CellPress
Current Biology
Magazine

Primer

Open science

Michael G. Bertram[1,2,3,*],
Josefin Sundin[4],
Dominique G. Roche[5,6],
Alfredo Sánchez-Tójar[7],
Eli S. J. Thoré[1,8,9] and Tomas Brodin[1]

The term 'open science' refers to a range of methods, tools, platforms and practices that aim to make scientific research more accessible, transparent, reproducible and reliable. This includes, for example, sharing code, data and research materials, embracing new publishing formats such as registered reports and preprints, pursuing replication studies and reanalyses, optimising statistical approaches to improve evidence assessment and re-evaluating institutional incentives. The ongoing shift towards open science practices is partly due to mounting evidence that studies across disciplines suffer from biases, underpowered designs and irreproducible or non-replicable results. It also stems from a general desire amongst many researchers to reduce hyper-competitivity in science and instead promote collaborative research that benefits science and society.

Core principles of open science

Open science aims to increase the accessibility, transparency, reliability and (re)usability of scholarly outputs (Figure 1). In addition, open science aligns with the principles of equity, diversity and inclusion, with the aim of opening the creation, evaluation and communication of scientific knowledge to marginalised scholars and societal actors beyond the traditional scientific community. This can be achieved by implementing open science practices throughout the research lifecycle, from study design to publication, and beyond.

Open methods involve transparently describing and publicly sharing protocols, materials and analytical code. They help validate research findings by facilitating reproducibility (same methods, same data) and replicability (same methods, different data). Preregistration — the process of archiving a study's hypotheses, design and methods prior to data collection or analysis — increases reliability by compelling authors to carefully design their study in advance. Early feedback on a preregistered methodology can be requested from experts or obtained via formal peer-review through a specific article type called 'registered reports'.

Open data are created by archiving research data in the open domain, ideally a public repository, and thereby making them freely accessible. Researchers sharing open data should follow the 'FAIR' sharing principles, which aim to make open data findable, accessible, interoperable and reusable. When datasets contain sensitive information, such as participant identifiers or the location of threatened species, authors can share anonymised or synthetic data, or publish only the associated metadata, making sensitive data findable but not readily accessible.

Open access publication and *open peer-review* make scientific articles and their peer-review history publicly available without restrictions, enhancing accessibility and transparency. Typically, scientific journals charge fees to academic libraries for accessing journals or to authors for publishing open access. However, authors can now readily archive accepted — but not copy-edited — manuscripts on preprint servers ('green open access') or publish their work in journals that are both free-to-read and free-to-publish ('diamond open access').

Open educational resources are materials (e.g. textbooks, courses) that are released under an open license, allowing them to be freely accessed, retained, remixed, revised, reused and redistributed (the 'five Rs') for teaching, learning and research. As with the practices listed above, open educational

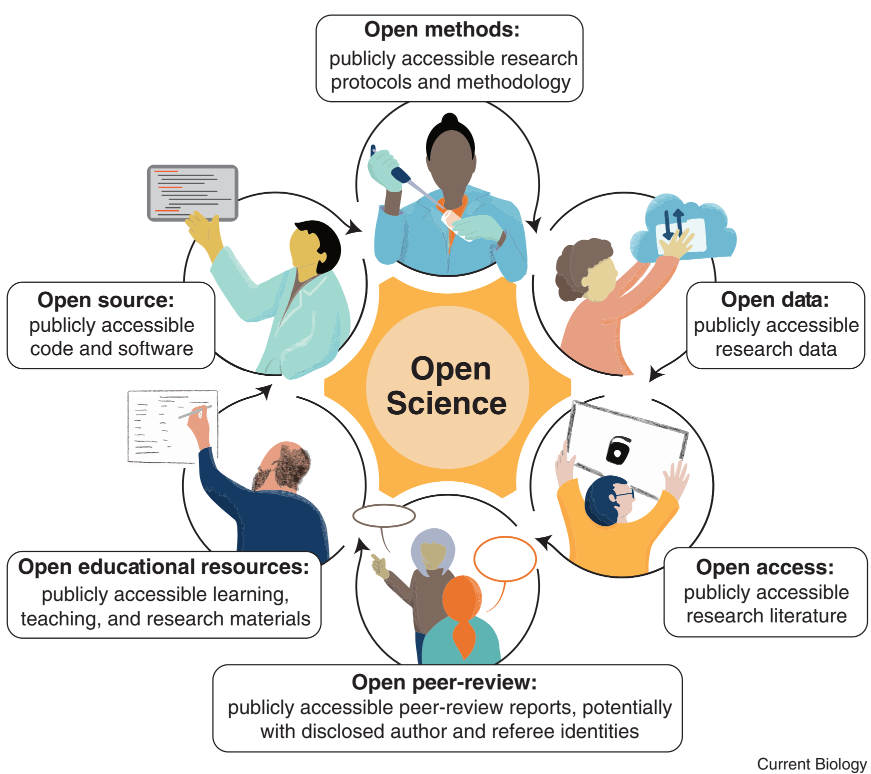

Figure 1. The core principles of open science.
Open science embraces principles to make science accessible, transparent and reliable, and thereby avoid common threats to reproducibility and replicability (e.g. questionable research practices, confirmation bias). The authors are grateful to Bertsy Goic (DrawInScience) for assisting with the creation of visuals.

CellPress

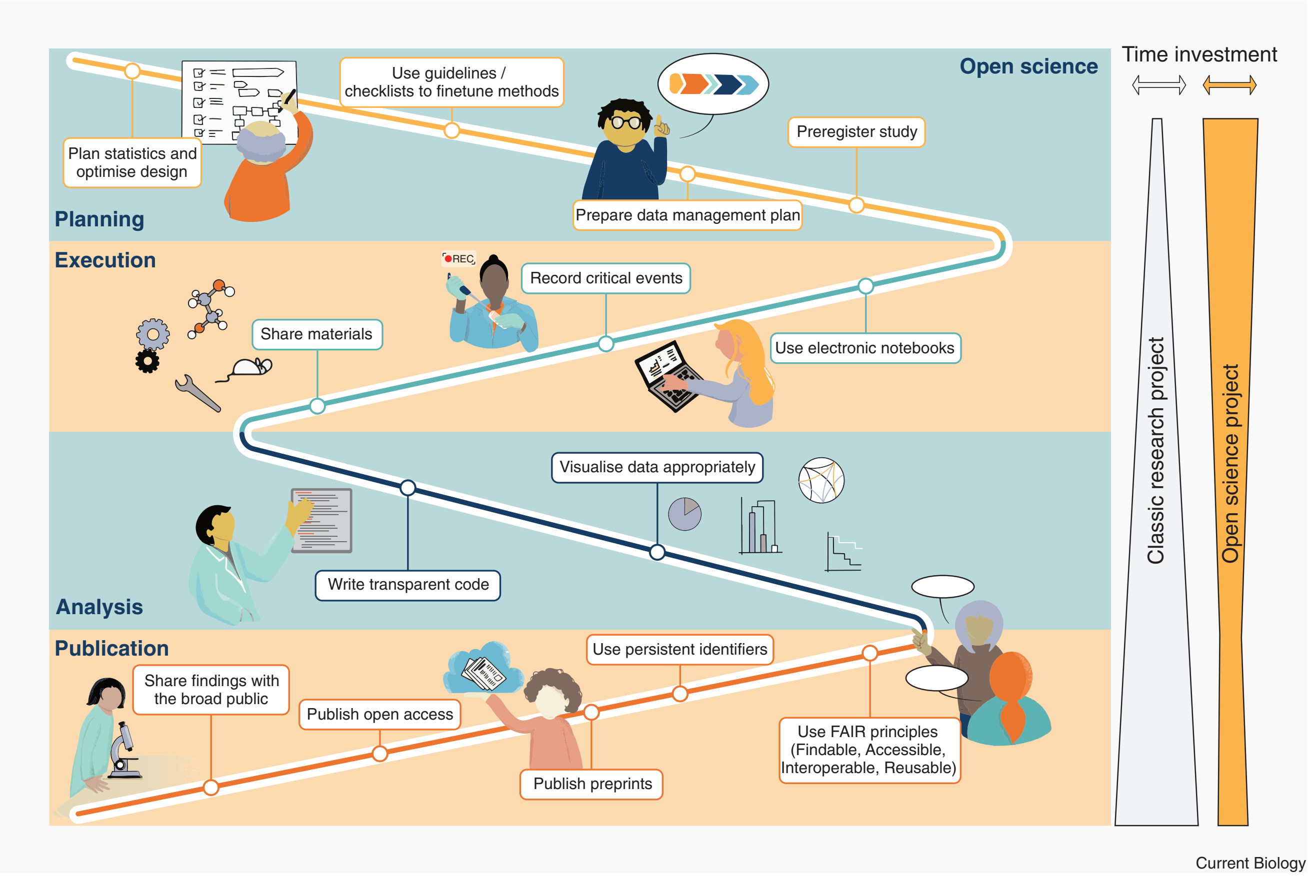

Figure 2. Examples of open science practices that can be implemented throughout the research lifecycle.
Implementing open science practices at various stages of a project (planning, execution, analysis and publication) helps maximise the impact of science. The different practices within each stage are not necessarily in strict chronological order. Typically, engaging in open science practices requires a greater time investment in the early stages of a research project (front-loaded) compared to classic research projects, which are often rear-loaded. Adapted from Diederich et al. (2022). The authors are grateful to Bertsy Goic (DrawInScience) for assisting with the creation of visuals.

resources generate broad and equitable benefits for scientists and society. Likewise, *open source software*, where the source code is freely available and the terms of use allow dissemination and adaptation, empowers a wider community of researchers to contribute and build upon existing work.

Barriers to open science

Open science has grown from grassroot origins towards becoming a policy priority for academic institutions and a standard working method for many researchers. However, widespread participation in open science is hampered by several barriers and constraints, including: a lack of awareness, knowledge and incentives; limitations related to resources and technology; and, intellectual property concerns, as well as social and cultural factors.

First, researchers may not be familiar with open science practices and how to implement them at different stages of a project's lifecycle. For example, there are many tools to help researchers share their materials, data and results, each with their own advantages and shortcomings. When scientists are not trained to navigate these options, it can be challenging to know which tool to use and how to make the best use of it. Lack of awareness and training stems, at least in part, from a lack of perceived benefit among researchers and weak institutional incentives. Open science practices are often not well-aligned with classic incentives for career advancement in academia, which often focus on prestige and productivity.

Second, researchers may be reluctant, or even unable, to embrace open science practices due to time and financial constraints. For example, writing a preregistration or submitting a registered report, pursuing replications and preparing datasets for open archiving all entail additional commitment. Likewise, the cost of publishing open access, particularly in prestigious journals, can be prohibitive. Lastly, participation in open science is impeded by technological barriers, such as a lack of access to infrastructure (e.g. sufficient internet access and speed to upload large files) or to (free-to-use) platforms for sharing research outputs.

Third, researchers may fear that publicly sharing research ideas and outputs could favour competitors, and that research data could be misused or misinterpreted by third parties. Sensitive data, as well as data and results related to innovations or patents, require careful consideration and adequate protection when deciding on public release. Researchers may also feel insecure about their scientific approaches and fear embarrassment or retaliation — for example, when signing an open peer-review. Furthermore, open science can be hampered by language barriers, particularly for non-native English speakers. Typically, all of these constraints and barriers to open science are intricately interconnected and can depend on a researcher's career stage and stability as well as gender and ethnic identity.

Consequences of not engaging in open science

Reliability and reproducibility are core tenets of the scientific method but are difficult to guarantee, or even assess, when researchers do not communicate their work transparently. These concerns are tightly connected: if materials and methods are not openly available and transparently described, the quality and robustness of a study cannot be fully assessed by the research community. A study may also be difficult or impossible to replicate. There can be many reasons for replication failure, such as honest error, bias (confirmation and publication bias), HARKing ('hypothesising after the results are known', or presenting unexpected results as *a priori* hypotheses), P-hacking (manipulating data or statistical analyses until statistically non-significant results become significant) and scientific misconduct (data falsification and fabrication). Open science practices cannot eradicate these issues but contribute

Table 1. Tools for open science practice.

Stage	Open science practice	Tools	Description
Planning	Use guidelines/checklists to fine tune methods	PREPARE	Guidelines to reduce waste, promote animal alternatives, and increase reproducibility of animal testing
		ARRIVE	Checklist of recommendations to improve the reporting of animal research
		CRED	Checklist of recommendations to improve the conducting and reporting of ecotoxicological research
	Plan statistics and optimise design	G*Power	Tool to compute statistical power
		InVivoStat	R-based tool for the statistical planning of animal research
	Preregister study	Registered Reports	Resources for publishing registered reports
	Prepare data management plan	DMPTool	Tool to create data management plans
Execution	Record critical events	protocols.io	Platform to share methods
		Open Science Framework	Open source project management tool to share methods, preregistrations and preprints
		Bio-protocol	Peer-reviewed journal dedicated to the publication of methods
		CIRS-LAS	Reporting system to record and analyse critical events in animal experimental research
	Use electronic notebooks	eLabFTW/open BIS	Electronic lab notebooks
	Share materials	Anishare	Database to share animals to reduce the number of animals in research
		AniMatch	Platform for exchange of organs and tissues of laboratory animals
		Addgene	Platform to share plasmids
Analysis	Visualise data	Interactive Dotplot	Tool to create interactive graphs
		PlotsOfData	Tool to visually and statistically compare experimental groups
		BoxPlotR	Tool to generate customised boxplot
	Write transparent code	R	Software environment for statistical computing and graphics
		Python	Programming language
		Jupyter Notebooks	Tool to arrange and share workflows across various programming language
	Share code	GitHub	Git hosting platform for managing code and tracking changes
		Zenodo/figshare/ Dryad	Public repositories

(*Table continued on next page*)

Table 1. *Continued.*

Stage	Open science practice	Tools	Description
Publication	Use FAIR principle	GoFAIR	Initiative to implement the FAIR data principles
	Use persistent identifiers	ORCID ID	Provides a persistent digital identifier to distinguish among researchers (Open Researcher and Contributor ID)
		Research Resource Identifiers	Portal to promote research resource identification, discovery, and reuse
	Publish preprints	arXiv	Preprint server for studies in various disciplines
		bioRxiv	Preprint server for studies in biology
		ChemRxiv	Preprint server for studies in chemistry
		EcoEvoRxiv	Preprint server for studies in ecology, evolution and conservation
		medRxiv	Preprint server for studies in medicine
	Publish open access	DOAJ	Platform to identify the open access policies of scientific journals (Directory of Open Access Journals)
		Sherpa Romeo platform	Platform to identify the open access policies of scientific journals
	Share findings with research community and the public	Twitter	Social media and social networking service
		Mastodon	Social media and social networking service
		LinkedIn	Business and employment-focused social media and networking service
		ResearchGate	Social media and social networking service for researchers

This list is not exhaustive but is intended as a starting point for researchers interested in adopting open science practices. Adapted from Diederich *et al.* (2022).

to lowering their occurrence and magnitude.

Open science is relevant to a broad range of stakeholders, including researchers, policymakers and the public. Without openness, most research outputs are inaccessible to stakeholders and the general public, which undermines public trust in science. Openness is not sufficient to ensure public trust, but the ability of peers and the public to scrutinise scientific claims increases their credibility. A lack of openness can also hinder the advancement of knowledge and limit the potential impact of scientific findings by hindering consensus building and integration across disciplines. Open science facilitates collaboration within and among disciplines, as well as a more efficient use of research resources in terms of funding, equipment, knowledge and time. Open science practices can also confer personal benefits to researchers. For instance, sharing research findings has been shown to increase citation rates, boost efficiency, foster collaboration and encourage community engagement.

Key resources for adopting open science practices

Open science means that transparency and sharing of research outputs are considered throughout the research lifecycle (Figure 2). Many tools exist to implement this across all stages of a study, from the early planning of a project, through its execution, data analysis and publication (Table 1).

Early sharing of information during the planning of a project enables peers to give *a priori* input on study design. First, there are many guidelines and checklists (e.g. PREPARE, ARRIVE, CRED) that can be consulted during the planning stage to optimise research methods, inform researchers about key methodological details to document and improve communication among collaborators. Second, rigorous statistical planning helps to optimise the design of an experiment, and can be facilitated by free-to-use tools to calculate sample size and statistical power (e.g. G*Power software, R packages such as InVivoStat). Study plans can also be preregistered, and many journals already offer to peer-review such preregistrations via registered reports. When such a report is provisionally accepted, the final study will be published regardless of its findings, actively combating publication bias. Rigorous planning of a study also includes preparing a data management plan for which there are several free tools (e.g. DMPTool) and checklists to comply with funder requirements and prevent common data management errors.

Many tools enable documenting the methodological details of a study. In addition to publishing methods in scientific manuscripts, detailed information can be shared and made more visible and citable (e.g. protocols.io, Open Science Framework). Alternatively, methods can be published in specific peer-reviewed journals, such as Bio-protocol. Furthermore, web-based reporting systems exist to record and analyse critical events during a study or potential failed attempts (e.g. Critical Incident Reporting System in Laboratory Animal Science [CIRS-LAS]). Notably, electronic laboratory notebooks can be used to document information throughout a study and make methods and data searchable and traceable. These can be fine-tuned to a researcher's specific needs and research area. Many electronic laboratory notebooks are commercially available or exist as open-source software (e.g. eLabFTW, open BIS). In addition, specific reagents, equipment, or other physical resources (e.g. animals, tissues) can be shared to promote reproducibility. Researchers who are willing to share materials can describe this in publications or use open-source software and databases to offer or seek materials (e.g. Anishare, AniMatch, Addgene).

Research data should be shared following the FAIR sharing principles as described in the GoFAIR guidelines. Moreover, various tools are freely available to help visualise data appropriately, including free web-based tools like Interactive Dotplot, PlotsOfData and BoxPlotR. Likewise, code should be made transparent and accessible. This is facilitated by free-to-use and open-source programming languages, such as R or Phython. Workflows across various programming languages can be arranged and shared using the free web-based tool Jupyter Notebooks. Data and code, including accompanying metadata and readme files, can be archived in open repositories (e.g. Zenodo, figshare, Dryad) and are citeable by means of persistent identifiers (commonly digital object identifier, DOI). Researchers who fear that their data or code could be used inappropriately can use embargoes, and some journals have anti-scooping policies. Open researcher and contributor ID (ORCID) and research resource identifiers can be used to correctly identify authors and research resources, respectively.

Full research manuscripts can be published on platforms that are freely accessible. Among others, the Directory of Open Access Journals (DOAJ) and the Sherpa Romeo platform allow users to identify the open access policies of scientific journals. Notably, manuscripts can be published as preprints on designated preprint servers (e.g. arXiv, bioRxiv, ChemRxiv, EcoEvoRxiv, medRxiv) before, but also after, formal publication in a scientific journal. This makes the study quickly available to peers and increases its visibility. Furthermore, research findings can be disseminated to peers and the wider public through various social media platforms (e.g. Twitter, Mastodon, LinkedIn), for example as narrative summaries, and through specialised platforms like ResearchGate.

Towards an open science community

Despite their benefits, open science practices can be challenging to implement and may appear overwhelming at first. To achieve an open science community, institutional support and incentives are needed at multiple levels that make open science possible (via suitable infrastructure), easy (via user-friendly processes), normative (through communities), rewarding (via incentives) and required (through policy) — as described in the Center for Open Science theory of research culture change model (www.cos.io/blog/strategy-for-culture-change).

Recent progress in open-source infrastructure and tools has made it easier for researchers to implement open science practices. Journals, funders and universities are changing their policies and incentives for researchers. However, even with these initiatives, normalising open science is only possible through communities of practice. Open science communities are bottom-up learning groups that allow researchers with different expertise levels to interact, support and learn from each other. Above all, these communities can help make open science normative by facilitating the ongoing development of open science infrastructure, innovation and policies. They also contribute to making open science practices more visible, which is a critical but sometimes neglected component in accelerating their adoption. To date, more than 200 grassroots open science networks exist worldwide, such as the ReproducibiliTea Journal Club Series, which helps researchers create local open science journal clubs and has already spread to 100 institutions across 22 countries. Another example is the framework for open and reproducible research training (FORRT), which aims to integrate principles of open and reproducible science into higher education.

Open science communities are often initiated by researchers in a bottom-up fashion. However, an increasing number of institutions support these initiatives by welcoming them, providing funding, identifying ambassadors within the institution and rewarding these efforts. Open science communities aim to include all researchers, but strategically targeting the early majority — individuals who are curious about open science practices, but have not yet adopted them — can facilitate the paradigm shift. Importantly, open science communities ideally extend beyond researchers and include students, staff, teachers and other stakeholders, and promote equity, diversity and inclusion. Initiatives such as the international network of open science and scholarship communities (INOSC) offer a starter kit to help researchers set up and foster a local open science community at their institution (www.startyourosc.com).

In sum, open science practices are not yet widely adopted but are nevertheless increasingly changing how science is conducted. This change is coming about not through top-down incentives and policies alone but rather in combination with bottom-up initiatives such as open science communities.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FURTHER READING

Allen, C., and Mehler, D.M.A. (2019). Open science challenges, benefits and tips in early career and beyond. PLoS Biol. *17*, e3000246.
Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., Fondberg, R., Heininga, V.E., Heunis, S., Koh, W.Q., Masselink, M., *et al.* (2021). Towards wide-scale adoption of open science practices: the role of open science communities. Sci. Public Policy *48*, 605–611.
Brandt, M.J., IJzerman, H., Dijksterhuis, A., Farach, F.J., Geller, J., Giner-Sorolla, R., Grange, J.A., Perugini, M., Spies, J.R., and van't Veer, A. (2014). The replication recipe: what makes for a convincing replication? J. Exp. Soc. Psychol. *50*, 217–224.

Current Biology
Magazine

Chambers, C.D. (2019). What's next for registered reports? Nature *573*, 187–189.
Diederich, K., Schmitt, K., Schwedhelm, P., Bert, B., and Heinl, C. (2022). A guide to open science practices for animal research. PLoS Biol. *20*, e3001810.
Gomes, D.G.E., Pottier, P., Crystal-Ornelas, R., Hudgins, E., Foroughirad, V., Sánchez-Reyes, L.L., Turba, R., Martinez, P.A., Moreau, D., Bertram, M.G., *et al.* (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. Proc. R. Soc. B *289*, 20221113.
Gownaris, N.J., Vermeir, K., Bittner, M.I., Gunawardena, L., Kaur-Ghumaan, S., Lepenies, R., Ntsefong, G.N., and Zakari, I.S. (2022). Barriers to full participation in the open science life cycle among early career researchers. Data Sci. J. *21*, 2.
Heise, V., Holman, C., Lo, H., Lyras, E.M., Adkins, M.C., Aquino, M.R.J., Bougioukas, K.I., Bray, K.O., Gajos, M., Guo, X. *et al.* (2023). Ten simple rules for implementing open and reproducible research practices after attending a training course. PLoS Comput. Biol. *19*, e1010750.
Higginson, A.D., and Munafò, M.R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. PLoS Biol. *14*, e2000995.
Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., and Ioannidis, J. (2017). A manifesto for reproducible science. Nat. Hum. Behav. *1*, 0021.
Muñoz-Tamayo, R., Nielsen, B.L., Gagaoua, M., Gondret, F., Krause, E.T., Morgavi, D.P., Olsson, I.A.S., Pastell, M., Taghipoor, M., Tedeschi, L. *et al.* (2022). Seven steps to enhance Open Science practices in animal science. PNAS Nexus *1*, pgac106.
O'Dea, R.E., Parker, T.H., Chee, Y.E., Culina, A., Drobniak, S.M., Duncan, D.H., Fidler, F., Gould, E., Ihle, M., Kelly, C.D., *et al.* (2021). Towards open, reliable, and transparent ecology and evolutionary biology. BMC Biol. *19*, 68.
Roche, D.G., O'Dea, R.E., Kerr, K.A., Rytwinski, T., Schuster, R., Nguyen, V.M., and Cooke, S.J. (2022). Closing the knowledge action gap in conservation with open science. Conserv. Biol. *36*, e13835.
UNESCO (2021). UNESCO Recommendation on Open Science. United Nations Educational, Scientific and Cultural Organization. Available online at https://unesdoc.unesco.org/ark:/48223/pf0000379949 (accessed February 14, 2023).

[1]Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden. [2]Department of Zoology, Stockholm University, Stockholm, Sweden. [3]School of Biological Sciences, Monash University, Melbourne, Australia. [4]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Drottningholm, Sweden. [5]Department of Biology, Carleton University, Ottawa, Canada. [6]Institut de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland. [7]Department of Evolutionary Biology, Bielefeld University, Germany. [8]Laboratory of Animal Ecology, Global Change and Sustainable Development, KU Leuven, Leuven, Belgium. [9]TRANSfarm - Science, Engineering & Technology Group, KU Leuven, Lovenjoel, Belgium.
*E-mail: michael.bertram@slu.se

Correspondence

Long-term spatial memory across large spatial scales in *Heliconius* butterflies

Priscila A. Moura[1], Fletcher J. Young[2,3], Monica Monllor[2], Marcio Z. Cardoso[1,4], and Stephen H. Montgomery[3,*]

Locating food in heterogeneous environments is a core survival challenge. The distribution of resources shapes foraging strategies, imposing demands on perception, learning and memory, and associated brain structures. Indeed, selection for foraging efficiency is linked to brain expansion in diverse taxa, from primates[1] to Hymenopterans[2]. Among butterflies, *Heliconius* have a unique dietary adaptation, actively collecting and feeding on pollen, providing a source of essential amino acids as adults, negating reproductive senescence and facilitating an extended longevity[3]. Several lines of evidence suggest that *Heliconius* learn the spatial location of pollen resources within an individual's home range[4], and spatial learning may be more pronounced at these large spatial scales. However, experimental evidence of spatial learning in *Heliconius*, or any other butterfly, is so far absent. We therefore tested the ability of *Heliconius* to learn the spatial location of food rewards at three ecologically-relevant spatial scales, representing multiple flowers on a single plant, multiple plants within a locality, and multiple localities. *Heliconius* were able to learn spatial information at all three scales, consistent with this ability being an important component of their natural foraging behaviour.

Heliconius establish 'traplines', foraging routes along which specific plants are repeatedly visited with high spatial and temporal regularity, initiated from a stable roosting site[3]. Traplines are highly individualistic, even among butterflies that share a roost, suggesting that memory is guiding their movements[5]. When translocated hundreds of meters, wild *Heliconius* (*erato* and *melpomene*) also quickly orientate towards, and return to, their site of origin[6]. However, many assumptions about spatial learning in *Heliconius* remain untested.

To experimentally validate spatial learning in *Heliconius* across spatial scales, we performed three experiments at increasing spatial distances. In the first experiment, freshly eclosed, insectary-reared *Heliconius erato phyllis* ($n = 44$) were trained to locate rewarding feeders on a 100 cm^2 4 x 4 grid of 3 cm flowers (Figure 1A). We trained butterflies to associate either *border* (Figure 1A, column A or D) or *centre* (B2/3, C2/3) flowers with a positive food reward. After five days of training, we recorded individual preference using empty feeders. Comparing individual preferences before and after training, we found that trained butterflies showed an increased preference for the position of positively rewarded flowers (Figure 1D; $\chi^2_1 = 4.908$, $p = 0.027$), with no difference in performance between *border* and *centre* groups ($\chi^2_1 = 0.097$, $p = 0.756$).

We subsequently increased the spatial scale by conducting a similar experiment in a ~3 m^2 two-armed maze (Figure 1B). Individuals ($n = 26$) were released centrally, and we recorded their naïve preference for foraging in the left or right arm. The butterflies were then trained to associate one arm with a positive reward before a final preference test. We again found an effect of training on foraging preference (Figure 1E; $\chi^2_1 = 22.949$, $p < 0.0001$). There was a significant interaction between the training effect and direction of the reinforced feeder ($\chi^2_1 = 11.875$, $p < 0.001$), with butterflies trained to the left showing greater fidelity to the learned spatial cue (Figure 1E). Nevertheless, both *left* ($n = 13$, $\chi^2_1 = 11.639$, $p < 0.001$) and *right* ($n = 13$, $\chi^2_1 = 9.254$, $p = 0.002$) groups showed a significant learning effect.

Comparing the results of experiment one and two we found a significant interaction between the training effect and experimental design, with a stronger learning effect in experiment two (Figure 1D,E; $\chi^2_1 = 31.661$, $p < 0.0001$). This could be consistent with superior learning performance with more spatially distant resources. We therefore performed a third experiment that approaches the spatial scale over which wild *Heliconius* forage between pollen resources. Using the Metatron facility in Ariège, France[7], we created ~60 m wide T-mazes by connecting four cages