On the use of directed acyclic graphs in behavioural ecology and evolution

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 graphical models to visualise hypotheses. DAGs are generally used in the field of causal inference and their use is spreading across different fields. However, in biology and especially in behavioural ecology and evolution, DAGs are still underutilised. Here, we point out why DAGs are such useful tools for these fields. Using concrete examples, we demonstrate that including DAGs in empirical studies is helpful for summarising all the important underlying assumptions about the ecology of the study species. With that, DAGs increase the readability and transparency of papers, which could help solve the replication crisis. Moreover, it makes the work of reviewers and meta-analysis researchers easier. Lastly, 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. With this paper, we hope to encourage all biologists to include DAGs in their empirical papers.

Keywords: causal inference, transparent science, science communication, bad controls

Introduction

Directed acyclic graphs (DAGs) are graphical models to visualise the different variables and their assumed effects on each other within a study system. In a sense, 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 are represented by variables as nodes, connected by arrows pointing towards assumed causal effects ('Directed'). Crucially, they are non-circular ('Acyclic'), that is, cause and effects do not feed back, and thus a hypothesis has to be made about which variable is cause and which one is consequence for the specific circumstances that are studied. DAGs are well-established in the field of causal inference and their use is increasing across different fields (e.g., in computing, environmental sciences and clinical psychology). 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 and evolution. 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).

Don't we already know this?

First, we want to acknowledge that DAGs are already a well-established concept (e.g. Pearl 2009; Pearl and Mackenzie 2018; McElreath 2020). They have been important in explanations of causal inference, and many papers using path-analyses or structural equation models are accompanied by a DAG. Yet, DAGs are still underutilised in biology and especially in behavioural ecology and evolution. To demonstrate this point, we analysed 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 = 119 research articles, volumes 210 - 215), and found that no article contained a DAG. 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.

Why should we use DAGs?

We argue that using DAGs in empirical biology research has two main benefits. First, DAGs can increase the transparency, readability and effectiveness of science communication, which could contribute to solving the replication crisis. Second, DAGs can help in understanding which variables should be included in statistical models. In this paper we will mostly discuss the first point. The second point is well-established (e.g. Pearl and Mackenzie 2018; McElreath 2020), but we will give a short overview of the use of DAGs for statistics in the context of behavioural ecology, and we will explain what bad controls are and that they are largely unknown or neglected in this field. Moreover, we will recommend key literature on the use of DAGs in statistics for enthusiastic readers who would like to learn more about DAGs and causal inference.

DAGs to increase transparency in scientific research

Hypotheses in biology 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, a lot of 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, because ecological knowledge on a study system can seem trivial for researchers of that study system. Yet, these underlying assumptions often influence the outcome of a study, or even the expectations of a study. Using a DAG in every empirical study can clarify most of these underlying assumptions. This can decrease confusion of the readers and reviewers 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, where the underlying ecological assumptions can have drastic effects on the evolutionary predictions (Fig. 1). 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*, Komdeur 1994; Acorn Woodpecker *Melanerpes formicivorus* Koenig 1981). When territory quality is low, helpers might consume the few resources that are available, therefore leaving less resources for 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, one would expect an optimal number of helpers per territory, depending 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 substantially increase the resource availability (or in another way increase the resource availability), helpers ‘cause’ the territory quality (e.g. cichlids *Neolamprologus obscurus,* Tanaka et al. 2018; cichlids *Neolamprologus pulcher*, Bruintjes et al. 2010). In this case, helpers decrease the negative effect of them consuming resources, as they also assure extra availability of resources. Therefore, an increase of the number of helpers often has a positive effect on survival offspring, 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

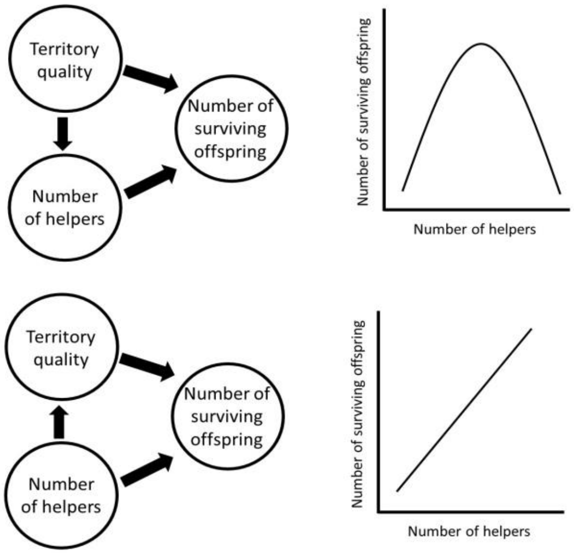

Figure 1: Two directed acyclic graphs (DAGs) about the effect of territory quality and number of helpers on the number of surviving offspring. Depending on the causal relationship between territory quality and number of helpers, the effect of number of helpers on number of surviving offspring changes.

unpredictable environments has received a lot of attention (e.g. Cockburn and Russell 2011; Jetz and Rubenstein 2011; Cockburn 2020). 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 to be 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 per individual available. Hence, territories with less helpers might in fact produce more surviving offspring and thus becoming less social might in fact 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.

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. 2). While it is evident that the direct effect of predation on survival is negative, this result might not be found or the result might be inflated when experimentally increasing/decreasing predator density, because experiments manipulate the total effect of predation 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 effect of predation on survival. In such cases, a DAG can clarify what the expected mechanisms 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), 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. 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 compact 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.

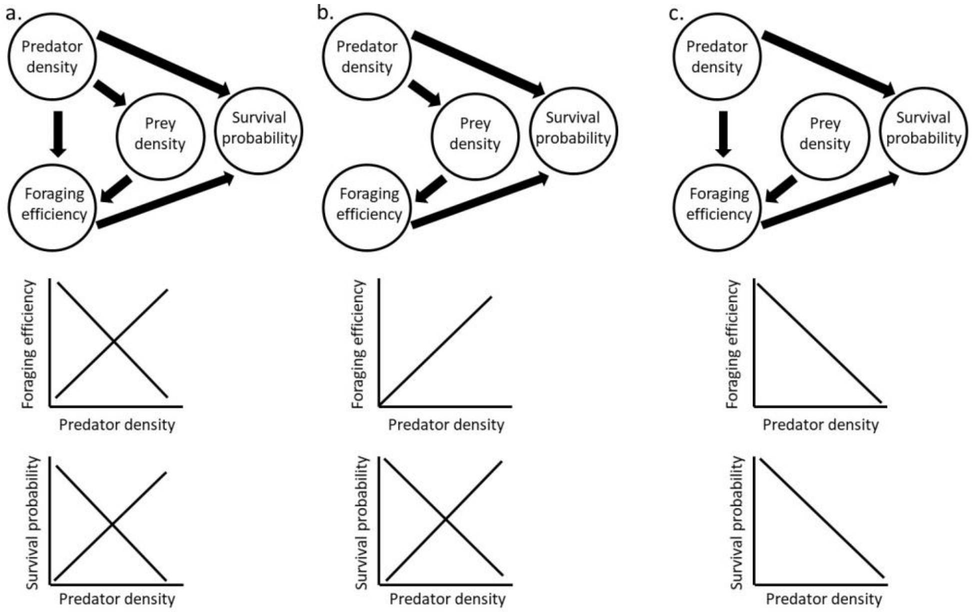

Figure 2: Three directed acyclic graphs about the effects of predator density, prey density and foraging efficiency of prey on the survival probability of prey individuals. a). Shows a causal structure where predator density both directly and indirectly affects foraging efficiency, and directly and indirectly affects 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). Depending on these small changes in the causal structure, different effects of predator density on foraging efficiency and survival probability can be expected. Additionally, depending on the causal structure, it is possible to have an exact expectation on the effects of predator density on foraging efficiency. Yet, in other cases, it depends on the strength and direction of each direct effect what the total effect of predation on foraging efficiency of prey and prey survival is, and therefore it is not possible to have an exact prediction about the strength and direction of the total effect until an analysis has been run.

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. 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. 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.

DAGs as justification for statistical models

DAGs have generally been used to justify which (control) variables should be included in or excluded from statistical models. Yet, in biology so far there seems to have been little notice of the concept of bad controls (Montgomery et al. 2018; Cinelli et al. 2022), and the main justification given for control variables to be included in a model is 'biological relevance'. While biological relevance of course is an important criterion (why include something in a statistical model that is irrelevant for the response variable), it should not be the only one. Especially in long-term study systems where many variables are measured over a long time, it can become tempting to add control variables without thinking about how this really affects the estimated effect in question.

Here, we will discuss two ways in which adding control variables can do harm rather than good, which are commonly called colliders and pipes (see Fig. 3; McElreath 2020; Cinelli et al. 2022). Collider variables are caused by both the predictor variable and the response variable. By adding such variables in a model as a control, a stronger correlation between the predictor and response variable is found (through this collider variable) than what is actually true (or a weaker correlation when the indirect effect has the opposite direction than the direct effect). This thus falsely inflates (or deflates) the estimated effect, sometimes leading to significant results for correlations that in reality are absent, or vice versa. Pipe variables are caused by the predictor variable and cause the response variable. 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. When the

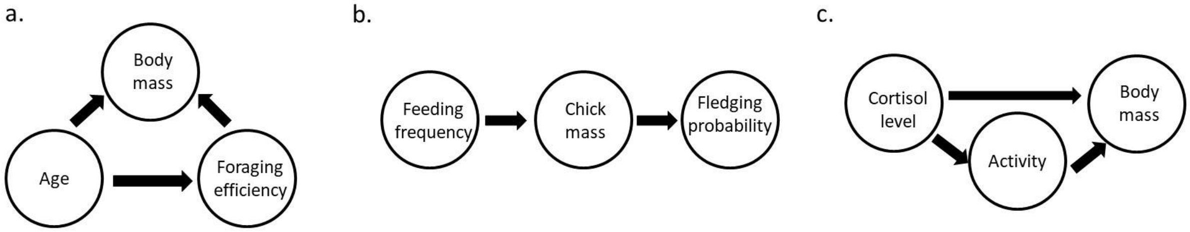

Figure 3: Three directed acyclic graphs showing a collider variable and two pipe variables. a). When a study tries to estimate the effect of age on foraging efficiency and the expected causal structure is like that in a)., it is important to not include body mass, because it is a collider. When body mass is added to the statistical model, a correlation between age and foraging efficiency is opened through body mass, therefore inflating (or deflating) the estimate of the correlation between age and foraging efficiency. A collider thus decreases the accuracy of the estimated effect and can result in wrong conclusions about the strength and direction of an estimated effect. b). When a study tries to understand the effect of feeding frequency on the probability that offspring fledge, and the causal structure is as in b)., 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. c). shows a study that manipulated the cortisol levels in animals and wants to know the effect of this on body mass (and the causal structure is as what is depicted). Often experiments try to estimate total effects of their manipulation. Yet, the manipulation also changes all the intermediate variables. Hence, 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 which part of body mass is caused by a direct effect of cortisol, and which part by an indirect effect of cortisol through activity, 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 assure there is an effect of cortisol on activity.

total effect of the predictor variable should be estimated (for example, in many medicine studies), 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). Path analyses are a subset of 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 (Wright 1934; Streiner 2005). SEMs can additionally be extended so that latent variables (i.e. variables that cannot directly be measured) could be included in a model (e.g. Busana 2021).

While the question whether to include a variable in a statistical model is much simpler in experiments, it nevertheless is 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 cause all individuals to not behave independently), 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. Hence, we would like to convince biologists that hypothesising certain correlations (with a DAG) is better than (implicitly) assuming that every variable is completely independent, unless explicitly (preferably again with a DAG) is hypothesised that variables are independent from each other. This saves research from unnecessary inflations or deflations of estimates caused by colliders and pipes, and thus makes our research better.

On tips how to build a DAG and use causal inference, we would like to refer to Laubach et al. (2021) and Arif and MacNeil (2022) for explanations with biology examples, to McElreath (2020) - including the accompanying youtube videos - for explanations on causal inference especially well-suited to beginners, and to Judea Pearl’s work (e.g. Pearl 1995; Pearl 2009; Pearl 2010; Pearl and Mackenzie 2018; Cinelli et al. 2022) for a more in depth understanding.

Conclusion

With this paper we would like to convince the reader that using DAGs in behavioural ecology and biology in general can be 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. Because of that, DAGs could contribute to solving the replication crisis and could at least make the work of reviewers and researchers doing meta-analyses easier. 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 hypothesising without any 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 found, and at least statistical mistakes caused by expected colliders or pipes can be prevented. Lastly, producing a DAG is a simple and quick exercise, so why would we not add DAGs to our papers?

Data accessibility statement

No data was used for this publication.

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.

Author contributions

Both authors contributed to the conceptualisation and the writing of this paper. Both authors have revised the manuscript and approve of its publication.

References

Arif S, MacNeil MA. 2022. Predictive models aren't for causal inference. Ecology Letters. 25(8):1741–1745. doi:10.1111/ele.14033.

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(5):1054–1063. doi:10.1111/j.1365-2435.2010.01715.x.

Busana M. 2021. Drivers of cooperative breeding and population dynamics in Seychelles warblers [PhD Thesis]. University of Groningen, Groningen, The Netherlands.

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

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. Vol. 52. Elsevier. p. 243–291.

Cockburn A, Russell AF. 2011. Cooperative breeding: a question of climate? Current Biology. 21(5):R195–R197. doi:10.1016/j.cub.2011.01.044.

Jetz W, Rubenstein DR. 2011. Environmental uncertainty and the global biogeography of cooperative breeding in birds. Current Biology. 21(1):72–78. doi:10.1016/j.cub.2010.11.075.

Koenig WD. 1981. Reproductive success, group size, and the evolution of cooperative breeding in the acorn woodpecker. The American Naturalist. 117(4):421–443. doi:10.1086/283726.

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.

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(1943):20202815. doi:10.1098/rspb.2020.2815.

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

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(3):760–775. doi:10.1111/ajps.12357.

Pearl J. 1995. Causal diagrams for empirical research. Biometrika. 82(4):669–710.

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

Pearl J. 2010. An introduction to causal inference. The International Journal of Biostatistics. 6(2). doi:10.2202/1557-4679.1203.

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

Streiner DL. 2005. Finding Our Way: An Introduction to Path Analysis. Canadian Journal of Psychiatry. 50(2).

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(9):152. doi:10.1007/s00265-018-2566-7.

Wright S. 1934. The method of path coefficients. The Annals of Mathematical Statistics. 5(3):161–215.