Detecting and quantifying social transmission using network-based diffusion analysis

Matthew J. Hasenjager, Ellouise Leadbeater & William Hoppitt*

Summary

1. Although social learning capabilities are taxonomically widespread, demonstrating that freely interacting animals (whether under wild or captive conditions) rely on social learning has proved remarkably challenging.
2. Network-based diffusion analysis (NBDA) offers a means for detecting social learning within such freely interacting groups. Its core assumption is that if a target behaviour is being socially transmitted, then its spread should follow the pattern of connections in a social network that reflects opportunities for social learning.
3. Here, we provide a comprehensive guide for using NBDA. We first present the types of questions that NBDA can address, as well as introduce its underlying mathematical framework. We then guide researchers through the process of: (i) selecting an appropriate social network to address different research questions; (ii) determining which NBDA variant should be used; and (iii) incorporating other variables that may impact asocial and social learning. We then discuss how to interpret the output of an NBDA model, as well as provide practical recommendations for model selection.
4. Throughout the manuscript, we highlight extensions to the basic NBDA framework. These include the incorporation of dynamic network structures to capture changes in social relationships during the diffusion process, and estimating information flow across multiple types of social relationship using a multi-network NBDA.

5. Alongside this information, we provide worked examples and tutorials demonstrating how to perform analyses using the newly developed *NBDA* package written in the R programming language.

1 Introduction

Over recent decades, a vast body of research has revealed that social learning capabilities are widespread across the animal kingdom (Heyes 1994; Hoppitt & Laland 2013), and that social transmission can result in culture-like phenomena in non-humans (Laland & Galef 2009). And yet, although social learning is predicted to be adaptive across diverse contexts (Rendell *et al*. 2010), demonstrating that animals rely on social learning in the wild has remained notoriously challenging (Reader 2004; Laland & Janik 2006; Schuppli & van Schaik 2019). The main difficulty lies in disentangling social and asocial influences on learning in contexts where animals are free to interact (or not) with each other and with their environment. For instance, a historically common approach has been the ethnographic method, in which social learning can be inferred as the cause of behavioural variation across populations only if genetic and ecological explanations (e.g. differential opportunities for individual learning resulting from spatially heterogeneous resources) can first be ruled out (Whiten *et al*. 1999). However, this conservative approach precludes investigation of how such factors may interact (Laland & Janik 2006; Wild *et al*. 2019), and may systematically underestimate the prevalence of socially transmitted behaviours in the wild (Schuppli & van Schaik 2019). As such, researchers have sought to develop alternative methods for inferring social learning in circumstances where only observational data are generally available.

Network-based diffusion analysis (NBDA) is just such an approach (Franz & Nunn 2009; Hoppitt, Boogert & Laland 2010). NBDA follows the assumption that individuals are more likely to learn from one another if they frequently associate or interact (Coussi-Korbel & Fragaszy 1995). Thus, social transmission is inferred if the spread of a novel behaviour pattern follows the connections in a social network that reflects opportunities for social learning (Hoppitt 2017). Moreover, rather than assume that a given behaviour diffuses entirely through either social or asocial processes, NBDA estimates the strength of social learning relative to asocial learning. NBDA therefore provides researchers with a means by which to evaluate the impact of different factors (e.g. genetic, phenotypic, ecological) on both social and asocial learning (Hoppitt, Boogert *et al*. 2010).

Since its initial development, NBDA has enabled investigation of social transmission across diverse taxa, including cetaceans (Allen, Weinrich, Hoppitt & Rendell 2013; Wild *et al*. 2019), primates (Hobaiter, Poisot, Zuberbühler, Hoppitt & Gruber 2014), songbirds (Aplin, Farine, Morand-Ferron & Sheldon 2012), and teleost fish (Atton, Hoppitt, Webster, Galef & Laland 2012; Hasenjager & Dugatkin 2017). Such studies have resulted in several extensions to the basic NBDA model, including the use of dynamic networks that take into account changes in social relationships over time (Hobaiter *et al*. 2014), inclusion of multiple network types to evaluate how social transmission may be influenced by different forms of social relationship (Farine, Aplin, Sheldon & Hoppitt 2015), and incorporation of learning tasks that involve multiple steps to complete (Atton *et al*. 2012). However, many of these extensions have not been publicly available until recently (Hoppitt, Photopoulou, Hasenjager & Leadbeater 2019), and references to their implementation are scattered across the literature. Here, we aim to provide a comprehensive and up-to-date resource for researchers interested in using NBDA, and to illustrate the use of the newly developed *NBDA* package for R (Hoppitt *et al*. 2019). In the Supporting Information, we also provide tutorial R scripts showing how to implement the analyses using the *NBDA* package.

2 Initial considerations

2.1 What types of questions can NBDA address?

There are typically two primary aims a researcher might have when employing NBDA. The first is to evaluate the strength of evidence for social transmission, and to quantify its impact. For example, Allen *et al*. (2013) used NBDA to investigate the spread of an innovative foraging behaviour—lobtail feeding—over a 27 year period through a population of humpback whales (*Megaptera novaeangliae*). Their analysis revealed that the lobtail technique did not diffuse through the population at random, nor through individual learning alone. Rather, the order in which whales acquired this behaviour was predicted by the strength of their social connections to knowledgeable individuals (i.e. those that had previously learned the lobtail technique). In other words, frequently associating with individuals that practised lobtail feeding provided whales with opportunities to learn this behaviour, such that an estimated 45–85% of whales that acquired lobtail feeding did so through social transmission (Allen *et al*. 2013).

The second main application of NBDA is to identify the typical pathways of information flow through a group. In this sense, networks represent hypotheses about how information is expected to spread. For instance, a researcher could compare NBDA models fitted with alternative networks—e.g. networks quantifying affiliative *versus* agonistic interactions—to determine which type of interaction or social relationship is most important in facilitating social learning (e.g. Kulahci *et al*. 2016). In addition to networks built from empirical data on social relationships , researchers can also construct theoretically derived networks that represent hypothesized pathways of information flow. For example, Atton, Galef, Hoppitt, Webster, and Laland (2014) presented novel foraging tasks to shoals of three-spined sticklebacks (*Gasterosteus aculeatus*) in which individuals were familiar with some shoal mates, but not with others. To determine whether familiarity facilitated information flow between sticklebacks, a binary network that indicated whether each pair of individuals was familiar (1) or not (0) was constructed, and included in an NBDA. This familiarity network was found to better predict the order in which sticklebacks both discovered and solved the foraging task than either a network of shoaling associations, or a homogeneous network in which all individuals were connected with a strength of 1. In other words, patterns of familiarity directed patterns of social learning within these shoals.

2.2 What types of data does NBDA require?

There are two main components of an NBDA model. The first component is data on the order or timing with which individuals acquire a behavioural trait of interest (i.e. diffusion data; Section 5). Under certain circumstances, such as in laboratory experiments or through use of automated tracking technology, researchers might have highly resolved data on the exact time that each individual first performed the target behaviour. In other cases, the available data might be much less detailed. For instance, it may only be possible to state that an individual first performed the trait at some point within a certain timespan. The resolution of the diffusion data determines which NBDA variants can be used, though other factors are also important when making this selection (see Section 5).

The second main component of the model is a social network (or networks) that is thought to reflect opportunities for social learning. The reasoning here is that if the target behaviour spreads through social transmission, then we would expect this diffusion to follow the pattern of connections in such a network. There are many types of social network a researcher could include in an NBDA. For example, association networks indicate the propensity for pairs of individuals to co-occur in space and time. Such a network might be used to estimate the probability that one individual's performance of the target behaviour is observed by another (Hoppitt 2017). Another possibility is to include networks that capture particular forms of interaction that are known or suspected to transmit information—e.g. honeybees (*Apis mellifera*) searching for a novel foraging site can acquire its spatial coordinates by following the waggle dances of successful foragers (Grüter & Farina 2009). The most appropriate choice of network will often depend on the research question(s) (see Section 4). A researcher can also include other predictor variables, such as sex, body size, or personality type, that may influence asocial and/or social learning (see Section 6).

3 The basic NBDA model

An understanding of the basic NBDA model is key to understanding and interpreting the various forms of NBDA and its extensions, so we first present the mathematical formulation of the model in its most fundamental form and explain it in some detail. The basic NBDA model can be expressed as

$$\lambda_i(t) = \lambda_o(t)\big(1 - z_i(t)\big)\left(s\sum_{j=1}^{N} a_{ij} z_j(t) + 1\right)$$

Eqn. 1

where $\lambda_i(t)$ is the rate at which individual *i* acquires the target behaviour as a function of time, $\lambda_o(t)$ is a baseline rate function, $z_i(t)$ is the 'status' of individual *i* at time t, (1 = informed; 0 = naïve), *N* is the number of individuals in the population, *s* is a parameter determining the strength of social transmission, and a_{ij} is the network connection from *j* to *i*. The NBDA model can be expanded in various ways beyond the model defined in Eqn. 1, which we describe and define in the Sections below.

The $\left(1 - z_i(t)\right)$ term ensures that only naïve individuals can learn, since when *i* is informed, $z_i(t) = 1$, so $\left(1 - z_i(t)\right) = 0$ and consequently $\lambda_i(t) = 0$. The rate at which a naïve individual acquires the target behaviour by social transmission is assumed to be proportional to $\sum_{j=1}^{N} a_{ij} z_j(t)$, the total connection of *i* to informed individuals at time *t*. Consequently, *s*, a parameter fitted to the data, estimates the rate of transmission per unit connection relative to the rate of asocial learning of the target behaviour. Depending on the type of network used, *s* can sometimes be interpreted in a more specific manner: e.g. the rate of social transmission from an informed to naïve individual during periods when they are associating, relative to the rate of asocial learning (see Section 4). *s* = 0 represents the case of no social transmission: the null model of interest if a researcher is quantifying the evidence for social transmission, in which the rate of acquisition is determined by the rate of asocial learning alone. We refer to models in which *s* is constrained to 0 as "asocial learning models" or "asocial" models, which should be taken as shorthand for a model with asocial learning *only*, since asocial learning is also occurring when *s* > 0. Finally, the baseline rate function, $\lambda_o(t)$, (terminology adapted from survival analysis; see Moore 2016) determines how the rate of learning generally changes over time. Different types of NBDA make different assumptions about the shape of $\lambda_o(t)$, as explained in Section 5.

4 Different types of networks

The social network can be thought of as the key predictor variable in an NBDA. In principle, one can use any type of social network that specifies a non-negative connection in each direction for each dyad. However, different types of network may be appropriate depending on the aim of the NBDA, and the exact meaning of the *s* parameter may vary depending on the type of network (Hoppitt 2017).

4.1 Networks for detecting and quantifying social transmission

When the goal of an NBDA is to simply to detect and quantify social transmission, there are many different types of social network a researcher can choose. For information on techniques for constructing empirical social networks, see (Croft, James & Krause 2008; Whitehead 2009; Farine & Whitehead 2015). The simplest network one could input is a binary or unweighted network, in which individuals that are socially connected share a link ($a_{ij} = a_{ji} = 1$), whereas those that are not remain unlinked ($a_{ij} = a_{ji} = 0$). In this case, *s* estimates the rate of social transmission from an informed individual to a socially connected naïve individual, relative to the rate of asocial learning.

Perhaps the most obvious choice for NBDA is an association network, where a_{ij} estimates the proportion of time *i* spends associating with *j*. Ideally, one would assume that individuals can only socially learn from one another when they are associating. For this assumption to be reasonable, the criterion for *i* to be recorded as associating with *j* has to be specified at the appropriate spatial scale. Individuals recorded as associating must be within observation distance, whereas individuals recorded as not associating must tend to be at a distance at which observation is impossible or unlikely (Hoppitt 2017). For example, Allen *et al*. (2013) used an association network to track the spread of a novel feeding technique through a population of humpback whales (*M. novaeangliae*). Since whales needed to be within two body lengths to be recorded as 'associating' and the study was conducted over an area of approximately 1000 square miles, the aforementioned assumption seems reasonable. In such cases, *s* can be interpreted as the rate of social transmission from an informed to naïve individual during periods when they are associating, relative to the rate of asocial learning (Hoppitt 2017).

In contrast, other studies on captive birds (Boogert, Reader, Hoppitt & Laland 2008; Boogert, Nightingale, Hoppitt & Laland 2014) have used a criterion based on proximity (e.g. nearest neighbour) within an enclosure of a few square meters, such that dyads observed as not associating are still within observation distance. We refer to the former as 'large-scale association networks' and the latter as 'small-scale association networks'. When using small-scale association networks, there is no guarantee that *s* can be interpreted in the same specific manner as for large-scale association networks. In other words, *s* may not necessarily provide the rate of social transmission during periods in which individuals are able to observe knowledgeable individuals. Rather, use of a proximity network represents the hypothesis that individuals are more likely to learn from demonstrators that they tend to be found near to than from those that are more spatially distant (see Section 4.2 for further discussion).

An alternative type of network is an observation network, where a_{ij} represents the number of opportunities *i* has had to observe *j* performing the target behaviour. Such a network is perhaps the most direct method for detecting and quantifying social transmission. If an observation network is to be used, it makes sense to use a dynamic (time-varying) version of the observation network, so we delay further discussion of observation networks until Section 4.4.

4.2 Networks for establishing the typical pathways of information transfer

Another aim a researcher might have is to elucidate the typical pathways of diffusion by comparing the fit of alternative NBDA models using different networks (Franz & Nunn 2009;

Hoppitt 2017). The result of this process would suggest the types of relationship that are important in providing the opportunity and/or motivation to observe and learn from others. For example, Kulahci *et al.* (2016) found in a study on ravens (*Corvus corax*) that a social network based on affiliative interactions, such as allopreening and food sharing, predicted the spread of a novel foraging behaviour better than networks based on aggressive interactions and proximity. Alternative models can be fitted and compared using Akaike's Information Criterion corrected for sample size (AICc, see Section 9) and whichever network best approximates the true pathway(s) of transmission is likely to be favoured (Hoppitt 2017). A researcher will often have the combined aim of detecting and quantifying social transmission, and can include an asocial model (s = 0) in the model comparison. If no network provides a substantially better fit than the asocial model, there is little evidence for social transmission following any of the networks included in the comparison. If there is evidence for social transmission, the best fitting model can be used to generate estimates of the strength of social transmission (s).

A number of types of networks might be included in such an analysis. For instance, proximity networks are derived from data on spatial relationships among individuals, with a common example being an association network that estimates the propensity of pairs of individuals to co-occur in space and time (Franks, Ruxton & James 2010; Farine & Whitehead 2015). If the criterion used for association is thought to approximate the conditions for observation, then s can be interpreted in the manner described for large-scale association networks in Section 4.1; i.e. the rate of social transmission from informed to naïve individuals during periods in which the latter can observe the former, relative to the rate of asocial learning. However, if proximity networks are collected on a small spatial scale, and thus may not fully encompass opportunities for observation, they rather represent the more general hypothesis that individuals in close proximity will tend to learn from one another more often than those that are more spatially distant (Hoppitt 2017). Interaction networks quantify the rate at which dyads interact, or show a specified type of interaction (e.g. allopreening, fights) (Croft, James & Krause 2008). When used in an NBDA, interaction networks represent the hypothesis that a particular interaction type predicts the rate at which individuals learn from one another. As such, they are not *a priori* preferable to proximity networks for an NBDA. Instead, interaction networks can be thought of as a competing set of hypotheses that can be compared empirically, both to one another and to proximity networks.

The estimate of *s* yielded from an interaction network or small-scale proximity network is more general and abstract than for large-scale association networks: *s* estimates the rate of social transmission from informed to naïve individuals per unit of network connection, relative to the rate of asocial learning. In such cases, *s* may be difficult to interpret biologically and may also not be comparable across networks with different scales. Such circumstances may make it difficult to gauge the importance of social transmission. A solution is to convert the estimate of *s* into the estimated proportion of learning events that occurred by social transmission as opposed to asocial learning (see Section 7.5).

4.3 Including transmission weights

The standard NBDA model implicitly assumes that all individuals perform the target behaviour at a similar rate once they have learned it. However, it may be that some individuals perform the behaviour more often, and thus socially transmit the behaviour more effectively, than those that perform it less frequently. If a researcher has a measure of the rate at which individuals performed the target behaviour during the course of the diffusion, this information can be included in the model as transmission weights, W_j, by replacing a_{ij} with $W_j a_{ij}$. Thus, the rate of transmission is assumed to be proportional to rate of performance. W_j should be an estimate of the rate at which the target behaviour is performed by *j* once it is informed, so ideally $W_j = n_j/(T - t_j)$, where n_j is the number of performances, *T* is the total time of the diffusion, and t_j is the time at which *j* acquired the target behaviour. *s* now estimates the rate of social learning per unit connection multiplied by performance rate, relative to asocial learning. However, we can be more specific if a_{ij} is a large-scale association network, and if we assume that *i* has a probability of learning the target behaviour from each observation. Since $W_j a_{ij}$ estimates the rate at which *i* observes *j* perform the target behaviour, *s* estimates the probability of learning each time a naïve individual observes an informed individual perform the behaviour, relative to the rate of asocial learning (Hoppitt 2017). Hoppitt (2017) suggests that if transmission weights are available, they should be included in the analysis if they improve model fit (i.e. decrease AICc, see Section 9).

4.4 Dynamic networks and observation networks

The basic NBDA model defined in Eqn. 1 assumes that the social network does not change over the course of the diffusion, i.e. that it is a 'static network'. However, under some circumstances, the structure of a network may undergo substantial changes during the diffusion process, e.g. as a result of demographic processes or shifting dominance ranks. By extending the basic NBDA model so that it can incorporate a time-varying or 'dynamic network', we can include these temporal changes in the analysis (Hobaiter *et al.* 2014). This is done simply by replacing a_{ij} (i.e. the connection from individual *j* to *i*) with $a_{ij}(t)$, the connection from individual *j* to *i* at time *t*:

$$\lambda_i(t) = \lambda_o(t)\big(1 - z_i(t)\big)\left(s\sum_{j=1}^{N} a_{ij}(t)z_j(t) + 1\right)$$

Eqn. 2

Therefore, the static network becomes a special case where $a_{ij}(t) = a_{ij}$.

However, we advise caution when considering whether to include an association or interaction network as a dynamic network in an NBDA. If the network is broken down into time periods that are too small, apparent changes in network structure may simply be the result of sampling error. In addition, by breaking up the network data into smaller chunks, estimates of connection strength may become less precise. Therefore, we suggest that researchers only use a dynamic association or interaction network if there is sufficient data in each time period to ensure precise estimates of network ties (Hoppitt & Farine 2018).

In contrast, it usually makes sense to use a dynamic observation network rather than a static observation network. If one wishes to detect and quantify social transmission, then ideally, the researcher would like a complete record of when the target behaviour was performed, by whom, and who observed each performance. It is possible to obtain data close to this level of resolution in cases where the target behaviour is only performed in a specific location (or locations) that can be monitored closely. For example, Hobaiter *et al*. (2014) used NBDA to analyse the diffusion of moss sponging—the use of pieces of moss to obtain water from holes in trees—in chimpanzees (*Pan troglodytes*). They were able to use a dynamic observation network because the initial spread of this behaviour was documented at only a single water hole. Researchers might obtain a similar level of resolution using an artificial foraging task that can be monitored closely (e.g. van de Waal, Renevey, Favre & Bshary 2010), or when information transfer is largely restricted to particular locations, such as the honeybee ‘dancefloor’ (Leadbeater & Hasenjager 2019; Box 4).

In a dynamic observation network, $a_{ij}(t)$ is the number of times *i* has observed *j* perform the target behaviour prior to *t*. In practise, it will usually be sufficient to specify the network only at the times at which each acquisition event occurred. The corresponding static observation network would be where a_{ij} gives the total number of times *i* observed *j* performing the behaviour. However, the latter network does not allow for the actual time course of observation and acquisition events. For example, imagine a group of three chimpanzees (A, B and C) learning moss sponging by social transmission (see Fig. 1). A learns how to moss sponge first, and is observed performing it three times by B, after which B learns this behaviour. Next, C observes A perform the moss sponging behaviour four times then learns the behaviour. The static observation network (taken from Event 3 in Fig. 1) would represent this pattern as $a_{A,B} = 3$ and $a_{A,C} = 4$. Thus, the static network predicts that C will learn before B, whereas in reality we would expect B to learn first, as would be predicted by the dynamic observation network.

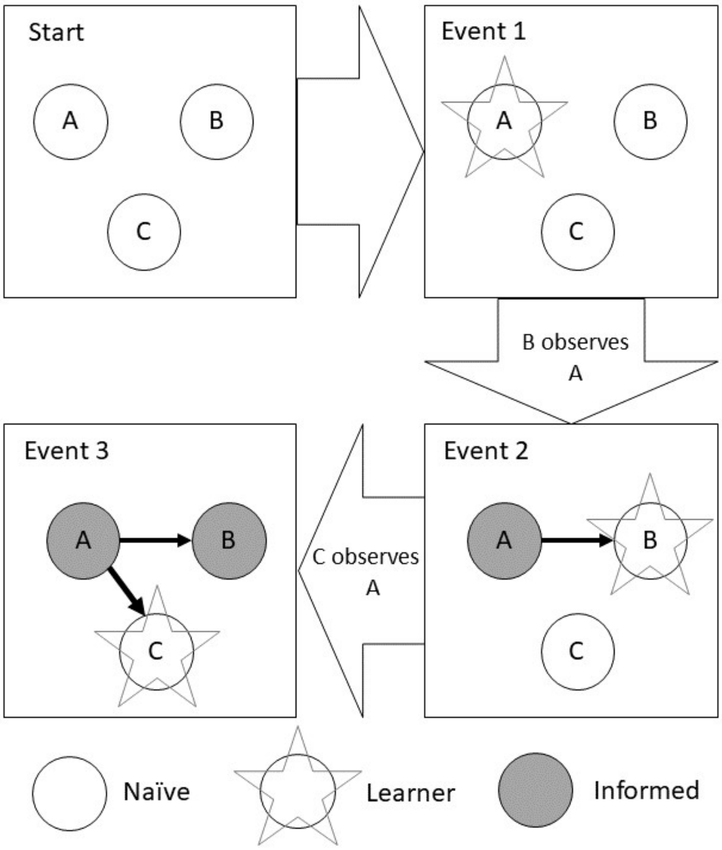


Fig. 1. An example showing the predictive power of a hypothetical dynamic observation network whereby three individuals (A, B and C) learn to perform a particular behaviour. Arrows represent social transmission events. See main text for explanation.

Use of a dynamic observation network has the advantage that it can infer social learning if the chance order in which individuals observe the behaviour predicts the order of diffusion, even if there is little or no underlying structure in the social network. Unfortunately, *s* does not have a straightforward interpretation (Hoppitt 2017), so we suggest that researchers obtain an estimate of the proportion of learning events that occurred by social transmission as an interpretable measure of its strength (see Section 7.5).

In contrast, it will often not make sense to include an observation network alongside association or interaction networks in a model comparison meant to establish the typical pathways of information transfer. The goal in such an analysis is, in part, to find a network that best approximates opportunities for observation and social learning. The observation network bypasses this approximation since it is a direct quantification of these opportunities. However, it may make sense for a researcher to compare models with different observation networks representing different types of observations (see Box 4 for an example) if they wish to know which of these pathways (or combination of pathways) best explains the diffusion data (see Section 9). See (Hoppitt 2017; Hobaiter *et al.* 2014) for further recommendations on using a dynamic social network.

4.5 Non-visual social learning and learning from products

Thus far we have assumed that social transmission of novel behaviour occurs when one individual observes another performing it. The term ‘observes’ should not necessarily be taken to mean restricted to the visual modality, but rather should be interpreted in a broad sense, where behaviour can be observed in any modality. Familiar examples include the many species that learn vocalizations by listening to others—e.g. whale song (Noad, Cato, Bryden, Jenner & Jenner 2000). The recommendations provided above should therefore be considered in light of the modality in question. For instance, a large-scale association network needs to reflect the scale over which social learning can occur—e.g. auditory cues may travel much further than visual ones. Furthermore, it is well documented that behaviour can be transmitted when a naïve individual encounters the products of an informed individual’s performance of that behaviour (e.g. Terkel 1995; Leadbeater & Chittka 2008). In such cases, the predictive power of a network in an NBDA will depend on the extent to which it approximates *i*’s opportunities to encounter the products of *j*’s behaviour. To date, we are aware of no uses of NBDA that are targeted at behaviour transmitted through product learning, nor through non-visual transmission, but these remain potential uses of the method.

5 Diffusion data and types of NBDA

In the context of NBDA, diffusion data refers to the pattern of spread of the target behaviour, and provides the response variable for the analysis. There are two main variants of NBDA: order-of-acquisition diffusion analysis (OADA), which takes as data the order in which individuals acquired the target behaviour, and time-of-acquisition diffusion analysis (TADA), which takes as data the times of acquisition. TADA can be further subdivided into a version that treats time as a continuous variable (continuous TADA or 'cTADA'), and a version that takes time as a discrete variable split into units (discrete TADA or 'dTADA'). Here, we first explain how a researcher should decide between the different variants.

5.1 OADA, cTADA, or dTADA?

The original form of NBDA was the dTADA (Franz & Nunn 2009), with the OADA and cTADA being proposed soon afterwards (Hoppitt, Boogert *et al*. 2010). All forms can be expressed in the form given in Eqn. 1 and 2. Choice of OADA versus cTADA versus dTADA depends on the diffusion data available and the assumptions one is willing to make about the baseline rate function, $\lambda_o(t)$. Here, we discuss the latter issue first.

OADA makes no specific assumptions about the shape of $\lambda_o(t)$, but only assumes that this function is the same for every individual in the diffusion (to understand why, see Box 2). In contrast, TADA requires the researcher to make assumptions about the form of $\lambda_o(t)$, and fit parameters controlling its shape. When these assumptions are met, TADA offers more statistical power than OADA (Hoppitt, Boogert *et al.* 2010). This is especially true when the network is highly homogeneous (i.e. when it is densely connected with relatively little variation in connection strength). Indeed, when the network is completely homogeneous—that is, if all possible connections exist and are of equal strength—OADA will be unable to distinguish social transmission from asocial learning since all orders of acquisition would be equally likely in both models.

In the simplest case, one can fit a TADA that assumes a constant baseline hazard rate of learning, $\lambda_o(t) = \lambda_o$, with an extra parameter, λ_o, fitted to the data (Franz & Nunn 2009; Hoppitt, Boogert *et al*. 2010). However, if the rate at which individuals learn asocially increases over time, this can cause a spurious positive result for social transmission in a TADA (Hoppitt, Kandler, Kendal & Laland 2010). For example, if a novel foraging task is presented to a group of animals, they might initially exhibit neophobic responses towards it; as this effect fades over time, the rate at which they learn to solve the task asocially will likely increase. Conversely, if $\lambda_o(t)$ decreases over time—e.g. if the resources necessary to learn the behaviour begin to deplete—this can reduce the power of TADA to detect social transmission. Fortunately, TADA can be modified to have a non-constant baseline rate. Any positive function can be specified for $\lambda_o(t)$. However, the *NBDA* package has two functions built-in which will be sufficient in most cases. One corresponds to a gamma distribution of latencies under asocial learning (Hoppitt, Kandler *et al.* 2010), and the other to a Weibull distribution of latencies (a common choice in survival analysis; Moore 2016). Both offer flexible modelling of $\lambda_o(t)$ with a shape parameter that allows for the possibility of systematically increasing, constant, and systematically decreasing baseline functions.

If instead $\lambda_o(t)$ fluctuates unpredictably, this can badly reduce the power of TADA, but OADA will remain unaffected (Hoppitt, Boogert *et al.* 2010). For example, if a field experiment is conducted in which a population of animals is presented with a foraging task, there may be many factors influencing the rate at which individuals in the population solve the task at any given time, such as weather conditions, predation risk, the presence of prey, or diurnal rhythms. In principle, if all the variables causing fluctuations in the baseline acquisition rate can be identified and included in the model (see Section 6), TADA could still be appropriate. However, OADA is a far easier option.

So what does this mean for a researcher choosing between OADA, cTADA and dTADA? If the researcher only has data on the order in which individuals acquired the behaviour, then OADA must be used (Fig. 2). However, if data on exact times of acquisitions are available, there is a choice between OADA and cTADA. If it is likely that $\lambda_o(t)$ fluctuates unpredictably, then OADA is again to be preferred. However, if the researcher is confident that the baseline rate function can be assumed to be constant or can be modelled as a potentially systematically increasing or decreasing function, then cTADA is to be initially preferred, since it offers more statistical power under these circumstances. In such cases, we recommend that models with both constant and Weibull (and/or gamma) baseline functions be fitted, and the best fitting baseline function be used to generate parameter estimates (see Section 9.2). However, if very different results are obtained from models with different baseline functions (e.g. strong support for asocial learning versus strong support for social transmission), it suggests that the analysis is dominated by the time course of events as opposed to the pattern of diffusion through the network. For an example of such a situation, see Tutorial 7 in the Supporting Information. In such cases, we recommend that researchers switch to OADA since it is invariant to the shape of $\lambda_o(t)$, and sensitive only to the pattern of diffusion through the network. The above recommendations are summarized in Fig. 2.

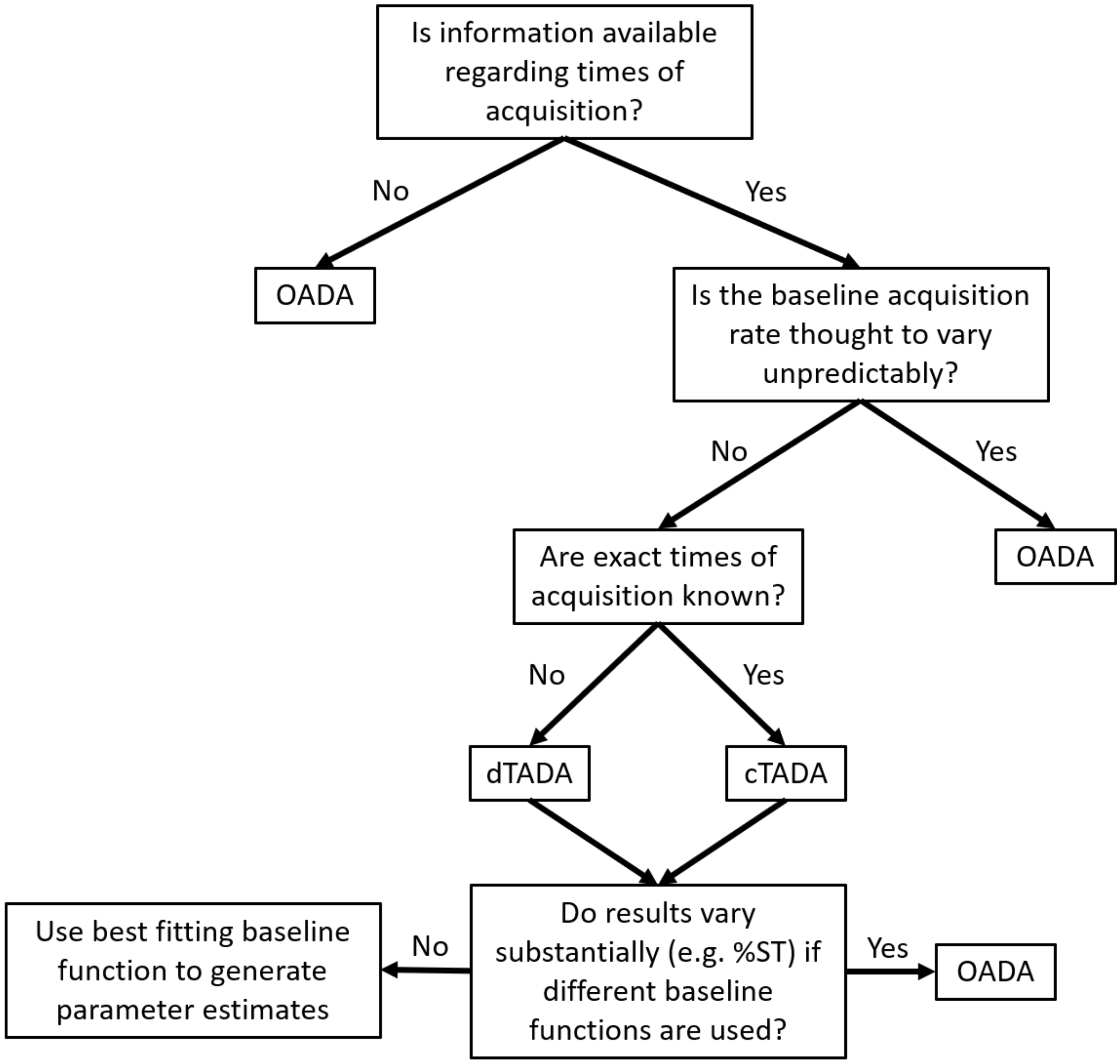


Fig. 2. Flowchart for selecting the appropriate NBDA model.

In other cases, some information on time of acquisition may be available, but the exact times are not known. One reason this could occur is if the population is sampled periodically, giving a temporal snapshot of who is informed at any given time. The researcher then knows only the time period in which each individual acquired the behaviour. The natural choice here is a dTADA, though if the sampling periods are sufficiently frequent, it may be possible to resolve the order of acquisition (a few ties can be accommodated, see Section 10.3). In such cases, there is a choice between OADA and dTADA, and the researcher can use the same reasoning for choosing between OADA and cTADA described above (Fig. 2). Another reason we might have inexact times of acquisition is if there is observation error in the recorded time of acquisition. Franz and Nunn (2010) find that this can inflate the false positive error rate for social transmission in a dTADA when the time units are small. Since the results of a dTADA converge on the results of cTADA as the time units get smaller, observation error will also inflate the false positive rate in a cTADA. However, by using dTADA with a sufficiently long time unit, the problem is alleviated. Franz and Nunn (2010) provide a rule of thumb that there should be ≥ 50% probability that an individual who has acquired the trait will be observed performing it in a given time unit. This can be checked by calculating the proportion of time units in which individuals are observed performing the behaviour following the time unit in which their performance was first observed.

If TADA is chosen, it is important that the times entered into the model are cumulative times that include only those periods during which it was possible for the animals to learn the behaviour. For example, imagine a foraging task presented to a group of animals at 9–10 a.m. each day. If individual A learns to solve the task 5 minutes into the session on the second day, it would be attributed 65 mins as its time of acquisition, since A could only solve the task when the task was available to be solved.

Note that in a TADA, while evidence for a model of social transmission over an asocial model supports the presence of social transmission, it does not necessarily constitute evidence that transmission follows the network provided. Therefore, we recommend researchers include an additional model (or set of models) in which the social network is replaced with a homogeneous network (connections of 1 for all dyads). If the homogeneous network is favoured over the measured social network (see Section 9), it implies either that transmission occurs homogeneously amongst the group, or, more likely, that the measured social network is substantially different from the real pathways of transmission (Whalen & Hoppitt 2016).

5.2 Modelling multiple diffusions

Thus far we have assumed that the researcher has data from a single diffusion, i.e. the spread of a single behaviour pattern through a single population or group. But a researcher can also combine data from multiple diffusions, such as the same foraging task presented to different groups, into a single NBDA model. There are a number of ways that this can be done. Let us first extend the NBDA model from Eqn. 2 to multiple diffusions:

$$\lambda_{il}(t) = \lambda_{ol}(t)\big(1 - z_{il}(t)\big)\left(s\sum_{j=1}^{N} a_{ijl}(t)z_{jl}(t) + 1\right)$$

Eqn. 3

Here, subscript *l* denotes the diffusion number (i.e. $\lambda_{il}(t)$ is the rate of acquisition for individual *i* in diffusion *l*).

The first option is to fit an OADA in which the shape of the baseline rate, $\lambda_{ol}(t)$, is unspecified and allowed to vary among diffusions. In this case, the analysis is sensitive only to the order within each diffusion. However, this approach ignores the possibility that the spread of behaviour 'takes off' at different times in different diffusions. For instance, imagine a study consisting of three diffusions in which everyone in group 1 learns in the first 5 mins, everyone in group 2 learns in the middle of the experiment, and everyone in group 3 learns at the end of the experiment. This pattern is consistent with innovations arising at different times in each group and rapidly spreading through each group, but would be ignored by the OADA described above, thus resulting in lower statistical power to detect social transmission.

A researcher could instead use a TADA if the assumptions are reasonable (see Section 5.1). In principle, one could fit a TADA with separate $\lambda_{0l}(t)$ fitted to each diffusion. However, this results in a rather complex model and we suggest this route be avoided (this option is not supported in the *NBDA* package). A preferable option is to assume that the shape of the baseline function is the same in all diffusions, $\lambda_{0l}(t) = \lambda_0(t)$. One can then control for the possibility of a different rate of asocial learning in each group by including a 'group' individual level variable (see Section 6). However, this requires a choice of baseline function, and as recommended above, if the results are not robust to this choice then OADA is to be preferred (Fig. 2).

A compromise is to assume that $\lambda_{0l}(t) = \lambda_0(t)$, but to leave $\lambda_0(t)$ unspecified. This amounts to fitting an OADA in which all diffusions are treated as a single diffusion. Thus, the order of acquisition is specified across all diffusions, but individuals from different diffusions are not connected in the network. More generally, a researcher can pool diffusions into 'strata', and assume that diffusions within the same 'stratum' have the same baseline rate function: i.e. $\lambda_{0l}(t) = \lambda_{0S(l)}(t)$, where *S*(*l*) is the stratum for diffusion *l*. In this case, the researcher treats each stratum as a single diffusion (again providing zero connections for dyads in different diffusions) (Hoppitt & Laland 2013). We refer to this model as a 'stratified OADA'. As with a TADA, a 'group' individual level variable can be included to control for the possibility that groups differ in their asocial acquisition rate.

In a multiple diffusion analysis using TADA or stratified OADA, comparing a network-based model of social learning to an asocial model does not test whether the diffusion follows the network within each group. For example, if everyone in each group learns homogeneously, the network provided to the analysis is likely to be a reasonable approximation of the pathway of learning, due to the zero connections between individuals in different groups. Therefore, the network model is likely to provide a better fit than the asocial model. To test whether the diffusion in fact follows the social network within each group, a researcher must fit an alternative model in which the connections within each group are set to 1 and, if using a stratified OADA, connections between groups are set to 0 (we term this network the 'group network'). If the social network provides a substantially better fit than the group network, this suggests that the social network approximates the pathways of learning within each group. If instead the group network is favoured over both the asocial model and the network model, then the researcher has evidence of social transmission within each group, but no evidence that transmission follows each group's social network.

So far we have assumed that researchers are analysing multiple diffusions on different sets of individuals. Alternatively, it could be that individuals are present in more than one diffusion, e.g. if different foraging tasks are presented to the same group. In such cases, the rate of acquisition for each individual is likely to be correlated across diffusions. This can be accounted for by including random effects. The *NBDA* package allows this to be done in an OADA using the *coxme* package (Therneau 2009), using the technique described by (Hoppitt, Boogert *et al*. 2010).

5.3 Seeded demonstrators

In many diffusion studies, some individuals start the diffusion already informed, often because they are trained to perform the target behaviour and 'seeded' in the diffusion. Such individuals are easily accounted for in an NBDA by simply setting status, $z_j(t)$, to 1 for all t > 0. The *NBDA* package easily allows for incorporating such information (see Tutorial 1 in the Supporting Information).

6 Individual level variables

NBDA can be expanded to include other predictor variables that might influence the rate of social transmission and/or asocial learning, termed 'individual level variables' (ILVs) (Hoppitt, Boogert *et al*. 2010). We expand Eqn. 2 to include the effects of *V* continuous or binary ILVs as follows:

$$\lambda_i(t) = \lambda_o(t)\big(1 - z_i(t)\big)\left(s\exp(\Gamma_i)\sum_{j=1}^{N} a_{ij}(t)z_j(t) + \exp(\mathrm{B}_i)\right)$$

$$\mathrm{B}_i = \sum_{k=1}^{V} \beta_k x_{k,i}$$

$$\Gamma_i = \sum_{k=1}^{V} \gamma_k x_{k,i}$$

Eqn. 4

Where $x_{k,i}$ is the value of the k^{th} variable for individual *i*, β_k is the coefficient of the effect of variable *k* on asocial learning, and γ_k is the coefficient of the effect of variable *k* on social transmission (see Section 7.1 for how these coefficients can be interpreted).

6.1 Why include ILVs?

The most obvious reason to include ILVs in an NBDA is if the researcher is interested in the effect those variables may have on asocial and/or social learning (see Box 3 for an example). Alternatively, one may wish to include a potentially confounding variable that might cause a spurious social transmission effect. This can occur if a variable is both correlated with the network and has an effect on asocial learning (Hoppitt, Boogert *et al.* 2010)—e.g. older individuals may tend to associate with one another and be more likely to acquire a novel foraging trait through asocial learning. Hoppitt, Boogert *et al*. (2010) showed that such confounds could be statistically controlled for by including the relevant ILV in the NBDA model.

6.2 Additive, multiplicative and unconstrained models

When NBDA was first extended to include ILVs, two variants were proposed (Hoppitt, Boogert *et al*. 2010). The additive model assumed that all ILVs affected only the rate of asocial learning, $\Gamma_i = 0$, whereas the multiplicative model assumed that all ILVs influenced both asocial learning and social transmission, and did so by the same amount—i.e. $\beta_k = \gamma_k$ for all *k*. Studies using this approach tended to include separate sets of additive and multiplicative models during model selection, and used AICc to choose between the two (see Section 9). However, this approach excludes the possibility that an ILV might have a different effect on social transmission and asocial learning, so we generally prefer fitting the 'unconstrained' model (Hoppitt & Laland 2013) in which β_k and γ_k are estimated independently. Nonetheless, for some variables it may make sense to assume *a priori* that they only operate on asocial learning ($\gamma_k = 0$), only on social transmission ($\beta_k = 0$), or that they affect asocial learning and social transmission the same amount ($\beta_k = \gamma_k$). Therefore, in the *NBDA* package the user can specify which variables affect social transmission, which affect asocial learning and which are assumed to have the same effect on each.

6.3 Entering ILVs

s is estimated relative to the baseline rate of asocial learning, which is the rate of asocial learning when all ILVs in the model are set to zero. As such, a researcher should attempt to enter ILVs in a way that makes interpretation of *s* most meaningful.

Continuous variables: We recommend centring all continuous variables (subtract the mean) such that they have a mean of zero. In this way, the baseline rate of asocial learning is set to the mean of all continuous variables. Dividing each variable by its standard deviation such that it is fully standardized (mean = 0, SD = 1) is also advisable since it improves the probability of model convergence.

Binary variables: for two level factors, such as sex, the most obvious way to code the variable is 0/1 (e.g. males = 0, females = 1) such that the estimated effect β_k or γ_k gives the difference on the log scale between the two levels (see Section 7.1). This means that the baseline asocial learning rate will be set to whichever level of the factor is set to zero. An alternative is to code the variable as -0.5/0.5 (e.g. males = -0.5, females = 0.5). Since the difference between the two levels is 1, the estimated effect still gives the difference on the log scale between the two levels, but the baseline asocial learning rate is set to the midpoint of the two levels. It may also be necessary to re-code binary variables once the analysis has been run to obtain interpretable estimates of *s* (see Section 7.4).

Factors: categorical variables with *F* > 2 levels can be broken down into *F* - 1 indicator variables in the same way as for a standard regression analysis. For example, if we have an 'age category' with adults, sub-adults and juveniles, this could be broken down into a variable 'juv' which takes the value 1 for juveniles and 0 for adults/sub-adults, and a variable 'sub' which takes the value 1 for sub-adults and 0 for adults/juveniles. In doing so, adults becomes the reference level (juv = 0, sub = 0) to which juveniles (juv = 1, sub = 0) and sub-adults (juv = 0, sub = 1) are compared. Whichever factor level is set as the reference is also the baseline rate of asocial learning. So in our example, *s* is estimated relative to the adult rate of asocial learning. Again, it may also be necessary to re-code factors once the analysis has been run to obtain interpretable estimates of *s* (see Section 7.4).

6.4 Time-varying ILVs

NBDA can be further expanded to include ILVs that vary over the course of the diffusion:

$$\lambda_i(t) = \lambda_o(t)\big(1 - z_i(t)\big)\left(s\exp\big(\Gamma_i(t)\big)\sum_{j=1}^{N} a_{ij}(t)z_j(t) + \exp\big(\mathrm{B}_i(t)\big)\right)$$

$$\mathrm{B}_i(t) = \sum_{k=1}^{V} \beta_k x_{k,i}(t)$$

$$\Gamma_i(t) = \sum_{k=1}^{V} \gamma_k x_{k,i}(t)$$

Eqn. 5

where $x_{k,i}(t)$ is the value of the *k*th variable for individual *i* at time *t.* For OADA, we only need to specify the value of each variable at the time of each acquisition event, and the *NBDA* package allows a user to do this. For a cTADA, time-varying ILVs can currently be specified such they change value only at the times of the acquisition events. For a dTADA, a value is specified for each time unit—that is, it is assumed that the value does not change within each time unit.

7 Interpretation of NBDA models

7.1 Individual level variables

Continuous variables: Note from Eqn. 4 and 5 that ILVs are modelled as having a linear effect on the log scale (as with most survival analysis models (Moore 2016) and any generalized linear models with a log link function). Therefore, $\exp(\beta_k)$ estimates the multiplicative effect of one unit increase in x_k on the rate of asocial learning, and $\exp(\gamma_k)$ estimates the multiplicative effect of one unit increase in x_k on the rate of social learning (i.e. incoming social transmission). If the variable has been standardized, the estimates give the effect of one SD increase in x_k. One can transform the effect back to the original scale by dividing β_k and γ_k by the SD for the unstandardized variable.

For example, imagine that we have an ILV ‘age’, which had a SD of 10 years. We standardized the variable and obtained the estimates $\beta_{age} = 1.5$ and $\gamma_{age} =$ -0.8. We can therefore estimate that for an increase in age of 1 SD (10 years), the asocial learning rate increases by a factor of exp(1.5) = 4.48x, whereas the rate of social learning decreases by a factor of exp(-0.8) = 0.45x. Or we can obtain our estimate on the scale of years: the rate of asocial learning increases by a factor of exp(1.5/10) = 1.16x per year, whereas the rate of social learning decreases by a factor of exp(-0.8/10) = 0.92x per year.

Binary variables: If a binary variable has been coded such that there is a difference of 1 between the levels (e.g. 1/0 or -0.5/0.5), $\exp(\beta_k)$ estimates the ratio of asocial learning rates between the two levels. Likewise, $\exp(\gamma_k)$ estimates the ratio of social learning rates between the two levels. For example, imagine we have an ILV 'sex' with -0.5 = male and 0.5 = female. We get β_{sex} = 1.8 and $\gamma_{sex} = -1.2$. Therefore, females are an estimated exp(1.8) = 6.05x faster than males at asocial learning and an estimated exp(-1.2) = 0.30x slower at social learning. Alternately, we can reverse the sign of the γ_{sex} coefficient and say that males are an estimated exp(1.2) = 3.32x faster than females at social learning.

Factors: coefficients can be interpreted in the same manner as binary variables in a pairwise manner. For our example in Section 6.3, imagine that we got $\beta_{juv} = 0.74$ and $\beta_{sub} = 0.32$. We can conclude that juveniles are an estimated exp(0.74) = 2.10x faster at asocial learning than adults and sub-adults are an estimated exp(0.32) = 1.38x faster at asocial learning then adults. To get the estimated difference between juveniles and subadults, we back-transform the difference between their coefficients, $\exp(\beta_{juv} - \beta_{sub})$: juveniles are an estimated exp(0.74-0.32) = 1.52x faster at asocial learning than sub-adults.

7.2 Social transmission (s)

In general, *s* is the rate of social transmission per unit connection, relative to the baseline rate of asocial learning, but may have a more specific interpretation depending on the network used (see Section 4). The baseline rate of asocial learning is obtained by setting all ILVs to zero (see Section 6.3).

For example, imagine that we have a large scale association network (see Section 4.1), a continuous ILV 'age' centred on zero, and a binary variable 'sex', coded as males = 0, females = 1, and we obtain an estimate of *s* = 3.2. We can conclude that the rate of social transmission from informed to naïve individuals during periods when they are associating was estimated at 3.2x the baseline rate of asocial learning (that is, the asocial learning rate for a male of average age).

7.3 Obtaining and interpreting confidence intervals

Confidence intervals (CIs) for a parameter give a plausible range for the real value of that parameter; that is, an X% CI is expected to contain the true value of a parameter on X% of occasions. CIs therefore should be obtained, reported and interpreted for any parameters of interest, including *s*. A common way to obtain CIs is take the maximum likelihood estimate ± 1.96 x the standard error, referred to as Wald confidence intervals. However, Wald CIs can be misleading if the uncertainty in the value of a parameter is asymmetrical, as is often the case for parameters in an NBDA. In particular, for *s* there is often more certainty in the lower limit than there is for the upper limit.

A preferred approach for dealing with such a scenario is to use the profile likelihood technique (Morgan 2010), which provides CIs reflecting any asymmetry in the certainty in a parameter (Fig. 3). The profile log-likelihood is the -log-likelihood for a specified value of the parameter, once all other parameters in the model have been optimized. If a specified value, *v*, for the parameter has a profile log-likelihood that is within 1.92 units of the minimum, then *v* falls within the 95% CI; this is because the 95% profile CI contains all values that would not be rejected at the 5% level in a likelihood ratio test (see Section 9.1). So to find the 95% CI, researchers need to plot the profile log-likelihood, draw a line at 1.92 units above the minimum value (which is also the -log-likelihood of the fitted model), and find the upper and lower crossing points (Fig. 3). Functions are provided in the *NBDA* package to facilitate this process (e.g. see Tutorial 1 in the Supporting Information).

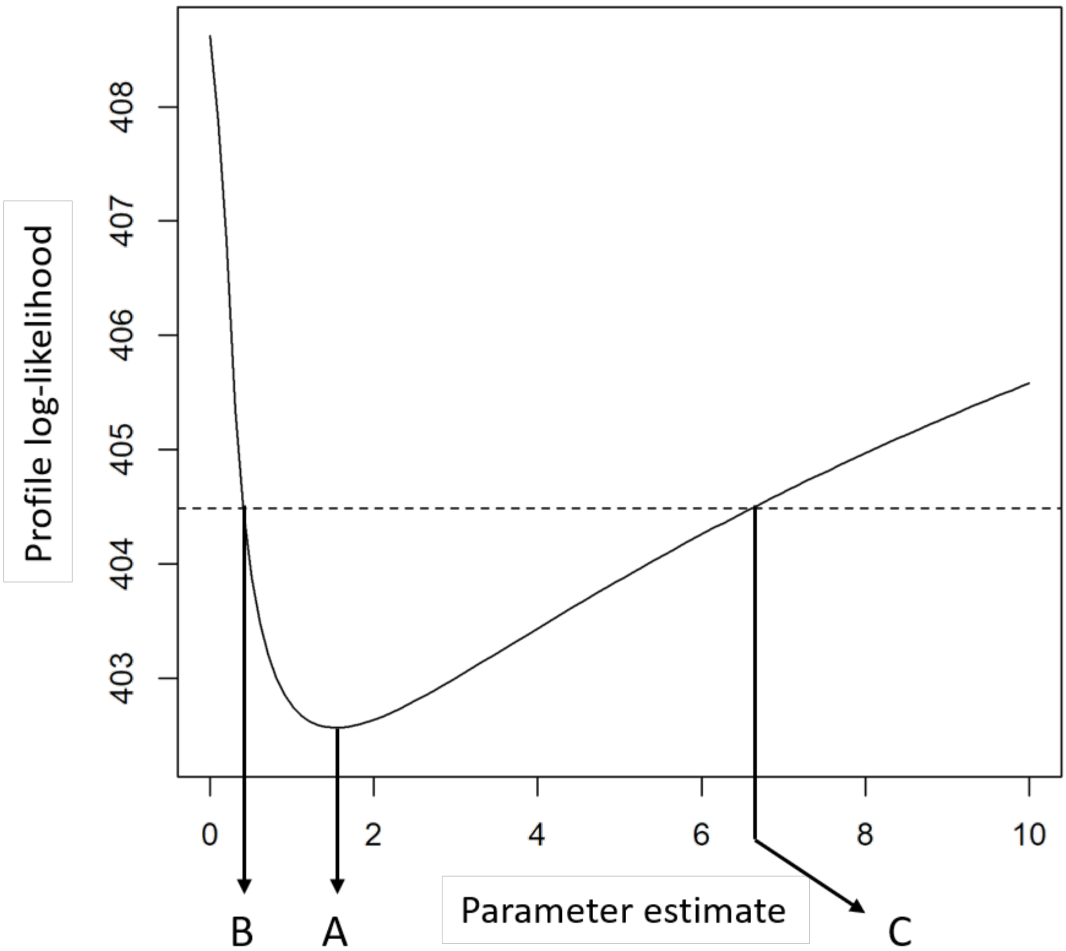


Fig. 3. Profile log-likelihood plot used for obtaining confidence intervals for parameters in which there is asymmetry in the uncertainty regarding their values. The profile log-likelihood is the -log-likelihood for a specified value of the target parameter once all other parameters in the model have been optimized. The lowest point of the curve (A) corresponds to the profile log-likelihood for the parameter value obtained from the fitted model. The dashed line indicates 1.92 units above this minimum -log-likelihood. Values that fall within this region are within the 95% CI. The values at which the curve crosses this dashed line indicate the lower (B) and upper (C) values for the 95% confidence interval. Here, the estimate from the fitted model is 1.54 (95% CI: 0.40, 6.61).

The CI for *s* allows the researcher to determine the level of information provided by their data about the importance of social transmission in their diffusion(s), as shown in Table 1.

Table 1. Interpreting 95% confidence intervals for *s*.

	Upper limit of 95% CI	
Lower limit of 95% CI	Low value	High value
0	Little or no social transmission	Weak or no evidence of social transmission, but cannot rule out an important effect either
Low value	A small effect of social transmission	Evidence of social transmission, but uncertain whether the effect is strong or weak
High value	Not possible	Strong evidence of social transmission that has an important effect in the diffusion

Whist it may sometimes be possible for the researcher to interpret the value of *s* directly, and thus determine what values should be considered ‘low’ or ‘high’, in many cases this will be difficult. In such cases researchers can transform the upper and lower limits of the 95% CI into upper and lower estimates of the percentage of events that occurred by social transmission (see Section 7.5).

Confidence intervals for the effects of ILVs can be interpreted in an analogous manner, but the parameter values should first be back-transformed as described in Section 7.1, after which, the point of no effect is exp(0) = 1. CIs for ILVs could also potentially include values in either direction (i.e. greater than and/or less than 1).

7.4 Dealing with large estimates for s

Note that sometimes very large estimates of *s* can be obtained, especially in an OADA, which can seem difficult to interpret. This also usually means that we cannot find an upper limit for the 95% CI for *s* (see Section 7.3). There are two main reasons that such large estimates can arise.

First, this can occur if an ILV has a very large positive coefficient. For example, let us assume that we have $\beta_{sex} = 14$ in our example above; this corresponds to females being an estimated 1,200,000x faster to learn asocially than males. This is probably because the only individuals that ever learned when their total connection to demonstrators, $\sum_{j=1}^{N} a_{ij}(t)z_j(t)$, was zero were female. This makes it logically plausible that only females can learn asocially, resulting in a profile log-likelihood for β_{sex}that flattens out to an asymptote as β_{sex} tends to infinity (see Box 4 Figure 1 for an example of this). This means that we can only set a lower estimate on β_{sex}, but it also impacts the estimated value for *s*. This is because *s* is being estimated relative to the asocial learning rate for males: since males are effectively concluded to have an asocial learning rate of 0, this pushes *s* up to an arbitrarily large value. This also means that we cannot obtain an upper limit for the 95% CI for *s*. We can solve this problem by simply re-parameterizing the model such that females are set to zero. We will then obtain a model output with $\beta_{sex} = -14$, but *s* will now be estimated relative to the (non-zero) female rate of asocial learning. This is now likely to yield an interpretable estimate for *s* and an upper limit for the 95% CI. In general, if large values of *s* are obtained and/or no upper limit can be found for the 95% CI, re-parameterize the model such that all the ILVs are set to zero at the point where they have their maximum effect size. The model may then yield a value of *s* that is more easily interpretable.

A second reason that large estimates of *s* can be obtained is if the diffusion follows the network very closely. The most extreme case is if the next individual to learn is always the one with the highest total connection to informed individuals, $\sum_{j=1}^{N} a_{ij}(t) z_j(t)$. In such cases, the profile likelihood for *s* will keep levelling out towards infinity—as far as the underlying logic of the NDBA model is concerned, these values of *s* are plausible. In such cases, one can only set a lower plausible limit on *s*, and report “s is estimated to be at least [insert lower 95% CI]”. However, we may be able to set an upper limit on the percentage of events that are estimated to have occurred by social transmission (see Section 7.5).

7.5 Estimating the percentage events occurring by social transmission

For some types of network, it is not easy to interpret *s* in an intuitive manner (see Section 4), and thus it can be difficult to get an idea of the importance of social transmission in the spread of the target behaviour. A solution is to convert *s* into an estimate of the proportion of learning events that occurred by social transmission (which we refer to as %ST). The probability that each event, *e*, occurred by social learning can be calculated as:

$$p_{social,e} = \frac{s \exp\left(\Gamma_i(t_e)\right) \sum_{j=1}^{N} a_{ij}(t_e) z_j(t_e)}{s \exp\left(\Gamma_i(t_e)\right) \sum_{j=1}^{N} a_{ij}(t) z_j(t_e) + \exp\left(\mathrm{B}_i(t_e)\right)}$$

Eqn. 6

where *i* is the individual that learned during event e, and t_e is the time at which event *e* occurred. This is the predicted relative rate of social transmission divided by the predicted total relative learning rate for *i* at the time of learning. The mean of $p_{social,e}$ across all events is then the estimated proportion of events that occurred by social transmission (%ST). One can obtain analogous estimates for the upper and lower limits of the 95% CI for *s*. For an example of how this may be achieved, see Tutorial 2 in the Supporting Information.

s and %ST quantify the importance of social transmission in subtly different ways, with the latter taking into account the connections of the network. For illustration, imagine two diffusions of the same behaviour, in two different populations A and B. An NBDA using a large-scale association network yields an estimate of *s* = 4 in population A and *s* = 2 in population B. However, because population B tends to have stronger associations than population A, we obtain an estimate that 50% of events occurred by social transmission in population A and 75% in population B. In population A, for every unit of time naïve individuals spent with informed individuals, social transmission occurred at double the rate than in population B. However, because individuals in population B associated more often, more individuals in population B are likely to have learned by social transmission.

8 Multiple network NBDA

The approaches described in Eqns. 1-5 assume that social transmission follows only a single pathway, represented by a single network (or a single type of network when modelling multiple diffusions). An alternative approach is to allow for the possibility that social transmission might follow more than one pathway, and do so at different rates (for an example, see Box 4). This situation can be modelled using a multiple network NBDA (Farine *et al.* 2015), expanding Eqn. 5 as follows:

$$\lambda_i(t) = \lambda_o(t)\big(1 - z_i(t)\big)\left(\exp\big(\Gamma_i(t)\big)\sum_n s_n \sum_{j=1}^{N} a_{n,ij}(t) z_j(t) + \exp\big(\mathrm{B}_i(t)\big)\right)$$

Eqn. 7

Where $a_{n,ij}(t)$ is the connection from *j* to *i* in network *n* at time *t*, and s_n is the rate of transmission per unit connection in network *n* (relative to the rate of asocial learning).

This model can be compared with those in which some or all of the *s* parameters are constrained. For example, comparison with a model in which $s_1 = s_2$ tests for a difference in transmission rate between network 1 and network 2. We could also consider models in which there is no transmission in a specific network, e.g. $s_1 = 0$, to test for evidence of social transmission on a specific pathway.

We can also estimate the percentage of events occurring by social transmission via a specific network *n*, $\%ST_n$ (see Section 7.5). We first expand Eqn. 6 to calculate the probability that each event occurred as a result of social transmission via network *n*:

$$p_{n,e} = \frac{s_n \exp\big(\Gamma_i(t_e)\big) \sum_{j=1}^{N} a_{n,ij}(t_e) z_j(t_e)}{\exp\big(\Gamma_i(t_e)\big) \sum_m s_m \sum_{j=1}^{N} a_{m,ij}(t_e) z_j(t_e) + \exp\big(\mathrm{B}_i(t_e)\big)}$$

Eqn. 8

We then take the mean value of $p_{n,e}$ across all events to obtain $\%ST_n$. s_n allows the comparison of the rate of transmission per unit connection in each network, and thus is sensitive to the scale of each network—e.g. if we divide network *n* by 2, the value of s_n will be doubled. In contrast, $\%ST_n$ is invariant to the scale of each network, but also takes into account the number and strength of connections in each network. See Farine *et al.* (2015)

for further discussion on how to quantify the influence of each network in a multi-network NBDA.

One potential use of multi-network NBDA is to break down association or observation networks into different pathways to test for biases in transmission. For example, to test for a rank bias in transmission we might break down an association network into two networks: network 1 containing the links from higher to lower ranks (and 0 connections elsewhere), and network 2 containing links from lower to higher ranks. We can then compare this model with one in which $s_1 = s_2$ in order to test for a rank bias—that is, are individuals more (or less) likely to learn from those with higher rank than those with a lower rank? Hoppitt (2017) provides further discussion of the potential uses of multi-network NBDA. Farine, Spencer and Boogert (2015) provide an excellent example of how a network can be broken down into a number of pathways to test hypotheses about social transmission. Wild *et al.* (2019) use multi-network NBDA in a slightly different way: to tease apart the effects of social transmission, shared environment and genetic relatedness on a foraging behaviour in bottlenose dolphins (*Tursiops aduncus*).

9 Model selection approaches

9.1 Model comparison

In the preceding sections, we have alluded to a number of different situations where the fit of two or more NBDA models needs to be compared in order to assess the evidence for competing hypotheses, including:

a) comparing a model of social transmission to an asocial model (s = 0) to quantify the evidence for social transmission (Section 4.1)

b) comparing a network-based model of social transmission to a model with a homogeneous network (Section 5.1) or with a group network (Section 5.2)

c) comparing models with different networks (Section 4.2) or different combinations of networks (Section 8) to ascertain which best approximates the pathways of transmission

d) comparing multi-network models with models in which some or all s_n are constrained (e.g. $s_1 = s_2$, or $s_1 = 0$; Section 8).

In some cases, the models to be compared are nested—i.e. one is a special case of the other model, with constraints imposed on one or more parameters (this is true for a and d unless different baseline rate functions are fitted in each model). When this is true, one can use a likelihood ratio test (LRT) to obtain a *P* value quantifying the evidence against the null hypothesis represented in the constrained model (Morgan 2010). Here, the test statistic is the likelihood ratio, which equals 2x the difference in the negative log-likelihoods between the two models (Box 2). The *P* value is then obtained from the upper tail of a chi-square distribution with degrees of freedom equal to the difference in the number of parameters in the two models. For example, imagine that we fit an OADA model with three networks and no ILVs, and wish to test the null hypothesis that the rate of transmission is the same in each network. We fit a null model with the constraint $s_1 = s_2 = s_3$. We obtain a difference in negative log-likelihoods of 3.23, giving a test statistic of 6.46. In the full model, three parameters are fitted (i.e. *s* is estimated separately for each network), whereas only one *s* parameter is fitted in the null model. So, we obtain a *P* value from the upper tail of a chi-square distribution with 2 df—the R command is: pchisq(6.46, df=2, lower.tail=F)—giving us *P* = 0.0396, i.e. reasonable evidence against the null hypothesis.

However, a LRT cannot be used to compare two or more non-nested models, such as models that contain different networks (e.g. b and c above). In such cases, one can use Akaike's Information Criterion (AIC) to compare the fit of models. A full explanation of the theoretical basis for AIC and a guide for its use can be found in Burnham and Anderson (2002). Burnham, Anderson and Huyvaert (2011) provide a succinct review of this topic. Here, we give a brief outline. AIC is calculated as $2x -log(L) + 2k$, where -log(L) is the negative log-likelihood for a model, and k is the number of parameters in that model. In practise, we recommend use of AICc, a version of AIC that corrects for sample size; the *NBDA* package provides AICc for fitted models, taking sample size to be the number of acquisition events.

Models with lower AICc are those that explain the data better after penalizing for the number of parameters in the model. The penalty imposed is not arbitrary; it is chosen such that the difference in AICc between any two models fitted to the same data estimates the difference in Kullback-Leibler (K-L) information. In turn, K-L information measures the extent to which the predicted distribution for the response variable differs from its true distribution. In other words, it estimates the information that is lost when moving from the true distribution to the model. Consequently, AICc provides a theoretically well justified measure of the relative fit of two or more models. We can transform the difference in AICc between two models (ΔAIC) to obtain the relative support for the two models, exp(ΔAIC/2). This value quantifies the ratio of probabilities that each model is the one with the best K-L information (termed the 'best K-L model').

For example, imagine that we fit a model with a proximity network (AICc = 382), and a model with a network quantifying the rate of grooming interactions (AICc = 373.5). Thus, these data suggest that the grooming network is a better approximation of the pathways of transmission than the proximity network, but how certain of this result can we be? It might just be a chance result of sampling error. The difference in AICc (ΔAICc) between these two models is 9.5, giving a relative support of exp(9.5/2) = 115.6. This means that the grooming network is 115.6x more likely to be a closer approximation to the transmission pathways than the proximity network, which we would take to be very strong support in favour of the grooming network.

If a researcher has a number of candidate models, they can list them in increasing order of AICc to show the order of preference in model fit (Box 4). They can then calculate the Akaike weight for each model as a measure of its support. To do this, one first calculates the AICc difference between each model, *i,* and the best model, $\Delta_i = AICc_i - AICc_{best}$. The Akaike weight for model *i* is then $\mathrm{w}_i = \exp\left(-\frac{1}{2}\Delta_i\right) / \sum_j \exp\left(-\frac{1}{2}\Delta_j\right)$, and can be interpreted as the probability that model *i* is the best K-L model in the set, accounting for sampling error.

9.2 Multi-model inference

If there are a number of ILVs to consider in addition to our competing hypotheses about social transmission, this complicates the model selection process. We could simply include all ILVs in all candidate models, but requiring these models to fit additional parameters may decrease the precision of our estimates for *s*. Ideally, we only want to include the variables for which there is evidence of an effect on asocial and/or social learning. The traditional approach to this would be to select the combination of ILVs that provides the best model fit, and base our inferences on that model. With modern computing power, it would even be feasible to fit all possible combinations of ILVs and select the lowest AICc as our best model. However, this approach inherently assumes we are *certain* that the best-supported model really is the best one (in the sense of minimizing K-L information loss). As we saw in Section 9.1, there is often substantial uncertainty due to sampling error over which model really is the best; this model selection uncertainty is quantified by the Akaike weight (Burnham & Anderson 2002; Burnham *et al*. 2011).

Multi-model inference is a set of tools that allows us to account for model selection uncertainty when we make our inferences (these tools are available in the *NBDA* package). The first such tool allows us to quantify the overall strength of evidence for a particular network (or combination of networks) by calculating the total Akaike weight for that network (otherwise simply known as "support" for that network). This is done by simply summing the Akaike weights, Σw_i, for all the models that contain the network. This value can be thought of as the probability that the best K-L model is one that includes the network of interest. We can obtain the support for all the networks (or network combinations) we are considering as an overall measure of the extent to which each one approximates the pathways of transmission. For this to be a fair comparison, a researcher needs to ensure that there are the same number of models for each network. However, if the same combinations of ILVs are considered for each network, this condition will be met. Support can also be obtained for an effect of each ILV on asocial and social learning rate in an analogous manner. We can also compare the overall fit of models with different baseline rate functions, or particular combinations of baseline function and network(s).

The question remains as to whether we can validly use Σw_i to measure the level of support for asocial models versus social models (i.e. models with a social transmission component). This depends on the set of models that we are considering. Imagine the case where we have an OADA with 3 ILVs and 2 networks, with only 1 of these networks included in any given model. An approach previously used was to consider additive models in which ILVs affected only asocial learning, and multiplicative models, in which ILVs affected both asocial and social learning by the same amount (see Section 6.2). There are 8 different combinations of the 3 ILVs, giving 8 multiplicative + 8 additive models with network 1, 8 multiplicative + 8 additive models with network 2, and 8 asocial models (since the additive and multiplicative models are the same when $s = 0$). If we compared the support for asocial models versus social models, we would be comparing 8 models against 32 models, giving an unfair and misleading picture of the support for social transmission. Instead, we should do a five way comparison of: (a) asocial learning only; (b) network 1 multiplicative; (c) network 1 additive; (d) network 2 additive; and (e) network 2 multiplicative. Within each category, there are 8 models. The Σw_i for each of the 4 categories of social models can be thought of as support for competing hypotheses about social transmission, which can also be compared with the support for asocial models.

However, in Section 6.2 we argued that using an unconstrained model was preferable to the additive versus multiplicative model approach described above. Recall that in the unconstrained model, the effect that each ILV has on asocial and social learning is estimated independently, allowing for the possibility that any ILV could have different effects on each type of learning. In our example above, this means that instead of having 2 sets of 8 combinations of ILVs, we now have 36 combinations of effects on asocial and social learning! This assumes that is plausible that any of our ILVs could affect social learning without affecting asocial learning. So in our example above, we now have 36 network 1 models, 36 network 2 models and 8 asocial models (since ILVs cannot affect social learning when $s = 0$). Thus, a three-way comparison of support would be unfair and misleading. We recommend that total Akaike weights are not generally used to quantify the relative support for asocial models versus social models where the unconstrained model is used. Researchers can use the total Akaike weights to select the best supported network(s), and then use confidence intervals on the *s* parameters (Section 7.3) to assess the strength of evidence against asocial learning ($s = 0$). However, if the asocial models have the greatest support despite the smaller number of models, this can be taken as evidence against social transmission.

Model-averaged estimates (MAEs) provide researchers with a means to estimate the values of parameters in a way that accounts for model selection uncertainty. MAEs are an Akaike weighted average of the parameter value in each individual model. Unconditional standard errors (USEs) can also be calculated as a measure of precision that accounts for both the uncertainty in the value of parameters among models, as well the within-model uncertainty quantified by traditional standard errors (SEs) (Burnham & Anderson 2002). Unfortunately, SEs cannot always be calculated for NBDA models, meaning that USEs across a model set can also not be calculated. Where SEs are only missing for a few models of low Akaike weight, we recommend replacing these with the Akaike weighted average SE across all other models, and calculate USE as usual to obtain an approximation. However, if SEs are missing for many models, or for models with high Akaike weight, we recommend omitting USEs.

For *s* parameters, we recommend obtaining MAEs and USEs that are conditional on the relevant network(s) being presented in the model. If a large number of networks are considered, then any given *s* parameter will be absent from the vast majority of models in the set, and MAEs and USEs will be misleading. Conditioning on the subset of models that contain a specific network reweights the Akaike weights such that they sum to 1 within the subset, and then carries out multi-model inference using those models. This is equivalent to asking ‘given that the best K-L model contains network *n*, what is our best estimate of *s*?” The MAE for an *s* parameter can still be misleading if there are some models in the set for which *s* is estimated arbitrarily large (see Section 7.4). Even if these models have a tiny Akaike weight, they can still badly skew the estimate of *s*. In such cases, we suggest that the model weighted median for *s* is obtained instead as an estimate that is robust to extreme estimates with low Akaike weight.

USEs provide a useful way of calculating unconditional 95% CIs for parameters that account for model selection uncertainty: one simply calculates MAE ± 1.96 x USE. However, these CIs can be misleading in cases when the profile likelihood is asymmetrical for the same reason Wald CIs can be (see Section 7.3). Burnham and Anderson (2002) suggest a method for inflating 95% profile likelihood intervals (Section 7.3) to account for model selection uncertainty. Instead of using a cut-off line 1.92 units above the minimum negative log-likelihood (Fig. 3), one elevates the cut-off line by a factor = $USE^2/(SE \text{ in best model})^2$. However, as noted above, USEs cannot always be obtained. Furthermore, in NBDA it is not uncommon for the inflation method to return a 95% CI for *s* that includes zero even when all the conditional 95% CIs exclude zero (so logically an unconditional 95% CI should also exclude zero). Therefore, instead of using the inflation method, we recommend obtaining the 95% CI conditional on the best model containing a parameter.

Since there is usually particular interest in determining how strong, at a minimum, social transmission is, we recommend assessing the robustness of the lower limit of the 95% CI to model selection uncertainty. This can be done by obtaining the 95% lower limit for all models containing the relevant *s* parameter and the corresponding estimate of %ST, and interpreting them. For example, if all these values are > 0, then the evidence for social transmission is robust to model selection uncertainty. We also suggest providing a model-averaged version of the value of %ST corresponding to the 95% lower limit, as a lower plausible limit on the importance of social transmission after accounting for model selection uncertainty (see Tutorial 7 in the Supporting Information for the relevant code).

10 Further extensions and considerations

10.1 Error and uncertainty in network structure

Hoppitt (2017) considers the effect of error in the measured social network, considering cases where there is random noise or systematic bias resulting in relative overestimates or underestimates of larger connections. No sources of error inflated the type 1 error rate, showing that a positive result for social transmission can be trusted even when the network may not be accurate. However, some sources of error tended to make estimates of s and %ST conservative. Researchers should bear this in mind when interpreting confidence intervals if network error is suspected. Another possibility is that some individuals in the population have limited network data. Wild and Hoppitt (2019) develop a procedure to determine which individuals, if any, should be dropped from the analysis.

10.2 Untransmitted social effects

When the target behaviour is constrained to be performed at a specific location, e.g. the solution to a foraging task, it may be that closely associated individuals are likely to encounter the task at the same time, purely as a result of being together, and thus solve at a similar time. This could result in a statistical pattern that looks like social transmission in an NBDA, referred to as an 'untransmitted social effect' (Atton *et al.* 2012; Hoppitt & Laland

2013). One way to control for this effect is to exclude the possibility that individuals that learned together close in time could have learned from one another, i.e. consider them to be ‘tied’ with regards to the incoming social information (Hoppitt, Boogert *et al*. 2010). Any remaining social transmission effect is then unlikely to be a result of an untransmitted effect. In a dTADA, such individuals can simply be included as learning in the same time period. A similar complication arises when using a dynamic observation network (see (Hobaiter *et al.* 2014; Hoppitt 2017) for discussion of this problem).

10.3 ‘True’ ties

Another type of tie arises if we are uncertain of the order in which one or more individuals learned the target behaviour; we term these ‘true ties’ (Hoppitt, Boogert *et al.* 2010). This could arise if the population’s behaviour is only sampled periodically, or because two individuals learn the behaviour so close together in time it is impossible to determine the order. This problem is easily dealt with in a dTADA, since the tied individuals are simply included as learning in the same time period. In cTADA, the tied individuals can be assigned the same learning time, and considered to be ‘tied’ in the sense described in Section 10.2, since it is implausible that social transmission occurred between two individuals who learned at approximately the same time. The problem is also conceptually easy to solve in an OADA. The likelihood for a tied event is simply the sum of likelihoods for all orders that are consistent with the observed tie—e.g. if individuals A, B and C are tied, we sum the likelihood for the six possible orders ABC, ACB, BAC, BCA, CAB and CBA (in the *NBDA* package, one simply specifies the true ties). This approach may be feasible if we have a small number of true ties involving only a few individuals. However, if we have true ties involving many individuals, calculation of the likelihood can take a prohibitively long time.

For example, a single 6-way tie requires calculating the likelihood for 6! = 720 possible orderings for the true tie. Therefore, if an OADA is preferred, then we suggest that researchers do all they can to resolve any true ties. If the computation remains infeasible, then a TADA must be used.

10.4 Bayesian NBDA

NBDA can be re-cast in a Bayesian framework, which has a number of potential advantages, such as easy inclusion of random effects and better methods for accounting for uncertainty in data. A Bayesian version of NBDA has been investigated and used by Whalen & Hoppitt (2016) and Nightingale, Boogert, Laland and Hoppitt (2014). However, there is not yet a user-friendly package to implement a Bayesian NBDA.

11 Conclusion

NBDA provides a flexible approach for detecting and quantifying the impact of social transmission on the spread of information and novel skills through animal groups, and for elucidating the typical pathways of information flow. With the widespread adoption of social network techniques in the field of animal behaviour, the data necessary for NBDA is likely to be increasingly available. Here, we have sought to guide interested researchers through the process of selecting the appropriate NBDA variant and network structure(s) for their research question, incorporating individual-level variables that may impact social and asocial learning, selecting amongst alternative models on the basis of their relative support, and interpreting model outputs in terms of their biological significance. NBDA may thereby help to achieve a greater understanding of the links between social structure and social learning dynamics within natural settings.

Box 1. Glossary

Asocial (or individual) learning: learning through trial-and-error or personal sampling of the environment. In the context of NBDA, this refers to learning the target behaviour independently of others, i.e. not through social transmission.

Asocial model: in the context of NBDA, a model in which the target behaviour is never learned through social transmission, i.e. learning is always asocial learning.

Diffusion data: data detailing the spread of a target behaviour pattern through a population or group of animals.

Individual-level variable (ILV): a variable that varies among individuals, and is included in an NBDA for its potential effect on the rate of asocial and/or social learning

Homogenous network?

Network-based diffusion analysis (NBDA): a statistical method for quantifying the influence of social transmission, mediated by one or more social networks, in the diffusion (or spread) of a target behaviour through a group of animals.

Order-of-acquisition diffusion analysis (OADA): a variant of NBDA that takes as data the order in which individuals acquired a target behavioural pattern (usually inferred from the time at which they first perform it).

Social learning: learning that is facilitated by observation of, or interaction with, another individual or its products (Hoppitt & Laland 2013 after Heyes 1994). Social learning can (but does not always) result in the social transmission of behaviour.

Social network: A mathematical description of social structure, in which nodes (usually representing individuals) are connected by edges (or ties) indicating some form of social relationship. It is formally represented as an adjacency matrix (Farine & Whitehead 2015).

Social transmission: occurs when the prior acquisition of a behavioural trait *T* by one individual *A*, when expressed either directly in the performance of *T* or in some other behaviour associated with *T*, exerts a lasting positive causal influence on the rate at which another individual *B* acquires and/or performs *T* (Hoppitt & Laland 2013).

Time-of-acquisition diffusion analysis (TADA): a variant of NBDA that takes as data the time at which individuals acquired a target behavioural pattern (usually inferred from the time at which they first perform it).

Box 2. Fitting a basic OADA using maximum likelihood

Here, we show how a basic OADA model, containing only a single parameter, *s*, is fitted to the data by maximum likelihood. Note that this process is carried out automatically by the *NBDA* package (Hoppitt *et al.* 2019) when fitting an OADA model, but it is useful for a researcher to understand how the model is fitted. Maximum likelihood works by finding the values of the parameters for which the observed data is most likely. This is done by first deriving a likelihood function that specifies the likelihood of the data for a given set of parameter values. For OADA, the likelihood for a single acquisition event, *E*, is:

$$L_E = \frac{\lambda_e(t_E)}{\sum_{l=1}^{N} \lambda_l(t_E)}$$

Where *e* is the individual that learns on event *E*, and t_E is the time immediately prior to event *E*. In other words, L_E is the probability that *e* would be the next individual to learn, which is the rate of learning for e at time t_E, divided by the sum of rates for everyone in the population, $\sum_{l=1}^{N}\lambda_l(t_E)$. If we define the relative rate of learning to be

$$R_i(t) = \frac{\lambda_i(t)}{\lambda_o(t)} = \left(1 - z_i(t)\right)\left(s\sum_{j=1}^{N} a_{ij}(t)z_j(t) + 1\right)$$

L_E reduces to:

$$L_E = \frac{\lambda_o(t)R_e(t_E)}{\lambda_o(t)\sum_{l=1}^{N} R_e(t_E)} = \frac{R_e(t_E)}{\sum_{l=1}^{N} R_e(t_E)}$$

Therefore $\lambda_o(t)$ drops out of the likelihood function. The likelihood function for the whole diffusion, *L*, is the product of the likelihoods for all acquisition events. In principle, the value of *s* could be chosen to directly maximise the likelihood. However, for computational stability, one equivalently takes the negative logarithm of the likelihoods for each event and adds them together, -log(*L*), then finds the value of *s* that minimizes -log(*L*), where:

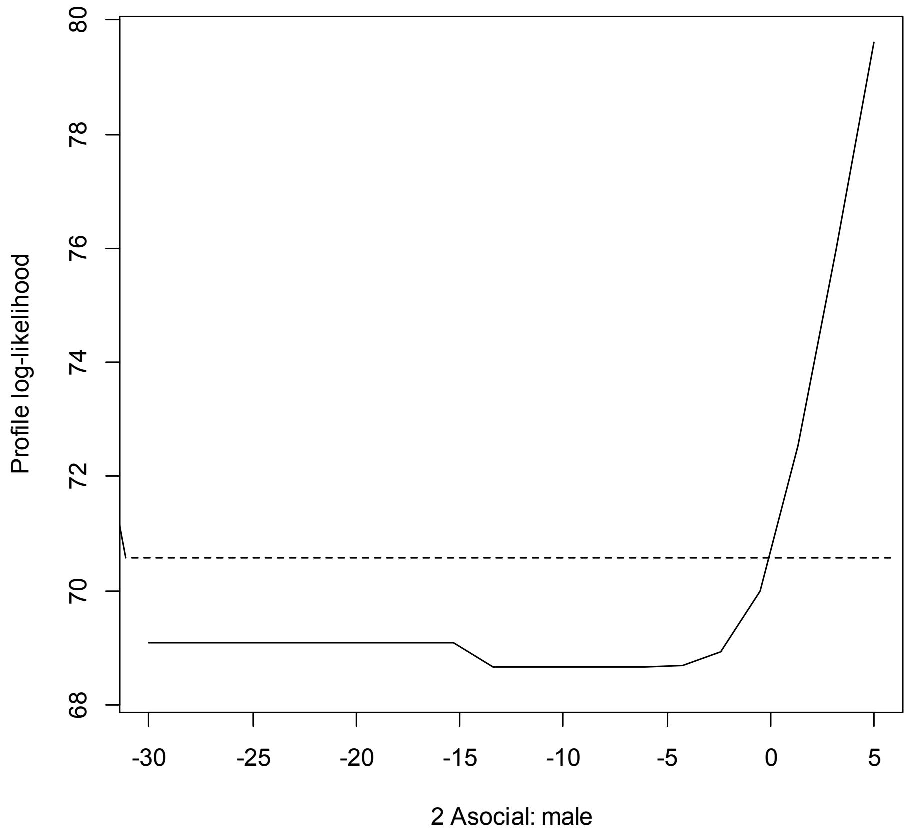
$$\log(L) = \sum_{E=1}^{D} log\left(R_e(t_E)\right) - \sum_{E=1}^{D} log\left(\sum_{l=1}^{N} R_e(t_E)\right)$$

This value of *s* is known as the maximum likelihood estimator for *s*, and the corresponding value of -log(*L*) is known as the negative log-likelihood (or -log-likelihood) for the model. When there is more than one parameter in the model, the optimization algorithm finds the combination of parameter values that minimizes -log(*L*). A review of the likelihood functions for NBDA, including cTADA and dTADA, is found in Hoppitt and Laland (2013).

Box 3. Fitting an OADA with individual level variables (ILVs)

A researcher will often wish to include ILVs in an NBDA model, either to investigate their impact on social and/or asocial learning, or to control for spurious social transmission effects (Section 6.1). Here, we illustrate how this can be done using the *NBDA* package; code for this example is found in Tutorial 2 in the Supporting Information. We generated a simulated social network of 30 individuals, as well as the order in which they acquired a target behaviour. We also have two ILVs: age (in years) and sex. To ease interpretation of *s* and to facilitate model convergence, we standardized age by first subtracting the mean and then dividing by the standard deviation. The *NBDA* package includes three options for how an ILV can affect learning: (a) additive models assume that an ILV impacts asocial learning only; (b) multiplicative models assume that an ILV impacts both asocial and social learning, and does so by the same amount; and (c) unconstrained models assume that an ILV differs in its effect on asocial and social learning (Section 6.2). We fit each of these three models to our simulated data. On the basis of AICc (Section 9.1), we find that the additive model is best supported. Box 3 Table 1 presents the parameter estimates, SEs, and 95% CI from this model. *s* estimates the rate of social transmission per unit of network connection, relative to the baseline rate of asocial learning. Here, this baseline rate is set as the asocial learning rate of a female of average age (Section 6.3). Because of asymmetry in the uncertainty for the values of some parameters (i.e. *s* and sex), 95% CI were obtained using profile likelihood techniques (Section 7.3). The asocial learning rate is estimated to decrease by exp(-1.027/SD(age)) = 0.62x per year of age. However, the 95% CI for age indicate that asocial learning rates may plausibly decrease by as much as 0.16x per year or increase by up to 1.35x. In other words, we can conclude that there is little evidence for a strong effect of age on asocial learning. Turning to sex, we find that females are estimated to be exp(19.84) = 4.13×10^{8} times faster than males at learning asocially! If we examine the profile log-likelihood, we find that it is very asymmetric (Box 3 Figure 1). In fact, it approaches an asymptote as the estimated effect moves towards negative infinity. This is because only females ever learned with zero network connections to informed individuals, meaning that it is plausible that only females *can* learn asocially (at least as far as the model is concerned). In this instance, we can only obtain the upper 95% CI at -0.001. So, we can conclude that females are at least exp(0.001) = 1.001x as fast as males at learning asocially.

Parameter	Estimate	SE	95% CI
Social transmission rate, *s*	2.97	3.92	0.40, 101.42
Age (years)	-1.03	1.03	-3.94, 0.65
Sex = ‘male’	-19.84	11618.17	-∞, -0.001



Box 4. Testing for social transmission across multiple pathways

It may be the case that a target behavior is socially transmitted across multiple pathways (i.e. network types), but at different rates in each. To test for this, one can input multiple networks into an NBDA and estimate a separate rate of social transmission (*s*) for each one. For example, honeybees (*A. mellifera*) can learn about foraging opportunities through multiple forms of interaction. Waggle dances performed by successful foragers provide the location of profitable foraging sites to naïve bees, while chemosensory information (e.g. food odor, nectar quality) can be obtained when receiving nectar during trophallaxis (reviewed in Grüter & Farina 2009). Even simply contacting other foragers with antennae can facilitate olfactory learning about food sources (Cholé *et al*. 2019). To assess the relative importance of these transmission pathways during recruitment of foragers to a novel foraging site, we recorded all interactions within the hive between trained demonstrator bees that collected food from a feeding station and a cohort of potential recruits that had never before visited that site. We also recorded the order in which these naïve bees successfully located the feeding station. To capture the temporal ordering of in-hive interactions between demonstrators and recruits, all three networks—i.e. dance following interactions, trophallactic exchanges, and antennal contacts—were input as dynamic, time-varying networks (see Section 4.4). Box 4 Table 1 provides the relative support for a candidate set of models. A comparison of models 2 and 3—either with a likelihood ratio test (χ^2_2= 11.12, P = 0.004) or on the basis of AICc—reveals that estimating *s* separately for each network type is favored over assuming a common transmission rate across all interaction types. However, in this instance, Model 1 which includes only the time-varying dance following network is clearly favored—w_1 = 0.94, indicating that there is very little uncertainty over the best model out of those considered here. That the temporal ordering of dance following interactions is key is revealed by Model 1 receiving exp(25.48/2) =

341124x as much support as the model that used the corresponding static observation network (Model 5). Finally, an asocial model (Model 4) that assumed that discovering the feeding station occurred through independent search alone received virtually no support. Model 1 yielded a very large estimate of $s = 9.94 \times 10^7$, most likely because the order of in which recruits discovered the feeding site followed the network of dance following interactions very closely (see Section 7.4). Converting this value into %ST suggests that following dances for the feeding station accounted for an estimated 100% (95% CI: 91.2%, + ∞) of the 16 recruitment events. The code for these models and analyses is found in the Supporting Information.

Model	*s* parameters	Network type (static/dynamic)	log(*L*)	*K*	AICc	ΔAICc	w_i
1	s_{Dance}	Dynamic	30.11	1	62.51	0	0.96
2	$s_{Dance} + s_{Trophallaxis} + s_{Antennation}$	Dynamic	30.11	3	68.23	5.72	0.05
3	$s_{(Dance + Trophallaxis + Antennation)}$	Dynamic	35.67	1	73.64	11.13	0.004
4	Asocial model ($s = 0$)	N/A	43.08	0	86.16	23.65	6.9×10^{-6}
5	s_{Dance}	Static	42.85	1	87.99	25.48	2.76×10^{-6}

Acknowledgements

This project has received funding from the European Horizon 2020 Research and Innovation Programme under ERC Starting Grant BeeDanceGap 638873.

References

Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013) Network-based diffusion analysis reveals cultural transmission of lobtail feeding in humpback whales. *Science*, **340**, 485-488. doi: 10.1126/science.1231976

Aplin, L. M., Farine, D. R., Morand-Ferron, J., & Sheldon, B. C. (2012) Social networks predict patch discovery in a wild population of songbirds. *Proceedings of the Royal Society B*, **279**, 4199-4205. doi: 10.1098/rspb.2012.1591

Atton, N., Galef, B. J., Hoppitt, W., Webster, M. M., & Laland, K. N. (2014) Familiarity affects social network structure and discovery of prey patch locations in foraging stickleback shoals. *Proceedings of the Royal Society B*, **281**, 20140579. doi: 10.1098/rspb.2014.0579

Atton, N., Hoppitt, W., Webster, M. M., Galef, B. G., & Laland, K. N. (2012) Information flow through threespine stickleback networks without social transmission. *Proceedings of the Royal Society B*, **279**, 4272-4278. doi: 10.1098/rspb.2012.1462

Boogert, N. J., Nightingale, G. F., Hoppitt, W., & Laland, K. N. (2014) Perching but not foraging networks predict the spread of novel foraging skills in starlings. *Behavioural Processes*, **109**, **Part B**, 135-144. doi: 10.1016/j.beproc.2014.08.016

Boogert, N. J., Reader, S. M., Hoppitt, W., & Laland, K. N. (2008) The origin and spread of innovations in starlings. *Animal Behaviour*, **75**, 1509-1518. doi: 10.1016/j.anbehav.2007.09.033

Burnham, K. P., & Anderson, D. R. (2002) *Model selection and multimodel inference: A practical information-theoretic approach.* 2 edn. New York, NY: Springer.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, **65**, 23-35. doi: 10.1007/s00265-010-1029-6

Cholé, H., Carcaud, J., Mazeau, H., Famié, S., Arnold, G., & Sandoz, J.-C. (2019) Social contact acts as appetitive reinforcement and supports associative learning in honeybees. *Current Biology*, **29**, 1407-1413. doi: 10.1016/j.cub.2019.03.025

Coussi-Korbel, S., & Fragaszy, D. M. (1995) On the relation between social dynamics and social learning. *Animal Behaviour*, **50**, 1441-1453. doi: 10.1016/0003-3472(95)80001-8

Croft, D. P., James, R., & Krause, J. (2008) *Exploring animal social networks.* Princeton, NJ: Princeton University Press.

Farine, D. R., Aplin, L. M., Sheldon, B. C., & Hoppitt, W. (2015) Interspecific social networks promote information transmission in wild songbirds. *Proceedings of the Royal Society B*, **282**, 20142804. doi: 10.1098/rspb.2014.2804

Farine, D. R., Spencer, K. A., & Boogert, N. J. (2015) Early-life stress triggers juvenile zebra finches to switch social learning strategies. *Current Biology,* **25**, 2184-2188. doi: 10.1016/j.cub.2015.06.071

Farine, D. R., & Whitehead, H. (2015) Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, **84**, 1144-1163. doi: 10.1111/1365-2656.12418

Franks, D. W., Ruxton, G. D., & James, R. (2010) Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology*, **64**, 493-503. doi: 10.1007/s00265-009-0865-8

Franz, M., & Nunn, C. L. (2009) Network-based diffusion analysis: a new method for detecting social learning. *Proceedings of the Royal Society B*, **276**, 1829-1836. doi: 10.1098/rspb.2008.1824

Franz, M., & Nunn, C. L. (2010) Investigating the impact of observation errors on the statistical performance of network-based diffusion analysis. *Learning & Behavior,* **38**, 235-242. doi: 10.3758/LB.38.3.235

Grüter, C., & Farina, W. M. (2009) The honeybee waggle dance: can we follow the steps? *Trends in Ecology and Evolution*, **24**, 242-247. doi: 10.1016/j.tree.2008.12.007

Hasenjager, M. J., & Dugatkin, L. A. (2017) Fear of predation shapes social network structure and the acquisition of foraging information in guppy shoals. *Proceedings of the Royal Society B*, **284**, 20172020. doi: 10.1098/rspb.2017.2020

Heyes, C.M. (1994) Social learning in animals: categories and mechanisms. *Biological Reviews*, **69**, 207-231. doi: 10.1111/j.1469-185X.1994.tb01506.x

Hobaiter, C., Poisot, T., Zuberbühler, K., Hoppitt, W., & Gruber, T. (2014) Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biology*, **12**, e1001960. doi: 10.1371/journal.pbio.1001960

Hoppitt, W. (2017) The conceptual foundations of network-based diffusion analysis: choosing networks and interpreting results. *Philosophical Transactions of the Royal Society B,* **372**, 20160418. doi: 10.1098/rstb.2016.0418

Hoppitt, W., Boogert, N. J., & Laland, K. N. (2010) Detecting social transmission in networks. *Journal of Theoretical Biology*, **263**, 544-555. doi: 10.1016/j.jtbi.2010.01.004

Hoppitt, W., & Farine, D. (2018) Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups. *Animal Behaviour,* **136**, 227-238. doi: 10.1016/j.anbehav.2017.08.029

Hoppitt, W., Kandler, A., Kendal, J. R., & Laland, K. N. (2010) The effect of task structure on diffusion dynamics: Implications for diffusion curve and network-based analyses. *Learning & Behavior,* **38**, 243-251. doi: 10.3758/LB.38.3.243

Hoppitt, W., & Laland, K. (2013) *Social learning: An introduction to mechanisms, methods and models.* Princeton, NJ: Princeton University Press.

Hoppitt, W., Photopoulou, T., Hasenjager, M., & Leadbeater, E. (2019) NBDA: a package for implementing network-based diffusion analysis. 0.7.10 edition. Available at: http://github.com/whoppitt/NBDA

Kulahci, I. G., Rubenstein, D. I., Bugnyar, T., Hoppitt, W., Mikus, N., & Schwab, C. (2016) Social networks predict selective observation and information spread in ravens. *Royal Society Open Science*, **3**, 160256. doi: 10.1098/rsos.160256

Laland, K.N., & Galef, B.G. (eds.) (2009) *The question of animal culture*. Cambridge, MA: Harvard University Press.

Laland, K. N., & Janik, V. M. (2006) The animal cultures debate. *Trends in Ecology and Evolution*, **21**, 542-547. doi: 10.1016/j.tree.2006.06.005

Leadbeater, E., & Chittka, L. (2008) Social transmission of nectar-robbing behavoiur in bumble-bees. *Proceedings of the Royal Society B*, **275**, 1669-1674. doi: 10.1098/rspb.2008.0270

Leadbeater, E., & Hasenjager, M. J. (2019) Honeybee communication: there's more on the dancefloor. *Current Biology*, **29**, R285-R287. doi: 10.1016/j.cub.2019.03.009

Moore, D. F. (2016) *Applied survival analysis using R.* Cham, CH: Springer International Publishing.

Morgan, B. J. T. (2010) *Applied stochastic modelling.* 2nd edn. Boca Raton, FL: CRC Press.

Nightingale, G., Boogert, N. J., Laland, K. N., & Hoppitt, W. (2014) Quantifying diffusion in social networks: a Bayesian approach. In J. Krause, D. James, D. W. Franks. & D. P. Croft (Eds.), *Animal social networks* (pp. 38-52). Oxford, UK: Oxford University Press,.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N. & Jenner, K. C. (2000) Cultural revolution in whale songs. *Nature,* **408**, 537. doi: 10.1038/35046199

Reader, S. M. (2004) Distinguishing social and asocial learning using diffusion dynamics. *Learning & Behavior*, **32**, 90-104. doi: 10.3758/BF03196010

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., … Laland, K. N. (2010) Why copy others? Insights from the social learning strategies tournament. *Science*, **328**, 208-213. doi: 10.1126/science.1184719

Schuppli, C., & van Schaik, C. P. (2019) Animal cultures: how we've only seen the tip of the iceberg. *Evolutionary Human Sciences*, **1**, 1-13. doi: 10.1017/ehs.2019.1

Terkel, J. (1995) Cultural transmission in the black rat: pine cone feeding. *Advances in the Study of Behavior,* **24**, 119-154. doi: 10.1016/S0065-3454(08)60393-9

Therneau, T. M. (2018) *coxme: Mixed effects Cox models.* R package version 2.2-10. Available at: http://CRAN.R-project.org/package=coxme

van de Waal, E., Renevey, N., Favre, C. M., & Bshary, R. (2010). Selective attention to philopatric models causes directed social learning in wild vervet monkeys. *Proceedings of the Royal Society B,* **277**, 2105-2111. doi: 10.1098/rspb.2009.2260

Whalen, A., & Hoppitt, W. J. (2016) Bayesian model selection with network based diffusion analysis. *Frontiers in Psychology,* 7, 409. doi: 10.3389/fpsyg.2016.00409

Whitehead, H. (2009) *Analyzing animal societies: Quantitative methods for vertebrate social analysis*. Chicago, IL: University of Chicago Press.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., … Boesch, C. (1999) Cultures in chimpanzees. *Nature*, **399**, 682-685. doi: 10.1038/21415

Wild, S., Allen, S. J., Krützen, M., King, S. L., Gerber, L., & Hoppitt, W. J. E. (2019) Multi network-based diffusion analysis reveals vertical cultural transmission of sponge tool use within dolphin matrilines. *Biology Letters*, **15**, 20190227. doi: 10.1098/rsbl.2019.0227

Wild, S., & Hoppitt, W. (2019) Choosing a sensible cut-off point: assessing the impact of uncertainty in a social network on the performance of NBDA. *Primates*, **60**, 307-315. doi: 10.1007/s10329-018-0693-4