space and time than between the correlates of a phenomenon (Sells et al. 2018)(Fig. 1). For instance, climate-envelope models are still commonly used to predict future species distributions (Beale et al. 2008) despite the fact that links between correlates often fail (Gutiérrez et al. 2014) and climate *per se* may not be the direct driver of distributions. In an example from our own group, predictions that fit observed data well in the region where the model was built completely failed when predicted to a new region only 250 km away (Betts et al. 2006). Although it is true that mechanisms can also exhibit non-stationarity, at least in these instances logic can be used to make informed decisions about whether or not causal factors are likely to hold in a new place or time.

Why Should You Have Hypotheses? (A Self-Interested Perspective)

We have already described two arguments for hypothesis use, both of which should have positive influences on reproducibility and therefore progress in science; (1) a priori multiple working hypotheses prevent attachment to a single idea, (2) hypotheses encourage exploration of mechanisms, which should increase the transferability of findings to new systems. Both these arguments have been made frequently in the eco-evolutionary literature for decades and should have positive influence on reproducibility and therefore progress in science (Elliott and Brook 2007, Loehle 1987, Rosen 2016, Sells et al. 2018) but our results show that such arguments have been lost on the majority of researchers. One hypothesis recently proposed to explain why "poor methods persist [in science] despite perennial calls for improvements" is that such arguments have largely failed because they do not appeal to researcher self-interest (Smaldino and McElreath 2016). In periods of intense competition for grants and top-tier publications, perhaps arguments that rely on altruism fall short. However, happily, there are at least four self-interested reasons that students of ecological and evolutionary science should adopt the hypothetico-deductive method.

1. Clarity and Precision in Research

First, and most apparent during our review of the literature, hypotheses force clarity and precision in thinking. We often found it difficult to determine the core purpose of papers that lacked clear hypotheses. One of the key goals of scientific writing is to communicate ideas efficiently (Schimel 2011). Increased clarity through use of hypotheses could potentially even explain the pattern for manuscripts using hypotheses getting published in higher impact journals. Editors are increasingly pressed for time and forced to reject the majority of papers submitted to higher impact outlets prior to detailed review (AAAS 2018). 'Unclear message' and 'lack of clear hypotheses' are top reasons a paper ends up in the editor's reject pile (Eassom 2018, Elsevier 2015). If editors have to struggle as often as we did to determine the purpose of a paper, this does not bode well for future publication. Clearly communication through succinctly stated hypotheses is likely to enhance publication success.

Hypotheses also provide crucial direction during study design. Nothing is more frustrating than realizing that your hard earned data cannot actually address the key study objectives, or rule out alternative explanations. Developing clear hypotheses and in particular multiple alternative hypotheses ensures that you actually design your study in a way that *can* answer the key questions of interest.

2. Personal Fulfillment

Second, science is more likely to be fulfilling and fun when the direction of research is clear, but perhaps more importantly, when questions are addressed with more than one plausible answer. Results are often disappointing or unfulfilling when the study starts out with a single biological hypothesis in mind (Symes et al. 2015) – particularly if there is no support for this hypothesis. If multiple working hypotheses are well crafted, something interesting and rewarding will result *regardless of the outcome*. This results in a situation where researchers are much more likely to enjoy the *process* of science because the stress of wanting a particular end is removed. Subsequently, as Chamberlin (1965) proposed, "the dangers of parental affection for a favorite theory can be circumvented" which should reduce the risk of creeping bias. In our experience reviewing competitive grant proposals at the U.S. National Science Foundation, it is consistently the case that proposals testing several compelling hypotheses were more likely to be well received – presumably because reviewers are risk averse and understand that ultimately finding support for *any* of the outcomes will pay off. Why bet on just one horse when you can bet on them all?

3. Intrinsic Value to Mechanism

Mechanism seems to have intrinsic value for humans — regardless of the practical application. Humans tend to be interested in acquiring understanding rather than just accumulating facts. As a species we crave answers to the question "why."

Indeed, it is partly this desire for mechanism that is driving a recent perceived "crisis" in machine learning, with the entire field being referred to as "alchemy" (Hutson 2018); algorithms continue to increase in performance, but the mechanisms for such improvements are often a mystery – even to the researchers themselves. "Because our model predicts well" is the unsatisfying scientific equivalent to a parent answering a child's "why?" with "because that's just the way it is." This problem is beginning to spawn a new field in artificial intelligence "AI neuroscience" which attempts to get into the 'black-box' of machine learning algorithms to understand how and why they are predictive (Voosen 2017).

Even in some of our most applied research, we find that managers and policy makers when confronted with a result (e.g., thinning trees to 70% of initial densities reduced bird diversity) want to know *why* (e.g., thinning eliminated nesting substrate for 4 species); if the answer to this question is not available, policy is much less likely to change (Sells et al. 2018). So, formulating mechanistic hypotheses will not only be more personally satisfying, but we expect it may also be more likely to result in real-world changes.

4. You Are More Likely To be Right

In a highly competitive era, it seems that in the quest for high publication rates and funding, researchers lose sight of the original aim of science: to discover a truth about nature that is transferable to other systems. In a recent poll conducted by *Nature,* more than 70% of researchers have tried and failed to reproduce another scientist's experiments (Baker 2016). Ultimately, each researcher has a choice; put forward multiple explanations for a phenomenon on their own or risk 'attachment' to a single hypothesis and run the risk of bias entering their work, rendering it unreproducible, and subsequently being found wrong by a future researcher. Imagine if Lamarck had not championed a single hypothesis for the mechanisms of evolution? Although Lamarck potentially had a vital impact as an early proponent of the idea that biological evolution occurred and proceeded in accordance with natural laws (Stafleu 1971), unfortunatly in the modern era he is largely remembered for his pet hypothesis. It may be a stretch to argue that he would have necessarily come up with natural selection, if he *had* considered natural selection, the idea would have emerged 50 years earlier, substantially accelerating scientific progress and limiting his infamy as an early evolutionary biologist. An interesting contemporary example is provided by Prof. Amy Cuddy's research focused on "power posing" as a means to succeed. The work featured in one of the most viewed TED talks of all time but rather famously turned out to be irreproducible (Ranehill et al. 2015). When asked in a TED interview what she would do differently now, Prof. Cuddy noted that she would include a greater diversity of theory and multiple potential lines of evidence to "shed light on the psychological mechanisms" (Biello 2017).

Conclusion

We acknowledge that formulating effective hypotheses can feel like a daunting and unwarranted hurdle for ecologists. However, we suggest that initial justifications for absence of hypotheses may often be unfounded. We argue that there are both selfish and altruistic reasons to include mechanistic multiple working hypotheses in your research: (1) testing multiple working hypotheses simultaneously makes for rapid and powerful progress which is to the benefit of all (Platt 1964), (2) you lessen the chance that confirmation bias will result in you publishing an incorrect but ‘sexy’ idea, (3) hypotheses provide clarity in design and writing, (4) research using hypotheses is more likely to be published in a high impact journal, and (5) you are able to provide satisfying answers to “why?” phenomenon occur. However, few current academic metrics appear to reward use of hypotheses. Therefore, we propose that in order to promote hypothesis use we may need to provide additional incentives (Edwards and Roy 2016, Smaldino and McElreath 2016). We suggest editors reward research conducted using principles of sound scientific method and be skeptical of research that smacks of data dredging, post-hoc, and single hypotheses. If no hypotheses are stated in a paper and/or the paper is purely descriptive, editors should ask whether the novelty of the system and question warrant this, or if the field would have been better served by a study with mechanistic hypotheses. Eleven of the top 20 ecology journals already indicate a desire for hypotheses in their instructions for authors – with some going as far as indicating “priority will be given” for manuscripts testing clearly stated hypotheses. Although hypotheses are not necessary in all instances, we expect that their continued and increased use will help our disciplines move toward greater understanding, higher reproducibility, better prediction, and more effective management and conservation of nature. We recommend authors, editors, and readers encourage their use (Box 2).

Box 1. **Definitions of Hypotheses and Associated Terms**

Hypothesis: an explanation for an observed phenomenon

Research Hypothesis: a statement about a phenomenon that also includes the potential mechanism or cause of that phenomenon. Though a research hypothesis doesn't need to adhere to this strict framework it is often best described as the "if" in an "if-then" statement. In other words, "if X is true" (where X is the mechanism or cause for an observed phenomenon) "then Y" (where Y is the outcome of a crucial test that supports the hypothesis). These can also be thought of as "**mechanistic hypotheses**" since they link with a causal mechanism. For example, trees grow slowly at high elevation because of nutrient limitation (hypothesis); if this is the case, fertilizing trees should result in more rapid growth (prediction).

Prediction: the potential outcome of a test that would support a hypothesis. Most researchers call the second part of the if-then statement a "prediction".

Multiple alternative hypotheses: multiple plausible explanations for the same phenomenon.

Descriptive Hypothesis: descriptive statements or predictions with the word 'hypothesis' in front of them. Typically researchers state their guess about the results they expect and call this the "hypothesis" (e.g., "I hypothesize trees at higher elevation will grow slowly").

Statistical Hypothesis: a predicted pattern in data that should occur if a research hypothesis is true.

Null Hypothesis: a concise statement expressing the concept of "no difference" between a sample and the population mean.

Box 2. **Recommendations for Improving Hypotheses Use in Ecology and Evolution**

Authors: Know that you are human and prone to confirmation bias and extremely effective at false pattern recognition. Thus, inductive research and single working hypotheses should be rare in your research. Remember that if your work is to have a real "impact", it needs to withstand multiple tests from other labs over the coming decades.

Editors and Reviewers: Reward research that is conducted using principles of sound scientific method. Be skeptical of research that smacks of data dredging, post-hoc, and single hypotheses. If no hypotheses are stated in a paper and/or the paper is purely descriptive, ask whether the novelty of the system and question warrant this, or if the field would have been better served by a study with mechanistic hypotheses. If only single hypotheses are stated, ask whether appropriate precautions were taken for the researcher to avoid finding support for a pet idea (e.g., blinded experiments, randomized attribution of treatments etc.). To paraphrase Platt (1964): beware of the person with only one method or one instrument, either experimental or theoretical.

Mentors: Encourage your advisees to think carefully about hypothesis use and teach them how to construct sound multiple, mechanistic hypotheses. Importantly, rather than demanding hypothesis use blindly as a requisite to science, explain why hypotheses are important to the scientific method, the individual and group consequences of excluding them, and the rare instances where they may not be necessary.

Policy makers/ media/ educators/ students/ readers: It almost goes without saying, but always read with skepticism; have a scrutinous eye out for single hypothesis studies and p-hacking. Reward multi-hypothesis, mechanistic, predictive science by giving it greater weight in policy decisions (Sutherland et al. 2013), more coverage in the media, greater leverage in education and more citations in reports.

Box 3: **When Are Hypotheses Not Useful?**

Of course, there are a number of instances where hypotheses might not be useful or needed. It is important to recognize these instances where hypotheses are not necessary to prevent the pendulum from swinging in direction where without hypotheses, research ceases to be science (Wolff and Krebs 2008). Below are several important types of ecological research where formulating hypotheses may not always be beneficial.

When the goal is prediction rather than understanding. Examples of this exception include species distribution models (Elith et al. 2008) where the question is not why species are distributed as they are, but simply where species are predicted to be. Such results can be useful in conservation planning (Guisan et al. 2013; see below). Another example lies in auto-recognition of species (Briggs et al. 2012) where the primary concern is getting identification right rather than the biological or computational reasons for correct predictions. In such instances, complex algorithms can be very effective at uncovering patterns (e.g., deep learning). A caveat and critical component of such efforts is to ensure that such models are tested on independent data. Further, if model predictions are made beyond the spatial or temporal bounds of training or test data, extreme caution should be applied (see Fig. 4).

When the goal is description rather than understanding. In many applications, the objective is to simply quantify a pattern in nature; for example, where on earth is losing forest at the most rapid rates (Hansen et al. 2013)? Further, sometimes so little is known about a system or species that formulating hypotheses is impossible and more description is necessary. In rare instances, an ecological system may be so poorly known and different to other systems that generating testable hypotheses would be extremely challenging. Darwin's observations while traveling on the Beagle are some of the best examples of such 'hypothesis generating' science; these initial observations resulted in the formulation of one of the most extensively tested hypotheses in biology. However, such novelty should be uncommon in ecological and evolutionary research where theoretical and empirical precedent abounds (Sells et al. 2018). In the field of biogeography there is the commonly held view that researchers should first observe and analyze patterns, and only then might explanations emerge ('pattern before process'); however, it has frequently been demonstrated that mechanistic hypotheses are useful even in disciplines where manipulative experiments are impossible (Crisp et al. 2011).

When the objective is a practical planning outcome such as reserve design. In many conservation planning efforts, the goal is not to uncover mechanisms, but rather simply to predict efficient methods or contexts for conserving species (Myers et al. 2000, Wilson et al. 2006). Perhaps this is the reason for such low prevalence of hypotheses in conservation journals (e.g., Conservation Biology).

Methods

Literature Analysis

To test for the presence of hypotheses and whether this was associated with grants, journal impact factor, and citation rates, we sampled the ecology and evolution literature using a stratified random approach. We selected 22 journals that have been in existence since 1990, 13 of which were from journals that focused more on general ecology, 6 more applied ecology journals, and 3 multidisciplinary science journals (*Science, Proceedings of the National Academies of Sciences, Nature*; see Fig. 3 for full list). Journals were also selected to cover a gradient in impact factor (ISI Web of Science impact factor range: 0.55 – 40.14). We randomly sampled articles from these journals in 5-year bins to ensure that the full date-range was covered (1990-2015). We removed articles in the following categories: editorials, corrections, reviews, opinions, methods papers. Once selected, articles were randomly distributed to this paper's authors for detailed examination.

Authors of this paper (MGB, ASH, DF, SF, DG, SH, HK, UK, KL, KM, JN, BP, JSR, TSS, JV, DZC) were given a maximum of 10 minutes to find research hypothesis statements within the abstract or introduction of papers. We chose 10 minutes to simulate the amount of time an editor pressed for time might spend trying to get the measure of the introductory material in an article. After this initial 10 minute period we determined: (1) whether or not an article contained at least one hypothesis, (2) whether hypotheses were mechanistic or not (i.e., did they claim to examine the mechanism for an observed phenomenon), (3) whether multiple alternative hypotheses were considered (sensu Chamberlain 1898), (4) whether hypotheses were 'descriptive' (that is, they did not explore a mechanism but simply stated the expected direction of an effect; we define this above as a "prediction" [Box1]). (5) We also examined all papers for funding sources and noted the presence of a national-level competitive grant (e.g., National Science Foundation, European Union, Natural Sciences and Engineering Research Council). Journal impact factor and individual article citation rates were gleaned directly from Web of Science.

We also tested whether the temporal trends in hypotheses in our sample mirrored the broader literature. For the same set of 22 journals in our sample, we conducted a Web of Science search for articles containing "Hypoth*" in the title or abstract. To calculate the proportion of articles with hypotheses (from 1990-2018), we divided the number of articles with hypotheses by the total number of articles (N=302,558).

Statistical Analysis

We used generalized linear mixed models (GLMMs) to test for change in the prevalance of various hypothesis types over time (mechanistic, multiple, any hypothesis). Presence of a hypothesis was modeled as dichotomous (0,1) with binomial error structure and 'journal' as a random effect to account for potential lack of independence among articles published in the same outlet. The predictor variable (i.e., year) was scaled to enable convergence. GLMMs were implemented in R using the lme4 package (Bates et al. 2018). Similarly, we tested for differences in hypothesis prevalence between basic and applied journals using GLMMs with 'journal' as a random effect.

We tested whether the presence of hypotheses influenced the likelihood of publiction in a high-impact journal using generalized linear models with a gaussian error structure. We used the log of journal impact factor (+ 0.5) as the response variable to improve normality of model residuals. We tested the association between major competitive grants and the presence of a hypotheses using logistic regression with 'hypothesis presence' (0,1) as a predictor and presence of a grant (0,1) as a response.

Finally, we tested whether hypotheses increase citation rates using linear mixed effects models (LME); presence of various hypotheses (0,1) were predictors in univariate models and average citations per year (log-transformed) was the response. 'Journal' was treated as a random effect, which assumes that articles within a particular journal are unlikely to be independent in their citation rates.

Acknowledgements

Funding from the National Science Foundation (NSF-DEB-1457837) to MGB and ASH supported this research. We thank Craig Loehle for thoughtful comments on the manuscript.

Author contributions

MGB and ASH conceived of the study, ASH designed and organized data collection, MGB analyzed the data and wrote the first draft of the manuscript, all authors collected data, provided input on design and key manuscript themes and edited the manuscript.

Data availability

On publication, data will be made publically available via Datadryad: https://datadryad.org//

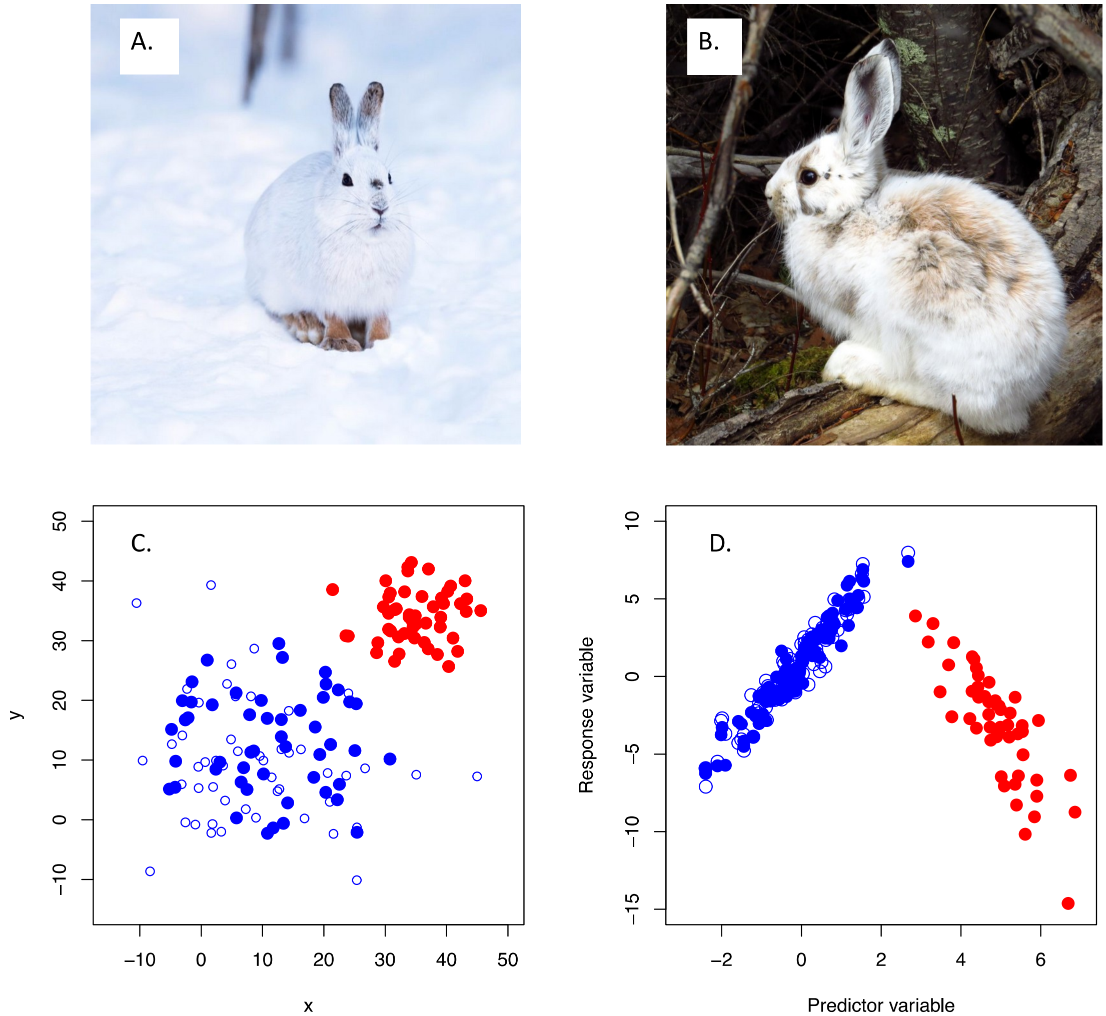

Fig. 1. Examples of non-stationarity between predictor and response variables resulting in poor reproducibility in simple predictive models. Imagine a machine learning model (e.g., random forest; [Elith et al. 2008]) that attempts to predict snowshoe hare survival as a function of hare color, but data have only been collected in winter. The model would predict extremely well within the temporal and spatial bounds of the data (white hares survive well in relation to brown hares). In a situation with no mechanistic hypotheses, a researcher studying snowshoe hares in winter might conclude, via correlations, that white hares (A) survive better than dark hares (B). On the other hand, a researcher testing the mechanism for hare survival would (ideally via experimentation) arrive at the conclusion that it is not the whiteness of hares, but rather blending with the background that confers survival (the camouflage hypothesis). Understanding mechanism results in model predictions being more robust to novel conditions. The machine learning model would be completely irreproducible in locations or months when there is no snow, whereas

the mechanistic understanding would be robust to such changes in space and time resulting in transferability to non-winter conditions. (C) Shows x and y locations for training a correlative model (blue filled circles) and testing it (blue open circles). Even if the model performs well on these independent test data, there is no guarantee that it will predict well to the red circles that are outside of the spatial bounds of the existing data. Non-stationarity (in this case caused by a nonlinear relationship between predictor and response variable) could result in correlative relationships shifting substantially if extrapolated to new times or places (D) Understanding drivers behind ecological patterns – via testing mechanistic hypotheses – reduces this risk.

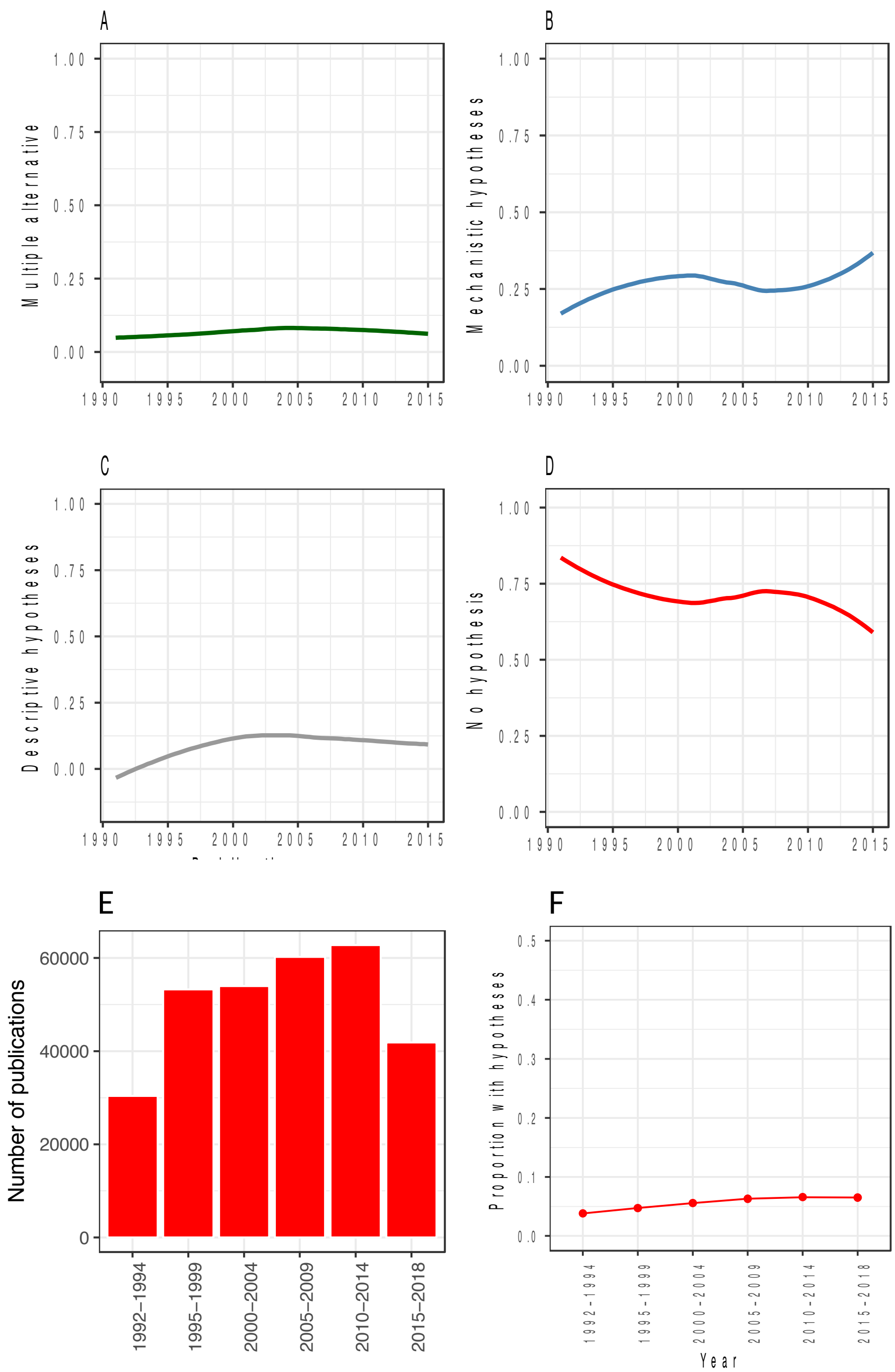

Fig. 2 Trends in hypothesis use from 1990-2015 from a sample of the ecological and evolutionary literature (A: multiple alternative hypotheses, B: mechanistic hypotheses, C: descriptive hypotheses [predictions], and D: no hypotheses present). We detected no temporal trend in any of these variables. Lines show loess smoothers with 95% confidence intervals. Dots show raw data. Total number of publications in ecology and evolution in selected journals

has increased (E) but use of the term ‘hypotheses’ in the title or abstracts of these 302,558 articles has remained flat, and at very low prevalence.

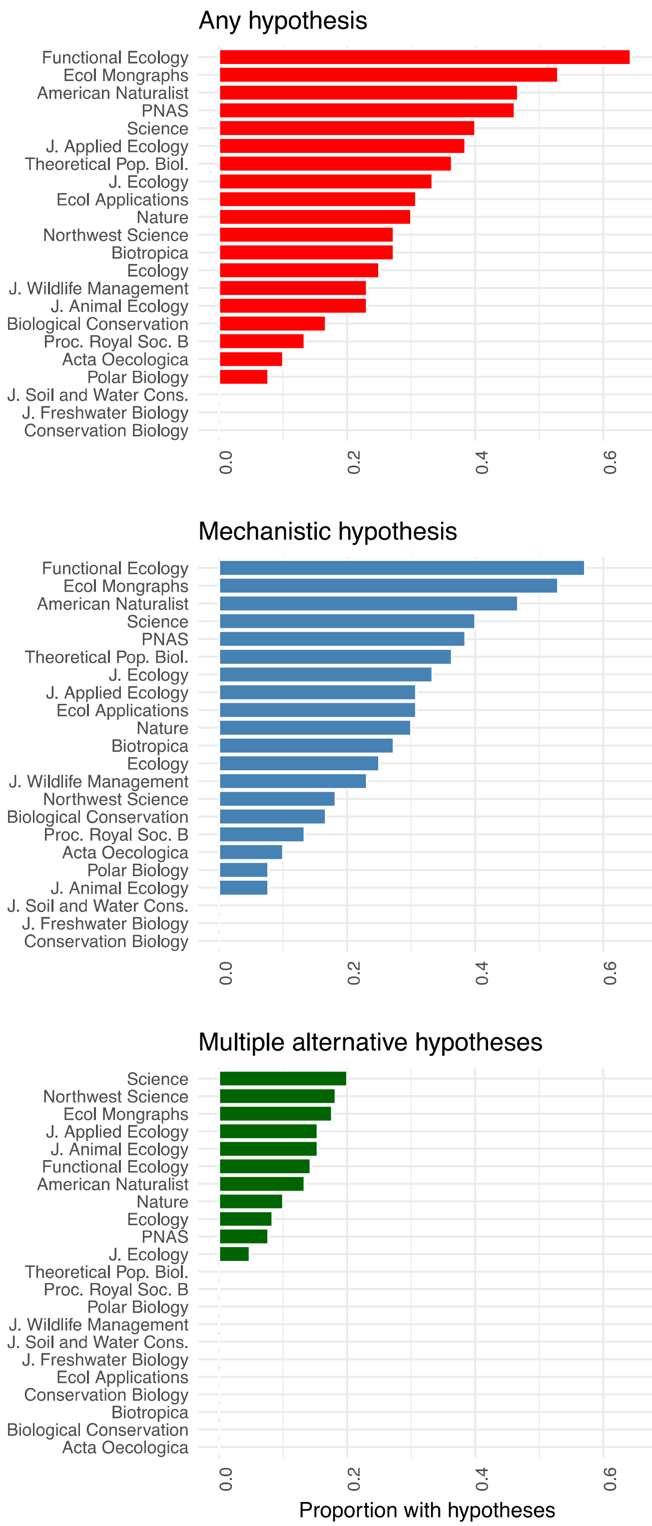

Fig. 3 Frequency distributions showing proportion of various hypotheses types across selected ecology and evolution journals. Hypothesis use varied greatly across publication outlets.

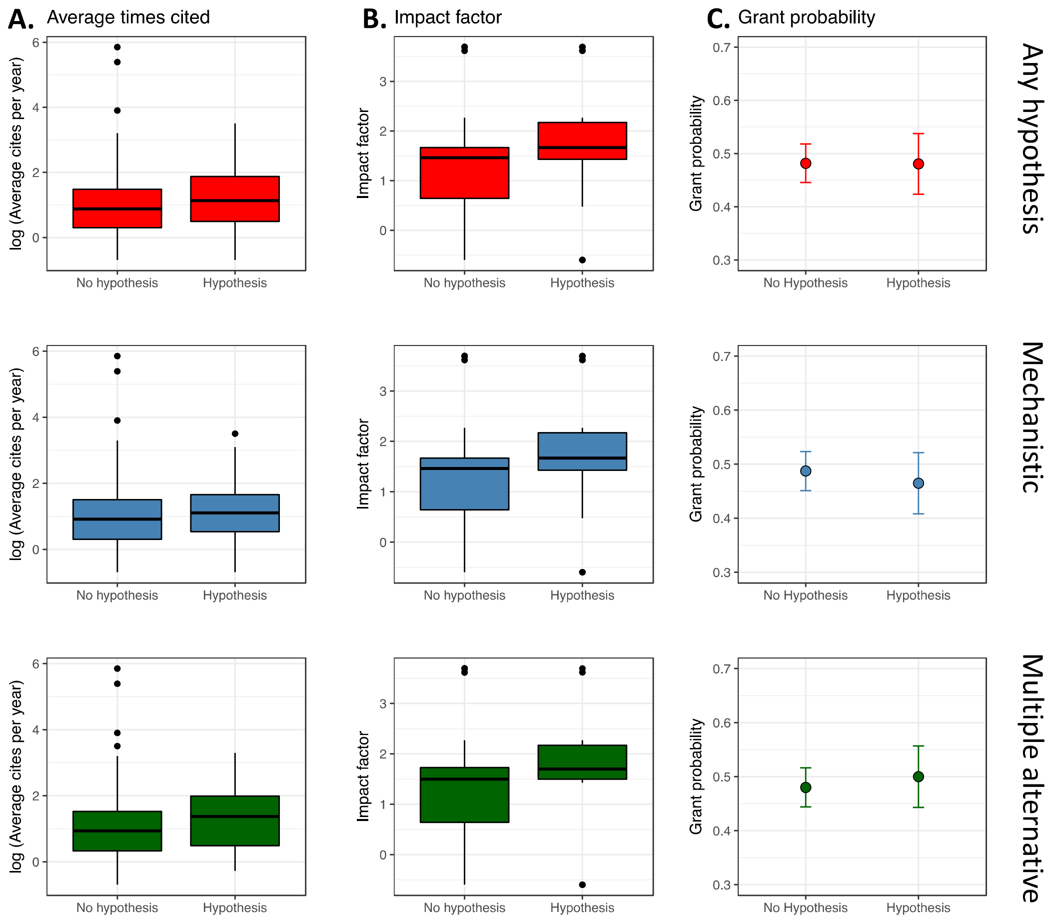

Fig. 4 Relationships between having a hypothesis (or not) and three commonly sought after scientific rewards (A: average times a paper is cited/year, B: journal impact factor, and C: the likelihood of having a major national competitive grant). We found no statistically significant relationships between having a hypothesis and citation rates or grants, but articles with hypotheses tended to be published in higher impact journals.

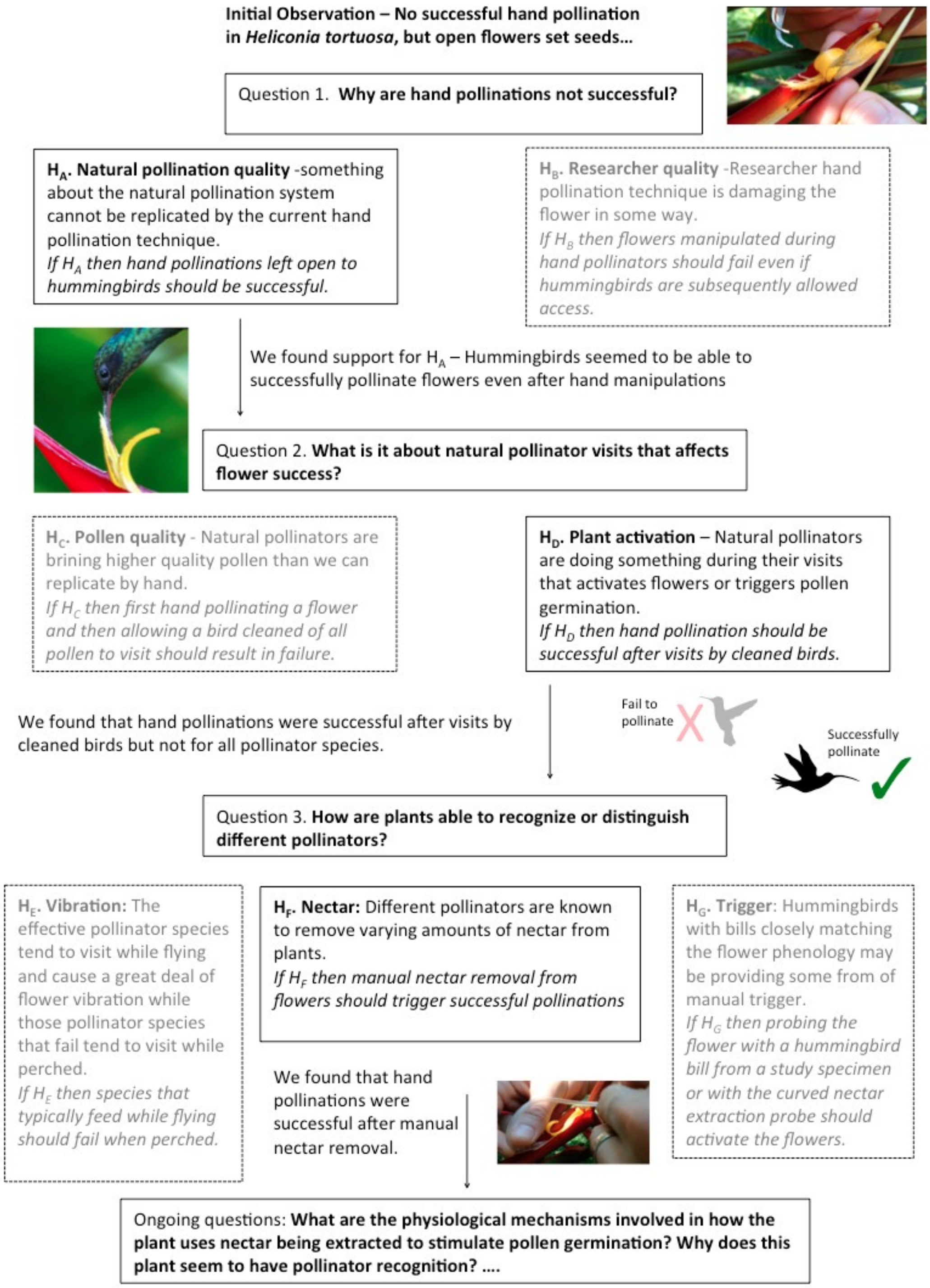

Fig. 5. Hypothesis generation is possible at all levels of organization, and does not need to get to the bottom of a causal hierarchy to be useful. As illustrated in this case study, using published work by the authors, support for a hypothesis at one level often generates a subsequent question and hypotheses at the next. After each new finding we had to return to the white board and draw out new alternative hypotheses as we progressed further down the

hierarchy. Supported hypotheses are shown in black and the alternative hypotheses that were eliminated are in grey. A single study is not expected to tackle an entire mechanistic hierarchy. In fact, we still have yet to uncover the physiological mechanisms involved in this phenomenon.

Literature Cited

AAAS. 2018. What percentage of submissions does Science accept? AAAS Science Contributors (http://www.sciencemag.org/site/feature/contribinfo/faq/index.xhtml - pct_faq).

Andrén H. 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: A review. Oikos 71: 355-366.

Ayala FJ. 2009. Darwin and the scientific method. Proceedings of the National Academy of Sciences 106: 10033–10039.

Ayres MP, Lombardero MJ. 2017. Forest pests and their management in the Anthropocene. Canadian Journal of Forest Research 48: 292-301.

Baker M. 2016. 1,500 scientists lift the lid on reproducibility. Nature 533: 452-454.

Bates D, Maechler M, Bolker B. 2018. 'lme4' Linear mixed-effects models using S4 classes. R Core team: https://cran.r-project.org/web/packages/lme4/lme4.pdf.

Beale CM, Lennon JJ, Gimona A. 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. Proceedings of the National Academy of Sciences 105: 14908-14912.

Betts M, Diamond T, Forbes GJ, Villard M-A, Gunn J. 2006. The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. Ecological Modeling 191: 197-224.

Biello D. 2017. Inside the debate about power posing: a Q & A with Amy Cuddy. Ideas.TED.com: https://ideas.ted.com/inside-the-debate-about-power-posing-a-q-a-with-amy-cuddy/.

Briggs F, Lakshminarayanan B, Neal L, Fern XZ, Raich R, Hadley SJK, Hadley AS, Betts MG. 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America 131: 4640-4650.

Brush SG. 1974. Should the history of science be rated X? Science 183: 1164-1172.

Chamberlin TC. 1965. The method of multiple working hypotheses. Science 148: 754-759.

Crisp MD, Trewick SA, Cook LG. 2011. Hypothesis testing in biogeography. Trends in Ecology & Evolution 26: 66-72.

Eassom H. 2018. 9 common reasons for rejection. Wiley: Discover the Future of Research: https://hub.wiley.com/community/exchanges/discover/blog/2018/2001/2031/2019-common-reasons-for-rejection.

Edwards MA, Roy S. 2016. Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environmental Engineering Science 34: 51-61.

Egler FE. 1986. "Physics envy" in ecology. Bulletin of the Ecological Society of America 67: 233-235.

Einstein A. 1920. Relativity: The Special and General Theory. Henry Holt and Company: 78 pages.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802-813.

Elliott LP, Brook BW. 2007. Revisiting Chamberlin: Multiple working hypotheses for the 21st century. BioScience 57: 608-614.

Elsevier. 2015. 5 ways you can ensure your manuscript avoids the desk reject pile. Elsevier Connect: https://http://www.elsevier.com/authors-update/story/publishing-tips/5-ways-you-can-ensure-your-manuscript-avoids-the-desk-reject-pile.

Fahrig L. 2003. Effects of habitat fragmentation on biodiversity. Annual Review of Ecology, Evolution, and Systematics 34: 487-515.

Glass DJ, Hall N. 2008. A brief history of the hypothesis. Cell 134: 378-381.

Golub T. 2010. Counterpoint: Data first. Nature 464: 679.

Guisan A, et al. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16: 1424-1435.

Gutiérrez IJ, D. TC, A. JJ, Weng-Keen W, M. SS, G. BM. 2014. Precipitation and winter temperature predict long-term range-scale abundance changes in Western North American birds. Global Change Biology 20: 3351-3364.

Hansen MC, et al. 2013. High-resolution global maps of 21st-century forest cover change. Science 342: 850-853.

Houlahan JE, McKinney ST, Anderson TM, McGill BJ. 2017. The priority of prediction in ecological understanding. Oikos 126: 1-7.

Hutson M. 2018. AI researchers allege that machine learning is alchemy. Science Posted in: Technology May 3, 2018.: doi:10.1126/science.aau0577.

Lack D. 1954. The Natural Regulation of Animal Numbers. Oxford: Clarendon Press.

Landy F. 1986. Stamp collecting versus science - validation as hypothesis testing. American Psychologist 41: 1183-1192.

Loehle C. 1987. Hypothesis testing in ecology: Psychological aspects and the importance of theory maturation. The Quarterly Review of Biology 62: 397-409.

May RM. 1981. The role of theory in ecology. American Zoologist 21: 903-910.

Mayr E. 2002. What Evolution Is. New York , USA: Basic Books.

Milner S. 2018. Newton didn't frame hypotheses. Why should we? Real Clear Science: Posted in: Physics Today April 25,2018. DOI:2010.1063/PT.2016.2013.20180424a.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853.

O'Neill RV, Johnson AR, King AW. 1989. A hierarchical framework for the analysis of scale. Landscape Ecology 3: 193-205.

Osborne Patrick E, Foody Giles M, Suárez-Seoane S. 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. Diversity and Distributions 13: 313-323.

Peters RH. 1991. A Critique for Ecology: Cambridge University Press, Nature.

Pettorelli N, et al. 2017. Remote sensing in ecology and conservation: Three years on. Remote Sensing in Ecology and Conservation 3: 53-56.

Platt JR. 1964. Strong inference. Science 146: 347-353.

Popper K. 1959. The Logic of Scientific Discovery. New York, USA: Basic Books.

Quinn JF, Dunham AE. 1983. On hypothesis testing in ecology and evolution. The American Naturalist 122: 602-617.

Ranehill E, Dreber A, Johannesson M, Leiberg S, Sul S, Weber RA. 2015. Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. Psychological Science 26: 653-656.

Romesburg HC. 1981. Wildlife science: Gaining reliable knowledge. The Journal of Wildlife Management 45: 293-313.

Rosen J. 2016. Research protocols: a forest of hypotheses. Nature 536: 239-241.

Schimel J. 2011. Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded: Oxford University Press.

Sells S, Bassing S, Barker K, C. Forshee S, C. Keever A, W. Goerz J, S. Mitchell M. 2018. Increased scientific rigor will improve reliability of research and effectiveness of management. Journal of Wildlife Management 82: 485-494.

Smaldino PE, McElreath R. 2016. The natural selection of bad science. Royal Society Open Science 3: Published 21 September 2016. DOI: 2010.1098/rsos.160384.

Stafleu F. 1971. Lamarck: The birth of biology. Taxon 20: 397-442.

Sutherland WJ, Spiegelhalter D, Burgman MA. 2013. Twenty tips for interpreting scientific claims. Nature 503: 335-337.

Symes LB, Serrell N, Ayres MP. 2015. A pactical guide for mentoring scientific inquiry. The Bulletin of the Ecological Society of America 96: 352-367.

Tewksbury JJ, et al. 2014. Natural history's place in science and society. BioScience 64: 300-310.

Thompson JN. 2005. The Geographic Mosaic of Coevolution. Chicago, USA: University of Chicago Press.

van Dijk D, Manor O, Carey LB. 2014. Publication metrics and success on the academic job market. Current Biology 24: R516-R517.

Voosen P. 2017. The AI detectives. Science 357: 22-27.

Weinberg R. 2010. Point: Hypotheses first. Nature 464: 678.

Wilson KA, McBride MF, Bode M, Possingham HP. 2006. Prioritizing global conservation efforts. Nature 440: 337.

Wolff JO, Krebs CJ. 2008. Hypothesis testing and the scientific method revisited. Acta Zoologica Sinica 54: 383-386.