Let's DAG in – How DAGs can help Behavioural Ecology be more transparent.

Mirjam J. Borger[1,*] & A. Ramesh[2,*]

[1]Department of Evolutionary Biology, Bielefeld University, Germany
[2]Institute of Ecology and Evolution, Bern University, Switzerland
*Authors contributed equally

Corresponding author address: Department of Evolutionary Biology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
Email addresses: mirjam.borger@uni-bielefeld.de; aparajitha.ramesh@unibe.ch

Abstract

Directed acyclic graphs (DAGs) are powerful tools for visualizing assumptions/hypothesis and causal inference. Although their use is becoming more widespread across various disciplines, they remain underutilized in behavioural ecology and evolution. Here, we point out why DAGs can serve as highly valuable tools in this field, particularly in the context of observational and field studies, which can feature many variables with complex relationships. Using concrete examples, we show that including DAGs into empirical studies helps clarify and summarise the key underlying assumptions, which are often implicit. With that, DAGs can be used to make researchers aware of bad controls and help them to explicitly think through the relationship between variables and their inclusion in statistical models. In addition, providing DAGs makes the work of reviewers and meta-analysis researchers easier, more rigorous and reliable. Overall, DAGs enhance understanding and transparency, ultimately improving study reproducibility and contributing to greater reliability and replicability across the field. With this paper, we hope to encourage all behavioural ecologists to include DAGs in their papers.

Keywords: causal inference, science communication, bad controls, reproducibility, best practices

Introduction

Directed acyclic graphs (DAGs) are graphical models to visualise the different variables and their assumed effects on each other within a study system. Indeed, DAGs are graphical representations of your hypothesis and form the cornerstone of your statistical model, by formalising the causal structure between variables underlying the hypothesis. DAGs represent variables as nodes, connected by arrows pointing towards assumed causal effects ('Directed'), see box 1 for an explanation of some important terminology and usage. Crucially, DAGs are non-circular ('Acyclic'), that is, cause and effects do not feedback, and thus a hypothesis has to be made about which variable is cause and which one is consequence for the specific circumstances that is being studied. DAGs are well-established tools of causal inference and their use is increasing across different fields (e.g., in econometrics, [1]; environmental sciences and ecology, [2,3]; epidemiology and clinical studies, [4–6] but sparsely used in others such as behavioural ecology. Here, we aim to demonstrate that DAGs not only provide a robust framework for statistical analyses but also enhance transparency and replicability in research. Moreover, by visualising and comparing different DAGs across various systems for similar questions, we can scrutinise the underlying causal structures, offering new insights and potentially driving innovative inquiries in behavioural ecology. Thus, we argue that there are more benefits to using DAGs in research, other than their role of formalising statistical models and avoiding common pitfalls (such as the inclusion of colliders or pipe variables). In order to encourage their incorporation, we also discuss how to get started by reviewing key terminology and causal structures (illustrated in Box 1), while also guiding researchers to essential literature.

Don't we already know this?

DAGs are already a well-established concept (e.g. [7–9]), playing a central role in causal inference. They frequently accompany studies employing path analyses or structural equation models. Yet, DAGs remain underutilised in biology and especially in behavioural ecology. To demonstrate this point, we reviewed 6 issues of the journal *Behavioral Ecology* (n = 123 original articles, volume 34, issues 4-6 and volume 35, issues 1-3), and 6 volumes of the journal *Animal Behaviour* (n = 122 research articles, volumes 210 - 215), and found that only one article contained a DAG (see the data availability statement for a link to our data). This sample includes some articles that might not benefit from the use of a DAG (e.g. theoretical biology papers), but most of these papers included a statistical analysis of empirical data that might benefit from the inclusion of a DAG, as we will argue below.

Current practices

Statistical modelling is generally performed for one of the two main reasons: prediction or causation. For example, prediction applies when trying to extrapolate population size of a conserved species in the next years based on current information, or when inferring how fast a virus will spread in a population; whereas causation applies to performing an experiment to infer if *x* causes *y* and if this is a small or large effect, or by trying to find associations between variables, which ultimately means that *x* and *y* are directly or indirectly causally linked in some way, to understand if their relation is worthy of further study. Most statistical models are optimised for the first goal ('the best-fitting model'), despite being commonly used for addressing the second. Behavioural ecology is a field that mostly focusses on the second goal. This is evident from the use of phrases such as "*x* increases performance of *y*" or "*x* seems to be driven by *y*". Articles often verbally describe causal models and assumptions in the introduction, leading to the main question or hypothesis, and to a certain extent these causal models and assumptions are further described in the methods. Using a formal mathematical equation, Structural Causal Model (SCMs) or tools other than DAGs to express underlying causal assumptions or structures are not the norm. In many cases, causal language is vague as researchers keenly identify that correlations are not causations in the absence of randomized controlled experiments, but the overall goal still remains to learn something about causation.

This is often reflected in a vaguely causal discussion section, which then acts as a starting point for researchers following up on similar topics, essentially leading to causal assumptions that are not coherent and principled. Moreover, in the methods it is often mentioned that additional variables are added to models as a 'control' in the analysis, while these variables are often not mentioned in the hypothesis, and their inclusion is often not justified statistically (e.g. 'we controlled for *z* in the analysis as we expect *z* to affect *y* or we controlled for *z* because we expect that *z* is a confounder of *x* and *y*). Some articles describe that only uncorrelated variables are added to models, but again often no justification for this addition is given (e.g. 'we were also interested in the effect of *z* on *y*, so we also included this in the model' or 'we wanted to increase the precision of the estimate of the effect of *x* on *y*, and therefore we have included *z* in our model'). Similarly, variables are often later omitted from the statistical models citing collinearity. Hence, although researchers think about their assumptions and hypotheses when performing their analyses, this is often vaguely explained and not principled when it comes to their actual analyses.

Why should we use DAGs?

DAGs have two main benefits in empirical biology research. First, DAGs have been well-established as justifications of statistical models (regression models, which are most commonly used) and enable us to take a principled approach to our analyses [8,9]. Even without mastering the mathematics and theory behind DAGs and causal inference, we can translate our knowledge about the system into a DAG and benefit from the rather simple rules of DAGs to understand which variables should be included in a statistical model and how to overcome potential biases. We will give a short overview of the use of DAGs for statistics in the context of behavioural ecology, and explain what bad controls are and that they are largely unknown or neglected in this field. Second, DAGs can increase the transparency, readability and effectiveness of science communication, which could contribute to solving the replication crisis.

<u>DAGs as justification for statistical models</u>

DAGs have generally been used to justify which (control) variables should be included in or excluded from statistical models. Yet, so far, there seems to have been little notice of the concept of bad controls in behavioural ecology [10,11]. In our review of the 245 research articles, we found that predominantly, variables were included 'to control for them' without specifying why this was necessary. The common justification given for control variables to be included in a model is 'biological relevance'. While biological relevance is an important criterion (why include something in a statistical model that is irrelevant for the response variable), it is not a clear justification for inclusion in a statistical model without considering the underlying causal assumptions. Especially in long-term study systems where many variables are measured over a long time, it can become tempting to add control variables without considering how this affects the estimated effect in question. In some cases, the addition of control variables can do harm rather than good, by falsely changing the estimate of the relation in question, while in other cases, controlling for a variable is necessary to obtain a correct estimate of an effect.

Two ways in which adding control variables to a statistical model can wrongly influence estimates are commonly called colliders and pipes (see box 1, Fig 1; [9,11]). Collider variables are caused by both the predictor variable and the response variable (or caused by descendants of the predictor and response variables). There is no causal path through a collider between the arrows pointing into it, i.e. predictor and response variable both cause the collider variable and are not associated to each other through the collider, as it is downstream to both of them. However, by including such a collider variable in a statistical model, a path between the predictor and response variable is opened through the collider variable, thus creating an association between the predictor and the response variable. The direction in which the estimate is changed, depends on the direction of the collider effect (i.e. the correlation will become more positive when the collider effect is positive, and more negative when the collider effect is negative).

Hence, the inclusion of colliders sometimes causes significant results for associations that were absent in reality, or vice versa. For example, when a study tries to estimate the effect of age on foraging efficiency and the expected causal structure is like that in Fig. 1A. Here, both age and foraging efficiency independently affect body mass. It is important to not include body mass, because it is a collider. When body mass is added to the statistical regression model, a causal path between age and foraging efficiency is opened through body mass, therefore inflating (or deflating) the estimate of the effect of age on foraging efficiency. In sum, collider bias can decrease the accuracy of an estimated effect and can result in wrong conclusions about the strength and direction of this estimated effect.

Pipe variables are caused by the predictor variable and cause the response variable. They are also called mediators. For pipes, it is important to realise whether it is important for the research question to estimate the total effect of the predictor variable on the response variable, or the direct effect of the predictor variable on the response variable (see Box 1). When the total effect of the predictor variable should be estimated (for example, in drug trials where the goal is to determine the effectiveness of a medicine), a pipe variable should not be included, as this would lead to false conclusions (e.g. about the effectiveness of a medicine). However, when a mechanism is studied and the influence of mediator variables (affected by the predictor variable) on the response variable is of interest, then pipe variables can be included. Yet, in such a case, it might be of interest to use path analyses or structural equation models (SEMs). Both types of models include a hypothesised causal network, and essentially incorporate multiple sub-models into one model. In other words, all arrows within a hypothesised causal network can be studied within a single model, and therefore more can be learned about how mechanistically an explanatory variable (directly and indirectly) can cause a response variable [12,13]. SEMs can additionally be extended so that latent variables (i.e. variables that cannot directly be measured) could be included in a model [9,14,15]. For example, when a study tries to understand the effect of feeding frequency on the probability that offspring fledge, and the causal structure is as in Fig. 1b, it depends on the biological question whether chick mass should be included in the statistical model, because chick mass is a pipe variable. When the goal is to estimate the total effect of feeding frequency on fledging probability, then chick mass should not be included. Yet, when the goal is to estimate only direct effects, then chick mass should be included. In this second case, a model will show no direct effect of feeding frequency, and a strong effect of chick mass on fledging probability. As another example, Fig. 1c shows a study that manipulated the cortisol levels in animals to estimate its effect on body mass (and we assume the causal structure as in Fig. 1c). The manipulation also changes all the intermediate variables, and hence, to estimate the total effect of the manipulation, including activity in a statistical model is wrong (it is a pipe), as it takes out the indirect effect of cortisol on body mass. However, when the experiment specifically wants to know how much of the change in the body mass is caused by a direct effect of cortisol, activity should be included in the statistical model. However, in such a case, it might be better to perform a path analysis, also including the direct effect of cortisol on activity, to ensure there is an effect of cortisol on activity.

Confounders, on the other hand, should always be included in the statistical model, and not including confounders may lead to spurious associations. Confounders causally affect both the response and the predictor variable. This means that there is a causal path between these variables through the confounding variable, unless the confounder is added to the statistical model. For example, Fig. 1d shows a DAG for a study testing the effect of vegetation density on boldness tendencies across populations in the field using measurements such as flight initiation distance. Temperature affects the vegetation densities, with higher temperature leading to denser vegetation, but also the boldness of the fish via their metabolic rates and therefore is a confounder. To estimate the effect of vegetation density on the boldness of fish, temperature should be included in the statistical model.

While the question of whether to include a variable in a statistical model is much simpler in experiments, it is nevertheless an important question for observational studies. Currently, many observational studies still hold the old belief that when two variables are not significantly correlated, they can both be included in a statistical model, in essence assuming that every variable in their DAG is completely independent and only affects the response variable. Yet, when being forced to articulate these assumptions, researchers might realise that effects are not independent. DAGs can thus be useful to aid the thought process and help in making hidden assumptions and relationships explicit. Moreover, assuming that many ecological variables are completely independent from each other seems illogical in field systems. We know that many factors in ecological systems affect each other (e.g. climatic effects might interact, a social environment can create non-independent data points across individuals), and in fact, we have dedicated whole research areas to it (e.g. systems ecology, community ecology). Additionally, the fact that two variables are only weakly (and potentially non-significantly) associated does not solve the issues that colliders and pipes cause. A weak collider variable could still inflate the estimated effect of the predictor variable on the response variable. Moreover, not including a confounder (which should correlate with the predictor and response variable) also decreases the accuracy of the effect in question. Hence, we would like to convince biologists that hypothesising certain associations (with a DAG) is better than (implicitly) assuming that every variable is completely independent, unless it is explicitly hypothesised (preferably again with a DAG) that variables are independent from each other. This saves research from unnecessary inflations or deflations of estimates caused by colliders, pipes, and confounders and thus makes our research better. Moreover, the addition of independent variables to models does not affect the accuracy of the estimation of the relationship in question, but only improves the precision of this estimate [11]. Whether that is necessary is up to the researcher, and might for example depend on how the data is analysed (e.g. Bayesian statistical results often already give information about precision).

DAGs to increase transparency in scientific research

Hypotheses in behavioural ecology are often broad and generalised, while we commonly test these hypotheses with much more specific and specialised (to one or a couple of species) statistical models. In this step from general to specific questions or hypotheses, numerous assumptions are made, usually based on the ecology of the study system. While researchers aspire to mention all these assumptions, it is easy to overlook some. This could stem from the fact that the ecological knowledge about a study system can seem trivial for researchers studying that system or from the fact that the complexity of a system can lead to certain assumptions being made implicitly or simply be overlooked. Yet, these underlying assumptions influence the expectations and also the outcome of a study. DAGs offer a simple graphical tool to clarify most of these underlying assumptions. This can decrease confusion among the readers and reviewers—and even the authors themselves—who often think about an overarching general hypothesis using the assumptions of the study system they work with / are familiar with.

For example, the evolution of cooperation and cooperative breeding is a well-studied topic in behavioural ecology [16], where the underlying ecological assumptions can have drastic effects on the evolutionary predictions (Fig. 2). In territorial species where territory size is more or less fixed, one would expect that acquiring helpers (subordinate individuals that help in raising offspring of dominant territory owners) depends on the territory quality. In other words, territory quality ‘causes’ the number of helpers (e.g. Seychelles warbler *Acrocephalus sechellensis*, [17]; Acorn Woodpecker *Melanerpes formicivorus*, [18]; see Fig. 2a). When territory quality is low, helpers might consume the few resources that are available, therefore leading to less resources for the offspring. When territory quality is high, there might be enough resources to sustain a certain number of helpers, but also for these helpers to increase feeding rates of offspring. In such species, we expect an optimal number of helpers per territory that depends on the quality of this territory, as there is a trade-off between the decreased resource availability due to helpers feeding themselves and the increased survival probability of offspring due to the assistance of these helpers. Alternatively, when helpers actively and substantially increase the size of the territory they reside on, and therefore increase the resource availability, helpers 'cause' the territory quality (e.g. cichlids *Neolamprologus obscurus,* [19]; cichlids *Neolamprologus pulcher*, [20]; see Fig. 2b). In this case, helpers decrease the negative effect of them consuming resources, as they also assure extra availability of resources. Therefore, an increase in the number of helpers often has a positive effect on offspring survival, independent of the territory quality before a helper was present. These underlying ecological patterns are often clearly mentioned when studying the effect of helper presence on offspring survival/number of offspring, but are sometimes overlooked when studying more complex questions about cooperative breeding. Yet, they might still have a large impact on the expected evolutionary patterns. For example, over the last years, the question whether cooperative breeding might buffer against harsh or unpredictable environments has received a lot of attention (e.g. [21–23]). This hypothesis could be consistent with study systems where the number of helpers influences territory quality, because in such a case, even in harsh environments, helpers might be able to improve territory quality enough for offspring to survive, while without helpers, this would not have been possible. Similarly, this hypothesis could be consistent with study systems where predation is the limiting factor of offspring survival instead of food availability, as in such a case, helpers might protect offspring from being preyed upon and increase their survival probability in that way, while there are enough resources for both helpers and offspring to consume. Yet, this buffering hypothesis seems illogical for species where the territory quality determines the number of helpers. In such a case, the competition for resources between helpers and offspring intensifies when conditions turn harsh, as there are now less resources available per individual. Hence, territories with fewer helpers might in fact produce more surviving offspring and thus become less social, might buffer against harsh environments. When the expected outcomes depend on the ecology of the species, a DAG of the different study systems can clarify why cooperative breeding might buffer against harsh environments in certain cooperative breeding species, but not in others, as the arrow between territory quality and number of helpers points in the opposite direction in the two cases.

Another hypothetical example to illustrate that DAGs can be helpful to increase transparency is the effect of predator density on foraging behaviour and survival of prey (Fig. 3). While it is evident that the direct effect of an increase in predator density on the survival of individual prey is negative, this result might not be found or found to a different extent than expected when experimentally changing the predator density, because experiments manipulate the total effect of predator density on survival, instead of the direct effect. Predation could, for example, have indirect effects on survival because it might also influence the foraging efficiency of individuals, because prey might forage less when predators are around, furthering the negative effect of predation on survival. Moreover, predation could have a negative effect on the population density of prey, which might in fact increase the foraging efficiency of individual prey, as there is less competition between the remaining individuals, and thus potentially increase their survival. A DAG can show which factors are expected to affect foraging efficiency and survival and could also explain why it is not so evident to form an expectation about the total effect of predation on survival. In such cases, a DAG can clarify what the assumptions are underlying the effect of predation on survival, and can explain why different patterns can be found in different species, or in different populations within the same species. Furthermore, a DAG can show how different mechanisms could lead to the same result. Comparing a population with predators that hunt during the foraging period of prey with one who forages at a different time (e.g. nocturnal predators, see Fig. 3a and b), shows that the negative effect of predation on foraging efficiency can be different between different study systems. In the case of nocturnal predators, there is only a positive effect of predation on foraging efficiency (through density regulation). However, the effect of predation on survival for nocturnal predators can still be anything from strongly positive to strongly negative, depending on the strengths of the indirect prey density effect and the direct predation effect. As another example, there could be a difference in the population size of two populations of the same species (see Fig. 3a and c). In large populations, the effect of predation on the prey density (and especially competition for food between prey) is likely much smaller than in small populations. Therefore, there might be almost no positive effect of predation on foraging efficiency in large populations, but potentially a strong positive effect of predation on foraging efficiency in smaller populations. Hence, in large populations, it might be expected that predation only has a negative effect on survival, while in small populations, the effect of predation on survival might be anything from positive to negative. DAGs highlight these small differences between study systems in a succinct way, and make it clear to the reader what exactly is expected in a study, even when it is not clear whether the total effect will be positive or negative.

Using DAGs thus makes it easier to follow exactly what the researcher is studying and how they expect the study system to work. This increases the readability and transparency of research for readers, and helps in improving the replicability of the (statistical) methods that were used. Next to that, DAGs can also help reviewers by adding clarity to their critiques. With a DAG, it is easier to separate whether authors have made a mistake in, or have a different opinion about, their statistical analysis or whether the reviewer and authors disagree conceptually on a question or its underlying assumptions, while in principle they agree on the statistical method. In our review of the 245 research articles, a model contained on average 6 variables (range = 2-23, median = 6). Given that there is a maximum of 15 causal arrows in a model with 6 variables, a misunderstanding about the exact hypothesis, including underlying assumptions, can easily happen. Incorporating a DAG can solve a lot of these issues. Moreover, in a recent report, 174 research teams were asked to analyse the same two datasets. The results in terms of effect sizes were strikingly variable, even presenting in opposite directions across research teams [24]. One of the main reasons was attributed to the substantial variation in the variables that were included as fixed and random effects in their statistical models. We believe that if researchers constructed a DAG prior to analysing their data, they would have a more principled approach to selecting variables for their analyses. This would shift the discussions towards why researchers had different underlying assumptions and their rationale for them. Instead of a crisis or caution about interpreting results in the field due to apparently valid statistical analyses leading to varied interpretations, the focus would be on the conceptual differences driving those choices. We argue that DAGs could thus improve the transparency about causal assumptions and the readability of an article, and with that potentially help solve the replication crisis, which is also facing this field [25,26]. When a DAG is included, it may become easier to replicate methodologies, including replicating the statistical decisions, which sometimes are not well described. In addition, it may become easier to conceptually replicate a study, where the same hypothesis or theory is tested in a different way (different contexts, different systems) to obtain generalizable results. DAGs aid this by identifying studies that incorporate the same underlying assumptions. Subsequently, DAGs can be used to distinguish studies with different underlying assumptions to start with and hence are not part of a conceptual replication (i.e. these studies where in fact not replicates, but researched a slightly different question, like in Fig. 2a and b). Moreover, DAGs can be a great addition to preregistrations, where they allow authors and peer reviewers to visualise the relationships clearer, leading to best methods for data analysis or take decisions on additional variables to measure. Ultimately, this helps reducing 'research waste' by enforcing better planning and reporting [27].

DAGs can also improve meta-analyses, as studies studying the same question but predicting the opposite given their underlying assumptions can be separated in a quantitative way, instead of combining studies to conclude apparent null effects. In recent times, meta-analysts in behavioural ecology suffer from the inability to delineate effects from complex statistical models, where it is not clear what variables are controlled for and why. This is especially true when the variables of interest are proxies (such as proxies of fitness, reproduction, parental care) or secondary variables that are not the response variable in the main statistical model, thus leading to exclusion of studies. In this case, DAGs also provide a more targeted inclusion criterion, in addition to improving the transparency of studies for effective meta-analyses. Lastly, DAGs can help new researchers (e.g. students, or people changing fields) to understand the key differences between study systems that are important for the questions they wish to study, without having to spend considerable amounts of time researching the ecology of all these species.

Where to start?

So far, we have provided examples, outlined basic structures and discussed how DAGs can make scientific discourse efficient and transparent. However, constructing a DAG is entirely based upon the expert subject knowledge of the researchers. After identifying the research question(s) and defining the estimand, it is important to include those relevant variables – the predictor and the response variables. Next, it is important to include all the common causes of our predictor and response variables. When introducing a new variable, its common causes must also be included. It is important to keep in mind that DAGs are not meant to be realistic depictions of your study system; rather, they aim to succinctly represent the research question and outline the underlying hypotheses. Consequently, mediators or mechanisms should only be included if they directly pertain to the research inquiries. For the keen reader or enthusiastic adopter of DAGs, we recommend referring to key literature on DAGs and causal inference. On tips how to build a DAG and use causal inference, we would like to refer to Laubach et al. [28], and Arif and MacNeil [29] for explanations with biological examples, to McElreath [9] - including the accompanying youtube videos - for explanations on causal inference especially well-suited to beginners, and to Judea Pearl's work [e.g. 7,8,11,15,30] for a more in depth understanding.

Drawing DAGs can be done by hand, but there is also an R package (see DAGitty; [31]) and website (https://www.dagitty.net/dags.html; [31]) available to help with this. This package also aids in identifying the conditional independencies and adjustment sets, to help with which variables to condition for, if your DAG is complex.

We would like to advise readers to draw DAGs before conducting their research (as it is a formalised version of their hypothesis), and they could be included in preregistrations if researchers like to do so. Moreover, while we think DAGs are valuable additions to manuscripts, we understand that their inclusion in a manuscript might be limited by length or figure restrictions, and we encourage researchers to include a DAG in their supplementary materials in such cases.

Limitations of DAGs

DAGs are useful and simple representations of causal assumptions, and we hope to have convinced the readers of their strengths by now. But they can come with a few drawbacks when put into practical use. First, systems in behavioural ecology are rarely simple, hence DAGs can become complex quite easily (but see [32] for an example of a more complex DAG). Decisions about the causal relationships in the DAG have to be made (e.g. should there be an arrow between two variables, and if so, in which direction), and it can be difficult to pick the 'right' answer. Indeed, there is no one true DAG, unless all causal relationships are well-known. However, in essence, this is an issue of knowledge about the system, and not one of drawing DAGs per se. In such cases, DAGs might help uncover which assumptions are still unknown and should be tested. Moreover, if a causal pathway is unknown, we argue that it is still better to have a hypothesis about it (which consequently can be tested) than to implicitly have an underlying assumption about a relationship (or lack thereof) between variables.

Second, DAGs still do not convey all information about the statistical analyses. For instance, DAGs cannot help make a decision or convey information about the following: Should variables be added as fixed or random effects? Is the expected relationship linear, quadratic or another shape, and is the relationship positive or negative? To summarise all this information, Structural Causal Models can be used, which are mathematical representations of a statistical model. Yet, some biologists in behavioural ecology tend to be a bit averse to math, and hence we argue that using a visual representation (a DAG) is a good start to transparently communicate causal assumptions in statistical models. Moreover, some of this information could potentially be added to DAGs, such as using colour coding or different shapes (e.g. green lines for hypothesised positive relationships, red lines for hypothesised negative relationships; random effects written in italics or contained in a square box instead of in a circle).

Third, causal inference and, therefore, to some extent DAGs, are unable to deal with interaction effects, where the effect of a variable is not on an outcome variable, but on the relationship between two variables. Some have suggested representing this by making an arrow point to another arrow (e.g. [33,34]), but the theory around the meaning of interactions in terms of causality is not well-developed.

Last, DAGs are acyclical, which means that causal inference so far is unable to include cyclical interactions or reciprocity. Yet again, we would argue that being unable to disentangle feedback loops between cause and effect is an inherent problem in many types of science and is not an issue of DAGs in itself. In some cases, time series analyses could help to disentangle these reciprocal interactions, if it is possible to find a time frame where *x* solely affects *y*, and another time frame where *y* solely affects *x* (e.g. timing of migration might cause timing of breeding, and timing of breeding might cause the timing of migration in the next year). In such a case, two DAGs could be produced for each time frame and these hypotheses could accordingly be tested in the correct time frames. However, when it is not possible to disentangle reciprocal interactions, DAGs cannot help in solving this issue. In such cases, researchers will have to decide which direction is more important for their question, and be cautious with interpreting their results.

Conclusion

With this paper we would like to convince the reader that using DAGs in behavioural ecology and beyond is beneficial. DAGs increase readability of papers because they show underlying assumptions that are not always mentioned. Exposing these underlying assumptions increases the transparency of research. DAGs can stand the test of time: even if statistical tools undergo changes or advances such that reading a model description to work out causal assumptions becomes more challenging, DAGs can offer a simple reporting standard and a common language for causal assumptions that justify statistical models. Thus, DAGs could contribute to solving the replication crisis and make the work of reviewers and researchers doing meta-analyses easier and more rigorous. Additionally, DAGs are extremely useful for considering which variables should be included in statistical models. Moreover, we think that a hypothesised causal structure is valuable (which often will be thought through when a DAGs is produced), because in our opinion, they are always better than implicitly hypothesising without adequate thought that every variable in a model is completely independent. DAGs, just as hypotheses, might be wrong, but by showing a DAG, these mistakes are easier to find, and at least statistical mistakes caused by expected colliders, pipes or confounders can be prevented.

Data accessibility statement

Data is stored on Zenodo and will be available once the manuscript is published.

Competing interests Statement

The authors declare to have no competing interests.

Acknowledgements

AR is funded by SNSF postdoctoral grant to Barbara Taborsky, University of Bern (grant no. 31003A_179208). MJB was funded by the German Research Foundation (DFG Project no. 316099922) as part of the CRC NC³ (SFB TRR 212). We would like to thank Christoph Netz, Ben Kawam, Euan Young, Jan Kreider, Jia Zheng, Frigg Speelman, Claire Tsui and Xia Zhan for all the discussions on causal inference. Moreover, we would like to thank Raphaël Scherrer for reading a previous version of this manuscript. Lastly, we thank Alfredo Sánchez-Tójar and the associate editor and two anonymous reviewers for their valuable comments on the manuscript.

Author contributions

MJB and AR conceptualised. MJB wrote the first draft, and both authors contributed to further writing and revising of the manuscript.

References

1. Kwon D-H, Bessler DA. 2011 Graphical methods, inductive causal inference, and econometrics: a literature review. *Comput Econ* **38**, 85–106. (doi:10.1007/s10614-010-9236-9)

2. Arif S, Massey MDB. 2023 Reducing bias in experimental ecology through directed acyclic graphs. *Ecology and Evolution* **13**, e9947. (doi:10.1002/ece3.9947)

3. Siegel K, Dee LE. 2025 Foundations and future directions for causal inference in ecological research. *Ecology Letters* (doi:10.1111/ele.70053)

4. VanderWeele TJ, Robins JM. 2007 Four types of effect modification. *Epidemiology* **18**, 561–568. (doi:10.1097/EDE.0b013e318127181b)

5. Byeon S, Lee W. 2023 Directed acyclic graphs for clinical research: a tutorial. *Journal of Minimally Invasive Surgery* **26**, 97–107.

6. Poppe L, Steen J, Loh WW, Crombez G, De Block F, Jacobs N, Tennant PWG, Van Cauwenberg J, De Paepe AL. 2025 How to develop causal directed acyclic graphs for observational health research: a scoping review. *Health Psychology Review* **19**, 45–65. (doi:10.1080/17437199.2024.2402809)

7. Pearl J. 2009 *Causality - Models, Reasoning, and Inference*. 2nd edn. Cambridge: Cambridge University Press.

8. Pearl J, Mackenzie D. 2018 *The Book of Why: the New Science of Cause and Effect*. First edition. New York: Basic Books.

9. McElreath R. 2020 *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. 2nd edn. London: Chapman & Hall.

10. Montgomery JM, Nyhan B, Torres M. 2018 How conditioning on posttreatment variables can ruin your experiment and what to do about It. *American J Political Sci* **62**, 760–775. (doi:10.1111/ajps.12357)

11. Cinelli C, Forney A, Pearl J. 2022 A crash course in good and bad controls. *Sociological Methods & Research* **0**, 1–34. (doi:10.2139/ssrn.3689437)

12. Wright S. 1934 The method of path coefficients. *The Annals of Mathematical Statistics* **5**, 161–215.

13. Streiner DL. 2005 Finding Our Way: An Introduction to Path Analysis. *Canadian Journal of Psychiatry* **50**.

14. Bollen KA. 1989 *Structural Equations with Latent Variables*. New York: Wiley.

15. Pearl J. 2010 An introduction to causal inference. *The International Journal of Biostatistics* **6**. (doi:10.2202/1557-4679.1203)

16. Davies NB, Krebs JR, West SA. 2012 *An Introduction to Behavioural Ecology*. 4th edn. Oxford, UK: Wiley-Blackwell.

17. Komdeur J. 1994 Experimental evidence for helping and hindering by previous offspring in the cooperative-breeding Seychelles warbler *Acrocephalus sechellensis*. *Behavioral Ecology and Sociobiology* **34**, 175–186.

18. Koenig WD. 1981 Reproductive success, group size, and the evolution of cooperative breeding in the acorn woodpecker. *The American Naturalist* **117**, 421–443. (doi:10.1086/283726)

19. Tanaka H, Kohda M, Frommen JG. 2018 Helpers increase the reproductive success of breeders in the cooperatively breeding cichlid *Neolamprologus obscurus*. *Behav Ecol Sociobiol* **72**, 152. (doi:10.1007/s00265-018-2566-7)

20. Bruintjes R, Hekman R, Taborsky M. 2010 Experimental global food reduction raises resource acquisition costs of brood care helpers and reduces their helping effort. *Functional Ecology* **24**, 1054–1063. (doi:10.1111/j.1365-2435.2010.01715.x)

21. Cockburn A, Russell AF. 2011 Cooperative breeding: a question of climate? *Current Biology* **21**, R195–R197. (doi:10.1016/j.cub.2011.01.044)

22. Jetz W, Rubenstein DR. 2011 Environmental uncertainty and the global biogeography of cooperative breeding in birds. *Current Biology* **21**, 72–78. (doi:10.1016/j.cub.2010.11.075)

23. Cockburn A. 2020 Can't see the "hood" for the trees: Can avian cooperative breeding currently be understood using the phylogenetic comparative method? In *Advances in the Study of Behavior*, pp. 243–291. Elsevier. (doi:10.1016/bs.asb.2019.11.002)

24. Gould E *et al.* 2025 Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biol* **23**, 35. (doi:10.1186/s12915-024-02101-x)

25. Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, Chee YE, Kelly CD, Gurevitch J, Nakagawa S. 2016 Transparency in ecology and evolution: real problems, real solutions. *Trends in Ecology & Evolution* **31**, 711–719. (doi:10.1016/j.tree.2016.07.002)

26. Yang Y, Van Zwet E, Ignatiadis N, Nakagawa S. 2024 A large-scale in silico replication of ecological and evolutionary studies. *Nat Ecol Evol* **8**, 2179–2183. (doi:10.1038/s41559-024-02530-5)

27. Purgar M, Klanjscek T, Culina A. 2022 Quantifying research waste in ecology. *Nat Ecol Evol* **6**, 1390–1397. (doi:10.1038/s41559-022-01820-0)

28. Laubach ZM, Murray EJ, Hoke KL, Safran RJ, Perng W. 2021 A biologist's guide to model selection and causal inference. *Proc. R. Soc. B.* **288**, 20202815. (doi:10.1098/rspb.2020.2815)

29. Arif S, MacNeil MA. 2022 Predictive models aren't for causal inference. *Ecology Letters* **25**, 1741–1745. (doi:10.1111/ele.14033)

30. Pearl J. 1995 Causal diagrams for empirical research. *Biometrika* **82**, 669–710.

31. Textor J, Van Der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH. 2017 Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int. J. Epidemiol.* **45**, 1887–1894. (doi:10.1093/ije/dyw341)

32. Borger MJ, Lee KGL, Weissing FJ, Richardson DS, Dugdale HL, Burke T, Pen I, Komdeur J. 2024 A path analysis disentangling determinants of natal dispersal in a cooperatively breeding bird. (doi:10.1101/2024.07.26.605387)

33. Attia J, Holliday E, Oldmeadow C. 2022 A proposal for capturing interaction and effect modification using DAGs. *International Journal of Epidemiology* **51**, 1047–1053. (doi:10.1093/ije/dyac126)

34. Bours MJL. 2023 Using mediators to understand effect modification and interaction. *Journal of Clinical Epidemiology* **163**, 117–121. (doi:10.1016/j.jclinepi.2023.09.005)

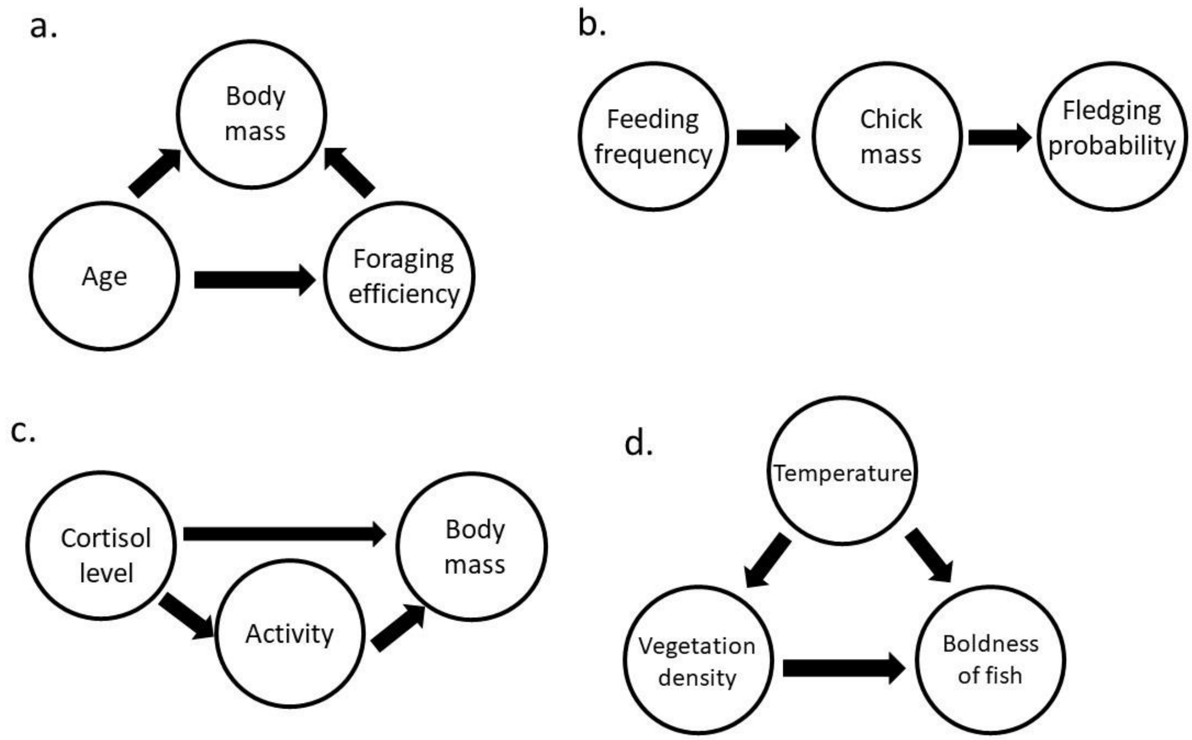


Figure 1: Four directed acyclic graphs showing collider, pipe and confounder variables. a) A DAG for a study trying to estimate the effect of age on foraging efficiency, where body mass is a collider for this relationship. b) A DAG for a study trying to estimate the effect of feeding frequency on fledging probability, where chick mass is a pipe variable. c) A DAG for an experiment trying to study the effect of cortisol on body mass, where activity is a pipe variable. d) A DAG for a field study estimating the impact of the density of vegetation (refuge) on the boldness of a fish, where the mean temperature influences the vegetation density as well as the boldness of fish (via metabolic rates), acting as a confounder.

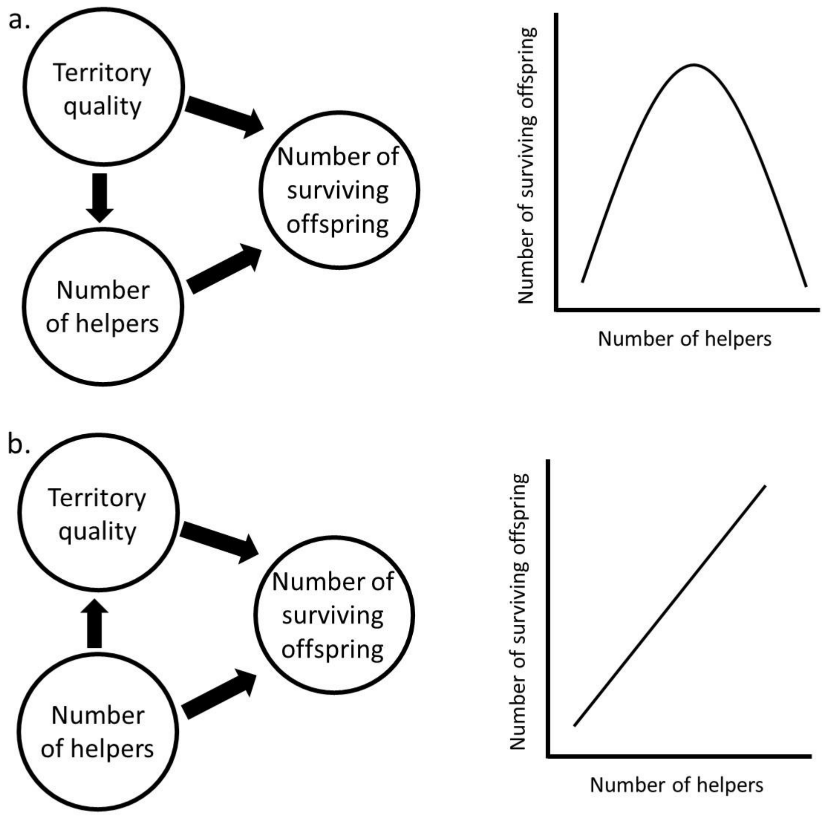


Figure 2: Two directed acyclic graphs (DAGs) about the effect of territory quality and number of helpers on the number of surviving offspring. a) Represents a case where territory quality causes the number of helpers, and b) represents a case where the number of helpers causes territory quality. Depending on the causal relationship between territory quality and number of helpers, the effect of number of helpers on number of surviving offspring changes.

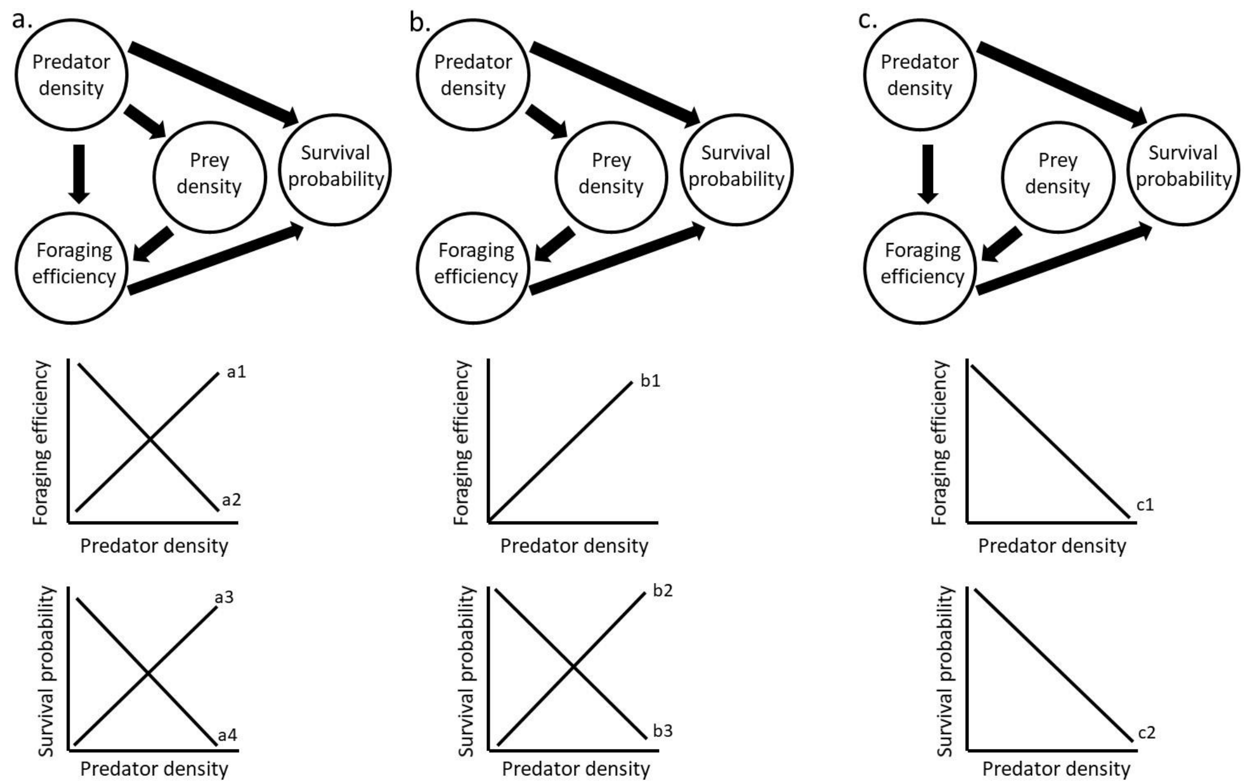


Figure 3: Three directed acyclic graphs about the effects of predator density, prey density and foraging efficiency of prey on the survival probability of prey individuals. a) shows a causal structure where predator density both directly and indirectly affects foraging efficiency, and directly and indirectly affects the survival probability of prey. b) shows the same causal structure but without a direct effect of predation on foraging efficiency (e.g. a nocturnal predator), and c) shows the same causal structure as a), but without a direct effect of predator density on prey density (e.g. a very large prey population). a1 to c2 represent the total effects that predator density could have on foraging efficiency and prey survival probability. In some cases, it is not possible to have a clear prediction and anything from a positive to negative relationship could be expected (e.g. a1 and a2), while in other cases, it is possible to have a prediction (e.g. b1). Whether it is possible to have a prediction about a total effect depends on the strength and direction of each of the direct and indirect effects that together cause the total effect. Predicting the direction of a total effect can be challenging, especially if (in)direct effects are expected to have opposite effects.

Box 1: The main ingredients of DAGs

DAGs are essentially composed of nodes, represented by circles, which are the variables and directed or single headed arrows depicting a causal relationship.

Terminology:

Here, we provide some terminology that is crucial for beginning to use DAGs and understanding the literature on causal inference using DAGs.

1. Estimand: The target quantity that is to be estimated in an analysis. This is related directly to our research question and is what we aim to calculate in our statistical analysis. In our simple examples below, the estimand is the direct effect of X on Y.
2. Direct effect: An effect that the change of a particular variable of interest (X) has on the outcome variable (Y). This means that we are only interested in the arrow that leads directly from X to Y and not the other arrows that emerge from X via another mediator to Y. This is achieved by blocking the other pathways that are not direct (see 5).
3. Total effect: The effect of our variable of interest (X) on our response variable (Y) via all the direct and indirect paths, but excluding the effects due to common causes or confounders (see Fundamental causal structures).
4. Conditioning: Also referred to as 'controlling', 'adjusting', 'stratifying' or 'partialling out' an effect, conditioning refers to the isolation of effects of the variable(s) of interest (X) on the outcome variable (Y) for a given value of the conditioned variable. This is often carried out by including the variable to condition as a covariate in the statistical model.
5. Blocking: Blocking a pathway means blocking a causal effect and therefore association among variables via that path. In the case of a confounder, we want to block the causal effect of the confounder by conditioning on it, while in the case of a collider (see Fundamental causal structures), the causal pathway is blocked when it is not conditioned on. Furthermore, in experimental designs, blocking can be achieved by randomization. For example, the effect of 'time of day' on an experimental outcome can be blocked by randomizing trials over the day or performing trials only at fixed time points in a day.

Fundamental causal structures:

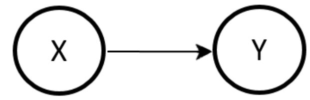
Here we will briefly touch upon the fundamental structures used in DAGs. In these examples, we are interested in the direct effect of X on Y (the estimand). We can represent this by Y~X. When we want to test this in our Generalised linear (mixed) models in R, we would write this as glm(y~x), which is the simplest model. For each of these cases, we will show the relationship between X & Y as a DAG. We here give examples of model statements in R, for GLM, as we would specify in the lme4 package (Bates et al. 2014). Note that we are showing the basic model structure as we use in R, without random effects or assumptions on the distribution of variables.

1. Total independence: X and Y are independent of each other and have separate underlying causal variables. Therefore, variation in X is independent of variation in Y.

1.

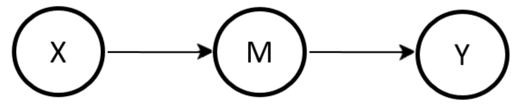
2. Dependence: X causes Y, or a change in the value of X causes a change in the value of Y. This takes the usual statistical model form, glm(y~x). Note that this need not necessarily be a linear relationship

2.



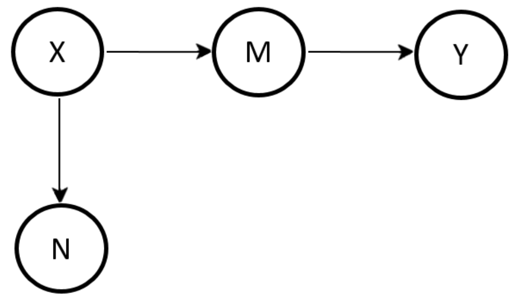
3. Pipe / Chain: Pipes are variables that are caused by x and are causing y. Pipes do not necessarily have to be one variable but can also be a chain of variables. In 3, M mediates the effect of X on Y. Conditioning on M leads to blocking of the causal pathway, making X and Y independent. That is, if we do glm(y~m+x), we will not find a relationship between x and y anymore. In this case, there are no direct effects of X on Y, but there is a total effect of X on Y mediated via M.

3.



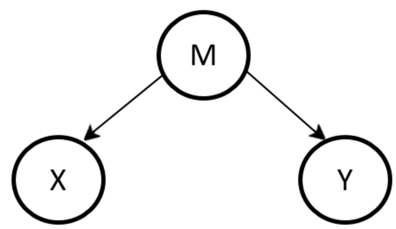
4. Descendant: When X causes M, M is also termed descendant of X and X is the ancestor of M. Similarly, in the diagram, N is also a descendant of X and Y is a descendant of M. Descendants are important because conditioning on descendants of a variable can have the same effect as conditioning on the variable, depending on the strength of their relationships. If we condition on N by glm(y~x+n), we may only partially uncover or not uncover the effect of X on Y. This property may be particularly crucial to consider when descendants are present and measured, but the ancestor is not measured, in the case of collider or confounder causal structures (described below).

4.



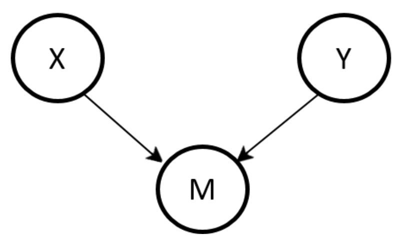
5. Fork / Confounder: Forks or confounders are variables that cause both X and Y. In 5, M is the confounder, which is the common causal ancestor that affects both X and Y, leading to a correlation. Conditioning on M leads to independence of X and Y. glm(y~x+m) is the correct form for this causal structure. In some cases, when a confounder variable is unobservable but a descendant of the confounder was measured, conditioning on the descendant is important to at least partially control for the effects of the confounder on X and Y, to avoid a confounding relationship between X and Y.

5.

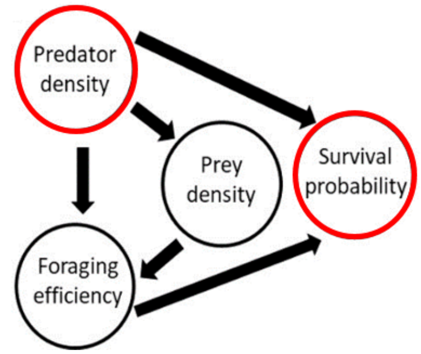


6. Collider: Colliders are variables that are caused by both X and Y. In 6, M is not the mediator but the common causal descendant of both X and Y. The causal path between X and Y are closed as long as M is not conditioned on. When a collider is added to a statistical model, a relationship between X and Y will be found (the causal path that leads through the collider variable). Hence, in this case, the correct model takes the form of glm(y~x). In cases where a collider has a descendant, it is also important not to condition on this descendant. Conditioning on a descendant of a collider will have the same effect of conditioning on the collider, depending on the strength of their relationships.

6.



Example:



Now, let us take the example DAG in Figure 2a. The question we are interested in is the direct effect of predator density on prey survival probability, which is our estimand. Our X is Predator density and Y is the survival probability. If we want to look at the direct effects of predator density on survival probability, without the effect it may have on foraging efficiency, we need to condition on 'Foraging efficiency', as this is a pipe variable. This will block all causal paths going to and emerging from 'Foraging efficiency'. It is important to note that labelling a variable as a confounder or a collider is always relative to the estimand and the structure of causal pathways between the predictor and the response variable. In this case, our model would look like

glm(Survival probability ~ Predator density + Foraging efficiency)

Predators reduce prey densities, leading to more efficient foraging and therefore increasing survival probability due to larger size, better health etc. Therefore, if we are interested in the whole ecological picture, we are interested in the total effect of 'Predator density' on 'Survival probability' (our new estimand). To achieve this, we should not condition on any variables, allowing all causal pathways between 'Predator density' and 'Survival probability'. This includes three pathways:

'Predator density' → 'Prey density' → 'Foraging efficiency' → 'Survival probability'
'Predator density' → 'Foraging efficiency' → 'Survival probability'
'Predator density' → 'Survival probability'

Our model in this case would look like

glm(Survival probability ~ Predator density)