OPEN ACCESS

CrossMark

Editorial

Introducing *Hypothesis Descriptions*

Daniel Mietchen[‡,§,|,¶,#], Jonathan M. Jeschke[‡,§,|], Tina Heger[‡,§,|,¤]

‡ Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
§ Freie Universität Berlin, Berlin, Germany
| Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany
¶ Institute for Globally Distributed Open Research and Education (IGDORE), Jena, Germany
# Ronin Institute of Independent Scholarship, Montclair, United States of America
¤ Technische Universität München, Munich, Germany

Corresponding author: Daniel Mietchen (daniel.mietchen@ronininstitute.org),
Jonathan M. Jeschke (jonathan.jeschke@fu-berlin.de),
Tina Heger (t.heger@wzw.tum.de)

Reviewable   v 1

Abstract

Hypotheses play a central role in the scientific process, yet the way they are introduced often leaves much room for interpretation, which makes it difficult to use them later on: to study and test them, to delineate their scope and to explore the relationships they have to other hypotheses or concepts, to datasets, methodologies or other resources. Here, we introduce a new article type in RIO that is dedicated to them: *Hypothesis Descriptions*. Such articles combine a specific verbal definition of a hypothesis with a concise description of its components and provide pointers to prior work as well as alignments with formal ways of knowledge representation, optionally including relevant nanopublications. With this format, we aim to facilitate the study of hypotheses in and of themselves, to improve their testability along with the documentation and interpretability of such tests, and to stimulate efforts towards standardization and automation in this space.

Keywords

formalized hypotheses, nanopublications, manuscript types

Motivation: Why a new article type for describing hypotheses?

A hypothesis is "[a]n assumption that

- is based on a formalized or nonformalized theoretical model of the real world and
- can deliver one or more testable predictions" (Heger et al. 2020, after Giere et al. 2006).

Hypotheses can arise at any step in a research cycle or even beyond, e.g. while observing a phenomenon or responding to a question, while exploring theoretical approaches to a problem, while reading or writing a manuscript, patent or proposal, while designing a data acquisition workflow, while curating, interpreting or integrating data or samples, or while incorporating new bits of information into an existing body of knowledge. Traditionally, few of these steps would be published on their own, and the publications resulting from a given research process may or may not contain all the hypotheses generated, explored or otherwise entertained on the way. Besides formal publications, there are various other channels through which hypotheses might enter scholarly discourse, including lectures or personal communications. Those hypotheses that were never communicated will essentially be forgotten, though any one of them might well be conceived independently by others, some of whom might eventually communicate them.

Many ways have been used to include a hypothesis - new or otherwise - in a publication. For instance, it could be in the title, which could be explicit or less so. Apart from the title, the hypothesis might be in any part of the publication and spelled out in detail. Across multiple publications, the same hypothesis (by any measure of sameness) might be referred to by one or more names (or even none), and the structures of different hypotheses might exhibit varying degrees of similarity. While hypotheses typically originate from a limited context, much of their appeal is in the extrapolation to new contexts, and much of their usage involves the delineation of their scope as well as aggregation of insights gathered from the study of multiple hypotheses. All of this could in principle be standardized.

On that basis, some research fields have developed practices that formalize the ways in which hypotheses (or certain aspects or variants thereof) are being stated, structured, delineated in scope, tested, or referred to. For instance, in formal logic and other branches of mathematics, conjectures are mathematical statements without a known proof, while theorems are mathematical statements that have been proven, and tools like lemmas and proof assistants can help in formalizing such statements while advancing from conjectures and proven lemmas to proofs of theorems (cf. Geuvers (2009)). This has matured to the point that conjectures, lemmas and theorems can be stated in a fully machine-actionable fashion (Buzzard 2020), that certain types of proofs can be generated (e.g. Nipkow (2001)) and verified (e.g. Avigad et al. (2007), Gonthier et al. (2013)) automatically, and that such automated systems routinely assist in both research (e.g. Cristiá and Rossi (2020)) and teaching (e.g. Villadsen et al. (2022)) within known theoretical limits (cf. Gödel (1930), From (2022)).

In other fields, formalization of hypotheses takes other forms. For instance, works in biological taxonomy typically contain taxon treatments (e.g. Douglas et al. 2023). These are sections that express - using highly standardized language in a highly standardized format (Agosti and Egloff 2009) - a hypothesis about how to delineate one taxon (e.g. a species, genus or subtribe) from others based on a set of materials and characteristics (cf. Härlin (2005), Kõljalg et al. (2020)). As these sets of materials (typically specimens or molecular sequences) and characteristics evolve, the need for taxon revisions may arise, i.e. modifications of the original taxon hypotheses, including potentially the creation of new ones (e.g. Srisonchai et al. 2018).

There are also fields in which there is less of a formal framework - if any - for expressing and handling hypotheses, which renders it more difficult to find pertinent hypotheses and work with them, including in automated fashions. This is the case, for instance, in some branches of ecology like invasion biology, where efforts are ongoing to map the landscape of existing hypotheses (Jeschke and Heger 2018, Enders et al. 2020, Jeschke et al. 2021). Such efforts would benefit from a more standardized approach to sharing hypotheses.

What could this look like? In short, the verbal definition of a hypothesis is translated into formulaic language, and that formalization of the hypothesis is then linked to existing knowledge by way of nanopublications annotated with standard identifiers (Bucur et al. 2023). Nanopublications are machine-readable assertions published in a standardized fashion and together with contextual and qualifying information, provenance and publication metadata (Groth et al. 2010). They are typically expressed via so-called semantic triples that combine three pieces of information and follow the basic form 'subject-relationship-object', with each element of the triple defined in a way that minimizes ambiguity. The current implementation of nanopublications in RIO is aligned with that for biodiversity publishing (cf. Penev et al. (2023)).

For instance, one of the assertions contained in Rodda and Savidge (2007) is that the brown tree snake (*Boiga irregularis*) is invasive to the Pacific island of Guam. In machine-friendly terms, this could be expressed with the following triple: "*Boiga irregularis*" as the subject, "Guam" as the object and "invasive to" as the relationship between subject and object. To further assist disambiguation, each of these three components would be expressed using suitable identifiers (e.g. the Wikidata identifiers Q900781 for "*Boiga irregularis*", Q16635 for "Guam" and P5588 for "invasive to") that point to additional pieces of information in the broader web of knowledge. The nanopublication approach can be applied to many different kinds of information, and the workflows we are establishing here for hypotheses, albeit demonstrated with an example from invasion biology, are applicable across many domains.

RIO is about communicating the research process all along the research cycle (cf. Mietchen et al. (2015)) and facilitating engagement with it (cf. Mietchen et al. (2021)). By providing a lightweight framework for expressing hypotheses in a standardized way useful to both humans and machines, we hope to support and encourage engagement with hypotheses in the context of a diverse set of research cycles.

Hypothesis Description manuscripts in RIO might well become a dedicated manuscript type eventually, but as long as the hypothesis-related workflows are still being ironed out, we suggest to use the existing manuscript type for *Research Idea* instead, as we have done in the example described below.

Any *Hypothesis Description* article should only have one target hypothesis, so as to avoid ambiguity and to facilitate the study of that particular hypothesis.

In the following, we will briefly outline the structure that we propose for *Hypothesis Descriptions*, provide preliminary instructions and an example as well as some further contextualization.

Components of a *Hypothesis description* article

In this section, we introduce the initial structure of a *Hypothesis description*. This structure is also represented in the *Hypothesis Description* template (Heger et al. 2024a) as well as in the example outlined in the next section. Community feedback is invited on all of these aspects.

In terms of front matter (title, authors, abstract, keywords, ethics, funding etc.), *Hypothesis Description* manuscripts will be handled mostly like any other manuscript, the exception being that the title should be prefixed with "Hypothesis Description" (not italicized) and otherwise just contain the name of the hypothesis in question.

We propose the following sections for the body of a *Hypothesis Description* manuscript (mandatory ones are bolded):

- an **introduction section** that provides context for the hypothesis, e.g. historic background for the hypothesis itself or for the concepts or relationships it contains;
- a **section with general information** about the hypothesis in question, about relevant research fields, literature and related hypotheses or relevant identifiers;
- a **section with one or more verbal definition(s)** of the target hypothesis;
- a section with a formalized representation of the target hypothesis and potentially its variants;
- an outlook section that could outline, for instance, foreseeable developments that build on the formal description of the target hypothesis, or suggest some ways of testing it;
- an acknowledgement section;
- a nanopublications section;
- a **reference section**.

The template provides instructions for each of the sections. The outlook section is optional, and for now, the machine-friendly version is too, since the workflows for that are still being developed.

An example: the Enemy Release Hypothesis

To illustrate how such *Hypothesis Description* papers can look like in practice, we accompany this editorial with an example (Heger et al. 2024b) from invasion biology that covers the Enemy Release Hypothesis (ERH). The ERH posits that when a given focal species is introduced into an area outside of its native range, there is a certain likelihood that its enemy species (or at least some of its enemies) will not be present in its non-native range, which would in turn increase the likelihood that the focal species becomes invasive in the new range.

Various textual definitions of the ERH have been stated in different scientific papers over time. They all relate to this overall idea, but differ slightly in their phrasing. Listing these differing definitions in a Hypothesis Description paper (see Table 1 in Heger et al. 2024b) can be useful for highlighting the differences, thus allowing researchers to make informed choices concerning which of them to use or to refer to. For example, taking a closer look at the exact formulation of the definitions reveals slight differences in specificity. Some definitions refer to 'invasion success' as the consequence of enemy release (e.g. Jeschke et al. (2012), Enders et al. (2018)), whereas others refer to 'increase in abundance and distribution' of the invader (Keane 2002, Daly et al. 2023).

The overall idea behind the ERH in fact is a rather complex mechanism, consisting of several elements: First, the process of transportation to a new area outside of the native range could lead to the situation that some enemies are 'left behind'. This is especially likely for those enemies that are specialized on the focal species. Second, the hypothetically reduced pressure by enemies in the new range could lead to a better performance of the invader. This complexity of the idea has led to the suggestion of naming a set of sub-hypotheses for the ERH (cf. Jeschke and Heger (2018) and Heger et al. (2020)). With the option to include formalized representations of hypothesis variants, a Hypothesis Description paper offers the opportunity to spell out these different elements even more explicitly. For instance, the variants could differ subtly or substantially in the way they specify the ecological context (e.g. terrestrial, freshwater or marine, pelagic or coastal), the focal group (e.g. grasses, rodents or one particular species of fish), the nature of the focal group's interactions with its enemies (e.g. host-parasite or predator-prey, specialist or generalist predator), or the mechanisms by which the release from enemies can lead to invasion success (e.g. whether it would primarily affect the establishment or spread stages of an invasion, and how).

Such a detailed definition of a hypothesis can be especially powerful when combined with the use of a controlled vocabulary, because this allows linking the used terms to definitions, while at the same time enhancing machine-actionability.

Both the listing of existing definitions and the formalized representation of hypothesis variants can enhance accuracy of scientific discussions around the respective hypothesis, and can allow a more reliable mapping of empirical evidence or experimental designs to hypothesized relationships. Likewise, meta-analyses that aggregate evidence from multipe

tests would profit from such more explicit and formalized definitions, because this would decrease the likelihood of misinterpretations and wrong assignments.

Integrating *Hypothesis Descriptions* into scientific workflows

Hypotheses can have a number of roles in scientific workflows. For instance, they can explain existing data or make predictions where data are missing. In principle, they can also be used to browse the scholarly literature by hypothesis (e.g. to see for which species, habitats or locations the Enemy Release Hypothesis has been tested), and a basic implementation of that is available via Scholia (Nielsen et al. 2017), e.g. for the Enemy Release Hypothesis*[1].

If hypotheses were properly integrated with metadata about the research questions, methods and datasets relevant to them, it would be easier to keep track of which hypotheses have been put to a test, for which ones (or which aspects or variants of them) confirmatory evidence is accumulating or lacking, and how the evidence regarding one hypothesis might affect others.

For these various roles of hypotheses, it is important that their respective scope is clearly delineated and communicated. We think that turning hypotheses into research objects in and of themselves that can be published, cited and versioned is a good step in this direction.

As is current practice for Research Ideas and other article types, *Hypothesis Description* manuscripts will be subject to peer review, which shall include associated nanopublications. Just like any other RIO publication, *Hypothesis Description* manuscripts can be updated, resulting in a new version that has its own unique identifier.

Conclusion

We believe that Hypothesis Description papers can have several merits. First, disclosing the different meanings of hypotheses and formalizing them as suggested above can enhance theory development. For example, Heger (2022) suggested representing the Enemy Release Hypothesis as a causal network graph. Future work can build on this and integrate the different causal variants of the ERH in a larger causal network describing hypothesized mechanisms of biological invasions. Second, linking explicit definitions and formalizations to entries in machine-readable resources like Wikidata will allow for assistance from automated tools when working with hypotheses. We therefore encourage others to publish similar papers on other hypotheses, in invasion biology and other domains.

Funding program

Conflicts of interest

The authors have declared that no competing interests exist.
Disclaimer: This article is (co-)authored by any of the Editors-in-Chief, Managing Editors or their deputies in this journal.

References

- Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2 (1). https://doi.org/10.1186/1756-0500-2-53
- Avigad J, Donnelly K, Gray D, Raff P (2007) A formally verified proof of the prime number theorem. ACM Transactions on Computational Logic 9 (1). https://doi.org/10.1145/1297658.1297660
- Bucur C, Kuhn T, Ceolin D, van Ossenbruggen J (2023) Nanopublication-based semantic publishing and reviewing: a field study with formalization papers. PeerJ Computer Science 9 https://doi.org/10.7717/peerj-cs.1159
- Buzzard K (2020) Proving Theorems with Computers. Notices of the American Mathematical Society 67 (11). https://doi.org/10.1090/noti2177
- Cristiá M, Rossi G (2020) Automated Proof of Bell–LaPadula Security Properties. Journal of Automated Reasoning 65 (4): 463-478. https://doi.org/10.1007/s10817-020-09577-6
- Daly E, Chabrerie O, Massol F, Facon B, Hess MM, Tasiemski A, Grandjean F, Chauvat M, Viard F, Forey E, Folcher L, Buisson E, Boivin T, Baltora-Rosset S, Ulmer R, Gibert P, Thiébaut G, Pantel J, Heger T, Richardson D, Renault D (2023) A synthesis of biological invasion hypotheses associated with the introduction–naturalisation–invasion continuum. Oikos 2023 (5). https://doi.org/10.1111/oik.09645
- Douglas H, Smith T, Bouchard P (2023) Palaearctic leaf beetle *Chrysolina fastuosa* (Coleoptera, Chrysomelidae, Chrysomelinae) new to North America. Biodiversity Data Journal 11 https://doi.org/10.3897/bdj.11.e103261
- Enders M, Hütt M, Jeschke J (2018) Drawing a map of invasion biology based on a network of hypotheses. Ecosphere 9 (3). https://doi.org/10.1002/ecs2.2146
- Enders M, Havemann F, Ruland F, Bernard-Verdier M, Catford J, Gómez-Aparicio L, Haider S, Heger T, Kueffer C, Kühn I, Meyerson L, Musseau C, Novoa A, Ricciardi A, Sagouis A, Schittko C, Strayer D, Vilà M, Essl F, Hulme P, van Kleunen M, Kumschick S, Lockwood J, Mabey A, McGeoch M, Palma E, Pyšek P, Saul W, Yannelli F, Jeschke J (2020) A conceptual map of invasion biology: Integrating hypotheses into a consensus network. Global Ecology and Biogeography 29 (6): 978-991. https://doi.org/10.1111/geb.13082

- From AH (2022) A Succinct Formalization of the Completeness of First-Order Logic. Proceedings of the 27th International Conference on Types for Proofs and Programs (TYPES 2021) https://doi.org/10.4230/lipics.types.2021.8
- Geuvers H (2009) Proof assistants: History, ideas and future. Sadhana 34 (1): 3-25. https://doi.org/10.1007/s12046-009-0001-5
- Giere R, Bickle J, Mauldin R (2006) Understanding scientific reasoning. 5th. Thomson Wadsworth, Belmont. [ISBN 978-0155063266]
- Gödel K (1930) Die Vollständigkeit der Axiome des logischen Funktionenkalküls. Monatshefte für Mathematik und Physik 37 (1): 349-360. https://doi.org/10.1007/bf01696781
- Gonthier G, Asperti A, Avigad J, Bertot Y, Cohen C, Garillot F, Le Roux S, Mahboubi A, O'Connor R, Ould Biha S, Pasca I, Rideau L, Solovyev A, Tassi E, Théry L (2013) A Machine-Checked Proof of the Odd Order Theorem. Interactive Theorem Proving163-179. https://doi.org/10.1007/978-3-642-39634-2_14
- Groth P, Gibson A, Velterop J (2010) The anatomy of a nanopublication. Inf. Serv. Use 30 (1-2). URL: https://dl.acm.org/doi/10.5555/1883685.1883690
- Härlin M (2005) Taxon names as paradigms: the structure of nomenclatural revolutions. Cladistics 19 (2): 138-143. https://doi.org/10.1111/j.1096-0031.2003.tb00302.x
- Heger T, Aguilar-Trigueros CA, Bartram I, Braga RR, Dietl GP, Enders M, Gibson DJ, Gómez-Aparicio L, Gras P, Jax K, Lokatis S, Lortie CJ, Mupepele A, Schindler S, Starrfelt J, Synodinos AD, Jeschke JM (2020) The Hierarchy-of-Hypotheses Approach: A Synthesis Method for Enhancing Theory Development in Ecology and Evolution. BioScience 71 (4): 337-349. https://doi.org/10.1093/biosci/biaa130
- Heger T (2022) What are ecological mechanisms? Suggestions for a fine-grained description of causal mechanisms in invasion ecology. Biology & Philosophy 37 (2). https://doi.org/10.1007/s10539-022-09838-1
- Heger T, Zarrieß S, Algergawy A, Jeschke J, König-Ries B (2022) INAS: Interactive Argumentation Support for the Scientific Domain of Invasion Biology. Research Ideas and Outcomes 8 https://doi.org/10.3897/rio.8.e80457
- Heger T, Mietchen D, Jeschke J (2024a) Template for a Hypothesis Description paper. Research Ideas and Outcomes 10 https://doi.org/10.3897/rio.10.e119808
- Heger T, Jeschke J, Bernard-Verdier M, Musseau C, Mietchen D (2024b) Hypothesis Description: Enemy Release Hypothesis. Research Ideas and Outcomes 10 https://doi.org/10.3897/rio.10.e107393
- Jeschke J, Gómez Aparicio L, Haider S, Heger T, Lortie C, Pyšek P, Strayer D (2012) Support for major hypotheses in invasion biology is uneven and declining. NeoBiota 14: 1-20. https://doi.org/10.3897/neobiota.14.3435
- Jeschke J, Heger T (2018) Invasion biology: hypotheses and evidence. CABI, Wallingford. https://doi.org/10.1079/9781780647647.0000
- Jeschke J, Heger T, Kraker P, Schramm M, Kittel C, Mietchen D (2021) Towards an open, zoomable atlas for invasion science and beyond. NeoBiota 68: 5-18. https://doi.org/10.3897/neobiota.68.66685
- Keane R (2002) Exotic plant invasions and the enemy release hypothesis. Trends in Ecology & Evolution 17 (4): 164-170. https://doi.org/10.1016/s0169-5347(02)02499-0
- Kõljalg U, Nilsson H, Schigel D, Tedersoo L, Larsson K, May T, Taylor AS, Jeppesen TS, Frøslev TG, Lindahl B, Põldmaa K, Saar I, Suija A, Savchenko A, Yatsiuk I, Adojaan K, Ivanov F, Piirmann T, Pöhönen R, Zirk A, Abarenkov K (2020) The Taxon Hypothesis

Paradigm—On the Unambiguous Detection and Communication of Taxa. Microorganisms 8 (12). https://doi.org/10.3390/microorganisms8121910
- Mietchen D, Mounce R, Penev L (2015) Publishing the research process. Research Ideas and Outcomes 1 https://doi.org/10.3897/rio.1.e7547
- Mietchen D, Penev L, Georgiev T, Ovcharova B, Kostadinova I (2021) Open science in practice: 300 published research ideas and outcomes illustrate how RIO Journal facilitates engagement with the research process. Research Ideas and Outcomes 7 https://doi.org/10.3897/rio.7.e68595
- Nielsen FÅ, Mietchen D, Willighagen E (2017) Scholia, Scientometrics and Wikidata. Lecture Notes in Computer Science237-259. https://doi.org/10.1007/978-3-319-70407-4_36
- Nipkow T (2001) More Church–Rosser Proofs. Journal of Automated Reasoning 26 (1): 51-66. https://doi.org/10.1023/a:1006496715975
- Penev L, Kuhn T, Pyle R, Mietchen D, Demirov S, Boyadzhieva I, Georgiev T (2023) Nanopublications for Biodiversity Go Live. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.110725
- Rodda G, Savidge J (2007) Biology and Impacts of Pacific Island Invasive Species. 2. Boiga irregularis, the Brown Tree Snake (Reptilia: Colubridae). Pacific Science 61 (3): 307-324. https://doi.org/10.2984/1534-6188(2007)61[307:baiopi]2.0.co;2
- Srisonchai R, Enghoff H, Likhitrakarn N, Panha S (2018) A revision of dragon millipedes I: genus *Desmoxytes* Chamberlin, 1923, with the description of eight new species (Diplopoda, Polydesmida, Paradoxosomatidae). ZooKeys 761: 1-177. https://doi.org/10.3897/zookeys.761.24214
- Villadsen J, From AH, Blackburn P (2022) Teaching Intuitionistic and Classical Propositional Logic Using Isabelle. Electronic Proceedings in Theoretical Computer Science 354: 71-85. https://doi.org/10.4204/eptcs.354.6

Endnotes

[*1] See https://scholia.toolforge.org/topic/Q85759287