A Critical Evaluation of the Relationship Intelligence Hypothesis

Submitted by Rebecca Claire Hooper to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences in July 2021

Number of pages: 233
Number of words (including all tables, legends and references): 66885

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

ABSTRACT

Cognitive abilities vary widely across the animal kingdom, and multiple hypotheses have been formulated to explain why. The Relationship Intelligence Hypothesis (RIH) posits that maintaining and managing high quality pair-bonds is a key driver of cognitive evolution. The RIH was first proposed more than a decade ago, and while its introductory paper has been widely cited, little work has been done to interrogate the RIH’s central predictions. In this thesis, I critically evaluate the RIH. First, in Chapters 2 – 4, I test central predictions of the RIH within one wild study system, the jackdaw (*Corvus monedula*). Specifically, I test whether (i) pairs have fully aligned fitness interests (Chapter 2); (ii) pairs use consolation as a mechanism through which to manage and maintain their bond (Chapter 3); (iii) pair-bond strength (1) varies between pairs, (2) is consistent within pairs, (3) positively correlates with socio-cognitive performance, and (4) positively correlates with reproductive success (Chapter 4). Finally, I test whether a commonly used method in the study of cognitive evolution - the comparative study of brain size – is methodologically robust (Chapter 5). I find that jackdaw partners do not always have fully aligned fitness interests, and that they do not use consolation to manage and maintain their pair-bond. However, pair-bond strength does vary between pairs, is consistent within pairs and is positively correlated with a measure of socio-cognitive performance, partner responsiveness. While I did not find a link between pair-bond strength and reproductive success, partners with stronger bonds were better able to adjust hatching synchrony to environmental conditions. Finally, I demonstrate that comparative studies of brain size are not methodologically robust. In Chapter 6, I integrate my findings with contemporary empirical and theoretical evidence in order to critically evaluate the RIH. I end by proposing future directions for the study of cognitive evolution.

Cover image: "Jackdaws" by Jyrki Salmi (licensed under CC BY-NC-SA 2.0)

CONTENTS

CONTENTS: Figures and Tables

Figures

Tables

LIST OF ACCOMPANYING MATERIAL

Chapter 2

CHAPTER_2_SupplementaryData.xlsx

CHAPTER_2_SupplementaryVideo1.mp4

CHAPTER_2_SupplementaryVideo2.mp4

CHAPTER_2_SupplementaryVideo3.mp4

Chapter 3

CHAPTER_3_SupplementaryVideo1.mp4

CHAPTER_3_SupplementaryVideo2.mp4

CHAPTER_3_SupplementaryVideo3.mp4

CHAPTER_3_SupplementaryVideo4.mp4

CHAPTER_3_SupplementaryVideo5.mp4

Chapter 4

CHAPTER_4_SupplementaryVideo1.mp4

Chapter 5

CHAPTER_5_SupplementaryData.xlsx

ACKNOWLEDGEMENTS

This PhD has taken a village, and I am indebted to so many people for helping to make it what it is. I have put off writing the acknowledgements until the very last minute because putting into words my gratitude for all the wonderful people who have helped me on this journey has seemed almost impossible. But, I will give it my best shot.

First of all, I'd like to thank my supervisor, Alex Thornton. I feel so privileged to have been supervised by Alex, whose passion for research, generosity with time, and thoughtfulness have made for the best possible environment in which to conduct a PhD. He has allowed me to grow as a scientist in many ways, both actively as he has guided my research, and passively as I have observed how he conducts research generally. I can only hope to emulate his approach to research that is both creative, big picture, blue sky thinking, and careful probing of the logic and details of the underlying ideas. Every time I have a chat with Alex I come away feeling reinvigorated about the things we research, and ready to face any challenges thrown my way. I hope that we have many years of collaboration ahead.

This PhD simply wouldn't exist without the hours of fieldwork put in by the whole Cornish Jackdaw Project (CJP) team. Guill McIvor will forever astound me with his encyclopaedic knowledge of the jackdaws, his grit and determination throughout every field season and his dedication to getting the best and most data possible. Victoria Lee was an incredible researcher to work alongside, and possibly the nicest human being I have ever met. Josh Arbon (aka Chinwag) has been a joy to work with – providing both stimulating science chats and lots of other chat besides (hence the nickname). Emily Cuff (aka Magoo – but she is not magoo at all) provided innumerable laughs, chats, pastries, and of course support through the odd accidental hangover. Emily also very kindly spent a summer helping me dissect naturally deceased chicks, for which I am sure I still owe her a few beers.

In my first field season, I had the absolute pleasure of working with Aude Pointecouteau, who joined the CJP as an intern. It was my first ever full season of fieldwork, and Aude was with me through every step of the way as we worked out the best way to do things. She was always a smiley presence at silly hours in the morning, and she deeply cared about the jackdaws and our research. I will always remember that season fondly. You'll be sad to hear I never got any better at keeping track of my pens though, Aude.

I would also like to thank the Master's students who I've had the pleasure to co-supervise during my PhD. In my second year, I had the joy of working with Ella Meekins. Ella and I ran the experiment in Chapter 3 together. We spent many very early mornings together, followed by long, cold days in the hide (somewhat eased by Ella's purchasing of battery-heated socks, an ingenious idea). We bonded over our love for animals and burritos, two things close to both of our hearts. Becky Brett was also a Master's student on the CJP who helped immeasurably with data collection for Chapter 5. Becky was brilliant to work with – super enthusiastic and always ready to tackle the challenges involved with the sort of data collection we were doing. Chapter 5 could not have been done without her hard work. I would also like to thank Luca Hahn for his commitment to video coding. Luca was a fantastic student who contributed massively to the data used in Chapter 4, and I have no doubt at all that he has a wonderful research career ahead of him.

I unfortunately do not have the space to list all of the students who have given their time and energy in data collection toward this PhD, but I would like to thank each and every one of them – even the ones who coded video years before I joined the CJP – for all of their hard work. I would particularly like to thank Emily Gilford, Rebecca Irwin, Anna Bowland and Angélica Bas Gomez, who were exceptional students. Emily and Rebecca put a lot of time and effort into lab work, and it was always a pleasure to spend time with them both. Anna and Angélica went above and beyond with video coding efforts, for which I am hugely grateful.

I must also extend my thanks to Paul Gluyas and his family for allowing us to conduct all sorts of weird and wonderful research on their farm. Bumping into Paul was always a highlight of the day, and I will forever hold a place in my heart for Pencoose – navigating carrying a ladder through hordes of curious calves, trying to befriend especially sassy horses, and accidentally chatting an hour away with Paul were wonderful ways to spend a spring. Also thank you to Odette Eddy and Saint Stithians Church, who allowed us to conduct research on their land.

I am not sure that I would have ended up where I am now if it wasn't for the support and mentorship of previous supervisors. Markus Eichhorn sparked my passion for research in the third year of my undergraduate degree. He made me think seriously about undertaking a PhD, and gave me guidance on the next steps to take. Also, if it weren't for him, I never would have acquired the nickname "Barnacle Beki". Andy Foote has been a fantastic mentor since my Master's project with him. Andy has always endeavoured to give me awesome opportunities – whether that was going to Iceland to hang out with orcas or having the opportunity to get involved with more killer whale research. He has been a wonderful mentor and colleague over the past few years, and I hope we carry on working together in the future. Katja Guschanski was also a brilliant supervisor, and my time as a student in her research group taught me a lot about what it means to be a good researcher. Finally, Lauren Brent was a fantastic supervisor whose passion for research was and is infectious. I am over the moon to be working with her again soon.

I could not possibly leave out the wonderful friends I have made in my PhD cohort. On the very first day of my PhD I met Aimée McIntosh and Laura Bulson, and they have remained incredible friends since. Between exploring the mountains of Majorca with Aimée, and camping in the wild of Dartmoor with Laura, we have had a whole lot of adventures packed in between our days doing research. I simply could not meet up with Ryan Imrie without planning some kind of trip, many of which we actually managed to do, and I must also thank Ryan for many fun science discussions at the pub. I particularly remember us debating cultural/genetic evolution while possibly being too tipsy to form coherent sentences. Also, Ryan was immeasurably helpful with lab work – showing me the ropes, problem-solving with me, and being amazingly generous with his time. Fraser Bell is a wonderful human and bird-whisperer who taught me a whole lot about ringing birds, and also about avoiding angry alpacas in dark fields. Ben Phillips is equally wonderful, a powerhouse of a researcher (although he would never admit to it) and a generally hilarious human (smart jopper). I couldn't possibly overlook Gin and Chinese nights with Tess Handby and Emily Cuff - a lovely way to spend a Friday night after a week of science-ing. There are just too many people to all thank personally, but getting to know and having the support of my entire cohort has truly been a wonderful experience.

My family have been a constant support throughout this PhD. I am pretty sure my mum's determination to solve any problem thrown her way and my dad's adoration and knowledge of the natural world have been the two biggest factors leading me to where I am today. Until midway into my PhD I also had the support of my wonderful nana and grandpa. My nana would ask the most left-field, interesting questions about birds, and always made me think in new ways about things. My grandad was and is always fascinated to hear about the jackdaws, and would regale me with stories of the antics of the jackdaws in his area whenever I saw him. So, thank you to all of my family for inspiration, belief and support throughout the last few years.

The biggest thanks of all go to Joe Wilde, who has been my rock throughout this PhD and the small matter of a global pandemic. There is no one on this planet I would rather have been stuck inside with for months at a time, and I quite simply could not have done this without him.

Finally, I would like to thank E.O. Wilson for writing the book "Letters to a Young Scientist". There were times when I worried I wasn't the right type of person to be a scientist, but this book made me realise that wasn't the case. The quote *"The ideal scientist thinks like a poet and only later works like a bookkeeper. Keep in mind that innovators in both literature and science are basically dreamers and storytellers"* made me realise that maybe the things I considered my weaknesses could actually be strengths.

AUTHOR’S DECLARATION

The work of this thesis was carried out under the supervision of Dr Alex Thornton. No part of this dissertation has previously been submitted for a qualification. All work is my own, except that specific below.

Chapter 2 CJP fieldwork assistants and video coding interns helped to collect and code video data. The CJP team ringed jackdaws and collected morphometric and blood data. Emily Cuff assisted in dissections of deceased chicks. Emily Gilford and Rebecca Irwin assisted with DNA extraction and PCR sexing. Pablo Capilla Lasheras and Alex Hayward offered advice on methods. Karen Moore conducted library construction and sequencing at Exeter Sequencing Services. Kathryn Maher assisted with bioinformatic analysis. Jochen Wolf and Matthias Weissensteiner allowed us access to a draft version of the jackdaw genome. Jisca Huisman offered valuable assistance with pedigree reconstruction. Guillam McIvor confirmed the ID of males involved in FEPC. Joseph Wilde helped with writing the behavioural data extraction script. Kathryn Maher, Karen Moore, Jisca Huisman, Guillam McIvor and David Hosken provided feedback on draft versions of the chapter.

This chapter is in preparation for publication as:

Hooper, R., Moore, K.M., Maher, K.H., McIvor, G.E., Hosken, D.J., Thornton, A. Ultimate drivers of forced extra-pair copulations in birds lacking penises: jackdaws as a case-study.

Chapter 3 Ella Meekins, a Master’s student with the CJP, created the playback audio sequences in Audacity, helped to run the playback experiments and coded ~half of the experimental video data. Angélica Bas Gomez, Anna Bowland, Coby Thompson-Knight, Emma Doyle, Lucy Penney and Sam Mosedale coded control videos. Erik Postma provided valuable feedback on the statistical analyses. Guillam McIvor, Ella Meekins, Aimée McIntosh and Joseph Wilde provided feedback on draft versions of the chapter.

This chapter has been published as:

Hooper R, Meekins E, McIvor G.E. and Thornton A. (2021) Wild jackdaws respond to their partner's distress, but not with consolation. R. Soc. open sci.8210253210253. http://doi.org/10.1098/rsos.210253

Chapter 4 Luca Hahn, a Master's student with the CJP, coded all videos filmed in the nest-building stage. Ella Meekins, Angélica Bas Gomez, Anna Bowland, Amy Hall, Coby Thompson-Knight, Emma Doyle, Emily Cuff, Lucy Penney, Gray Wirtanen, Sam Mosedale, Luca Hahn and Jonathan Hill coded incubation-stage videos. Life history data were collected by the CJP team. Neeltje Boogert, Josh Arbon and Guillam McIvor provided feedback on draft versions of the chapter.

This chapter is in preparation for publication as:

Hooper, R., Hahn, L.G., McIvor, G.E., Thornton, A. Pair-bond strength is repeatable and related to partner responsiveness in a wild corvid.

Chapter 5 Becky Brett, a Master's student with the CJP, helped to collate brain and body size data, and collected associated social and ecological variables. Resources from Tom Currie's workshop on PGLS modelling were used for the analysis.

This chapter is in preparation for publication as:

Hooper, R., Brett, B., Thornton, A. Further methodological caveats of comparative studies of brain size.

CHAPTER 1: Introduction

Background Cognition can be defined as an individual's ability to acquire, process, store and use information from its environment (Shettleworth 2010b). The building blocks of cognition – neurons (Dicke & Roth, 2016; G. Roth & Dicke, 2005) – are energetically intensive to use (Herculano-Houzel, 2011; Sterling & Laughlin, 2015). Indeed, 20% of a brain's energy usage is through the firing of neurons (Sterling & Laughlin, 2015). Yet neuron number varies vastly across the animal kingdom (Montgomery et al., 2018; Sterling & Laughlin, 2015). *Caenorhabditis elegans* survive and thrive with only 302 neurons (White et al., 1986), while humans have around 86 billion (Azevedo et al., 2009). Given their cost, for neurons to be maintained by selection they must offer some adaptive advantage. The fundamental advantage of possessing neurons is neatly summed up by Sterling and Laughlin, 2017, who state that for an organism to survive in the world, it must "*exchange amounts of information with its external environment*". Without the ability to do this to some degree, survival (and therefore, reproduction) would not be possible. But why is there so much variation in neuron number across species? A simple answer is that some species must exchange more information with their environment than others. Why this is the case, however, remains unclear, despite an array of hypotheses (Wartel et al., 2019). Arguably the most influential of these hypotheses, in terms of historically dominant narrative and research effort (Rosati, 2017), is the Social Intelligence Hypothesis (SIH) (Dunbar, 1998; Humphrey, 1976; Jolly, 1966).

First developed in the 1950s by Chance and Mead (1953), the underpinnings of the SIH were later expanded by Jolly (1966) and Humphrey (1976). The central premise of the hypothesis is that successfully navigating a "complex" social landscape is both informationally (and thus, cognitively) demanding and positively related to fitness outcomes. Thus, "complex sociality" generates selection on cognitive ability. Accordingly, we would expect to see a relationship between measures of social complexity and measures of cognitive ability. However, both social complexity and cognitive ability are difficult to quantify. Early investigations of the SIH thus used group size (i.e., quantity of social relationships) as a proxy of social complexity, and a measure of brain size as a proxy of cognitive ability (Dunbar, 1992, 1995). The idea that brain size should increase with measures of social complexity was dubbed the Social Brain Hypothesis (SBH) and received empirical support in primates. Specifically, a strong positive relationship was found between relative neocortex size and group size in primates (Dunbar, 1992, 1995). The neocortex is involved in higher-order cognitive functions (Funahashi, 2001). This result thus implies that individuals may require more sophisticated cognitive ability in order to navigate larger groups. However, this finding has proved contentious. While some studies do support that brain size and group size are positively correlated in primates (Barton, 1996; Dunbar, 1992, 1995; Dunbar & Bever, 1998; Shultz & Dunbar, 2007), others have failed to find a robust relationship (DeCasien et al., 2017; Powell et al., 2017; Wartel et al., 2019). Moreover, the same relationship is not found in non-primate taxa. Instead, pair-bonding and brain size appear to be positively correlated in many non-primates, such as birds, ungulates, bats and mammalian carnivores (Dunbar & Shultz, 2007; Emery et al., 2007; Shultz & Dunbar, 2010). The discovery of this relationship led to the formulation of the Relationship Intelligence Hypothesis (RIH) (Emery et al., 2007). The RIH can be considered a sub-branch of the SIH, and proposes that maintaining and managing a pair-bond is cognitively demanding (Emery et al., 2007). Thus, the relevant measure of “social complexity” in the RIH is the *quality* of bonds (Emery et al., 2007). The hypothesis was specifically formulated to explain cognitive evolution in birds; however, it should in principle apply to other taxa.

“Social complexity” is central to both the SIH and the RIH, but what exactly does it mean? There is no straightforward definition (Lukas & Clutton-Brock, 2018). Some researchers consider societies where individuals have highly aligned fitness interests as complex (e.g.(Abbot & Chapman, 2017; Anderson & McShea, 2001; Emery et al., 2007)). These societies tend to be characterised by, for example, cooperation, reproductive suppression, alloparental care and division of labour (Lukas & Clutton-Brock, 2018). This would lead to the conclusion that eusocial, cooperatively breeding and genetically monogamous pair-bonded species are particularly complex. On the other hand, some researchers consider societies where individuals have conflicting interests as complex (e.g.(Bergman & Beehner, 2015; Byrne & Whiten, 1988; Freeberg et al., 2012)). These societies tend to be characterised by, for example, low levels of relatedness, differentiated relationships and coalition formation (Lukas &

Clutton-Brock, 2018). When proponents of the SIH refer to social complexity, they tend to be referring to the latter, such that a society characterised by many individuals with conflicting interests generates cognitive challenges (Dunbar, 1995, 1998; Humphrey, 1976). The RIH, however, focusses more on the former.

In the original formulation of the RIH, Emery et al. (2007) propose that the informational challenges (i.e. “complexity”) of long-term pair-bonding arises because partners have aligned interested, and thus are highly invested in one another and the maintenance of the pair-bond (Emery et al., 2007). The RIH was formulated using corvids as an exemplar taxon. As in primates, corvids have large brains relative to their body size (van Horik et al., 2012) and despite smaller brains than non-human primates, have a similar number of neurons (Olkowicz et al., 2016). They also show similar cognitive performance to non-human primates (Güntürkün & Bugnyar, 2016). Corvids tend to form long-term, monogamous pair-bonds (Emery et al., 2007). Emery et al. (2007) assume that these long-term pair-bonds tend to be genetically as well as socially monogamous, meaning that partners have fully aligned fitness interests. However, evidence for genetic monogamy is limited. Moreover, whether non-sexual relationships beyond the pair-bond also influence the informational demands faced by partners is unclear. Thus, to what degree interests of long-term partners align is uncertain, yet must be understood in order to fully interrogate the informational demands of pair-bonding, and to evaluate the RIH.

But what precisely is it about the pair-bond that is thought to generate informational demands? Tracking and responding to subtle behavioural changes of the partner is considered to be a key challenge (Emery et al., 2007). However, tracking and responding to the behaviour of others does not necessarily require a great deal of information processing. For instance, flocking and shoaling species with a range of cognitive abilities are capable of tracking and responding to the fine-scale movements of multiple neighbours (Ballerini et al., 2008; Herbert-Read et al., 2011; Kotrschal et al., 2018). Indeed, selecting for guppies (*Poecilia reticulata*) with bigger brains and better cognitive performance did not improve their shoaling ability, indicating that the cognitive mechanisms involved in this behaviour are relatively simple (Kotrschal et al., 2018). However, Emery et al. (2007) invoke behaviours that potentially require more sophisticated information-processing. For example, they suggest that individuals may manage their pair-bond by engaging in third-party post-conflict affiliation with their partner, possibly as a mechanism through which to decrease the stress levels of the partner (Emery et al., 2007). Such behaviour, referred to by some researchers as consolation (Pérez-Manrique & Gomila, 2018), has been argued to require empathy (de Waal, 2010). Whether non-humans show empathy is, however, controversial. For instance, many studies that invoke consolation as an explanation of observed behaviours do not rule out alternative explanations, such as self-soothing of the putative “consoler” or solicited affiliation (Adriaense et al., 2020; Pérez-Manrique & Gomila, 2018). Thus, whether consolation actually occurs in non-humans is unclear. In addition, the majority of studies of consolation are conducted on captive or semi free-ranging species. While such studies are valuable, social interactions differ drastically between wild and captive populations (Inoue & Shimada, 2020; Pacheco & Madden, 2021). Thus, it is difficult to understand whether and when a behaviour would naturally occur, and where it is adaptive, without testing for it in an ecologically relevant environment (Boesch, 2020; Cauchoix et al., 2020; Thornton & Lukas, 2012; Webster & Rutz, 2020). Whether wild animals manage their relationships using strategies that require empathy is therefore not well-known.

The SIH and the RIH propose that cognitive evolution occurs because forming and maintaining social bonds requires socio-cognitive ability, and social bond quantity and/or quality positively influences fitness outcomes (Dunbar, 1998; Emery et al., 2007; Humphrey, 1976). As a result, both make some fundamental “Darwinian” predictions. In particular, they predict that the ability to form social bonds is (i) variable between individuals, (ii) consistent and heritable, and (iii) related to positive fitness outcomes. For the SIH, which focusses on both the quantity and quality of social bonds, there is some empirical support for each of these predictions. Many social taxa have differentiated social bonds (Aplin et al., 2015; Bergman & Beehner, 2015; Braun & Bugnyar, 2012; Heathcote et al., 2017; Seyfarth & Cheney, 2012; M. J. Silk & Hodgson, 2021; Weiss et al., 2021), and an individual’s propensity to form social bonds has been shown to be consistent and heritable in some mammalian taxa (Aplin et al., 2015; Brandl et al., 2021; Brent et al., 2013; Finkenwirth & Burkart, 2017; Koski et al., 2012;

Massen & Sterck, 2013; Mitani, 2009; Silk et al., 2010a, 2010b; Stanley et al., 2018). For instance, some female chacma baboons (*Papio hamadryas ursinus*) form stronger, more enduring bonds than others (J. B. Silk et al., 2010b), while rhesus macaques (*Macaca mulatta*) have consistent and heritable social network positions (Brent et al., 2013)**.** Moreover, individuals with more and/or stronger bonds have been shown to have higher fitness across a broad range of taxa (Cameron et al., 2009; McFarland et al., 2017; Schülke et al., 2010; Silk et al., 2010b; Silk & Hodgson, 2021; Spoon et al., 2006). For example, more socially integrated female feral horses have a higher foal birth rate and survival rate (Cameron et al., 2009), and chacma baboons that form stronger social bonds have increased survival (J. B. Silk et al., 2010b)**.** Thus, there is evidence across a range of taxa that the ability to form social bonds varies between individuals, is consistent and heritable within individuals, and results in differential fitness. However, empirical support for the RIH – focussing specifically on the quality of the pair-bond – is scarcer. Indeed, while pair-bond strength has been shown to vary between pairs in some species (Boucherie et al., 2018; Elie et al., 2011; Firth et al., 2018; Sparks, 1964; Spoon et al., 2004, 2006), whether an individual's ability form a strong pair-bond is consistent and heritable is generally unknown (but see Kralj-Fišer et al., 2007, who show that partner proximity is repeatable between wild greylag geese, *Anser anser*). Furthermore, while several studies have demonstrated that the length of the pair-bond and the familiarity of partners (both likely to be facets of pair-bond strength) influence fitness outcomes (Black, 2001; Culina et al., 2020; Naves et al., 2007; Sánchez-Macouzet et al., 2014; Van De Pol et al., 2006), only one study has directly interrogated whether pair-bond strength influences fitness (Spoon et al., 2006). Spoon et al. (2006) demonstrated that captive cockatiel (*Nymphicus hollandicus*) pairs with stronger bonds raise more chicks to independence. However, no research has yet been undertaken to test how pair-bond strength influences reproductive success in an ecologically relevant environment. In sum, the RIH has scarce support for its fundamental "Darwinian" predictions, and current available evidence for both the SIH and the RIH is from a range of disparate taxa. Testing multiple predictions of the RIH within a single wild study system is thus a crucial next step in evaluating this hypothesis.

Study system In this thesis, I examine predictions of the RIH within a single wild study system, the jackdaw (*Corvus monedula*). Jackdaws are a socially monogamous, colonially breeding species that form long-term pair-bonds and stay with their partner throughout the year (Cramp & Perrins, 1994; Henderson et al., 2000; Kubitza et al., 2015; Liebers & Peter, 1998; Ling et al., 2019; Röell, 1978; Wechsler, 1989). While pair-bonded adults affiliate with other flock members, the majority of affiliative behaviours are directed toward their partner (Kubitza et al., 2015). During the breeding season, both members of the pair cooperate to build the nest (Hahn et al., 2020; Röell, 1978), and the female incubates the eggs while her partner brings her food (Röell, 1978). Both parents care for the chicks until fledging, and the pair tend to have only one breeding attempt per year (Cramp & Perrins, 1994; Röell, 1978). In multiple populations, jackdaws have been shown to have no to low rates of extra-pair paternity, implying that they are often genetically as well as socially monogamous (Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021). However, rates of extra-pair paternity can vary substantially between populations (e.g. Birkhead & Moller, 1998; Gray, 1996; Hatchwell, 1988; Westneat, 1992). Thus, genetic monogamy should not be taken for granted in any untested population. Given that long-term pair-bonds are the foundational unit of jackdaw social life, they are an excellent species in which to test key predictions of the RIH (and, in parallel, the SIH).

In this thesis, all field-based studies were conducted on three wild populations of jackdaws, which form the Cornish Jackdaw Project (CJP). All sites are based in Cornwall, UK. Site X is based on the university campus, a semi-urban environment (50°17′32″N; 5°11′96″W). Size Y is based in a village churchyard and adjacent fields (50°11′26″N, 5°10′51″W). Site Z is based in a working farm (50°11′56″N, 5°10′9″W). Since 2013, the CJP has ringed over 3000 jackdaws with unique colour combinations and monitored the reproductive success of ~85 pairs per year, who breed in project-owned nest-boxes. Reproductive success data includes clutch size, hatching synchrony, fledgling number and fledgling mass. All nestlings where the eldest in the clutch reaches 25 days old are also ringed with unique colour combinations. During ringing of chicks and adults, body condition data is collected, and blood samples are taken for molecular sexing. Each nest-box has an internal shelf that allows for the placement of a

CCTV camera. Thus, along with detailed data on who occupies nest-boxes and what their reproductive outcomes are, detailed behavioural data is also recorded from within the nest-box.

Thesis framework In Chapter 2, I examine the mating system of jackdaws in our three study populations. Specifically, I ask: are jackdaws in these populations genetically monogamous? The RIH assumes that corvids with long-term pair-bonds are genetically monogamous (Emery et al., 2007), and thus that they have fully aligned fitness interests. This has implications for the cognitive challenges faced by individuals within a pair-bond. Therefore, understanding exactly how aligned the fitness interests of jackdaw partners are is essential to elucidating the informational demands potentially faced by partners. Beyond quantifying rates of extra-pair paternity, I also investigate the extra-pair mating behaviour of jackdaws in detail, shedding further light on behaviours that may introduce conflict of interest between partners. I then conduct a literature review to attempt to elucidate the ultimate reasons that this behaviour occurs.

In Chapter 3, I test for consolation between jackdaw partners. Consolation between jackdaws would suggest that partners may be capable of sympathetic or empathetic concern, and that they use this behaviour to maintain and manage the pair-bond. As discussed previously, former tests of consolation tend to be (i) unable to rule out alternative explanations for behaviour and (ii) undertaken on captive or semi free-ranging species, where the ecological relevance of the behaviour is unclear. I therefore designed an experiment on jackdaw pairs that ruled out alternative explanations of behaviour, and that was conducted in an ecologically relevant setting.

In Chapter 4, I interrogate four key predictions of the RIH (and in tandem, the SIH) within jackdaws. Specifically, I ask whether bond strength is (i) variable between pairs, (ii) consistent within pairs, (iii) positively correlated with socio-cognitive performance and (iv) positively correlated with fitness outcomes. This is the first time that multiple key predictions of either the SIH or RIH have been tested within a single study system, and indeed the first time that prediction (iii) has been directly tested in a non-human species.

In Chapter 5, I ask whether comparative brain size studies, such as those used to show that elements of sociality correlate with brain size, are a methodologically robust approach to interrogating cognitive evolution. Given that both the SIH and RIH are formulated from the results of such studies, it is vital to know if they are methodologically sound.

In Chapter 6, the General Discussion, I integrate the results of each data chapter into a review of the RIH, which leverages contemporary theoretical and empirical results to critically evaluate the hypothesis. I conclude the thesis by examining differing approaches to the study of cognitive evolution, asking why these differences arise, and proposing future directions in the study of cognitive evolution.

Note: Each data chapter is written as an independent piece of work. Certain information is therefore repeated, particularly in Methods sections. Apologies for any inconvenience this causes.

CHAPTER 2: Ultimate drivers of forced extra-pair copulations in birds lacking penises: jackdaws as a case-study

Abstract

Forced copulation is common, presumably because it can increase male reproductive success. Forced extra-pair copulation (FEPC) has been reported in birds, many of which lack a penis and thus are thought to require female cooperation for successful fertilisation. Why FEPC persists, despite a presumed lack of siring success and likely non-negligible costs (FEPCs often involve violent conflict) is unknown. Using the jackdaw (*Corvus monedula*) as a case study, we (i) use SNPs to quantify maximum extra-pair paternity rate through FEPC and (ii) evaluate support for seven hypotheses interrogating why FEPC exists in species without a penis. We then collate evidence for FEPC across penis-lacking birds. We find that FEPC in jackdaws is probably a relic from a time when it led to siring success, which if true, makes its persistence enigmatic given its presumed costs. Across birds lacking a penis, FEPC is taxonomically widespread, yet very little is known about its evolution and maintenance. Broader implementation of the approach used here will shed light on why this widespread sexual behaviour persists. Additional work is necessary to fully understand whether a penis is needed for paternity through forced copulation, and to quantify costs of FEPC incurred by the male and the victim female.

Introduction

Copulations that are unsolicited and resisted by females (forced copulations), are common across the animal kingdom (Clutton-Brock & Parker, 1995). Forced copulation exists as either a discrete or plastic reproductive tactic (Gross, 1996; Kustra & Alonzo, 2020). Most commonly, forced copulation is a plastic, condition-dependent strategy where poor quality males, unable to acquire mating through female choice, make the “best of a bad job” (Eberhard, 1982; Thornhill, 1981; Thornhill & Palmer, 2000). For instance, small male swordtails (*Xiphophorus multilineatus*) force copulations with females, while larger (more attractive) males court females who then cooperatively copulate (Rios-Cardenas et al., 2018). Similarly, in orangutans (*Pongo* spp.), females mate cooperatively with dominant (generally older) males, while subordinate (generally younger) males often force copulations (Galdikas, 1985; Schürmann

& van Hooff, 1986; Utami et al., 2002). Forced copulation would be evolutionarily favoured if it resulted in paternity (Eberhard, 1982; Rios-Cardenas et al., 2018; Thornhill, 1981), and it has been shown to enhance siring success in several species (e.g., Brekke et al., 2012, 2013; Hogg & Forbes, 1997). For species with internal fertilisation, it has been suggested that an intromittent organ (e.g. a penis) is necessary for forced copulations to result in siring success, because without a penetrative organ, female cooperation is required to achieve fertilisation (Briskie & Montgomerie, 2001; Fitch & Shugart, 1984; Gowaty & Buschhaus, 1998). Males of most taxa have penises, but 97% of bird species do not (Briskie & Montgomerie, 1997). Puzzlingly, forced copulations have been reported in several bird species that do not have penises (Gladstone, 1979; Gowaty & Buschhaus, 1998; McKinney et al., 1984; McKinney & Evarts, 1998; Westneat & Stewart, 2003). Here, we evaluate possible explanations for the existence of forced copulations in species lacking a penis.

The majority of bird species are socially monogamous (Lack, 1968), but forced extra-pair copulation (FEPC) is common in many species (Gladstone, 1979; Gowaty & Buschhaus, 1998; McKinney et al., 1984; McKinney & Evarts, 1998; Westneat & Stewart, 2003). FEPC is most commonly observed in bird species with penises capable of penetrating the female cloaca (Adler, 2010; Briskie & Montgomerie, 2001; Briskie & Montgomerie, 1997; McKinney & Evarts, 1998), and is most often adopted as part of a mixed reproductive strategy (i.e., males have a monogamous partner yet still engage in FEPC: McKinney & Evarts, 1998). In these species, males probably increase their reproductive success through FEPC (McKinney & Evarts, 1998). For FEPC to persist, benefits for the male (e.g. siring success) must exceed costs. Male costs are difficult to quantify, but are likely to be non-trivial. Males are at risk of injury during FEPC, either due to fights with the female or her partner (Gill et al., 2020; Gladstone, 1979), energy must be expended to engage in FEPC (Adler, 2010), and males engaging in FEPC leave their own partner unguarded and thus risk losing paternity there (Sorenson, 1994). Additionally, in species that lack a penis, the prevailing, although largely untested, viewpoint is that female cooperation (i.e., where the female positions herself so as to allow cloacal contact: McKinney & Evarts, 1998) is needed for fertilisation to occur (Briskie & Montgomerie, 2001;

Fitch & Shugart, 1984; Gowaty & Buschhaus, 1998). If FEPC cannot or rarely results in siring success, this generates the question: why do birds without a penis engage in FEPC if this is costly and will not result in paternity?

There are seven different, non-mutually exclusive hypotheses for the evolution and maintenance of FEPC (Table 1). (1) FEPC is maintained by selection due to direct fertilisation benefits. (2) FEPC is an evolutionary 'relic'. It is likely that ancestrally, all birds had penises (Montgomerie & Briskie, 2007). The relic hypothesis suggests that while FEPC ancestrally provided paternity benefits, it now persists as a non-adaptive vestigial behaviour. (3) FEPC is a pleiotropic by-product of another trait that is selectively advantageous for males (Thornhill & Palmer, 2000). For example, high testosterone levels may be advantageous but result in hyper-sexual drive, leading to FEPC (Davis, 2002). The remaining hypotheses posit that FEPC is selectively favoured, even though it does not directly result in paternity. These are (4) the 'Creation Of a Dangerous Environment' (CODE) hypothesis (Gowaty & Buschhaus, 1998), (5) the 'territory signalling' hypothesis (outlined in Gill et al., 2020), (6) the 'reproductive suppression' hypothesis and (7) the sperm quality hypothesis. The CODE hypothesis posits that males use FEPC to create an environment of fear for females. This is beneficial for males because it selects for females to seek male protection, and to trade this protection for copulation (Gowaty & Buschhaus, 1998). The 'territory signalling' hypothesis stemmed from observations that male jackdaws (*Corvus monedula*) produce loud copulation calls both when mating with their own partner, and when engaging in FEPC, suggesting they serve a territorial function (Gill et al., 2020), whereby males attempt to claim future territory by forcing copulation with the resident female and signalling this behaviour through loud copulation calls (Gill et al., 2020). Gill *et al.* (2020) also raise the possibility that FEPC may suppress the reproductive success of the target female. We label this the 'reproductive suppression' hypothesis, whereby males with a mated partner ('paired males') use FEPC as a tool to suppress the reproductive success of their neighbours, thus decreasing competition for themselves and/or their offspring. Finally, FEPC could serve to enhance sperm quality. Baker & Bellis (1993) suggested that masturbation evolved as a mechanism to maintain ejaculate quality through sperm turnover. Following this logic, males could be forcing copulation with non-partners to ensure only high-quality sperm fertilises their partner's eggs. Indeed, hatching success and chick quality has been shown to decrease significantly for pairs whose eggs are fertilised by older sperm in black-legged kittiwakes (*Rissa tridactyla*) (R. H. Wagner et al., 2004), suggesting this is a feasible idea. Here, we use genomic and behavioural examination of jackdaws (*Corvus monedula*) combined with a systematic literature review to interrogate the reproductive consequences of FEPC, evaluate support for possible explanations (Table 1), highlight the breadth of FEPC in birds lacking a penis, and underscore where future work would be best focused to aid in our understanding of the function of FEPC.

Jackdaws are long-term monogamous, colonially breeding corvids (Cramp & Perrins, 1994; Henderson et al., 2000; Liebers & Peter, 1998; Röell, 1978). Pairs have one brood per year, and the female fertile period is highly synchronous (Röell, 1978). Divorce is rare and pairs typically form in their first year and stay together for life (Cramp & Perrins, 1994; Röell, 1978). While cooperative extra-pair copulations have never been observed in jackdaws, in one population 82% of monitored females experienced at least one FEPC event during the breeding season (Gill et al., 2020). However, multiple studies, including that which found high rates of FEPC, show no or very low rates of extra-pair paternity (Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021). Rates can vary drastically between populations though (e.g. Birkhead & Moller, 1998; Gray, 1996; Hatchwell, 1988; Westneat, 1992), highlighting the need for population-specific genetic pedigrees to be built. We use single nucleotide polymorphism (SNP) data from double digest restriction site-associated DNA (ddRAD), a relatively new and under-utilised method for pedigree reconstruction (Brouwer & Griffith, 2019), to build a genetic pedigree of our populations. We then compare this with our detailed social pedigree, allowing us to estimate the rate of extra-pair paternity in our populations, and thus elucidate the maximal rate at which FEPC leads to fertilisation. Following this, we use fine-scale behavioural data to evaluate support for each of the six hypotheses outlined above. Finally, we expand our taxonomic focus to all species lacking penises. Although there have been several reviews of FEPC in the past (Gladstone, 1979; Gowaty & Buschhaus, 1998; McKinney et al., 1984; McKinney & Evarts, 1998; Westneat & Stewart, 2003), the most recent was almost 20 years ago and numerous further studies have reported FEPC behaviour in species lacking a penis since. We therefore collated a list of species where FEPC has been reported, in order to evaluate the breadth of FEPC behaviour across birds lacking penises, and to identify patterns of where FEPC occurs.

Table 1 Hypotheses for the evolution of FEPC in birds lacking intromittent organs, associated predictions and whether the prediction is supported, refuted or unknown in jackdaws. Y = yes, supported. N = no, not supported. U = unknown. * indicates that this is what current evidence suggests, but more research is needed.

Hypothesis		**Adaptive?**	**Directional Predictions**	**Support in Jackdaws**
Immediate fertilisation enhancement	FEPC is adaptive because it leads to fertilisation	Adaptive	FEPC results in fertilization success	Unlikely due to no to low rates of extra-pair paternity
Immediate fertilisation enhancement: Relic	FEPC used to be adaptive; however, it is not costly enough to have been selected out of the population	Non-adaptive	Fertilization success from FEPC may be absent or very low	Y
			FEPC occurs closer to the target female's fertile period	Y
			Males increase mate-guarding when their mate is fertile	Y
By-product	FEPC is a by-product of a different, beneficial trait	Non-adaptive	Fertilization success from FEPC may be absent or very low	Y
			Tendency to engage in FEPCs co-occurs with another trait, for example testosterone levels	U
			The trait with which FEPCs co-occur provides a fitness advantage for males	U
Creation of a dangerous	FEPC creates a dangerous environment for	Adaptive	Fertilization success from FEPC may be absent or very low	Y

environment (CODE)	females which fosters male mating advantage via social monogamy		Female fertility has no effect on FEPC behaviour	N
			Both paired and unpaired males attempt copulation	U
			Females' vulnerabilities to male aggression against them varies, so that e.g. smaller females have increased vulnerability to male aggression	N
			Males direct aggressive copulations at unmated or unguarded females	Y
			Male mate-guarding is not correlated with their partner’s fertility	N
Territory signalling	Males engage in FEPCs to claim territory	Adaptive	Fertilization success from FEPC may be absent or very low	Y
			Males advertise FEPC	Y
			Males who engage in FEPC are more likely to acquire territory	U
			Females in more desirable territories are more likely to be subject to FEPC	U
			FEPC occurrence is not correlated with the target female’s fertility	N
			Male mate-guarding is not correlated with their partner’s fertility	N
Reproductive suppression	Males engage in FEPCs to suppress the	Adaptive	Fertilization success from FEPC may be absent or very low	Y

	reproductive success of neighbours whose offspring would compete for resources with their own		Paired neighbours are more likely to engage in FEPC than other males	Y*
			FEPC results in fitness costs for the pair whose female experienced FEPC	U
			FEPC occurrence is not correlated with the target female’s fertility	N
			Male mate-guarding is not correlated with their partner’s fertility	N
Sperm turnover	Males engage in FEPCs to improve sperm quality before mating with their partner	Adaptive	Fertilization success from FEPC may be absent or very low	Y
			Males engage in FEPC more when their partner is fertile	N*
			Engaging in FEPC increases sperm quality/fertilisation success	U
			FEPC occurrence is not correlated with the target female’s fertility	N
			Male mate-guarding is not correlated with their partner’s fertility	N

Methods

Ethics This study was carried out following the ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching (Buchanan et al., 2012). Subjects were not captured for the purpose of this study, but had been previously captured and ringed by Cornish Jackdaw Project team members licensed by the British Trust for Ornithology and UK Home Office (project licence 30/3261).

Study sites From 2013-2019, the Cornish Jackdaw Project has collected behavioural and life history data from three sites in Cornwall, UK (Site X 50°17′32″N; 5°11′96″W; Site Y 50°11′26″N, 5°10′51″W; Site Z 50°11′56″N, 5°10′9″W). Across these field sites, ~85 nest-boxes were monitored throughout each breeding season, starting from when the jackdaws started building nests in March to when the chicks fledged in June. Nest-box owners were ringed after being caught through ladder trapping or using trapdoors at the nest-box. Non-box-owning birds were also ladder-trapped and ringed for individual identification. Nest-box chicks were ringed when the eldest in the clutch was 25 days old. During ringing of chicks and adults, tarsus length (an indicator of structural size; Rising & Somers, 1989) was recorded for each individual, and blood samples were collected by licenced members of the team and used for molecular sexing (as described by Griffiths *et al.,* 1998)*,* and pedigree reconstruction.

Pedigree analysis ***Blood/liver sampling*** For chicks that survived until 25 days old, blood samples for DNA extraction were collected at ringing. In 2018 and 2019, nest-boxes with chicks <25 days old were monitored closely so that deceased chicks could be collected before removal from the nest by parents. Deceased chicks were frozen at -20℃ as soon after death as possible, and tissue samples from the liver were collected for DNA extraction.

DNA extraction DNA was extracted from blood using Thermo Scientific GeneJET Whole Blood Genomic DNA Purification, and from liver using QIAGEN DNeasy Blood & Tissue Kit, both following manufacturer’s protocols. To validate sexes of social parents, and to sex offspring, we used PCR with P2P8 primers (Griffiths et al., 1998), and assigned sex visually following gel electrophoresis. All genetic and social sexes (i.e., sex assigned following behavioural observation) of parents matched. Samples were quality controlled for molecular weight and salt/protein contamination using gel electrophoresis with a 1kb ladder and Nanodrop, respectively. Samples showing DNA degradation or considerable salt/protein contamination were re-extracted. All samples were standardised to 10ng/µl in 40µl.

Sample selection Samples for genomic sequencing were selected based on fully sampled family units (i.e. broods where both social parents had also been sampled), site (sample number was proportional to site size), and quality of

DNA extraction (some liver samples had many bands below 1kb and high levels of protein contamination; these samples were therefore dropped from the sequencing pool and replaced with high quality samples from different individuals). Five duplicate samples (from the same DNA extraction) were included in the sequencing pool in order to estimate approximate sequencing error rate. We also included known social half-siblings (known through re-pairing of the social parent) so we could examine whether our final analysis had the power to detect half-sib relationships, which is key to estimating extra-pair paternity rates. A female with no known offspring or parents was also included, because she had been observed for several years to associate with a pair who were included (along with their offspring) in the sample selection. We also had video evidence of this female laying an egg in that pair’s nest-box (see Methods: behavioural data).

Our final sample for sequencing consisted of 188 individuals (plus five duplicates) across two sites. 149 individuals were included from Site YZ, comprising 113 (plus two duplicate) offspring from 47 broods and 22 sibships. Two parent samples were also duplicates. From Site X, 43 individuals were included, which comprised of 33 (plus one duplicate) offspring from 13 broods and six sibships. Quality control of DNA extracted from liver samples revealed a moderate level of degradation in many samples. Therefore, only six samples from deceased chicks were included in the analysis; all other samples were extracted from blood.

RadSeq library preparation Selected DNA samples were sent to Exeter Sequencing Services for library preparation and sequencing. The ddRADseq library preparation has been modified from Peterson et al., 2012, to enable multiplexing using Nextera dual indexes (Illumina) together with Pst1 and EcoRI restriction digest. Nextera indexing enables demultiplexing of 192 samples using the same pipelines as other Illumina projects. To this end, the adapter sequences have been modified from the Illumina Nextera/amplicon protocol. 16 oligonucleotides make up eight different P1 adapters for ligation to the Pst1 overhang and 2 oligonucleotides make up a single P2 adapter for ligation to the EcoR1 overhang. P1 adapters that ligate to the Pst1 overhang have six bases 5’ to the TGCA overhang which forms an in-line barcode and ensures sufficient complexity at the start of Read 1 for cluster identification. In ddRADseq libraries, multiple identical fragments resulting from the double digest cannot be separated from those generated by PCR during indexing. Therefore, the P2 adapter has six bases (YRYRYR) added 3' to the EcoR1 overhang which acts as a unique molecular identifier (UMI) and enables PCR duplicates to be removed during analysis. This UMI would be expected to have a complexity of 4096, which should be sufficient to identify PCR duplicates in the context of ddRAD, where only a limited amount of the genome is expected to be present each the library. Modification to the Illumina sequencing library design made for Pst1 EcoR1 RADseq is illustrated in Figure S1. Libraries were size selected using a double SPRI bead-based method resulting in fragments between 450bp and 800bp. Pooled libraries were 150 paired-end sequenced on the Novaseq 6000 (Illumina Corp. USA). There was high variability in read count per sample following the first round of sequencing, so a second round was undertaken on 174/190 samples. Both rounds underwent identical library preparation and sequencing protocols, and fastq files were therefore combined following sequencing and before data processing.

Bioinformatic pipeline Adapters were removed and bases were filtered for a mean Phred score of 20 from the 3' end using fastp 0.20.1 (Chen et al., 2018). Reads shorter than 75 base-pairs post-filtering were removed. Clone filter from the software STACKS v2.54 (Rochette et al., 2019) was used to filter PCR clones based on identical UMIs. Following clone filtering, UMIs were removed using cutadapt v1.18 (Martin, 2011) and reads were run through process radtags in STACKS. Reads were then aligned to the jackdaw reference genome (GenBank accession JABDSK000000000; Weissensteiner et al., 2020) using GSNAP/GMAP v2020-10-27 (Wu & Nacu, 2010) specifying a maximum of ten mismatches (-m 10), an indel penalty (-i 2) and turning off terminal alignments (–min-coverage = 0.95). Only reads that aligned uniquely were retained (-n 1, --quiet-if-excessive) (Paris et al., 2017).

Following alignment, we retained samples with >1 million aligned reads. These were run through the ref_map module of STACKS, specifying a minor allele frequency of 0.15 (--min-maf 0.15), retaining one random SNP per read (--write-random-snp) and retaining only SNPs present in >50% of samples (-r 0.5). Finally, we used PLINK v1.9 (Purcell et al., 2007) to discard loci in linkage disequilibrium. To do this, we considered all individuals without sampled parents as ‘founders’ for Site Y + Site Z, two sites with substantial overlap and assumed gene flow, and Site X, a separate site with little to no observed overlap with Site Y + Site Z. We tested for linkage disequilibrium using --indep, evaluating 50 SNP windows, five SNPs at a time with a variance inflation factor cutoff of 2 (Levine et al., 2019), and filtered out SNPs in linkage disequilbrium from our final datasets.

Pedigree reconstruction We used Sequoia v2.1.2 (Huisman, 2017) in R v4.0.2 (R Core Team, 2017) for pedigree reconstruction. Sequoia is an R package specifically designed for reconstructing multigenerational pedigrees using SNP data. Unlike traditional pedigree reconstruction software such as Colony (Jones & Wang, 2010), Sequoia does not require a candidate group of parents for each cohort, which would be difficult to accomplish with our data given multiple overlapping generations. Sequoia also explores a wider range of relationships than many other pedigree reconstruction software and runs quickly relative to alternatives (Flanagan & Jones, 2019).

Before pedigree reconstruction, SNPs and individuals were filtered to be scored for at least 30% of individuals or SNPs, respectively. Stricter filtering is typically recommended (Jisca Huisman, *personal communication*), but our SNP data was characterised by a high degree of missingness. Initial exploration showed that this level of filtering minimised the number of obvious assignment errors while maximising the number of individuals included.

To reconstruct the pedigree, we first estimated the association between age difference and relationship type with function *MakeAgePrior*(), using known and estimated birth years and the social pedigree. We then used the function *sequoia*(), specifying *Tassign* = 2 for conservative assignment of relatives, *Tfilter* = -4 to prevent filtering out of true relatives, *Complex* = “simp” to not explicitly consider double relationships, and *UseAge* = Extra. The genotyping error rate was set to 0.09, based on an estimation from four duplicated samples (see Results).

Following pedigree reconstruction and to further resolve relationships, we used *CalcPairLL*() with a flat age prior to explore relationship log likelihoods between all pairs of individuals (see Supplementary Materials for an extended explanation). From *CalcPairLL*(), the Log Likelihood Ratio (LLR) of full-sibship versus unrelated (FS/U), half-sibship versus unrelated (HS/U) and parent-offspring versus unrelated (PO/U) was calculated. LLR(FS/U) and LLR(HS/U) for assigned full-siblings and assigned unrelated pairs were plotted against values for known social half-siblings (where a social parent is shared through a re-pairing event) to allow evaluation of the likelihood that individuals unassigned to their social sibship were the product of extra-pair paternity or maternity (half-siblings with their social sibship), or egg-dumping (unrelated to their social sibship) (Jisca Huisman*, personal communication*; see Supplementary Materials).

Behavioural data ***Data extraction*** We fitted internal cameras with microphones inside nest-boxes during the 2014, 2015, 2018 and 2019 breeding seasons. Cameras were set to record starting from between 6 and 11 (n = 575 recordings), with a small subset set to record in the afternoon (n = 8 recordings; 13:30 - 14:30). Videos were recorded at three different stages of the breeding season: the nest building stage, the incubation stage, and the nestling stage, with one to three videos per pair per stage (total number of pairs = 130). The nest-building stage was filmed when building had begun and the floor of the nest-box with lined with material (Hahn et al., 2020); the incubation stage was filmed 2 – 10 days post the female’s fertility window, and the nestling stage was filmed at 4 – 10 (nestling 1) and 17 – 23 (nestling 2) days following the first egg hatching. Note that videos were filmed for various research purposes, and in-line with these protocols, no videos were filmed during the female’s fertile period (5 days pre-clutch initiation to the penultimate lay date; (Birkhead & Møller, 1993; Gill et al., 2020)).

Videos were coded by undergraduate and Master’s students, each of whom underwent a training period. Using a standardised ethogram with capacity for descriptive notes, video code was recorded using either Microsoft Excel (2014 - 2015) or the Behavioral Observation Research Interactive Software, BORIS (Friard & Gamba, 2016). The collated coded data, including notes, was searched for key terms that could potentially implicate an extra-pair intrusion event: “intr*”, “attack*”, “copul*”, “fight”, “aggress*”, “defen*”. All videos that contained any of these key terms in their coded data were re-watched by RH, to verify whether an extra-pair copulation had occurred. The confirmation of an extra-pair copulation event was based on the identity of the intruding male, or, if the individual was unidentifiable, on the pattern of behaviour observed relative to known extra-pair copulation events (see Supplementary Materials and Results). All extra-pair copulations (and associated male IDs) identified by RH were independently checked and confirmed by GM.

For fine-sale behavioural analyses, used to test specific predictions of each hypothesis presented in Table 1, all videos in the incubation stage that contained extra-pair copulation, and a subset of incubation-stage videos that did not contain extra-pair copulation, were re-coded by RH using a detailed behavioural ethogram (Table S3).

Statistical analysis We used GLMMs to investigate (i) differences in fine-scale behaviours between within-pair and extra-pair copulations, and (ii) fine-scale behavioural patterns of jackdaws in relation to FEPC, in order to test predictions of the hypotheses presented in Table 1. Statistical analyses were conducted using glmmTMB v1.0.1 (Brooks et al., 2017) in R v4.0.2 (R Core Team, 2017). For fine-scale behavioural models, the response variable was the number of seconds for which the behaviour occurred. For models comparing copulation length and copulation visit length between FEPC and within-pair copulation, no offset was necessary; for all other behavioural models an offset of either video duration (seconds) was used or, for models exploring time males spent with the females, the sum of seconds the female spent in the nest-box. The most appropriate error structure was selected based on model diagnostic plots (using DHARMa) and lowest AIC. Nbinom2 (negative binomial with a quadratic parameterisation) was selected for all models in Table 2 and Table 3 (behavioural models), except Table 3 model 2 which used nbinom1 (negative binomial with a linear parameterisation). For non-behavioural correlates of FEPC, logistic regressions with a binomial error structure and a probit link function were used. Pair ID was included as a random effect in all models, except for models comparing within-pair copulation and FEPC copulation length and copulation visit length, where male ID was the random effect. Models were validated by testing for normality of residuals, overdispersion and zero inflation using DHARMa (Hartig, 2020). Where zero inflation occurred, a zero inflation term was included in the model (Table 2, model 2; Table 3, model 1). Influential points were identified as those that that were more than four times the mean Cook’s Distance (Cook, 1979). Where full models and those with influential points removed are qualitatively identical, the latter are presented. Where a key result differs between models (n = 3 models: Table 3, models 1, 2 and 3), both are reported in the main text while the model without influential points is reported in the associated table, with additional detail in the Supplementary Materials. The model comparing within-pair copulation and FEPC visit lengths did not perform well with the inclusion of extreme values (n = 4, see Supplementary Material), so these were removed for this model (results remained qualitatively identical). The model investigating the relationship between female structural size and FEPC did not perform well with influential points excluded so the full model is presented, again with further detail in the Supplementary Materials. Note that sample size (n videos and n pairs) per model varies because (i) some information (such as tarsus size) is missing for some individuals, and covariates per model vary (Tables 2-4; Supplementary Material) and (ii) a varying number of influential points (min = 0 pairs, max = 5 pairs) were removed per model.

FEPCs beyond jackdaws To understand the prevalence of FEPCs in birds lacking penises, we collated species reported to engage in FEPC from previous reviews (Gladstone, 1979; Gowaty & Buschhaus, 1998; McKinney et al., 1984; McKinney & Evarts, 1998), and used Google Scholar to search for additional reports of FEPC, using the terms “bird” + “forced copulation” or “rape” or “FEPC” or “force” + “copulation”. We excluded species with penises, and included only species where active female resistance to attempted male copulation has been reported. For example, species where extra-pair males are known to chase females but attempted copulation has not actually been observed (e.g. tree swallows *Tachycineta bicolor*; Venier, Dunn, Lifjeld, & Robertson, 1993) were not included. If species were included in previous review papers but the primary literature did not describe female resistance to male copulation attempts, these species were not included. If the primary source describing FEPC could not be accessed (e.g. in cases of personal communication between review authors and researchers), then these species were included. Given that extra-pair behaviour can differ drastically between captive and wild populations (e.g. Griffith et al., 2010) we only included species where FEPC has been recorded in the wild. In order to identify broad-scale patterns in FEPC occurrence, we also collected information on species’ mating system, whether they are colonial breeders, the identity of the male engaging in FEPC (e.g. if he was known to be a neighbour), and whether unforced extra-pair copulation has also been observed.

Results

Pedigree analysis ***Bioinformatic pipeline*** Sequencing of ddRAD libraries resulted in 1161498224 raw reads across 190 samples (two samples failed to sequence), with 6113148.5 ± 3507048 reads per sample. Following adapter trimming and quality filtering, 87.28 ± 6.93% of reads per sample were retained, and following clone filtering, 82.56 ± 5.22% of filtered reads per sample were retained. Alignment in GMAP/GSNAP resulted in an average alignment of 66.33 ± 8.48% reads (2888614 ± 1647162). 26 samples had fewer than 1 million aligned reads and were not taken forward in the analytical pipeline. Following SNP discovery and filtering out of SNPs in linkage disequilibrium, 901 SNPs were retained for Site YZ and 770 SNPs were retained for Site X.

Sequencing error rate Before filtering out SNPs in linkage disequilibrium, we calculated approximate sequencing error rate using our five duplicate samples. One duplicate sample had an anomalously high sequencing error of 21.9%, and this duplicate was discarded from the analysis. The other four samples showed an error rate of 7.19%, 8.32%, 9.49% and 10.34%, giving an approximate error rate of ~9%.

Pedigree reconstruction *Site YZ:* 901 SNPs and 117 individuals were loaded into Sequoia. Following 30% missingness filters, 757 SNPs and 93 individuals were retained (17 parents, 76 offspring, 44 broods, 19 sibships). Pedigree reconstruction with *sequoia*() and inspection of pairwise relationships (calculated from *CalcPairLL*(); see Supplementary Materials), both of which were blind to the social pedigree, found that 67 of the 76 offspring clustered into their known social sibships and were the offspring of their social parent(s), where their social parent(s) was retained in the analysis.

The nine offspring who did not have convincing support for full sibship with their social sibships were investigated in detail, and a full discussion of each case is included in the Supplementary Materials. In brief, there were two likely cases of egg-dumping (where non-resident females lay eggs in a nest-box, and the resident male is not the father), two ambiguous cases of egg-dumping *or* half-sibship within a social sibship, and five likely cases of half-sibship within a social sibship. Half-sibship within a social sibship can be the result of either extra-pair maternity, where non-resident females lay an egg sired by the resident male in his nest-box, or extra-pair paternity, where resident females lay an egg sired by a non-resident male in the nest-box. Of the five likely cases of half-sibship we identified, two are highly likely to be the result of extra-pair maternity through an unusual 'follower' relationship, where an extra-pair female was often observed with a nest-box-owning pair, and was recorded on video laying an egg in their nest-box. This extra-pair female is likely to be the mother of two individuals from this nest-box, across two different years (2017 and 2019), with the resident male likely to be the father. This same male was also likely to have sired another chick in his nest-box whose mother was neither the resident female nor the known follower female, indicating that the resident male sired offspring with a total of three females, all while maintaining his pair-bond with one partner. One further case of likely half-sibship was identified to most likely be the result of extra-pair paternity, although we could not rule out extra-pair maternity, while the remaining case was unresolved (i.e. either extra-pair maternity or paternity). Of the two ambiguous cases of egg-dumping or half-sibship, one was most likely the result of either egg-dumping or extra-pair maternity, while the final case was unresolved (i.e. could be the result of egg-dumping, extra-pair maternity or extra-pair paternity).

Fertilisation as a result of FEPC would most likely be reflected as extra-pair paternity within a social sibship. Extra-pair maternity as a result of fertilisation through FEPC is considered highly unlikely, given that females would need to (i) know the box and (ii) take the risk of entering the box of the male who attacked her and fertilised her egg in order to lay that egg, and it is not immediately obvious why all this would occur. Furthermore, across eight years of field observations, we have not observed a nest-box-owning, breeding female enter another pair's nest-box around the date of laying. In contrast, we have observed a follower female laying her and the resident male's eggs inside his nest-box, and we therefore suggest that cases of extra-pair maternity in this population are most likely due to unobserved cases of follower females. Therefore, to calculate the potential rate of fertilisation through FEPC, we considered only extra-pair paternity. Including unlikely but potential cases of extra-pair paternity, this population has a maximal potential rate of fertilisation through FEPC of 3/76 (3.95%).

Site X: 770 SNPs and 38 individuals were loaded in Sequoia. Following 30% filters, we retained all SNPs and 37 individuals (9 parents, 28 offspring, 13 broods, six sibships). We identified no cases of egg-dumping and one potential case of half-sibship within a social sibship, which was most likely the result of extra-pair maternity, although could potentially be a case of extra-pair paternity instead (see Supplementary Materials). Thus, the maximal potential rate of extra-pair paternity through FEPC is considered to be 1/28 (3.57%).

Behavioural data analysis Of 1786.37 hours of nest-box video filmed and coded across 575 videos and 130 pairs, 87 within-pair copulations were observed. 40.22% (n within-pair copulation = 35) of these were in the nest-building (pre-fertile) stage (301.84 hours of video), 59.77% (n within-pair copulation = 52) occurred during the incubation (post-fertile) stage (708.51 video hours), and none occurred in the nestling-rearing (post-fertile) stage (776.01 video hours). In total we recorded 17 extra-pair copulation events (n = 5 events at Site X, n = 8 events at Site Y, n = 4 events at Site Z). Of these, one extra-pair copulation event was in the nest-building stage and 16 were in the incubation stage, observed across 14 videos (two videos contained two extra-pair copulation events). No videos in the nestling-rearing stage contained extra-pair copulation.

For the fine-scale behavioural analysis of videos in the incubation stage, all videos containing extra-pair copulation (n = 14 videos) and 87 randomly selected videos were re-coded by the lead author using the fine-scale behavioural ethogram (Table S3), equating to 290.28 hours of recording and 63 jackdaw pairs.

Intruding male ID In 13 of the 17 extra-pair copulation events, we could establish with certainty that the intruding male was not the resident male (see Supplementary Material). In six extra-pair copulation events, the intruder was identifiable due to a full set of visible rings, and in one case the identity was ambiguous (two possible identifications) due to one unclear ring. In five cases of confident identification, the intruder was a paired adult male who owned a neighbouring nest-box and had a partner either in the process of laying or incubating eggs (in two of these instances, the intruder was the same male forcing copulation with different neighbouring females). In the sixth case of confident identification, the male was known to be an adult but we had no further information about him. In the case of ambiguous identification, the male was either a paired adult neighbour or an adult male for whom we had no further information. Four of the five males who we could confidently identify had bred in nest-boxes the year preceding the extra-pair copulation. Two of these males had failed to fledge any offspring despite eggs being laid, one had built a nest with his partner but she did not lay eggs, and the final fledged two offspring (for comparison, the average reproductive success in that year was 1.85 ± 0.88 chicks per pair).

FEPC and within-pair copulation comparison 100% of extra-pair copulation events co-occurred with female nest-box defence and/or prolonged female aggression (pecking/kicking the head and body of the intruder, or turning around and fighting with claws) that occurred throughout the copulation attempt (Supplementary Video 1). We did not observe any instance of female copulation solicitation (bending the tail upward and to the side, often accompanied by horizontal tail shaking; Gill et al., 2020); no extra-pair copulations appeared to be completed by the intruder (wing flapping of less than two seconds indicated an aborted attempt; Gill et al., 2020), and none were associated with affiliative behaviours (allopreening, contact sitting/standing, see Table S3). In two cases, the resident male returned during the extra-pair copulation event; in one case the intruder immediately left the nest-box and in the other a fight ensued between the resident male and the intruder for ~22s until the intruder left (Supplementary Video 2).

The fine-scale behavioural dataset, which included videos only from the incubation stage of the breeding season, included 14 videos containing 16 extra-pair copulation events and 24 videos containing 32 within-pair copulations (across 29 male visits). In contrast to extra-pair copulation, no within-pair copulation events were preceded by nest-box defence, and 100% of within-pair copulations were associated with male affiliation toward the female prior to copulation. However, no copulations were solicited by the female. Only 28.13% of within-pair copulations appeared to be completed, and 46.67% were associated with female aggression (turning to her partner with an open beak or pecking at her partner's beak/head). In contrast to extra-pair copulations, female aggression during within-pair copulations was brief and caused the male to cease his copulation attempt. While the length of copulation was not significantly different between within-pair copulations and FEPCs (β = -1.38, SE = 2.51, χ^2 = 0.30, 95% CI [-6.29,3.54], P = 0.58), the length of the visit in which the copulation occurred was significantly longer in within-pair copulations than FEPCs (β = 1.50, SE = 0.20, χ^2 = 60.25, 95% CI [1.12,1.87], P = <0.0001; Figure 1a). See Supplementary Video 3 for an example of within-pair copulation.

Considering the female's nest-box defence, prolonged attack behaviours, and the lack of the intruder's pre-copulatory routines (specifically, affiliation towards the female), it appears that all identified extra-pair copulation events were indeed forced (FEPC) rather than cooperative.

Resident male behaviour and FEPC (Table 2) All FEPC events occurred when the resident male was absent from the nest-box. Females were more likely to experience FEPC when their partners spent less time with them at the nest-box in the period preceding the FEPC (measured as time the pair spent together controlled for overall time the female spent at the nest-box; β = -0.92, SE = 0.31, χ^2 = 8.81, 95% CI [-1.53,-0.31], P = 0.003; Figure 1b), relative to the behaviour of males whose partners did not experience FEPC. Females were also more likely to experience FEPC when their partners spent less time in vigilance at the nest-box (β = -1.11, SE = 0.50, χ^2 = 4.95, 95% CI [-2.09,-0.13], P = 0.03) in the period preceding the FEPC, relative to the behaviour of males whose partners did not experience FEPC.

Resident female behaviour and FEPC (Table 3) With influential points removed, females spent significantly more time in vigilance in videos where FEPC occurred relative to videos without FEPC (β = 0.73, SE = 0.36, χ^2 = 4.03, 95% CI [0.02,1.45], P = 0.045, Figure 1c). Females spent significantly more time in vigilance before the FEPC than after it (β = -0.61, SE = 0.25, χ^2 = 6.99, 95% CI [-1.10,-0.12], P = 0.01), suggesting that external cues alerted them to the possibility of danger, and that their increased vigilance did not protect them from being attacked. Females spent significantly less time incubating in videos with FEPC (β = -0.07, SE = 0.03, χ^2 = 4.61, 95% CI [-0.14,-0.01], P = 0.03).

Table 2 Resident male behaviour and FEPC. The response variable was number of seconds engaged in the behaviour, with an offset of video length for model 1, and female time in the nest-box in model 2. Pair ID was included as a random effect in both models, and both used an nbinom2 error structure. A zero inflation term was included in model 2. Both models have influential points removed; key results (FEPC in video; days since female fertile) were qualitatively identical with their inclusion. Bold indicates significant results.

Model n	Response	n videos	n pairs	Fixed effects	β	SE	χ^2	95% CI (lower)	95% CI (upper)	P-value
1	Time resident male is vigilant	73	41	**FEPC in video (yes)**	**-1.109**	**0.498**	**4.951**	**-2.085**	**-0.132**	**0.026**
				Video starting time	**-0.407**	**0.098**	**17.141**	**-0.599**	**-0.214**	**0.00003**
				Male tarsus	**-0.231**	**0.112**	**4.226**	**-0.451**	**-0.011**	**0.040**
				Female tarsus	**0.277**	**0.100**	**7.643**	**0.081**	**0.473**	**0.006**
				Male's minimum age	**0.324**	**0.109**	**8.866**	**0.111**	**0.537**	**0.003**
				Year(15)	**-0.319**	**0.844**	**11.953**	**-1.973**	**1.335**	**0.008**
				Year(18)	**0.311**	**0.913**		**-1.478**	**2.100**	
				Year(19)	**-1.148**	**1.011**		**-3.131**	**0.834**	
				Days since female fertile	-0.052	0.091	0.334	-0.230	0.125	0.563
2	Time resident male is with female	69	40	**FEPC in video (yes)**	**-0.918**	**0.309**	**8.811**	**-1.525**	**-0.312**	**0.003**
				Video starting time	**-0.116**	**0.057**	**4.065**	**-0.228**	**-0.003**	**0.044**
				Male tarsus	**-0.252**	**0.070**	**12.939**	**-0.389**	**-0.115**	**0.0003**
				Female tarsus	**0.220**	**0.069**	**10.129**	**0.085**	**0.356**	**0.001**
				Male's minimum age	**0.181**	**0.064**	**8.036**	**0.056**	**0.306**	**0.005**
				Year(15)	**0.842**	**0.536**	**14.596**	**-0.209**	**1.893**	**0.002**
				Year(18)	**1.065**	**0.556**		**-0.026**	**2.155**	
				Year(19)	**0.094**	**0.617**		**-1.117**	**1.304**	
				Days since female fertile	-0.052	0.057	0.829	-0.163	0.060	0.363

Without the removal of influential points (n = 4, n = 1 and n = 4 pairs for each model respectively; see Supplementary Materials for further details), there was no significant difference in either vigilance model (no FEPC versus FEPC: β = 0.49, SE = 0.37, χ^2 = 1.74, 95% CI [-0.24,1.21], P = 0.19; pre-FEPC versus post-FEPC: β = -0.35, SE = 0.28, χ^2 = 1.52, 95% CI [-0.90,0.20], P = 0.22) or incubation (β = -0.01, SE = 0.04, χ^2 = 0.02, 95% CI [-0.09,0.07], P = 0.87).

Non-behavioural predictors of FEPC The structural size of the female and her partner were unrelated to the female's risk of experiencing FEPC (female tarsus size: β = -0.18, SE = 0.12, χ^2 = 2.09, 95% CI [-0.15, 0.06], P = 0.15; male tarsus size: β = -0.05, SE = 0.21, χ^2 = 0.05, 95% CI [-0.45, 0.36], P = 0.83). However, the closer to her fertile window the female was, the more likely she was to experience FEPC (β = -0.45, SE = 0.21, χ^2 = 4.56, 95% CI [-0.53, -0.04], P = 0.03, Figure 1d). It is possible that the greater risk of FEPC closer to the fertile

Table 3 Resident female behaviour and FEPC. The response variable was number of seconds engaged in the behaviour, with an offset of video length. Pair ID was included as a random effect in model 1 and 3, while video ID was used as a random effect in model 2. Model 1 and 3 used an nbinom2 error structure; model 2 used an nbinom1 error structure. A zero inflation term was included in model 1. All models have influential points removed. Key results differed with their inclusion; see Results and Supplementary Material. Bold indicates significant results.

Model n	**Response**	**n videos**	**n pairs**	**Fixed effects**	**β**	**SE**	**χ^2**	**95% CI (lower)**	**95% CI (upper)**	**P-value**
1	Total time resident female is vigilant	73	41	**FEPC in video (yes)**	**0.732**	**0.365**	**4.028**	**0.017**	**1.448**	**0.045**
				Video starting time	0.028	0.075	0.144	-0.118	0.175	0.704
				Female tarsus	**0.212**	**0.092**	**5.308**	**0.032**	**0.393**	**0.021**
				Male tarsus	-0.071	0.101	0.493	-0.268	0.127	0.482
				Female's minimum age	-0.048	0.098	0.235	-0.240	0.145	0.628
				Year(15) Year(18) Year(19)	0.606 0.106 0.125	0.928 0.984 1.065	2.339	-1.213 -1.823 -1.963	2.424 2.035 2.213	0.505
				Days since female fertile	0.036	0.068	0.277	-0.097	0.168	0.598
2	Time pre-FEPC/post-FEPC resident female is vigilant	24	12	**Pre/post-FEPC (post)**	**-0.609**	**0.250**	**5.937**	**-0.119**	**-0.609**	**0.015**
				Video starting time	**1.971**	**0.746**	**6.986**	**3.433**	**1.971**	**0.008**
				Female's minimum age	**1.022**	**0.501**	**6.930**	**3.656**	**2.036**	**0.041**
				Year(15) **Year(18)**	**2.036** **0.027**	**0.827** **1.208**	**1.254**	**2.394** **0.140**	**0.027** **-0.187**	**0.031**
				Days since female fertile	-0.187	0.167	**4.157**	2.005	1.022	0.263
3	Total time resident female incubates	69	41	**FEPC in video (yes)**	**-0.073**	**0.034**	**4.612**	**-0.139**	**-0.006**	**0.032**
				Video starting time	-0.007	0.007	0.920	-0.020	0.007	0.338
				Female tarsus	-0.002	0.006	0.135	-0.015	0.010	0.713
				Male tarsus	**-0.019**	**0.007**	**7.378**	**-0.033**	**-0.005**	**0.007**

Female's minimum age	-0.004	0.008	0.211	-0.019	0.012	0.646
Year(15)	**-0.149**	**0.080**	**8.530**	**-0.306**	**0.009**	**0.036**
Year(18)	**-0.083**	**0.085**		**-0.250**	**0.083**	
Year(19)	**-0.067**	**0.090**		**-0.244**	**0.110**	
Male-to-female food-sharing	-0.001	0.005	0.044	-0.011	0.009	0.833
Days since female fertile	-0.005	0.006	0.555	-0.017	0.008	0.456

window could result from changes in resident male behaviour; for instance if males left their partner alone for longer periods. However, we found no evidence that males changed either vigilance (β = -0.05, SE = 0.09, χ^2 = 0.33, 95% CI [-0.23, 0.13], P = 0.56; Table 2, model 1) or time with the female (β = -0.05, SE = 0.06, χ^2 = 0.83, 95% CI [-0.16, 0.06], P = 0.36; Table 2, model 2) relative to number of days post-fertile.

FEPCs beyond jackdaws We identified 48 bird species lacking penises where FEPC has been reported, nine of which were not included in previous reviews (see Supplementary Data). These species spanned 23 families and 11 orders, out of a total of 22 bird orders that lack penises. Laridae (gulls), Ardeidae herons) and Corvidae (corvids) had the largest number of species recorded to engage in FEPC (n = 9, 7, 7 respectively). Of all included species, 98% of those with known mating systems (n species = 47) are always or primarily socially monogamous (n species = 46), while 75% (n species = 36) are always or sometimes colonial breeders. In 20 of these 28 species, the male was recorded in at least one instance to be a neighbour of the target female. In 19 of these species, the male was recorded in at least one instance to be a paired individual, while in four species the male was recorded in at least one instance to be unpaired. Just under half of the species recorded (48%, n species = 23) engage in unforced extra-pair copulation (i.e., where the female does not resist *or* actively seeks extra-pair copulation with non-partner males) as well as FEPC.

No studies have directly investigated whether FEPC can lead to fertilisation in species lacking penises, but there is strong evidence to suggest FEPC can lead to fertilisation in the stitchbird (or hihi; *Notiomystis cincta*) (Brekke et al., 2013; Ewen et al., 1999), and ambiguous evidence in American crows (*Corvus* *brachyrhynchos*) (Townsend, 2009). In contrast, the limited evidence available indicates that FEPC does not (or rarely) results in fertilisation in jackdaws (Gill et al., 2020), western gulls (*Larus occidentalis*) (Gilbert, Burke, & Krupa, 1998) and black-legged kittiwakes (*Rissa tridactyla*) (Helfenstein et al., 2004).

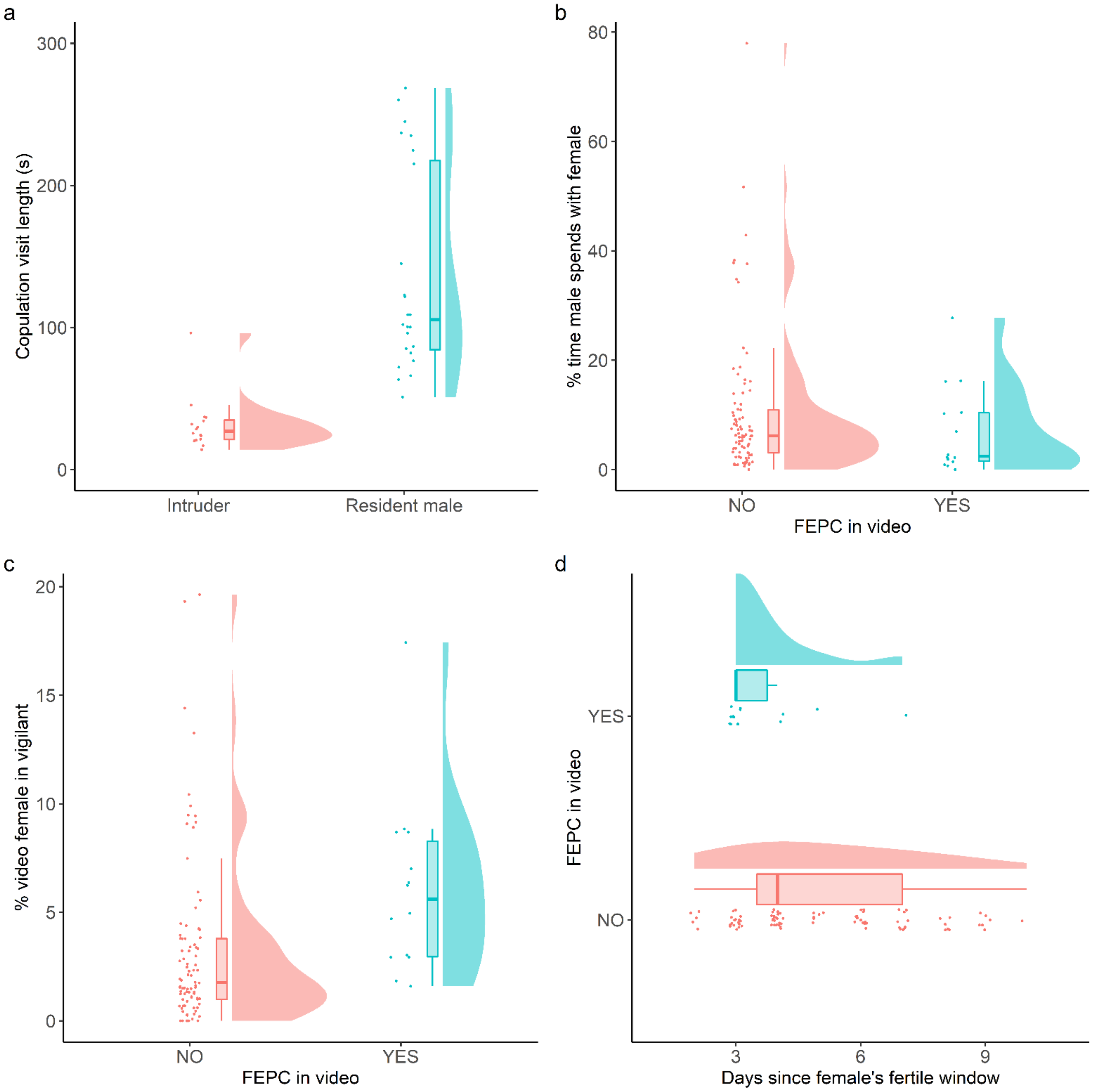


Figure 1a: length of visits (in seconds) in which copulation occurred for intruders (FEPC) and resident males (within-pair copulation). Note that copulation visits of >300s (all resident males) have been removed for visualisation purposes; **b**: percentage of time the resident male spends with the female, out of total female time in the nest-box, for videos with and without FEPC; **c**: percentage of video in which the female engages in vigilance across videos with and without FEPC; **d**: days since the female's fertile window and whether FEPC occurs.

Discussion

If species lacking a penis cannot achieve fertilisation through FEPC, then why forced copulation occurs in these species is puzzling. Generally, we do not know whether FEPC leads to paternity, but in jackdaws it seems that it does not, or rarely, results in siring success. Hence, explanations unrelated to siring success are needed to explain this behaviour. We now discuss our findings in jackdaws, the general pattern across penis-lacking birds, and identify knowledge gaps before recommending future research directions.

Despite multiple detailed observational studies of wild jackdaw populations (Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Röell, 1978; Turjeman et al., 2021), females have never been seen to engage in cooperative extra-pair copulation. Thus, the rate of extra-pair paternity is likely to reflect the siring success of FEPC. We found that paternity through FEPC was 0-4%, indicating that FEPC either never or rarely results in paternity. Paternity studies of multiple wild jackdaw populations have revealed similarly low levels of extra-pair paternity (Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021), even in populations where 82% of females experienced at least one FEPC event (Gill et al., 2020). While some previous studies did not observe behaviours within the nest-box, several authors also report behaviours in-line with FEPC (i.e., non-resident birds entering a nest-box with a lone female inside: Chen, R., *personal communication*; Henderson et al. 2000; Liebers & Peter, 1998). Together, these results indicate that while FEPC is not unusual in jackdaws, achieving siring success through FEPC is rare. This suggests that forced copulation may not be maintained via paternity benefits in jackdaws, which is at odds with the vast majority of species (Gross, 1996).

Given that FEPC provides a low probability of paternity, why do males engage in this behaviour? This is especially perplexing considering the potential costs of FEPC. While costs are inherently difficult to quantify in a wild and long-lived species, FEPC costs to males appear to be non-trivial in jackdaws. Specifically, both we and Gill et al. (2020) observed fights between females and the male attempting FEPC. Moreover, if the resident male returned during the FEPC, escalated fighting occurred. Such conflict is likely to result in injury given that pecking and kicking occurred for prolonged periods in a confined space, and this must also involve energy expenditure. In addition to these apparent costs, because males engaging in FEPC only ever entered a nest-box when the resident male was absent, they must have invested time monitoring the victim's nest-box (also see Gill et al. 2020). Furthermore, given that all but one of the identifiable males were breeding neighbours whose partner was laying or incubating eggs, these males traded time foraging for themselves and their partner (males food-share with females while they incubate: Röell, 1978) for time engaging in FEPC. Moreover, when seeking FEPC, males left their mates alone at the nest-box, thus making them vulnerable to FEPC (and the associated injury, stress and energy expenditure) while males were away. Taken together, it seems likely males incur a cost as a result of FEPC attempts.

Of the proposed hypotheses (Table 1) explaining why FEPC occurs, the 'relic' hypothesis is most supported. This posits that FEPC originally delivered paternity gain, and persists as a vestigial behaviour even though it is no longer advantageous. Specifically, we found that FEPC is more likely to occur closer to the female's fertile period, independent of the resident male's behaviour. This suggests that intruders try to access females when they are fertile, implying that the behaviour has evolved for paternity enhancement. Indeed, resident males increase their mate-guarding efforts when their partner is fertile (Gill et al., 2020), which suggests cuckoldry avoidance has evolved in males (Stockley, 1997). Furthermore, three of the four identifiable males for whom we had breeding information in the year previous to their FEPC attempt, had failed breeding attempts in that year. This suggests that males may once have adopted a "best of a bad job" strategy, where they tried to increase reproductive success by pursuing fertilisation outside of the pair-bond. Such strategies are commonly observed in species where forced copulation leads to fertilisation (Gross, 1996). However, it is also possible that we are still seeing a "best of a bad job" tactic, because we cannot rule out that FEPC may sometimes, albeit rarely, result in fertilisation. Thus, we cannot totally exclude the possibility that the "immediate fertilisation enhancement" hypothesis explains the persistence of FEPC in jackdaws, although we consider it unlikely. If the relic hypothesis explains the persistence of FEPC, a question remains: if the benefit of FEPC is low, yet it is a costly behaviour, why does it persist? One explanation for why FEPC persists is that it is selectively neutral, i.e. the benefits are low but the costs are also low, and it is therefore not selected against. Given the costs outlined above, we suggest this is unlikely. However, the field would benefit immensely from explicit quantification of putative costs.

We did not find strong support for any other of the hypotheses. The CODE hypothesis suggests that FEPC is adopted by males to create an ‘environment of fear’ for females, who evolve the counter-strategy of pairing with a male partner for protection, and in return offer access to copulation and fertilisation (Gowaty & Buschhaus, 1998). While we found support for the prediction that males direct FEPC attempts towards unguarded females (Gowaty & Buschhaus, 1998), because of putative costs of escalated fighting with resident males, it could be argued that targeting unguarded females would be expected regardless of the ultimate causation of FEPC. We did not find support for three other predictions of the CODE hypothesis, namely that FEPC is unrelated to the victim female’s fertility state, that female vulnerability to male aggression predicts where FEPC occurs (Gowaty & Buschhaus, 1998) and that male mate-guarding is not associated with his partner’s fertility (Gill et al., 2020). On balance, we suggest that the CODE hypothesis probably does not explain FEPC in jackdaws. We also found no evidence that males engage in FEPC when their own partner is fertile (in order to fertilise their partner with high quality sperm), a key prediction of the sperm turnover hypothesis. However, this is tentative since three of five males engaged in FEPC when their partner was post-fertile, and two did so when their partner was fertile. Thus, further investigation into this hypothesis is needed. Likewise, key predictions of the by-product hypothesis (FEPC as a pleiotropic behaviour) and the territory signalling hypothesis (males gain territory as a result of FEPC) could not be tested (see Table 1). Further work is also necessary to investigate these ideas.

We found ambiguous support for the reproductive suppression hypothesis. This hypothesis suggests that males target neighbouring females to decrease the reproductive success of direct competitors. Reproductive suppression has previously been demonstrated in birds. For example, in superb lyrebirds (*Menura novaehollandie*) and white-winged choughs (*Corcorax melanorhamphos*), individuals destroy nests to suppress the reproductive success of direct competitors (Austin et al., 2019; Heinsohn, 1988). Conspecific egg destruction may also occur in acorn woodpeckers (*Melanerpes formicivorus*) and Mexican jays (*Aphelocoma ultramarina*) (Mumme et al., 1983;

Trail et al., 1981) for the same reason. In this context, the reproductive suppression hypothesis predicts that (i) females incur costs as a result of FEPC, (ii) this negatively influences their fitness outcomes, and (iii) FEPC is undertaken by neighbouring males. Females may incur injury during conflict with the male (Gill et al., 2020), and given that FEPC occurs both during egg-laying and incubation (our populations; Gill et al., 2020) eggs may also be damaged (Gill et al., 2020). We could not quantify costs of injury to females, but we did find a behavioural cost to FEPC. Specifically, females increased vigilance prior to the FEPC event, probably because they could hear or see an intruding male in close vicinity to the nest-box, and decreased incubation. Interrupted incubation may lead to a sub-optimal adjustment of hatching synchrony to environmental conditions, which increases the likelihood of brood failure (Lack, 1966; Parejo et al., 2015; Wang & Beissinger, 2011). Considering up to 82% of females may experience FEPC (Gill et al. 2020), this apparently short-term incubation disruption could have major implications for reproductive success. However, since we only filmed each nest-box for a few hours and did not know the distribution of FEPC across the breeding season, we could not test whether fitness was impacted by FEPC. This should be the subject of future work. Additionally, our data indicated that FEPC is undertaken by neighbouring males. Of the FEPC males we could identify, four of five were paired, breeding neighbours. Although this lends support to the reproductive suppression hypothesis, it must be noted that males may simply be more likely to engage in extra-pair behaviour with their neighbours due to proximity, and that this pattern is also seen in species where unforced extra-pair copulation results in fertilisation (Mayer & Pasinelli, 2013). Two other predictions of this hypothesis, that FEPC is independent of the female’s fertile period, and that male mate-guarding is not associated with female fertility, were not met (Gill et al., 2020). In sum, this hypothesis deserves further interrogation, but it would be premature to conclude this is occurring in jackdaws at present.

It is also possible that FEPC can result in fertilisation in specific ecological conditions. That is, benefits and costs (to males and females) vary spatially or temporally. For example, in resource-poor years, resident males may mate-guard less due to increased foraging effort (Gill et al., 2020), while females may not have the energy to resist intruding males and simply accept copulations as a means of cost minimization (Westneat & Stewart, 2003). Consistent with this, despite low rates of extra-pair paternity in jackdaws (Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021), rates do vary between populations (Turjeman et al., 2021).

While we tentatively suggest the relic hypothesis is the best supported explanation of FEPC in jackdaws, much remains unknown despite intense study of jackdaw FEPC. FEPC in species lacking penises has generally not been well quantified, potentially because it is considered rare (Adler, 2010; Zduniak et al., 2016). However, we found reports of FEPC in 48 bird species from 11 orders (of a total of 22 orders where species do not have penises). This is likely to be vastly under-representative because in many species copulation does not occur in the open and is therefore difficult to observe (Sheldon, 1994; Westneat & Stewart, 2003), and our criterion for inclusion in the FEPC database was fairly conservative. In any case, FEPC in species lacking penises can be considered taxonomically widespread.

Despite a general lack of fine-scale quantification of FEPC behaviour across species, we were able to elucidate some broad patterns. Most species with FEPC tend to be socially monogamous, and the majority are always or sometimes colonial breeders. This is consistent with previous work (Gowaty & Buschhaus, 1998), however there is a caveat: most studies on extra-pair matings have been conducted on colonially breeding species. Within colonial species, it appears that neighbouring males frequently initiate FEPCs. However, whether this pattern is due to specific targeting of neighbours, or just proximity and ease of access, is not clear. While it might be expected that FEPC would be more common in species where males are larger and can more easily overpower females with lower injury risk, many species with FEPC show little sexual size dimorphism. Indeed, it stands to reason that females would incur a higher cost of resistance if males were much larger, and thus may instead accept extra-pair copulation in order to minimise costs. In just under half the species identified, unforced extra-pair copulation was also recorded. This is also likely to be an underestimate, given that unforced extra-pair copulations are probably less obvious to observers, so may be more easily missed than FEPC. In species with unforced extra-pair copulation, rejected extra-pair copulation attempts could lead to FEPC. Nevertheless, based on current knowledge, it seems that FEPC does also occur in a range of species where unforced extra-pair copulation does not occur (see Supplementary Data).

Within the Corvidae family, species showed similar FEPC behaviours to jackdaws. We found reports of FEPC in six species, in addition to jackdaws: American crows (*Corvus brachyrhynchos)*, carrion crows (*Corvus corrone*), common ravens (*Corvus corax*), northwestern crows (*Corvus* caurinus), rooks (*Corvus frugilegus*) and yellow-billed magpies (*Pica nuttalli*) (Birkhead, 1979; Gill et al., 2020; Kilham, 1984; Kramer, 1932; Roskaft, 1983; Townsend et al., 2009; Verbeek, 1972; Verbeek & Butler, 1999; Wittenberg, 1968). In four of these species, the initiators of FEPC were known in at least some cases to be paired, neighbouring males (Gill et al., 2020; Kramer, 1932; Roskaft, 1983; Verbeek, 1972; Wittenberg, 1968); in all cases where the location of the FEPC was recorded, FEPC always or mostly occurred at the female's nest site (Birkhead, 1979; Gill et al., 2020; Kilham, 1984; Kramer, 1932; Roskaft, 1983; Townsend et al., 2009; Verbeek & Butler, 1999; Wittenberg, 1968), and all FEPC events occurred while the resident male was absent and the female was alone (Birkhead, 1979; Gill et al., 2020; Kilham, 1984; Kramer, 1932; Roskaft, 1983; Townsend et al., 2009; Verbeek, 1972; Verbeek & Butler, 1999; Wittenberg, 1968). Thus, FEPC behaviour appears to be somewhat conserved within these closely related taxa.

While no studies have directly investigated whether FEPC increases siring success in species without a penis, evidence suggests FEPCs are successful in the stitchbird, where rates of FEPC correlate with the rate of extra-pair paternity (Brekke et al., 2013; Ewen et al., 1999). Despite lacking a penis, this species is unusual in that the male cloacae rotates and swells to 400% of their original size during the breeding season, and a unique face-to-face copulatory position is adopted during FEPC (Low et al., 2005). Together, these behavioural and physiological adaptations are thought to aid male stitchbird fertilisation success during FEPC (Brekke et al., 2013; Low et al., 2005). There is also some evidence to suggest FEPC increases siring success in American crows, where the rate of observed FEPC correlates with the rate of extra-pair paternity (Townsend, 2009). However, this is uncertain: the authors note that unobserved unforced extra-pair copulation may have led to extra-pair fertilisation instead (Townsend et al., 2009). Meanwhile, data from western gulls (*Larus* *occidentalis*) (Gilbert et al., 1998), black-legged kittiwakes (*Rissa tridactyla*) (Helfenstein et al., 2004) and jackdaws (here; Gill et al., 2020) suggest that FEPC does not, or rarely, results in siring success. The fact that FEPC may result in fertilisation in some species lacking penises, while it appears not to in others, again highlights the need for in-depth study of this behaviour.

Here, we investigated forced copulations in jackdaws to evaluate seven hypotheses as to why it occurs (Table 1). We find the rate of extra-pair paternity through FEPC to be between 0 and 4%. This indicates either no or a low rate of siring success through FEPC. Of the hypotheses presented, we tentatively find the most support for the relic hypothesis, which suggests that FEPC persists because at one time it led to siring success. Given that FEPC is likely to be costly, this suggests that the male must receive some benefit of engaging in FEPC. While we do not find any convincing support for the alternative hypotheses, and thus cannot identify what these benefits may be, we urge further study given that some key predictions could not be tested. Moreover, while FEPC is likely to be costly, this is uncertain and must be tested in future studies. Finally, we find that FEPC is widespread across birds lacking penises, and that the ecology of the behaviour appears to be highly conserved within at least one closely related group, the Corvidae family. However, very little is known about FEPC in most species, and almost nothing about why it has evolved and persisted. Understanding the ultimate function of FEPC is therefore an important component of wider research into sexual behaviour. We suggest that the framework presented here, where specific predictions of multiple hypotheses are tested using both genetic and behavioural data, will help us understand the ultimate underpinnings of this behaviour.

Supplementary Material

Library construction

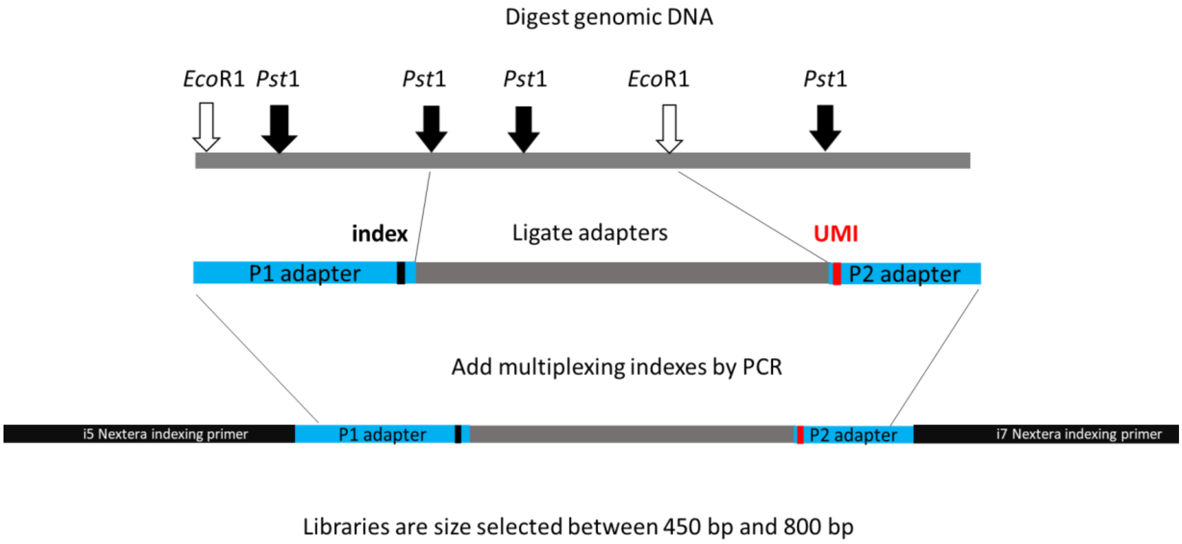


Figure S1 Schematic of *Pst*1 *EcoR*1 ddRADseq library generation

Pedigree reconstruction

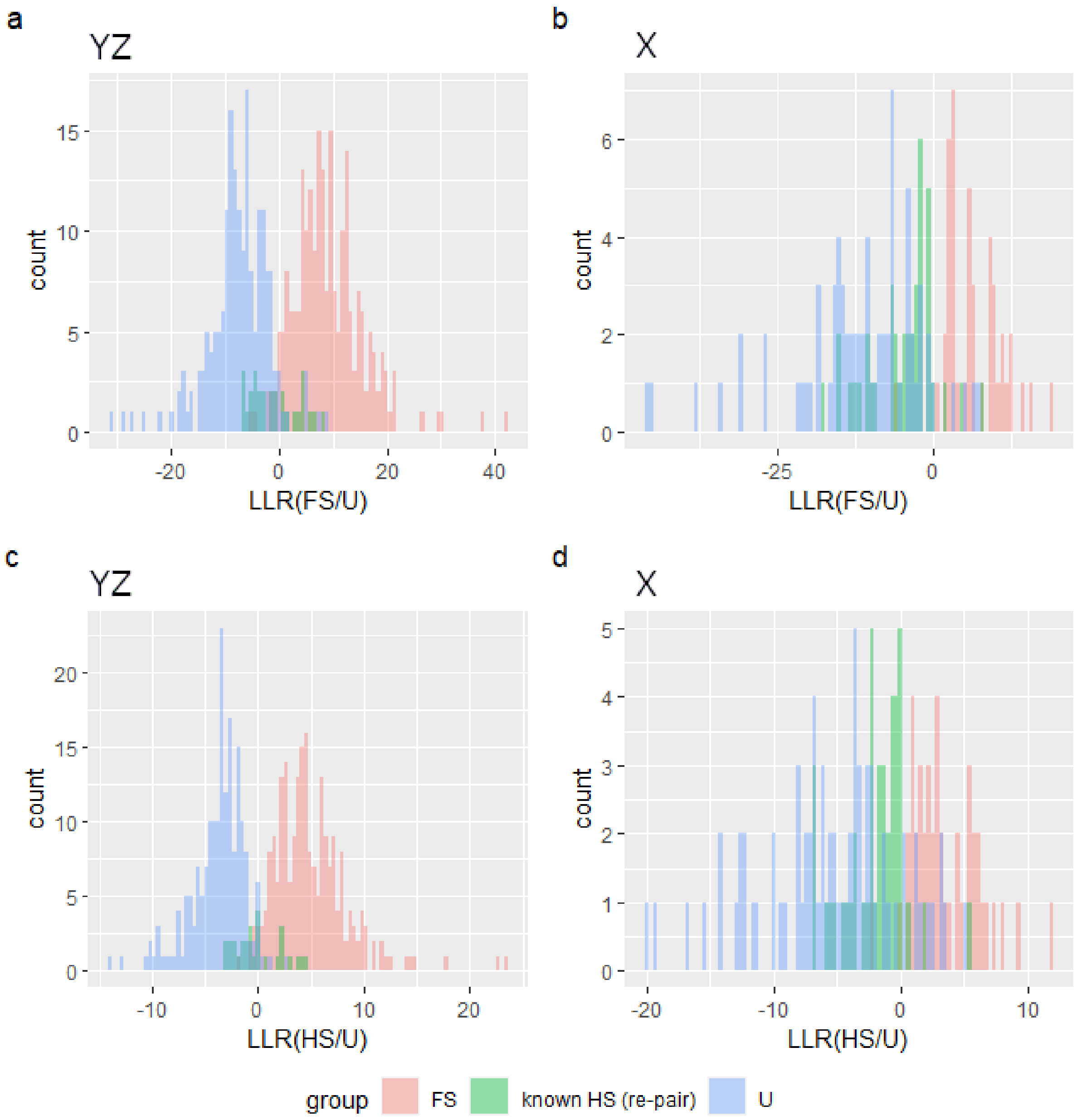


Figure S2 Pairwise log likelihood ratios (LLRs) for FS (full sibling)/U (unrelated) and HS (half-sibling)/U, per population (Site YZ and Site X).

Pedigree reconstruction was undertaken in Sequoia, implementing an integrated approach of pedigree reconstruction using *sequoia*() and detailed examination of pairwise relationship log likelihood ratios (LLRs) using *CalcPairLL*(). While an age prior, based on known ages of offspring and estimated ages of parents, was implemented in pedigree reconstruction with *sequoia*(), *CalcPairLL*() was run with a flat age prior (i.e. no ages were specified). This is because we were specifically interested in verifying parent-offspring, full-sibling and half-sibling relationships within known social family units using *CalcPairLL*(), and given our social pedigree, an age prior was not necessary to elucidate patterns of relatedness. For example, high support for LLR(full sibling (FS)/unrelated (U)) between social siblings indicates full sibship, while high support for LLR(FS/U) between a social parent and its offspring indicates a parent-offspring (PO) relationship (given the same relationship coefficient, 0.5, between full siblings and parent-offspring).

Of 76 offspring at Site YZ, there was strong support that all but nine individuals clustered into their social sibships and with their social parents (where social parents were included in the analysis). Of 28 offspring at Site X, one did not cluster into its social sibship/with its social parents. To understand whether these individuals were the result of egg dumping or extra-pair fertilisation, we first explored pairwise LLRs for full sibling versus unrelated (FS/U) and half-sibling versus unrelated (HS/U) for known social half-siblings (where a parent had re-paired). This allowed for a per-population identification of the range of LLR(FS/U, HS/U) that known half-siblings exhibit. Figure S2a and b show LLR(FS/U) for Site YZ and Site X, respectively, while Figure S2c and d show LLR(HS/U) for Site YZ and Site X, respectively.

Relationships inferred from Figure S2 for individuals who were not assigned as full siblings with individuals in their social sibship are presented in Table S1 (YZ) and Table S2 (X). Note that although there is some overlap between the LLR(FS/U,HS/U) distributions for known social half-siblings and the LLR(FS/U,HS/U) distributions for full siblings, sequoia() correctly assigned social half-siblings as *not a full sibling* 95.77% of the time (3 of 71 pairwise social half-sib relationships were assigned as full siblings). In addition, the known social half-siblings who were incorrectly assigned as full siblings showed clear patterns of half-sibship with multiple members of their social half-sibship (i.e. were not assigned as full siblings with them *and* had an LLR(FS/U, HS/U) within the bounds of half-sibship). Thus, integrating three pieces of information: (i) sequoia's assignment, (ii) pairwise LLR(FS/U, HS/U) and (iii) patterns of LLR(FS/U,HS/U) within the whole sibship, known social half-siblings could be identified in every case. Using this methodology for all individuals in the analysis, it is likely that all (or almost all) cases of half-sibship are identified.

Table S1: Site YZ Log likelihood ratio (LLRs) thresholds of full sibling (FS) versus unrelated (U) and half-sibling (HS) versus U, and the corresponding most likely relationship.

LLR	VALUE	RELATIONSHIP
FS/U	> 8	Full sibling
	-8 – 8	Half-sibling or unrelated
	< -8	Unrelated
HS/U	> 4	Full sibling
	-4 - 4	Half-sibling or unrelated
	< -4	Unrelated

Table S2: Site X Log likelihood ratio (LLRs) thresholds of full sibling (FS) versus unrelated (U) and half-sibling (HS) versus U, and the corresponding most likely relationship.

LLR	VALUE	RELATIONSHIP
FS/U	> 5	Full sibling
	-20 - 5	Half-sibling or unrelated
	< -20	Unrelated
HS/U	> 5	Full sibling
	-8 – 5	Half-sibling or unrelated
	< -8	Unrelated

Site YZ

J2862: potential egg dumping

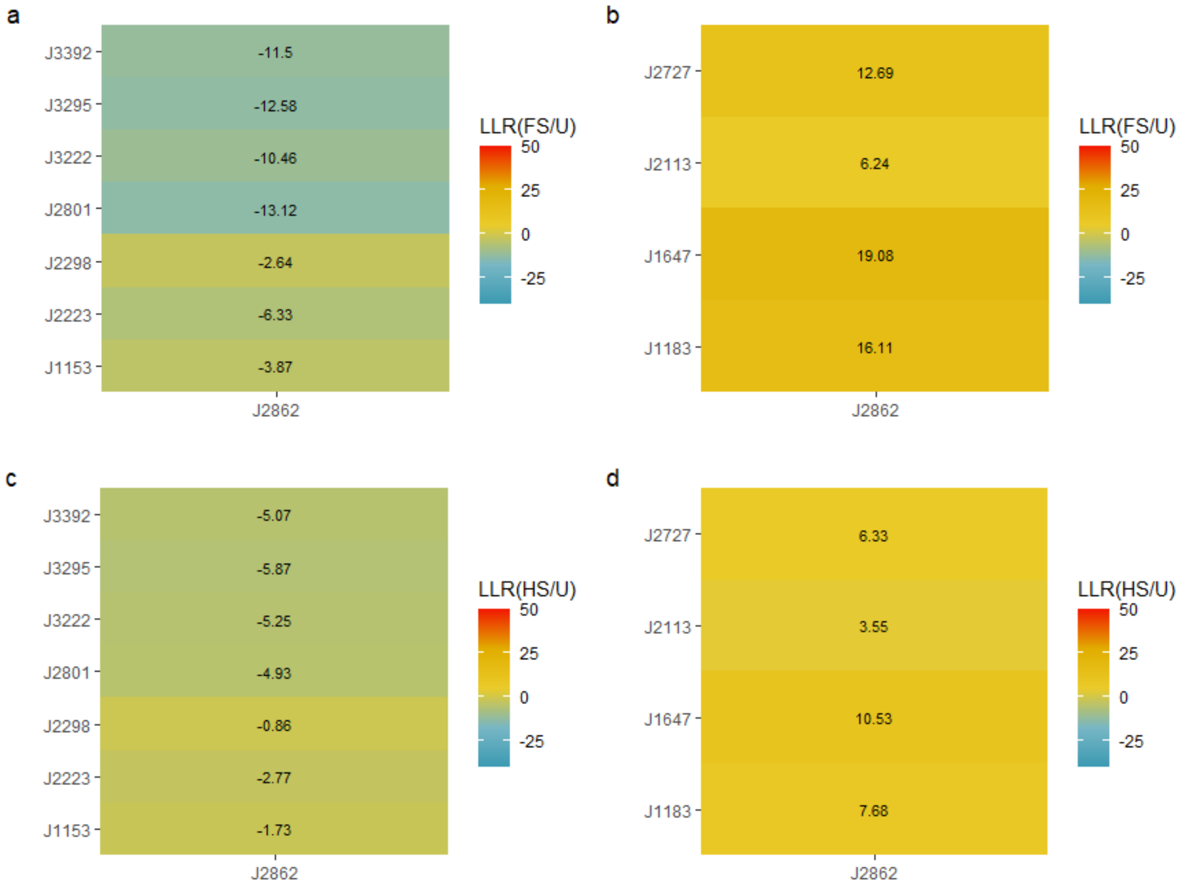


Figure S3 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios (LLRs) for J2862's social sibship and **b** for its assigned sibship. **c** and **d** show half-sibling/unrelated LLRs.

J2862 was not assigned to its social sibship, and instead was assigned to the sibship of a neighbouring pair. While neither sibship had any sampled parents (due to strict data filtering), we found substantial support that J2862 was unrelated to its social sibship and full siblings with individuals in the group it was assigned to by Sequoia (see inset LLRs and corresponding Table S1).

From our field data, we can confirm that (i) based on clutch size, lay and hatch dates, a second (non-resident) female is likely to have laid eggs in the nest-box J2862 is suggested to have been dumped into; however that (ii) the inferred genetic mother is unlikely to be J2862’s actual mother because of anomalous pigmentation patterns on this egg relative to the rest of her brood. Thus, the combined field and genetic data do not present a clear picture, and it must be considered that J2862 may be a mistaken identity due to a sample mix-up.

J2700: egg dumping

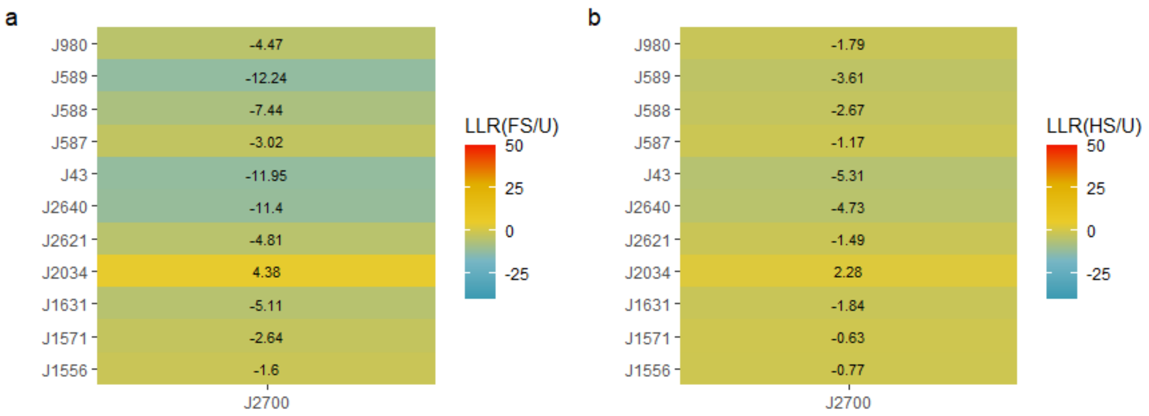
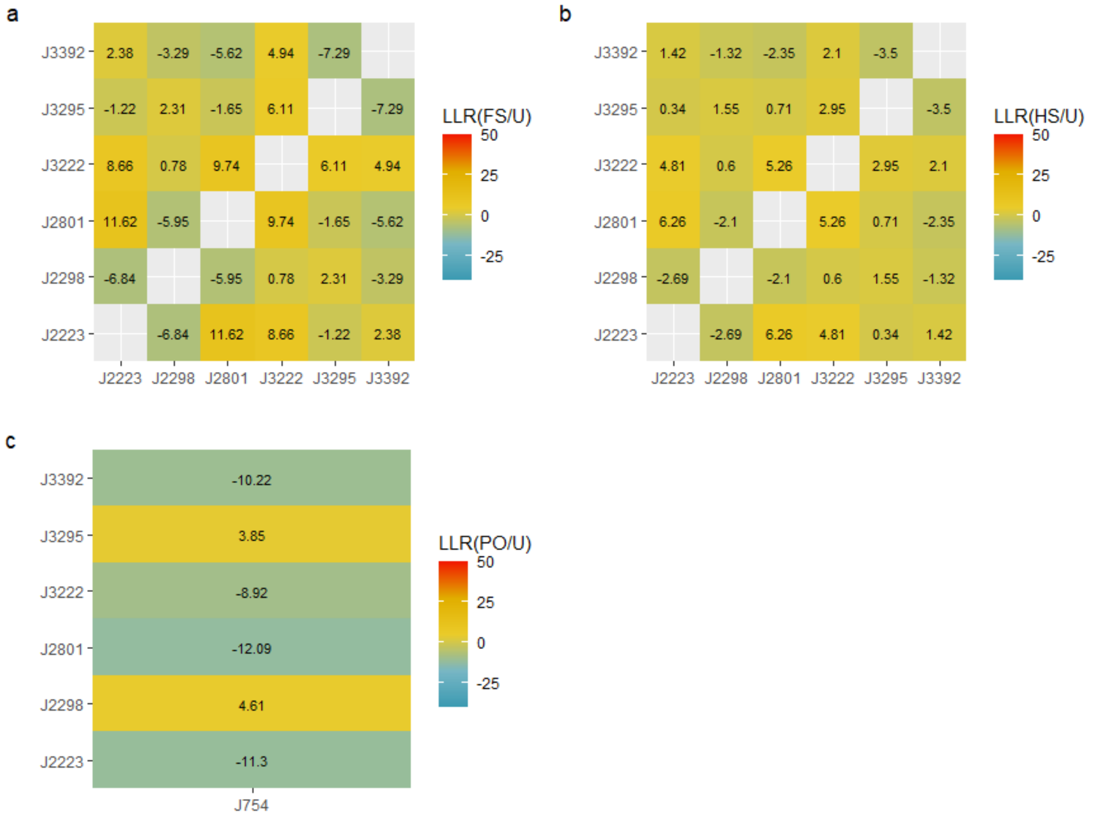


Figure S4 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios (LLRs) for J2700's social sibship and **b** half-sibling/unrelated LLRs.

J2700 was not assigned to any sibship. It is very unlikely to be a full or half-sibling with multiple social siblings, and has no strong support for full sibship with any (see inset LLRs and corresponding Table S1). J2700's social mother was also included in the analysis, and J2700 is more likely to be unrelated than her offspring (LLR(PO/U) = -5.57, where PO = parent-offspring and negative values indicate that U is more likely than PO). Taken together, this evidence suggests that J2700 was egg-dumped into this nest by an unsampled female. Field data supports this conclusion.

J3295, J2298 and J3392: extra-pair maternity



a

	J2223	J2298	J2801	J3222	J3295	J3392
J3392	2.38	-3.29	-5.62	4.94	-7.29	
J3295	-1.22	2.31	-1.65	6.11		-7.29
J3222	8.66	0.78	9.74		6.11	4.94
J2801	11.62	-5.95		9.74	-1.65	-5.62
J2298	-6.84		-5.95	0.78	2.31	-3.29
J2223		-6.84	11.62	8.66	-1.22	2.38

b

	J2223	J2298	J2801	J3222	J3295	J3392
J3392	1.42	-1.32	-2.35	2.1	-3.5	
J3295	0.34	1.55	0.71	2.95		-3.5
J3222	4.81	0.6	5.26		2.95	2.1
J2801	6.26	-2.1		5.26	0.71	-2.35
J2298	-2.69		-2.1	0.6	1.55	-1.32
J2223		-2.69	6.26	4.81	0.34	1.42

c

	J754
J3392	-10.22
J3295	3.85
J3222	-8.92
J2801	-12.09
J2298	4.61
J2223	-11.3

Figure S5 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios for the social sibship and **b** half-sibling LLRs; **c** shows the LLR of J754 (the ‘follower’ female) being the genetic parent of each individual (PO = parent-offspring).

Neither of the social parents of this social sibship were retained for analysis post-filtering; thus, relationships are inferred by patterns of sibship and relationships with the known ‘follower’ female associated with the social parents, J754.

J754 was observed laying an egg in the nest-box owned by the social parents, so she was strongly suspected to be the genetic mother of at least some of this sibship. Indeed, we found some support that J754 is J3295’s mother (2017; LLR(PO/U) = 3.85) and J2298’s mother (2019; LLR(PO/U) = 4.61). LLR(FS/U) and LLR(HS/U) values do not refute that J3295 and J2298 could be full siblings (see inset LLRs and corresponding Table S1), but equally there is no strong support. Nevertheless, the fact that every pairwise LLR with other individuals in the sibship are within the range of half-sibship, while none show strong support for being unrelated, suggest that J3295 and J2298 are the offspring of the social father, J2402, and are thus half-siblings with their social siblings.

Based on LLR(FS/U)s and LLR(HS/U)s, J3392 may also be a half-sibling of all other individuals, including J3295 + J2298, but it is highly unlikely to be a child of J754 (PO/U < -10). Given strong support for full sibship between J2223, J2801 and J3222 (see inset LLRs and corresponding Table S1), these are likely to be the offspring of the social parents. If J3392 is unrelated to J754, but a half-sibling of everyone else, including J3295 and J2298, then they most likely share a father in common: the social father, J2402.

It is important to note that we cannot absolutely rule out that there are cases of egg dumping here, but given that no pairwise combination shows a strong indication that a pair are unrelated, egg dumping is considered to be unlikely.

J1153: extra-pair paternity (or maternity)

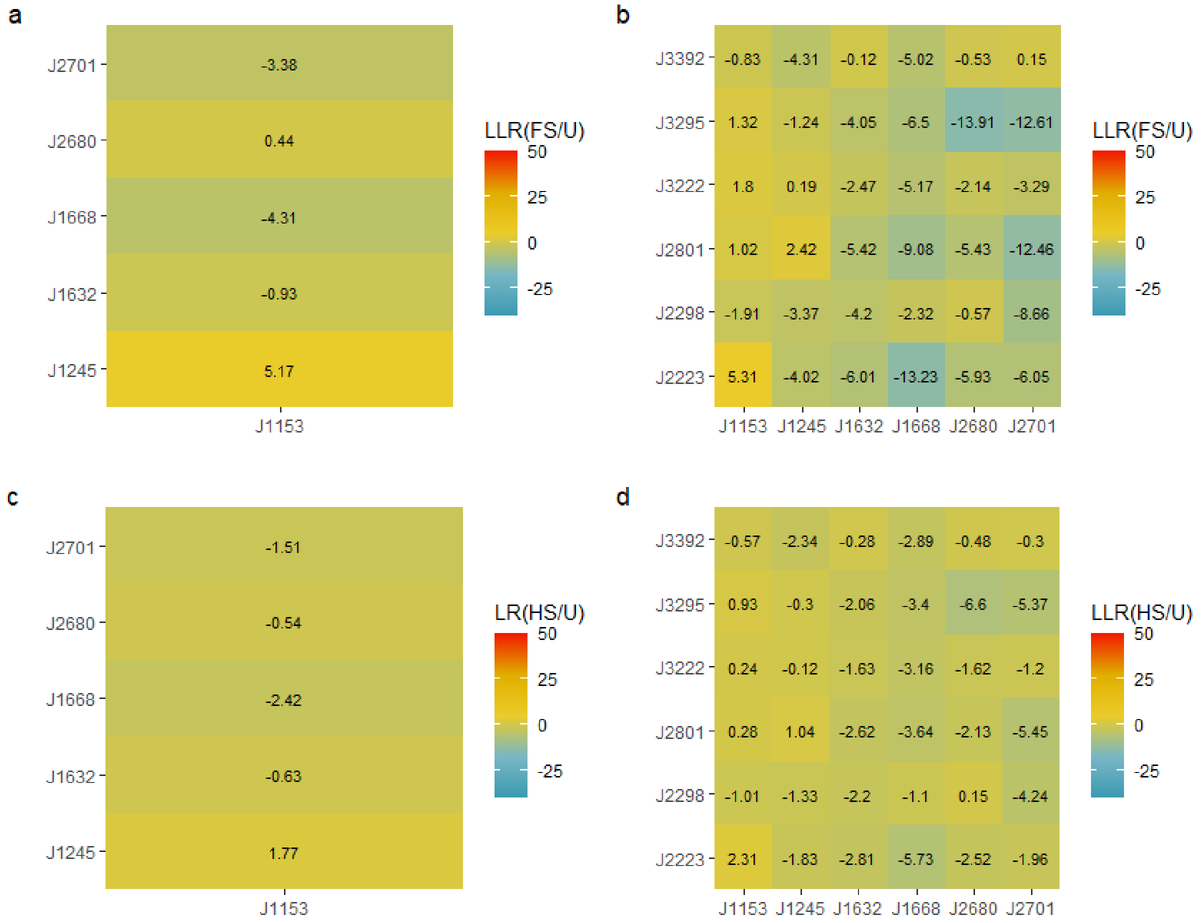


Figure S6 Heatmaps with inset values of full sibling/unrelated pairwise log likelihood ratios (LLRs) for **a** J1153's social sibship and **b** J1153's assigned sibship, plus its social sibling's relationships with its assigned siblings; **c** and **d** show half-sibling/unrelated LLRs.

J1153's LLR(PO/U) with its social father shows little support that they are parent-offspring (-2.43). Moreover, J1153 has LLR(FS/U) and LLR(HS/U) values that are in-line with half-sibship with every member of its social and its assigned sibship, and no strong support for any unrelated combinations (see inset LLRs and corresponding Table S1). Although this could be assumed to be a case of half-sibship, where J1153 is sired by the father of its assigned sibship, the picture becomes more complicated when assessing J1153's social sibship's relationships with its assigned sibship. Indeed, it appears that there is a general pattern of weak relatedness between the two sibships, which makes inference more difficult. Nonetheless, the data suggest that J1153 is unlikely to be a full sibling of its social sibship, and most likely to be a maternal half-sibling (although we cannot rule out that it is a paternal half-sibling), while its relationship as half-sibling with its assigned sibship remains speculative.

J2084: extra-pair maternity (or paternity)

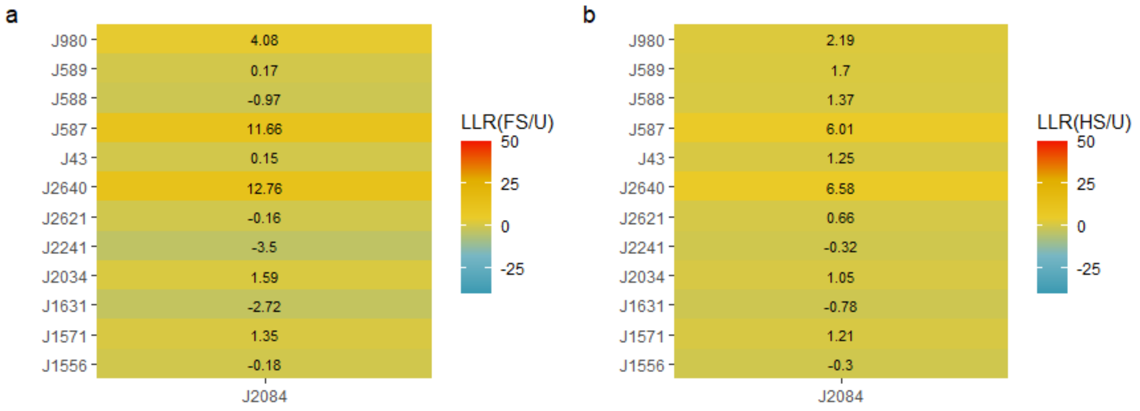


Figure S7 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios (LLRs) for J2084's social sibship and **b** half-sibling/unrelated LLRs.

J2084 has no strong support as the offspring of its social mother, J243 (LLR(PO/U) = -4.73). However, it has strong support as full sibling with two of its social sibship (who themselves are likely to be the offspring of J243 and full siblings with other members of the sibship), and LLR(FS/U,HS/U) values within the range of half-sibship, with no strong support for unrelatedness (see inset LLRs and corresponding Table S1). Therefore, the most likely scenario is that J2084 is a paternal half-sibling of its social sibship (the product of extra-pair maternity), although we cannot rule out extra-pair paternity.

J2144: egg dumping or extra-pair maternity

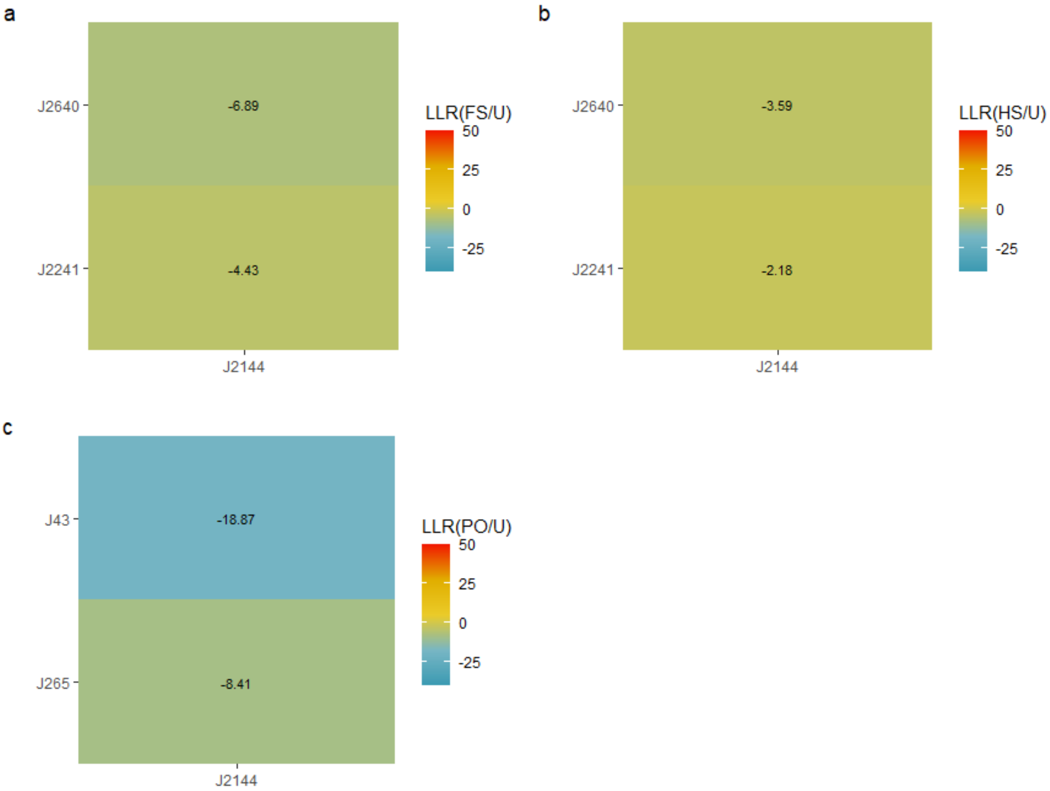


Figure S8 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios for**J2144's** social sibship and **b** half-sibling LLRs; **c** shows the LLR of J2144 having a parent-offspring relationship with its social parents, versus unrelated.

J2144 is highly unlikely to be the offspring of J43 based on LLR(PO/U), and has an ambiguous relationship with J265, implying egg dumping or half-sibship. FS/U with its social siblings lend support to either of these conclusions (see inset LLRs and corresponding Table S1), although the negative LLR(PO/U) with both parents means egg-dumping is the most likely scenario.

J2141/J2035: egg dumping or extra-pair maternity/paternity

No social siblings or parents of these two social siblings were retained in the analysis post-filtering. These individuals have an LLR(FS/U) of -7.7 and HS/U is -2.84, which indicates either egg dumping or half-sibship through extra-pair maternity or extra-pair paternity.

Site X

J2018: extra-pair maternity or paternity

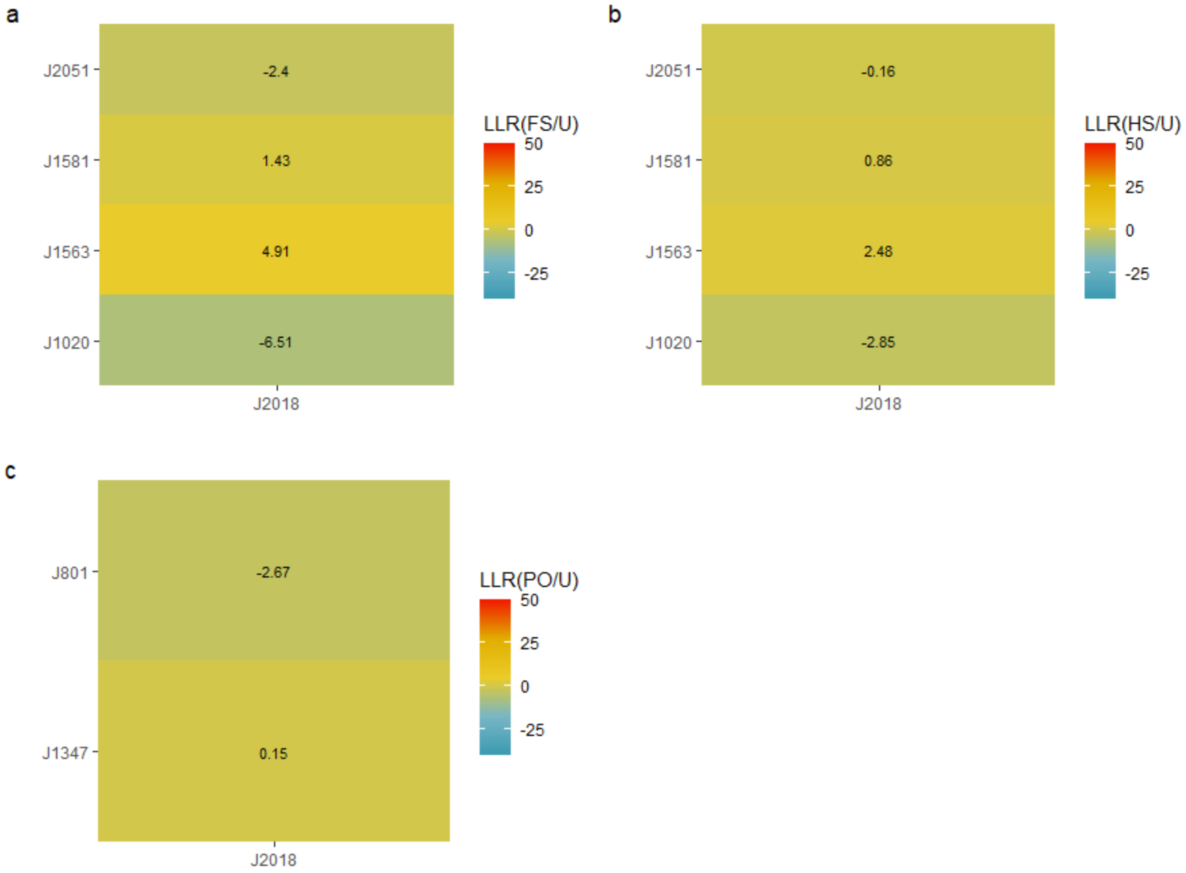


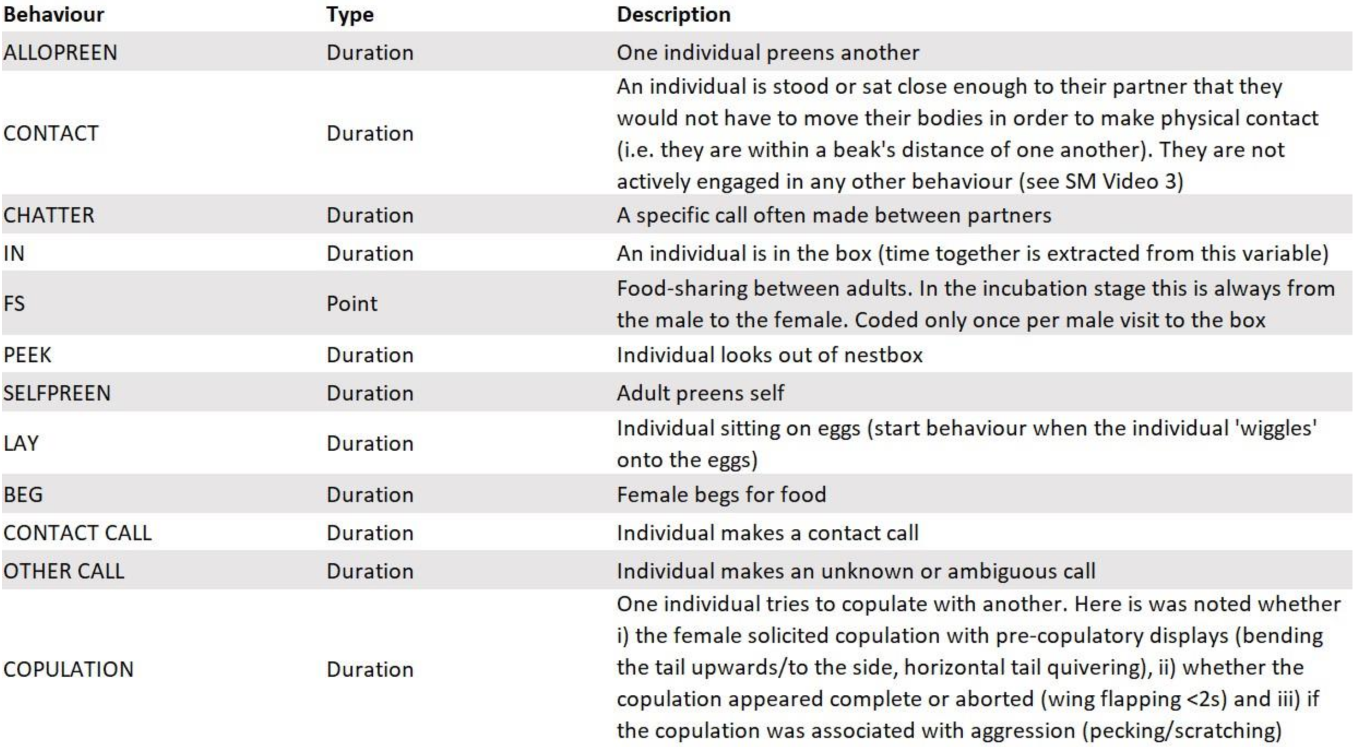
Figure S9 Heatmaps with inset values of **a** full sibling/unrelated pairwise log likelihood ratios for J2018’s social sibship and **b** half-sibling LLRs; **c** shows the LLR of J2144 having a parent-offspring relationship with its social parents, versus unrelated.

J2018 has an ambiguous relationship with both social parents, and ranges that fall into potential half-sibship with all its social offspring (see inset LLRs and corresponding Table S2). Given moderate support for full sibship (with J1563), where LLR(FS/U) approaches the boundary of strong confidence in full sibship, the most likely scenario is that J2018 is half-siblings with its social siblings (full sibling support would be unlikely if egg dumping had occurred); however, it is unclear as to whether this is due to extra-pair maternity or paternity.

Behavioural data analysis

Ethogram

Table S3 The behavioural ethogram used by the lead author to code fine-scale behaviours



Behaviour	Type	Description
ALLOPREEN	Duration	One individual preens another
CONTACT	Duration	An individual is stood or sat close enough to their partner that they would not have to move their bodies in order to make physical contact (i.e. they are within a beak's distance of one another). They are not actively engaged in any other behaviour (see SM Video 3)
CHATTER	Duration	A specific call often made between partners
IN	Duration	An individual is in the box (time together is extracted from this variable)
FS	Point	Food-sharing between adults. In the incubation stage this is always from the male to the female. Coded only once per male visit to the box
PEEK	Duration	Individual looks out of nestbox
SELFPREEN	Duration	Adult preens self
LAY	Duration	Individual sitting on eggs (start behaviour when the individual 'wiggles' onto the eggs)
BEG	Duration	Female begs for food
CONTACT CALL	Duration	Individual makes a contact call
OTHER CALL	Duration	Individual makes an unknown or ambiguous call
COPULATION	Duration	One individual tries to copulate with another. Here is was noted whether i) the female solicited copulation with pre-copulatory displays (bending the tail upwards/to the side, horizontal tail quivering), ii) whether the copulation appeared complete or aborted (wing flapping <2s) and iii) if the copulation was associated with aggression (pecking/scratching)

FEPC versus within-pair copulation

In 13 of 16 cases of suspected FEPC, the male could be identified as definitely not the resident male. This could be ascertained either through certain identification of the intruding male using ring combinations, or because the intruding male was unringed while the resident male was ringed (or vice versa). In the remaining three cases, behaviour clearly indicated an FEPC event. FEPCs were characterised by (i) female nest-box defence, (ii) no affiliative behaviour (allopreening or contact sitting/standing) and (iii) prolonged female attack. This was in contrast to within-pair copulations, none of which involved nest-box defence, all of which involved affiliative behaviour and none of which involved prolonged female aggression.

Covariate inclusion

For non-behavioural correlates of FEPC (models presented in Table 3), two models were run to test for the effect of structural size on the probability of FEPC because models performed poorly when both male and female structural size were included in one model. Similarly, all potential covariates except from number of days since the female was fertile were removed due to model performance issues. Both final models performed well.

In models presented in Table 2, minimum male age is used as a covariate, while minimum female age is used as a covariate in models presented in Table 3. This is because when subsetting to data where both the female and male's exact ages are known, all three variables are completely or highly correlated (female age and male age: df = 11, Pearson's correlation coefficient = 1, p < 0.0001; female age/male age and years together: df = 11, Pearson's correlation coefficient = 0.910, p < 0.0001). Therefore, the most biologically meaningful variable was selected for each model (for male behaviour, male age; for female behaviour, female age).

Because partners were of the same age in every case where exact age was known for both individuals, where one individual in a pair had a missing age we inputted their partner's age as an estimate (exact age = 0 (not known)).

WPC vs FEPC copulation visit length

Copulation visit length varied greatly, from a minimum of 14 seconds to a maximum of 2338 seconds. Only four visits, all of which by resident males (not intruders) were >500 seconds. Models were significantly over-dispersed with the inclusion of these four points according to DHARMa's dispersion test. With these points removed, the model performed well. The model presented in the main text is therefore the model with these points excluded.

Models where full results are presented

We identified influential points (four times the mean Cook's distance, as described in the methods) and removed them from our models. In the majority of cases this made no qualitative difference to results. However, models presented in Table 3 had differing results when influential points were removed. In Table 3, model 1 (female vigilance as a response to FEPC/no FEPC in a video), four influential points were identified. All were females from non-FEPC videos, three of which were in the *top* 5% for time spent in vigilance. Both the full model and the model with influential points removed are presented in the main text.

In Table 3, model 2 (female vigilance pre- and post-FEPC), one influential point was identified. This was a female who spent 0.03% of her time pre-FEPC in vigilance, and 3.17% of her time post-stressor in vigilance. With this female included, there was no significant difference in female behaviour pre- or post-stressor. With her removal, females were significantly more vigilant pre- relative to post-stressor.

In Table 3, model 3 (female incubation and FEPC/no FEPC in a video), four influential points were identified. Again, all were females from non-FEPC videos, three of which were in the *bottom* 5% of time spent incubating. As before, both the full model and the model with influential points removed are presented in the main text.

In the model exploring female structural size and FEPC, the removal of influential points resulted in violations of model assumptions. We therefore present the full model in the main text. Days since the female's fertile period was included as a covariate, and showed qualitatively identical results to those presented for the model exploring male structural size (that females are more likely to experience FEPC closer to their fertile period).

CHAPTER 3: Wild jackdaws respond to their partner’s distress, but not with consolation

Note: *a version of this chapter has been published as* Hooper R, Meekins E, McIvor G.E. and Thornton A. (2021) Wild jackdaws respond to their partner's distress, but not with consolation. R. Soc. open sci.8210253210253. http://doi.org/10.1098/rsos.210253

Abstract

Individuals are expected to manage their social relationships to maximise fitness returns. For example, reports of some mammals and birds offering unsolicited affiliation to distressed social partners (commonly termed ‘consolation’) are argued to illustrate convergent evolution of prosocial traits across divergent taxa. However, most studies cannot discriminate between consolation and alternative explanations such as self-soothing. Crucially, no study that controls for key confounds has examined consolation in the wild, where individuals face more complex and dangerous environments than in captivity. Controlling for common confounds, we find that male jackdaws (*Corvus monedula*) respond to their mate’s stress-states, but not with consolation. Instead, they tended to decrease affiliation and partner visit rate in both experimental and natural contexts. This is striking because jackdaws have long-term monogamous relationships with highly interdependent fitness outcomes, which is precisely where theory predicts consolation should occur. Our findings challenge common conceptions about where consolation should evolve, and chime with concerns that current theory may be influenced by anthropomorphic expectations of how social relationships should be managed. To further our understanding of the evolution of such traits, we highlight the need for our current predictive frameworks to incorporate the behavioural trade-offs inherent to life in the wild.

Introduction

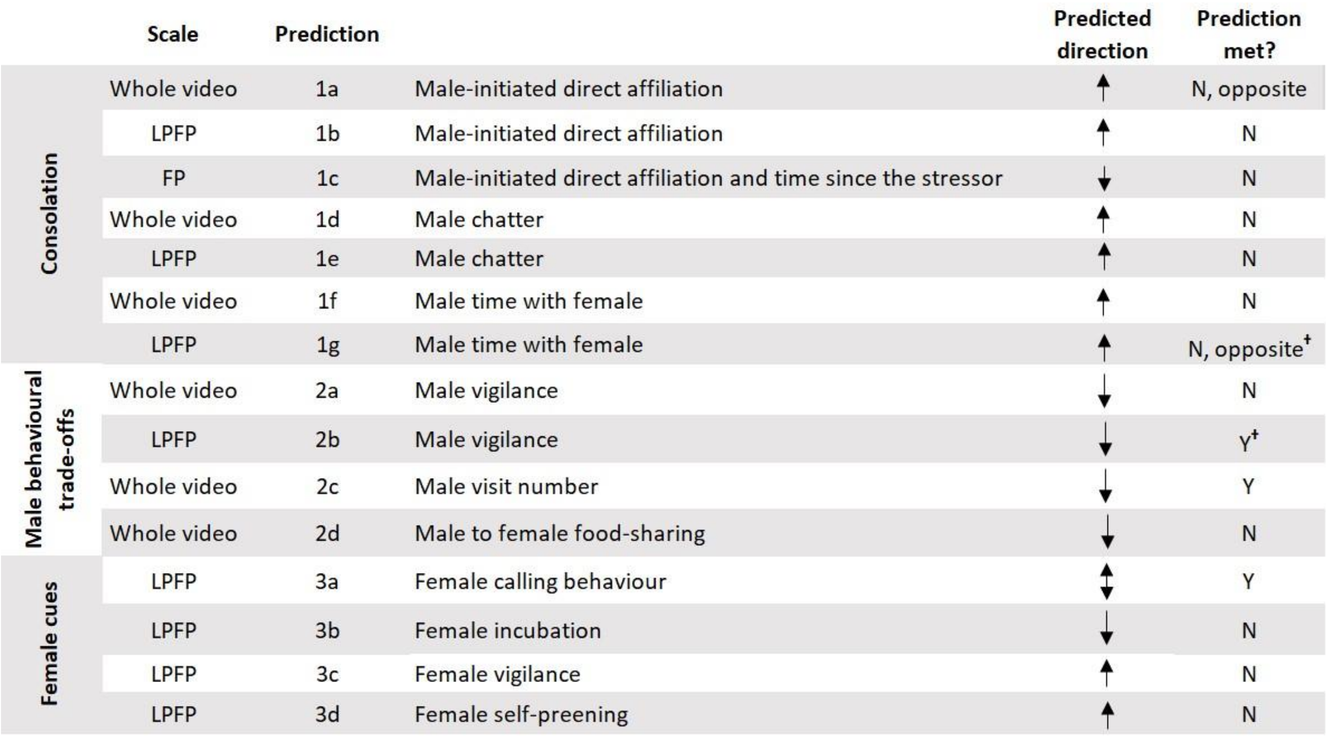
In species where social bonds differentially influence fitness outcomes, individuals should strategically manage and maintain relationships in order to maximise fitness returns (de Waal & Aureli, 1997; Kummer, 1978). One proposed mechanism through which individuals do this is via prosocial behaviour motivated by empathetic and/or sympathetic concern, such as consolation (Adriaense et al., 2020; Pérez-Manrique & Gomila, 2018). Consolation is said to occur when an individual directs unsolicited affiliation towards a distressed individual in order to alleviate their negative emotional state (Burkett et al., 2016). It is therefore considered to be ‘other-oriented’ behaviour (*sensu* (Pérez-Manrique & Gomila, 2018)). To date, consolation has been studied almost exclusively in the context of Post-Conflict Matched-Control (PC-MC) experimental designs. In these studies, researchers measure affiliative behaviour directed from bystanders toward individuals involved in a conflict, and compare this to baseline levels of affiliation (Adriaense et al., 2020; de Waal & van Roosmalen, 1979; Pérez-Manrique & Gomila, 2018). Results from these studies suggest that consolation may occur in a range of species, including great apes (Cordoni et al., 2006; de Waal & van Roosmalen, 1979; Palagi & Norscia, 2013), monkeys (Call et al., 2002; Palagi et al., 2014), canids (Quervel-Chaumette et al., 2016) and corvids (Fraser & Bugnyar, 2010a). The presence of consolation-like behaviours in such evolutionarily disparate lineages has been used to justify claims of convergent socio-cognitive evolution across divergent taxa (Güntürkün & Bugnyar, 2016). However, whether these findings truly reflect other-oriented behaviour is open to question.

Most study designs include potential confounds that hinder robust conclusions about the occurrence of consolation. First, the putative ‘consoling individual’ typically witnesses the stressor, which in the majority of studies is a conflict between groupmates. Given that not only being involved in (Wascher et al., 2010), but also witnessing conflicts has been found to increase physiological indicators of stress (Dezecache et al., 2017; Kano et al., 2016), the ‘consoler’ may simply engage in affiliation to reduce its own distress (Adriaense et al., 2020) (‘self-soothing’). Some studies attempt to control for this by recording behavioural proxies of distress in the ‘consoler’ (e.g. (de Waal & Aureli., 1996)), but in the absence of physiological measurements it is not possible to entirely eliminate the possibility of self-soothing. This confound is overcome in several studies of rodents and companion animals by ensuring that the ‘consoler’ is blind to the stressor (e.g. (Burkett et al., 2016; Li et al., 2019; Quervel-Chaumette et al., 2016)); however, the role of consolation in natural (as opposed to captive or domestic) environments remains unclear. Second, whether affiliation is solicited by the distressed individual (for example through the initiation of the affiliative contact or through specific signals (de Waal & Aureli., 1996)) is not always measured (e.g.(Burkett et al., 2016)), even though this would rule consolation out as an explanation of the observed behaviour. These two potential confounds represent behaviours that are proximately, and in the first case ultimately, different to consolation. A third potential confound is that directing affiliation toward an individual previously involved in a conflict may function to protect the ‘consoler’ from redirected aggression (Koski & Sterck, 2009; Palagi & Cordoni, 2009; Schino & Marini, 2012), to reconcile former opponents (Wittig & Boesch, 2010) and/or to strengthen and advertise alliances (Das et al., 1998). Here, affiliative behaviour may appear proximately identical to consolation but its ultimate function differs (Adriaense et al., 2020).

A further important limitation to our current understanding of consolation is that the vast majority of studies have been performed on captive or semi free-ranging populations. Wild individuals are subject to a more dangerous and complex ecological landscape than their captive counterparts (Pritchard et al., 2016), and there is growing recognition that behavioural and cognitive phenotypes measured in captivity may not reflect those employed by animals in the wild (Boesch, 2020; Cauchoix et al., 2020; Thornton & Lukas, 2012; Webster & Rutz, 2020). Indeed, in the few cases where consolation has been studied on both captive and wild populations of the same species, results are often inconsistent (Cordoni et al., 2006; Pérez-Manrique & Gomila, 2018), suggesting that the costs and benefits of consolation differ between contexts. Additionally, experiments that alter resource availability in captive populations have found that different levels of resource competition influence patterns of post-conflict affiliation (Sima et al., 2016). Wild animals are likely to face higher levels of resource competition, greater constraints on their activity budgets (e.g.(Inoue & Shimada, 2020)) and thus substantially different trade-offs (for example between investment in affiliation versus foraging) compared to their captive counterparts. Consequently, field studies are critical to determine if consolation plays a role in animal societies under natural conditions. To our knowledge no studies of consolation that control for the aforementioned confounds have been conducted on wild animals, yet only by studying consolation in the wild can we interrogate its adaptive value and thus understand the ultimate reasons it has evolved (Pritchard et al., 2016; Thornton & Lukas, 2012).

Here, we test whether wild jackdaws (*Corvus monedula*) exhibit consolation towards social partners while explicitly controlling for the potential confounds of self-soothing, solicitation and alternative conflict-related motives such as protection from redirected aggression. Jackdaws are a highly social member of the corvid family and form lifelong, monogamous pair-bonds where partners have almost completely interdependent fitness (Gill et al., 2020; Röell, 1978). Individuals in a partnership are therefore highly valuable to each other, in terms of the potential influence they have on one another's fitness outcomes (Kummer, 1978). This is therefore precisely the context in which investment in relationship management and maintenance should occur (Kummer, 1978), and thus where consolation would be expected (Adriaense et al., 2020; Kummer, 1978; Pérez-Manrique & Gomila, 2018). There is evidence that post-conflict third-party affiliation occurs in captive corvids (e.g.(Fraser & Bugnyar, 2010a; Logan, Emery, et al., 2013; Seed et al., 2007; Sima et al., 2018)), but robust conclusions about consolation (i.e. truly *other-oriented* affiliative behaviour (Pérez-Manrique & Gomila, 2018)) cannot be drawn due to the aforementioned caveats. Our experimental design overcomes these potential caveats by (i) ensuring that the potential 'consoler' was blind to the stressor, (ii) measuring fine-scale female behaviour to rule out female solicitation, and (iii) testing our predictions in a context where protection from redirected aggression and reconciliation of former opponents were not relevant. To do this, we exposed incubating females to a stressor while their partners were absent from the area and compared behaviours of both the male and the female in the pre- and post-stressor period. Under natural conditions, female jackdaws are occasionally subjected to violent, forced extra-pair copulations (FEPCs) by intruders while their partners are absent from the nest (Chapter 2; Gill et al., 2020). FEPC attempts are likely to be highly stressful, as females almost always resist them by defending the nest-box vigorously and attacking the intruder (see Supplementary Materials Section 1, Chapter 2 and Gill et al., 2020), but genetic analyses show they do not result in fertilisations (Gill et al., 2020), so do not present a risk of lost paternity to the male partner. For our experimental stressor, we simulated an FEPC event by exposing females alone at the nest-box to a playback of an unknown male landing on the nest-box. We supplemented this experimental data with data from naturally-occurring FEPC events, where females were subjected to forced copulations while their partners were absent. We predicted that (1) male jackdaws would console their stressed partners upon their return, but that (2) investing in consolation would result in trade-offs with other behaviours, such as vigilance, and (3) that males would respond to subtle but detectable changes in female behaviour. See Table 1 for a summary of predictions and whether they were supported.

Table 1 Predictions concerning consolation, behavioural trade-offs and female cues: all predictions relate to the post-stressor period in comparison to the pre-stressor period, except from prediction 1c. Arrows indicate the predicted direction of the effect. ǂ indicates a result specific to the experimental dataset only. LPFP = the Last Pre-stressor First Post-Stressor male visit; FP = First Post-stressor male visit.



	Scale	Prediction		Predicted direction	Prediction met?
Consolation	Whole video	1a	Male-initiated direct affiliation	↑	N, opposite
	LPFP	1b	Male-initiated direct affiliation	↑	N
	FP	1c	Male-initiated direct affiliation and time since the stressor	↓	N
	Whole video	1d	Male chatter	↑	N
	LPFP	1e	Male chatter	↑	N
	Whole video	1f	Male time with female	↑	N
	LPFP	1g	Male time with female	↑	N, opposite†
Male behavioural trade-offs	Whole video	2a	Male vigilance	↓	N
	LPFP	2b	Male vigilance	↓	Y†
	Whole video	2c	Male visit number	↓	Y
	Whole video	2d	Male to female food-sharing	↓	N
Female cues	LPFP	3a	Female calling behaviour	↕	Y
	LPFP	3b	Female incubation	↓	N
	LPFP	3c	Female vigilance	↑	N
	LPFP	3d	Female self-preening	↑	N

Methods

Ethics statement The experiment received ethical approval from the University of Exeter Bioscience Ethics Committee (eCORN001858) and followed ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching56. All birds involved in this study had been previously captured and ringed by British Trust for Ornithology and UK Home Office (project licence 30/3261) licenced researchers.

Study population The experiment was performed during the 2019 breeding season on 30 wild pairs of jackdaws in two study sites in Cornwall, UK (Site Y 50°11′26″N, 5°10′51″W; 14 pairs and Size Z 50°11′56″N, 5°10′9″W; 16 pairs). All jackdaws involved in the experiment were nest-box residents with rings allowing individual identification. Each pair was involved in one trial only.

Experimental stressor Internal CCTV cameras were placed in 30 target nest-boxes during the nest-building stage of the breeding season. Dummy speakers were attached externally to nest-boxes at least 14 days prior to experimentation, to ensure habituation before they were swapped with FoxPro Fury remote-controlled loudspeakers on the night preceding or dawn of the experimental day. All experiments were conducted after 10 ± 3 days of clutch initiation, when the clutch was complete, the female was no longer fertile (Gill et al., 2020; Henderson et al., 2000) and incubation had started. Targeting the pair during incubation meant the female was almost always in the nest-box, while the male visited frequently to provide her with food (termed ‘food sharing’). This allowed us to expose the female to the stressor while the male was foraging, with the knowledge that he would soon return to feed to the female. Experimenters used either a hide or a car as cover while running the experiment. If a hide was used, the hide was erected at least 12 hours before the experiment so that the pair had time to adjust to its presence. No acclimatisation period was needed when cars were used as these populations of jackdaws are habituated to their presence.

Speakers were pre-loaded with an audio sequence that aimed to simulate the sound of a non-partner male landing on the nest-box. We expected that this would be a stressful event for females given that female jackdaws are known to be subjected to forced extra-pair copulations (FEPCs) during incubation (regardless of fertility state) by non-partner males while alone in the nest-box, and based on video footage from our own populations and other study sites(Gill et al., 2020), FEPCs are almost always violently resisted by females (see Supplementary Materials Section 1). The audio sequence contained 20 seconds of jackdaw feet walking on a nest-box, followed by the contact call of one of four males from a different and non-overlapping population. Sequences were created in Audacity(Audacity Team, 2019). The four sequences were set to a volume that simulated natural landing and call volumes; we did this by recording calls directly outside the nest-box and the sound of feet landing on the nest-box from the interior of the box. We then adjusted the volume of the playback so that the decibel level received inside the nest-box from the playback matched these natural decibel levels. Sequences were assigned randomly to each pair.

A maximum of four experiments were run per day, starting from 8:00 – 9:40 or 13:30 – 14:30. Experimenters recorded internal pre-stressor footage for 1.5 hours while observing the nest-box from the hide or car. If we noticed that the female or male appeared hesitant to enter the nest-box in the pre-stressor period (e.g. they repeatedly returned to the area but did not enter the nest-box, suggesting that they may not have habituated to the presence of the hide), we called off the experiment and re-ran it on a following day. After 1.5 hours, we waited until the male had left the area before remotely triggering the speaker. Following the playback, we continued internal video recording for a further 1.5 hours. We also recorded external video footage during the playback.

Natural stressor Internal CCTV cameras were used to record the incubation stage of pairs in previous years (2014, 2015, 2018). As in our experimental data, these videos were filmed when the clutch was complete, the female was no longer fertile and incubation had started. In six of these videos (recorded at 6±1 days after clutch initiation, starting between 7:00 and 10:40AM) females were subjected to FEPCs (the stressor), a pre-stressor period and a post-stressor period had been filmed, the partner male was absent for the stressor, and both pre- and post-stressor periods contained visits by the partner male. There was no overlap between pairs in the experimental and natural data. This data was processed and coded identically to experimental data.

Video coding Videos were cut into three sections: pre-stressor, stressor and post-stressor, and cuts were then randomly labelled as cut 1, 2 or 3. The coder was thus blind to which treatment they were coding, but the three cuts from each video were always coded by the same coder. In one video from the natural dataset there were two FEPC events in close succession; here, the pre-stressor period refers to before the first FEPC, while the post-stressor period refers to the period following the second FEPC. Experimental and natural videos were coded in BORIS v7.4.6 (Friard & Gamba, 2016) using a detailed behavioural ethogram (see Supplementary Materials Section 2). EM coded the behaviours

‘IN’, ‘LAY’, ‘PEEK’ and ‘FS’ for 22 experimental videos. RH coded the full ethogram for six experimental and six natural videos, and coded ‘CONTACT’, ‘ALLOPREEN’, ‘CHATTER’, ‘SELFPREEN’ and ‘CALL’ in all videos. These behaviours are sometimes subtle and difficult to distinguish. RH had two years of experience recognising and coding these behaviours, hence why they were coded by RH only. For all behaviours coded by both RH and EM, interrater reliability (IRR) was calculated from 92 minutes of overlapping coded data using Cohen’s kappa with a three second time window. See Supplementary Materials Section 2 for further details of the ethogram and IRR results (all values were >0.8, considered to be a high strength of agreement (Landis & Koch, 1977)).

Data extraction All behaviours were measured as durations (in seconds), except for male visit number (measured as a count) and food-sharing (measured as a binary variable per male visit).

The affiliative behaviours measured were contact, allopreen, chatter and time together (see Supplementary Materials Section 2 for a detailed ethogram). Contact, allopreen and time together (analogous to proximity) are often used as measures of affiliation in corvids (Fraser & Bugnyar, 2010a; Logan, Emery, et al., 2013; Logan, Ostojić, et al., 2013). Contact and allopreen, which are non-overlapping behaviours, were summed and treated as one measure of affiliation, which we refer to as ‘direct affiliation’. This is because allopreen was too sparse to model independently (n = 6 across three different pairs), but too important to exclude from the analysis given that it is a direct and active form of affiliation. Only male-initiated direct affiliation was included in analyses. If male-initiated affiliation was preceded by female-initiated affiliation within a single male visit, this behaviour was excluded from analysis. This is because female-initiated affiliation may solicit male affiliation. However, potential female solicitation through affiliation only occurred in one instance (an allopreening event followed by contact). Chatter, a call that pairs often make when engaging in other affiliative behaviours (Chapter 4), was not incorporated in the measure of direct affiliation because (i) its affiliative function is ambiguous (e.g. it sometimes occurred when an individual was alone in the nest-box), and (ii) it was sometimes difficult to ascertain whether it was male- or female-initiated. Time together was modelled separately as a coarser-grained measure of affiliation. Because the female was incubating eggs and thus almost always in the nest-box (females occupied the box for 91.21% ± 5.92 of the video length), time spent together was principally under male control and thus can be interpreted as a male-initiated behaviour in this context.

Statistical analyses Statistical analyses were undertaken using glmmTMB v1.0.1 (Brooks et al., 2017) in R v4.0.2 (R Core Team, 2017). All model tables can be found in Supplementary Materials Section 3.

Sample size and subsets Of the 30 experimental pairs, three pairs were excluded. One pair was excluded due to equipment failure, one was excluded because the male did not enter the nest-box in the pre-stressor treatment (thus meaning we had no control for his within-box behaviour), and one was excluded because the male did not enter the nest-box in the post-stressor treatment.

In all successful experimental trials (n = 27), females left the nest-box upon hearing the playback. Our experimental design was intended to test for changes in the behaviour of males that returned to their distressed partner in the nest-box. However, in nine of 27 cases males arrived at the nest before their partner returned. These cases are analysed and presented separately in Supplementary Materials Section 4. With the removal of pairs where the male returned before the female the final full sample size for analyses presented in the main text was 24 (18 experimental pairs and six natural pairs), except for chatter where two experimental pairs were excluded because audio recording failed.

Fine-scale female behaviours were coded but given that we found no consolation these were not analysed for evidence of solicitation. Instead, we analysed a subset of this data to understand which cues males may have been using to inform their behavioural response. To ensure that males could only be responding to a change in the female’s behaviour, as opposed to other cues (e.g. olfactory cues from an FEPC event, or the cue of the female being outside the nest-box), we analysed data only for females who had been exposed to the experimental stressor and returned to the nest-box before their partner (n = 16). From this subset we also removed data for four females who interacted with their partner outside of the nest-box (see Supplementary Materials Section 5; no affiliative behaviours were observed in these interactions), resulting in a final sample size of 12.

For every analysis incorporating both data types (experimental and natural), we modelled the interaction between data type and treatment. If there was a significant interaction, we subset the data into natural and experimental and present both analyses. Details of all interaction models can be found in Supplementary Materials Section 3.

Model structure and validation Response variables were the summed duration or counts of behaviours per treatment. For durations, values were rounded to the nearest second unless they were <0.5, in which case they were rounded to 1 so as not to create false zeros. Fixed effects varied depending on the model, but treatment (pre-stressor/post-stressor; or in the case of the control data first-half/second-half) was always included, and where applicable pair identity was included as a random effect. AIC of models with different error structures (Poisson and Negative Binomial with linear (‘nbinom1’) and quadratic (‘nbinom2’) parameterisations) were compared and the model with the lowest AIC was selected. In cases where ΔAIC was not >2 between models with different distributions (this was often the case with ‘nbinom1’ and ‘nbinom2’ distributions), model plots were inspected using DHARMa (Hartig, 2020) and the best performing model based on model diagnostics was selected. All final models showed uniformity of residuals and no significant levels of zero-inflation or over-dispersion. Goodness of fit of final models was tested by comparing AIC to a null model (i.e. models with no predictor variables); models with AICs that were lower than null by ≥2 were considered to be better than null (Thomas, 2015). Results presented in the main text are based on models with influential points excluded (where influential points were identified as those that were more than four times the mean Cook’s Distance). See Supplementary Materials Section 3 for all model details, with and without influential points. Note that the exclusion of influential points did not qualitatively change conclusions drawn from results, except for male visit rate at the scale of the whole video, which was significant only after removing one influential point (see Supplementary Materials Section 3); female chatter which was no longer significant after the removal of one influential point (see Supplementary Materials Section 3), and male chatter in the control data, which was significant after the removal of two influential points (see Supplementary Materials Section 6). Diagnostic plots indicated that some generalised mixed models did not perform well on the subset of data where males returned before the female (n = 9); we therefore analysed these subsets with paired t tests or, if the assumption of normally distributed differences between pairs did not hold, Wilcoxon matched pairs signed rank tests (see Supplementary Materials Section 4).

Scales of analysis Direct affiliation, chatter, time together and male vigilance were analysed at the scale of the whole video and the Last Pre-stressor First Post-Stressor male visit (LPFP) scale. Male visit number and male food-sharing were analysed at the scale of the whole video only. Possible female behavioural cues (begging, chatter, self-preen, vigilance and incubation) were analysed at the LPFP scale only. This is because if behavioural changes due to stress do occur, they should be strongest when female stress levels were still high, i.e. in the immediate post-stressor period.

At the scale of the whole video an offset of the time the female spent inside the nest-box (in seconds) was included in models for behaviours that required female presence (direct affiliation, time together, food-sharing). For behaviours that occurred independently of female presence (male chatter, male vigilance, male visit number), video duration was included as the offset. At the LPFP scale we did not include an offset for models of male behaviour. This is because males were free to spend as much time as they chose with the female during these visits: their behaviours were constrained neither by female presence nor video duration. For analyses of female cues, we included an offset of length of the male’s visit.

Control data For 18 of the experimental pairs, internal CCTV footage had been recorded in previous years (2015, 2018) and no FEPC was captured. In these data, as in the experimental/natural data, the female was not fertile and had completed her clutch (videos were filmed 8±3 days post clutch-initiation, starting between 7:37 and 8:37). This data was used as a control to confirm that setting up internal CCTV recording was not the cause of any patterns observed, and ensure that the patterns observed were not present when no stressor occurred (for example as an artefact of changing affiliation levels throughout the day). Control videos were coded by several coders across different years, and coding was completed either in BORIS or manually in Excel. RH re-coded each section of video where the male and female were in the nest-box together. Post-coding, data from each video were trimmed to 10800 seconds and assigned as first-half (up to 5400 seconds) or second-half (5400 – 10800 seconds). Where a behaviour overlapped this split, the split was adjusted to occur after the behaviour had finished. The differences in ‘video length’ of each split were controlled statistically. Analyses of direct affiliation, time together, chatter, vigilance, visit number and food-sharing were conducted on control data at the scale of the whole video. Model details and full results can be found in Supplementary Materials Section 6.

Results

In the natural dataset, females were trapped in the nest-box by the intruding male and remained in the nest-box following the FEPC. In the experimental treatment, all females responded to the stressor by immediately leaving the nest-box (see Supplementary Video 1). They re-entered the nest-box on average 6.55 ± 4.21 minutes later and appeared hesitant upon re-entry (e.g. spending prolonged periods of time outside the box and inspecting the interior of the nest-box before re-entering; see Supplementary Video 2).

For males that returned to the nest-box after their partner had returned, we found no evidence for consolation in either experimental or natural data at the scale of the whole video (i.e. comparing the entire pre-stressor vs post-stressor periods). Instead, in both data types, males significantly decreased direct affiliation (contact and allopreen; see Methods, Supplementary Materials Section 2 and Supplementary Videos 3 and 4), chatter (a call often made between partners during affiliative contact; see Supplementary Materials Section 2 and Supplementary Video 5) and visit rate after the stressor (Figure 1, Table 2). There was also a non-significant tendency for males to spend less time with the female post-stressor (Table 2), but no change in food-sharing rate (Figure 1; Table 2) or vigilance (Table 2). The significant changes detected were not an artefact of naturally occurring behavioural changes throughout the day or a response to the initial video set-up, as control data (data from videos

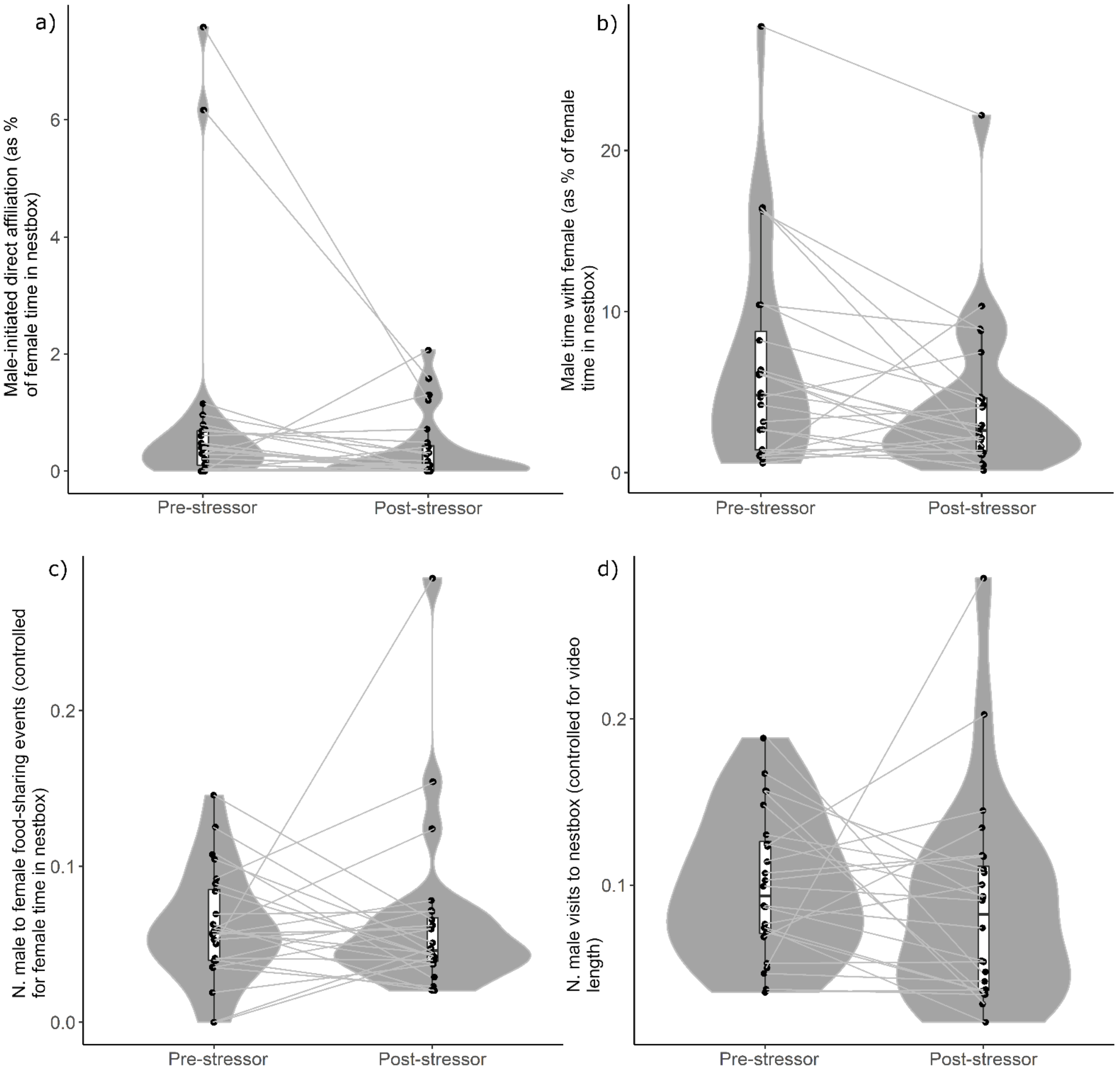


Figure 1 Behaviour changes across the pre- and post-stressor period, at the scale of the whole video. 1a shows the percentage of female time in the nest-box spent in male-initiated direct affiliation; b shows the percentage of female time in the nest-box that the male spent with the female; c shows the number of food-sharing visits made by the male, controlled for female time in the nest-box; d shows the male's visit number, controlled for video length. Grey ties connect the same individuals.

where no stressor occurred, see Methods) found no significant changes in behaviour (see Supplementary Materials Section 5), except from chatter, where control data showed a significant decrease over time after the removal of two influential points (β = -0.62; SE = 0.19; χ^2 = 10.38; p = 0.001).

We also examined behaviours at a finer scale, where data were subsetted to the Last male visit Pre-stressor and the First male visit Post-stressor (LPFP). At this scale, there was no significant change in male-initiated affiliation and chatter (Table 2) and no effect of time since the stressor on direct affiliation (β = -0.0003; SE = 0.0003; χ^2 = 1.49; p = 0.22). In the experimental dataset, males spent significantly less time with the female (see Table 2), while there was no significant change in behaviour in the natural dataset (Table 2). Thus, although male post-stressor responses differed slightly depending on context, in neither case did males console their partners. For males in the natural data there was no change in vigilance at the LPFP scale, while males in the experimental data significantly decreased vigilance in their first post-stressor visit (Table 2).

Despite not consoling their partners, males did change their behaviour post-stressor. We therefore examined fine-scale female behaviour that might be used by males to inform their behavioural change. To do this, we analysed a subset of data where the male could not have responded to any other cues to inform his behaviour, such as olfactory cues left by an intruding male in the natural dataset (see Methods). We found a small but significant decrease in female chatter post-stressor (all seven females who chattered in the last pre-stressor male visit decreased chatter in the first post-stressor male visit, by an average of -5.89% ± 2.48 of male visit time; β = -1.66; SE = 0.73; χ^2 = 5.26; p = 0.02). This was not robust to the removal of one influential point (β = -1.78; SE = 1.02; χ^2 = 3.08; p = 0.08); however, this model was better than a null model containing no predictors ($AIC_{null-full}$ = 4.86). We also detected a change in the rate of calling but not the duration of female begging calls (eight of 12 females increased while only two decreased the number of begging calls made post-stressor; β = 1.98; SE = 0.78; χ^2 = 6.46; p = 0.01; model better than null ($AIC_{null-full}$ = 3.57); see Supplementary Materials Section 3 for all other female behavioural results). These results suggest that there may be subtle alterations in female behaviour post-stressor, and that males may attend to these to inform their own behavioural changes.

For males who returned to the nest-box before the female, no aspect of male behaviour showed a significant pre- versus post-stressor change (see Supplementary Materials Section 4).

Discussion

Corvids are thought to have comparable cognitive abilities to primates (Emery & Clayton, 2004; Güntürkün & Bugnyar, 2016) and exhibit similar social behaviours (e.g.(Adriaense et al., 2019; Emery, 2004; Greggor et al., 2016; Lee, Régli, et al., 2019; Massen et al., 2014)), including third-party affiliation (Fraser & Bugnyar, 2010a; Logan, Emery, et al., 2013; Seed et al., 2007; Sima et al., 2018). Furthermore, jackdaw pairs form lifelong bonds and, unlike most birds, are almost entirely genetically monogamous (Chapter 2; Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021), despite FEPCs occurring during the breeding season (see Supplementary Materials Section 1, Chapter 2 and Gill et al., 2020). They consequently have some of the highest levels of fitness interdependence between mated partners within the animal kingdom. The pair-bond thus represents the most valuable bond in jackdaw society, in terms of fitness consequences (Kummer, 1978). Following arguments in the field that individuals should actively manage and maintain valuable relationships (de Waal & Aureli, 1997; Kummer, 1978), we predicted that jackdaws would show consolation toward their distressed partner, as a mechanism through which to maintain their valuable bond (Adriaense et al., 2020; de Waal & Aureli, 1997; Kummer, 1978; Pérez-Manrique & Gomila, 2018). However, when controlling for the potential confounds introduced by previous studies, we found no evidence that wild jackdaws console their distressed partners in either an experimental or natural context.

After their partner had experienced a stressor, male jackdaws generally responded by decreasing affiliation towards their partner and reducing visit rates. We suggest that rather than consoling their partners, males may instead have responded to female distress with a form of self-protection. Given that the stressor was unknown to the male, decreased visit rate may be a generalised response to a potential threat within a confined space (the nest-box); for example, the threat of escalated conflict with an intruding male (Gill et al., 2020). Males, who do not incubate the eggs but must feed the incubating female so that she does not leave the eggs (Röell, 1978), may be able to decrease visits to the nest-box during the incubation stage without directly influencing fitness consequences, as long as they maintain food-sharing rates.

Table 2 Each row represents a separate model and shows the estimate, standard error, χ^2 value, confidence intervals and p-value for the effect of treatment (post-stressor relative to pre-stressor) on the response variable of interest. All models included pair identity as a random effect and an offset of either video duration or female time in the box (see Methods for details). Full model details (without and with influential points) can be found in the Supplementary Materials (SM) Section 3. LPFP is the Last Pre-stressor/First Post-stressor male visit. Where the response variable is followed by (natural) or (experimental), a significant interaction of data type was observed (see Supplementary Materials Section 3), and the data was therefore analysed separately. Bold indicates significant results.

	Response variable	Prediction (see Table 1)	number of pairs	Post-stressor relative to pre-stressor						Model better than null
				Estimate	SE	χ^2	CI (2.5%)	CI (97.5%)	P-value	
Whole video	**Male-initiated direct affiliation**	**1a**	**23**	**-0.818**	**0.311**	**6.900**	**-1.428**	**-0.208**	**0.009****	**Y**
	Male chatter	**1d**	**21**	**-1.319**	**0.506**	**6.792**	**-2.310**	**-0.330**	**0.009****	**Y**
	Time together	1f	23	-0.434	0.229	3.594	-0.883	0.015	0.058	Y
	Male vigilance	2a	24	-0.544	0.288	3.565	-1.109	0.021	0.059	Y
	Male visit number	**2c**	**23**	**-0.270**	**0.137**	**3.866**	**-0.539**	**-0.001**	**0.049***	**Y**
	Male food-sharing	2d	22	-0.080	0.184	0.187	-0.440	0.281	0.666	Y
LPFP	Male-initiated direct affiliation	1b	23	-0.656	0.484	1.840	-1.604	0.292	0.175	N
	Time together (natural)	1g	5	0.537	0.317	2.877	-0.084	1.158	0.090	N
	Time together (experimental)	**1g**	**16**	**-1.101**	**0.285**	**14.960**	**-1.658**	**-0.543**	**<0.001*****	**Y**
	Male vigilance (natural)	2b	6	0.932	0.676	1.903	-0.392	2.256	0.168	N
	Male vigilance (experimental)	**2b**	**17**	**-0.899**	**0.454**	**3.932**	**-1.788**	**-0.010**	**0.047***	**Y**

Our results therefore suggest that males may have responded to post-stressor cues by strategically adjusting their behaviour so as to minimise their own exposure to a potential threat while maintaining behaviours that have a direct impact on reproductive fitness (food-sharing). An alternative explanation is that males decreased visits and affiliation with the female in order to invest in increased mate-guarding outside of the nest-box; however, we did not observe any behaviours to suggest this was the case (see Supplementary Materials Section 5).

Our results indicate that despite not consoling them, males do attend to subtle behavioural cues from their partner and use these to inform their behaviour. For example, males may have used subtle changes in female calling behaviour, such as decreased chattering or increased frequency of begging, as cues for behavioural change. However, the directionality of this result is unclear, and female behaviour may instead have changed in response to male behaviour. Other subtle cues that we did not measure, such as breathing rate(Carere & Van Oers, 2004), may also have been detected by the male. It is possible that male responses may have been mediated through emotional contagion (emotional state-matching between individuals (de Waal & Preston, 2017)), which is sometimes argued to be a form of empathy (de Waal & Preston, 2017); however, our current data do not allow us to address this. In future, the use of non-invasive methods to quantify physiological stress-states (e.g.(Ikkatai & Watanabe, 2015; Jerem et al., 2018)) may allow researchers to determine whether the stress responses of individuals mirror those of social partners that have experienced a stressor.

There are three potential explanations for the absence of consolation in this study. The first is that jackdaws do not engage in consolation. We found a significant and similar change in male behaviour following the female’s exposure to both a severe natural stressor, where the female was engaged in direct, violent conflict with another individual, and an experimental stressor, where the female left the nest-box immediately after the stressor and showed hesitancy upon return. This implies that both stressors elicited a similar response in females. Nevertheless, although males did alter their behaviour toward their partners post-stressor, we found no evidence to support the occurrence of consolation. This is at odds with previous studies on captive corvids, where consolation-like behaviours have been observed in post-conflict contexts (e.g.(Fraser & Bugnyar, 2010a; Logan, Emery, et al., 2013)). The absence of consolation in our study, where key confounds are controlled, therefore raises the possibility that the consolatory behaviours observed in captive populations may differ, in terms of their proximate underpinnings, or ultimate function, from true consolatory behaviour. To rule out alternative explanations of consolation, future work would benefit from explicitly addressing potential confounds and measuring the physiological stress state of study subjects. Until we can rule out alternative explanations and build a more robust understanding of the taxonomic distribution of consolation, claims of convergent socio-cognitive evolution (Güntürkün & Bugnyar, 2016) must be interpreted with caution. Given that the jackdaw pair-bond is arguably one of the most 'high value' (Kummer, 1978) relationships in the animal kingdom, the apparent absence of consolation seems to contradict theoretical predictions. Our results therefore raise fundamental questions as to whether predictions about where we should expect non-human sympathetic and empathetic concern to occur are unduly influenced by an anthropomorphic view of how animals should manage social relationships (Barrett, 2017). More broadly, our results chime with concerns that our current predictions in the field of animal sociality and cognition may be inappropriately influenced by anthropomorphic perspectives (Barrett, 2017; Heyes, 2012; Penn, 2011; Shettleworth, 2010a).

A second potential explanation of our results is that although consolation was not detected at the population-level, it may occur in some jackdaws but be highly inter-individually variable (Boogert et al., 2018; Thornton & Lukas, 2012). Jackdaws have been found to show substantial inter-individual variability in socio-cognitive behaviour (Lee, McIvor, et al., 2019), and it is possible that although the majority of males do not offer consolation to their distressed partner, a minority do. To further interrogate this hypothesis, an experimental design with repeated measures of individual responses to partner distress would be necessary (Boogert et al., 2018). Although future work in the study of consolation may benefit from such an experimental design, the value of understanding consolation beyond the population-level must be carefully weighed with welfare costs of repeated stressor exposure.

A final, non-mutually exclusive, explanation of our results is that jackdaws do not console in this specific context. We found that at the population-level and in an ecologically relevant setting, male jackdaws do not console their stressed partner and that this was consistent across both experimental and natural datasets. Nevertheless, we cannot eliminate the possibility that males may console their partners in other contexts. Indeed, in the few experimental cases where males returned to the nest-box before rather than after the female returned, males appeared to show a different pattern of behaviour (see Supplementary Materials Section 4). While these cases still provided no evidence for consolation, they demonstrate that even subtle differences in context can have detectable effects. Thus, in addition to more robust methodologies, we suggest that the field would benefit from testing for consolation across multiple ecologically relevant contexts within the same study species. Given that different contexts generate distinct trade-offs at an individual level, to understand a behaviour's ultimate function we require theory that incorporates the context-specific costs and benefits incurred by the individual performing the behaviour (Mcauliffe & Thornton, 2015; McNamara, 2013). For example, in the context of this study, for males who detect stress in their partner without knowing the source of the stressor, the fitness costs of staying in a potentially dangerous location to console the female may outweigh any benefits gained through offering consolation. Formal theoretical approaches that evaluate the adaptive value of responding to another individual's state while incorporating ecologically relevant trade-offs would form the basis for a more robust predictive framework than verbal arguments alone (McNamara & Leimar, 2020; Smaldino, 2020), and we suggest such models would be invaluable in furthering the field. Together, implementing robust methodologies that explicitly control for common confounds and formalising our predictions as to where and when consolation should be ultimately advantageous will allow for a deeper understanding of non-human sympathy and empathy.

Supplementary Material

SECTION 1: Forced extra-pair copulations (FEPCs) in jackdaws

SECTION 2: Behavioural ethogram

SECTION 3: Main text model details

SECTION 4: Analysis of cases where male returns before the female

SECTION 5: External data

SECTION 6: Analysis of control data

SECTION 1: Forced extra-pair copulations (FEPCs) in jackdaws

FEPCs are unsolicited copulations that are actively resisted by females. Internal nest-box video footage has revealed that FEPCs occur in jackdaw populations, both in our three study populations in Cornwall, UK (Chapter 2) and in a study population in Bavaria, Germany (Gill et al., 2020). FEPCs occur when the female is alone at the nest-box in both the fertile and post-fertile phase of the breeding season. However, jackdaws are almost entirely genetically monogamous (Chapter 2; Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021). In all populations, the majority of females responded to an intruding male with active resistance through nest-box defense (rushing to the nest-box entrance at the sound of a possible intruder) and direct attack (pecking and kicking). For example, in 97% of FEPC intrusions in the Bavarian population (n = 73), females defended the nest-box against the intruder, and in the two cases where this did not occur the female was either sleeping or had her back to the entrance when the intruder approached the box (Gill et al., 2020). In the Cornish populations, females engaged in nest-box defense and/or direct attack of the intruding male in every case (Chapter 2). FEPCs and the associated conflict may lead to injury or mortality of eggs/chicks and/or the female (Chapter 2; Gill et al., 2020). FEPCs are therefore highly likely to induce stress in female jackdaws.

SECTION 2: Behavioural ethogram

Table S1 The behavioural ethogram used in BORIS v7.4.6 to code video data. IRR is Inter-Rater Reliability calculated using Cohen's Kappa with a 3 second time window. Both coders were blind to treatment (pre- or post-stressor

Behaviour	**Type**	**Description**	**Coder**	**IRR**
ALLOPREEN	Duration	One individual preens another (see SM Video 4)	RH	NA
CONTACT	Duration	An individual is stood or sat close enough to their partner that they would not have to move their bodies in order to make physical contact (i.e. they are within a beak's distance of one another). They are not actively engaged in any other behaviour (see SM Video 3)	RH	NA
CHATTER	Duration	A specific call often made between partners (see SM Video 5)	RH	NA
IN	Duration	An individual is in the box (time together is extracted from this variable)	EM/RH	0.99
FS	Point	Food-sharing between adults. In the incubation stage this is always from the male to the female. Coded only once per male visit to the box	EM/RH	1
PEEK	Duration	Individual looks out of nestbox	EM/RH	0.86
SELFPREEN	Duration	Adult preens self	RH	NA
LAY	Duration	Individual sitting on eggs (start behaviour when the individual 'wiggles' onto the eggs)	EM/RH	0.99
BEG	Duration	Female begs for food	RH	NA
CONTACT CALL	Duration	Individual makes a contact call	RH	NA
OTHER CALL	Duration	Individual makes an unknown or ambiguous call	RH	NA

SECTION 3: Main text model details

Presented here are the full details of each model presented in the main text. For each model, an interaction between data type (experimental and natural) and treatment was investigated first. If this was significant (α < 0.05), the data was split into experimental and natural for further analysis. If it was non-significant, data were analysed together. For each model, the error structure with the lowest AIC of “Poisson”, “nbinom1” (negative binomial with a linear parameterisation) and “nbinom2” (negative binomial with a quadratic parameterisation), were selected. These three error structures were investigated because the response variable is a count (number of seconds for which the behavior occurred) in a given time period (video duration/length of time female in box; included as an offset). The models presented here are the final i) interaction and ii) main effects model (with and without influential points) per prediction. Where models with influential points are not presented, no influential points were identified in the main effects model. No models were overdispersed or zero-inflated. Note: ‘n.obs’ in the model tables is the number of observations, but this should be halved to obtain the true sample size as there were two observations per pair (one pre-stressor, one post-stressor) for all models except those in 1c.

Prediction 1a: There will be an increase in male-initiated direct affiliation post-stressor (whole video)

Interaction model

```
model1 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
                TREATMENT*DATA_TYPE +
                VIDEO_STARTS + offset(log(F_IN_DUR)) # F_IN_DUR = duration female is in the box
              + (1|PAIR), family = nbinom1, data = whole_video_male_last)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-4.368 **	(1.451)
Treatment (post-stressor)	-0.821 *	(0.324)
Data type (natural)	-0.170	(0.679)
Video starts	-0.115	(0.146)

Treatment (post-stressor) * Data type (natural)	-0.479	(0.670)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model4 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
          TREATMENT + DATA_TYPE +
          VIDEO_STARTS + offset(log(F_IN_DUR) + (1|PAIR), family = nbinom1, data =
          whole_video_male_last)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-4.322 **	(1.397)
Treatment (post-stressor)	-0.880 **	(0.330)
Data type (natural)	-0.345	(0.621)
Video starts	-0.112	(0.141)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-4.669 ***	(1.273)
Treatment (post-stressor)	-0.818 **	(0.311)
Data type (natural)	-1.043	(0.653)
Video starts	-0.078	(0.127)
n.obs	46	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1b: There will be an increase in male-initiated direct affiliation post-stressor (LPFP)

Interaction model

```
model11 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
  TREATMENT*DATA_TYPE + VIDEO_STARTS + (1|PAIR),
```

	Estimate (significance)	Standard Error
Intercept (mean)	1.487	(1.502)
Treatment (post-stressor)	-0.758	(0.534)
Data type (natural)	-0.534	(0.892)
Video starts	0.090	(0.139)
Treatment (post-stressor) * Data type (natural)	-0.188	(1.337)
n.obs	48	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Main effects model

```
model14 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
                TREATMENT + DATA_TYPE + VIDEO_STARTS + (1|PAIR),
                family = nbinom1, data = LPFP_after)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	1.504	(1.496)
Treatment (post-stressor)	-0.788	(0.492)
Data type (natural)	-0.601	(0.763)
Video starts	0.089	(0.139)
n.obs	48	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	0.459	(1.654)
Treatment (post-stressor)	-0.656	(0.484)
Data type (natural)	-0.525	(0.816)
Video starts	0.124	(0.153)
n.obs	46	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1c: Male-initiated direct affiliation will negatively correlate with time since the stressor occurred (FP)

Interaction model

```
model17 <- glmmTMB(DIRECT  AFFILIATION  DURATION ~
  TOGETHER  LATENCY_POST_PLAYBACK*DATA_TYPE +
VIDEO_STARTS,
  data = LPFP_post_after,family = nbinom1)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-1.369	(1.927)
Time until pair reunite	-0.0004	(0.0003)
Data type (natural)	1.035	(1.592)
Video starts	0.309	(0.193)
Time until pair reunite*Data type (natural)	-0.008	(0.008)
n.obs	24	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model22 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
            TOGETHER_LATENCY_POST_PLAYBACK + DATA_TYPE +
            VIDEO_STARTS,
           data = LPFP_post_after,family = nbinom1)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-1.134	(1.885)
Time until pair reunite	-0.0004	(0.0003)
Video starts	0.290	(0.189)
Data type (natural)	-0.716	(1.171)
n.obs	24	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1d: There will be an increase in male chatter post-stressor (whole video)

Note: chatter at the scale of the whole video is not necessarily male-initiated, as it is was often difficult to ascertain whether the male or the female initiated a chatter bout.

Interaction model

```
model23 <- glmmTMB(MALE_CHATTER_DURATION ~ TREATMENT*DATA_TYPE +
VIDEO_S
TARTS + (1|PAIR)+ offset(log(VIDEO_DURATION)), data = whole_video_Nona_last, family =
```

	Estimate (significance)	Standard Error
Intercept (mean)	-5.493 **	(2.077)
Treatment (post-stressor)	-0.842	(0.614)
Video starts	-0.194	(0.220)
Data type (natural)	0.806	(0.796)
Treatment (post-stressor)*Data type (natural)	-0.413	(0.980)
n.obs	44	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model26 <- glmmTMB(MALE_CHATTER_DURATION ~ TREATMENT +
VIDEO_STARTS + DAT
A_TYPE + (1|PAIR) + offset(log(VIDEO_DURATION)), data = whole_video_Nona_last,
family = nbi
nom1)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-5.411 **	(2.079)
Treatment (post-stressor)	-1.007 *	(0.482)
Video starts	-0.199	(0.221)
Data type (natural)	0.675	(0.743)
n.obs	44	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	3.320	(2.250)
Treatment (post-stressor)	-1.319 **	(0.506)
Video starts	-0.213	(0.235)
Data type (natural)	0.480	(0.821)
n.obs	42	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1e: There will be an increase in male chatter post-stressor (LPFP)

Note: chatter at this scale was analysed at a finer scale than at the whole video. The body movements of the individuals were scrutinized to try and ascertain which individual initiated a chatter bout. There were only 5 instances of male-initiated chatter, all in the pre-stressor period.

Prediction 1f: There will be an increase in time the male spends with the female post-stressor (whole video)

Interaction model

```
model33 <- glmmTMB(TIME_TOGETHER ~ TREATMENT*DATA_TYPE + VIDEO_STARTS +
offset(log(F_IN_DUR)) + (1|PAIR), family = nbinom2, data = whole_video_male_last)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-1.053	(0.692)
Treatment (post-stressor)	-0.431	(0.259)
Data type (natural)	0.553	(0.397)
Video starts	-0.212 **	(0.067)
Treatment (post-stressor) * Data type (natural)	0.077	(0.516)
n.obs	48	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Main effects model

```
model36 <- glmmTMB(TIME_TOGETHER ~ TREATMENT + DATA_TYPE + VIDEO_STARTS +
offset(log(F_IN_DUR)) + (1|PAIR), family = nbinom2, data = whole_video_male_last)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-1.063	(0.689)
Treatment (post-stressor)	-0.412	(0.225)
Data type (natural)	0.591	(0.305)
Video starts	-0.212 **	(0.067)
n.obs	48	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-1.100	(0.614)
Treatment (post-stressor)	-0.434	(0.229)
Data type (natural)	0.310	(0.294)
Video starts	-0.206 ***	(0.061)
nobs	46	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Prediction 1g: There will be an increase in time the male spends with the female post-stressor (LPFP)

Interaction model

```
model38 <- glmmTMB(TIME_TOGETHER ~ TREATMENT*DATA_TYPE +
VIDEO_STARTS + (1|PAIR), family = nbinom2, data = LPFP_after)
```

	Estimate (significance)	Standard Error
Intercept (mean)	5.994 ***	(0.921)
Treatment (post-stressor)	-1.141 ***	(0.325)
Video starts	-0.160	(0.089)
Treatment (post-stressor)*Data type (natural)	1.597 **	(0.613)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model: natural data

```
# natural
model41 <- glmmTMB(TIME_TOGETHER ~ TREATMENT + VIDEO_STARTS + (1|PAIR), family =
nbinom2, data = LPFP_after_nat)
```

	Estimate (significance)	Standard Error
Intercept (mean)	6.291 **	(1.970)
Treatment (post-stressor)	0.537	(0.324)
Video starts	-0.353	(0.236)
n.obs	12	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	4.722 ***	(1.058)

Treatment (post-stressor)	0.537	(0.317)
Video starts	-0.202	(0.126)
n.obs	10	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model: experimental data

```
# experimental
model44 <- glmmTMB(TIME_TOGETHER ~ TREATMENT + VIDEO_STARTS + (1|PAIR), family =
nbinom2, data = LPFP_after_exp)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	5.850 ***	(1.000)
Treatment (post-stressor)	-1.163 ***	(0.348)
Video starts	-0.141	(0.097)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	5.059 ***	(0.729)
Treatment (post-stressor)	-1.101 ***	(0.285)
Video starts	-0.088	(0.070)
n.obs	34	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2a: There will be a decrease in male vigilance post-stressor (whole video)

Interaction model

```
model47 <- glmmTMB(MALE_PEEK_DURATION ~ TREATMENT*DATA_TYPE + VIDEO_STARTS +
offset(log(VIDEO_DURATION))
```

	Estimate (significance)	Standard Error
Intercept (mean)	-2.057	(1.148)
Treatment (post-stressor)	-0.612	(0.331)
Data type (natural)	1.143	(0.614)
Video starts	-0.278 *	(0.116)
Treatment (post-stressor) * Data type (natural)	0.263	(0.646)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model50 <- glmmTMB(MALE_PEEK_DURATION ~ TREATMENT + DATA_TYPE + VIDEO_STAR
TS + (1|VIDEO_DURATION), family = nbinom2, data = whole_video_male_last)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-2.077	(1.146)
Treatment (post-stressor)	-0.544	(0.288)
Data type (natural)	1.276 *	(0.525)
Video starts	-0.279 *	(0.116)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2b: There will be a decrease in male vigilance post-stressor (LPFP)

Interaction model

```
model56 <- glmmTMB(MALE_PEEK_DURATION ~ TREATMENT*DATA_TYPE + VIDEO_STAR
TS + (1|PAIR), family = nbinom2, data = LPFP_after)
```

	Estimate (significance)	Standard Error
Intercept (mean)	6.530 ***	(1.480)
Treatment (post-stressor)	-1.350 **	(0.511)
Data type (natural)	-0.791	(0.776)
Video starts	-0.369 *	(0.145)
Treatment (post-stressor) * Data type (natural)	2.157 *	(0.919)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model: natural data

```
model59 <- glmmTMB(MALE_PEEK_DURATION ~ TREATMENT +
VIDEO_STARTS + (1|PAIR), family = nbinom2, data = LPFP_after_nat)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	7.770 **	(2.892)
Treatment (post-stressor)	0.932	(0.676)
Video starts	-0.629	(0.353)
n.obs	12	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model: experimental data

```
# experimental
model62 <- glmmTMB(MALE_PEEK_DURATION ~
TREATMENT + VIDEO_STARTS + (1|PAIR), family = nbinom2, data = LPFP_after_exp)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	6.269 ***	(1.650)
Treatment (post-stressor)	-1.382 *	(0.553)

Video starts	-0.337 *	(0.161)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	5.219 ***	(1.455)
Treatment (post-stressor)	-0.899 *	(0.454)
Video starts	-0.274 *	(0.138)
n.obs	34	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2c: There will be a decrease in visit number post-stressor (whole video)

Interaction model

	Estimate (significance)	Standard Error
Intercept (mean)	-5.849 ***	(0.433)
Treatment (post-stressor)	-0.135	(0.155)
Data type (natural)	0.219	(0.199)
Video starts	-0.124 **	(0.045)
Treatment (post-stressor) * Data type (natural)	-0.257	(0.293)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model68 <- glmmTMB(M_IN ~ TREATMENT + DATA_TYPE +
offset(log(VIDEO_DURATION))
             + (1|PAIR), family = poisson, data = whole_video_male_last)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-5.811***	(0.430)
Treatment (post-stressor)	-0.208	(0.131)
Data type (natural)	0.109	(0.157)
Video starts	-0.125 **	(0.045)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-5.938 ***	(0.436)
Treatment (post-stressor)	-0.270 *	(0.137)
Data type (natural)	0.177	(0.160)
Video starts	-0.114 *	(0.045)
n.obs	46	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2d: There will be a decrease in male food-sharing post-stressor (whole video)

```
model74 <- glmmTMB(M_FS ~ TREATMENT*DATA_TYPE + VIDEO_STARTS + offset(log(F_IN_DUR)) +
(1|PAIR), family = poisson, data = whole_video_male_last)
```

Interaction model

	Estimate (significance)	Standard Error
Intercept (mean)	-7.072 ***	(0.505)
Treatment (post-stressor)	-0.002	(0.201)
Data type (natural)	0.349	(0.256)
Video starts	-0.047	(0.051)

Treatment (post-stressor)*Data type (natural)	-0.228	(0.365)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Main effects model

```
model74 <- glmmTMB(M_FS ~ TREATMENT + DATA_TYPE + + VIDEO_STARTS +
offset(log(F_IN_DUR)) + (1|PAIR), family = poisson, data = whole_video_male_last)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-7.034 ***	(0.501)
Treatment (post-stressor)	-0.072	(0.168)
Data type (natural)	0.244	(0.197)
Video starts	-0.048	(0.051)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points model

	Estimate (significance)	Standard Error
Intercept (mean)	-7.520 ***	(0.528)
Treatment (post-stressor)	-0.080	(0.184)
Data type (natural)	0.455 *	(0.210)
Video starts	-0.014	(0.052)
n.obs	48	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 3a: Females will change calling behaviour post-stressor (LPFP)

Begging: Duration

```
model90 <- glmmTMB(BEG_F ~ TREATMENT + VIDEO_STARTS + offset(log(TIME_TOG)) + (1|PAIR),
data = calls_after_exp, family = nbinom2)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-3.047	(2.056)
Treatment (post-stressor)	1.294	(0.836)
Video starts	0.082	(0.193)
n.obs	24	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	5.872	(5.869)
Treatment (post-stressor)	1.544	(0.873)
Video starts	-0.946	(0.652)
n.obs	22	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Begging: Count

```
model92 <- glmmTMB(BEG_COUNT_F ~ TREATMENT + VIDEO_STARTS +
offset(log(TIME_TOG)) + (1|PAIR), data = calls_after_exp, family = nbinom2)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-4.270 *	(1.996)
Treatment (post-stressor)	1.874 *	(0.755)
Video starts	0.128	(0.188)
n.obs	24	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error

Intercept (mean)	2.469	(4.245)
Treatment (post-stressor)	1.982 *	(0.780)
Video starts	-0.637	(0.458)
n.obs	22	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Chatter: Duration

```
model97 <- glmmTMB(CHATTER_F ~ TREATMENT + VIDEO_STARTS +
offset(log(TIME_TOG)) + (1|PAIR), data = calls_after_exp, family = poisson)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-5.802 *	(2.881)
Treatment (post-stressor)	-1.663 *	(0.725)
Video starts	0.196	(0.291)
n.obs	24	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	22.622	(16.068)
Treatment (post-stressor)	-0.933	(1.043)
Video starts	-3.084	(1.845)
n.obs	22	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 3b: Females will decrease incubation post-stressor (LPFP)

Only 7 values are not 100%: 5 females do not incubate for 100% of the male's last pre-stressor visit and 2 females do not incubate for 100% of the male's first post-stressor visit.

Prediction 3c: Females will increase vigilance post-stressor (LPFP)

There is only one case where the female peeks while the male is also in the nest-box and it is in the last male visit pre-stressor.

Prediction 3d: Females will increase self-preening post-stressor (LPFP)

There are no cases of female self-preening while the male is also in the nest-box.

SECTION 4: Analysis of nine cases where male returns before the female

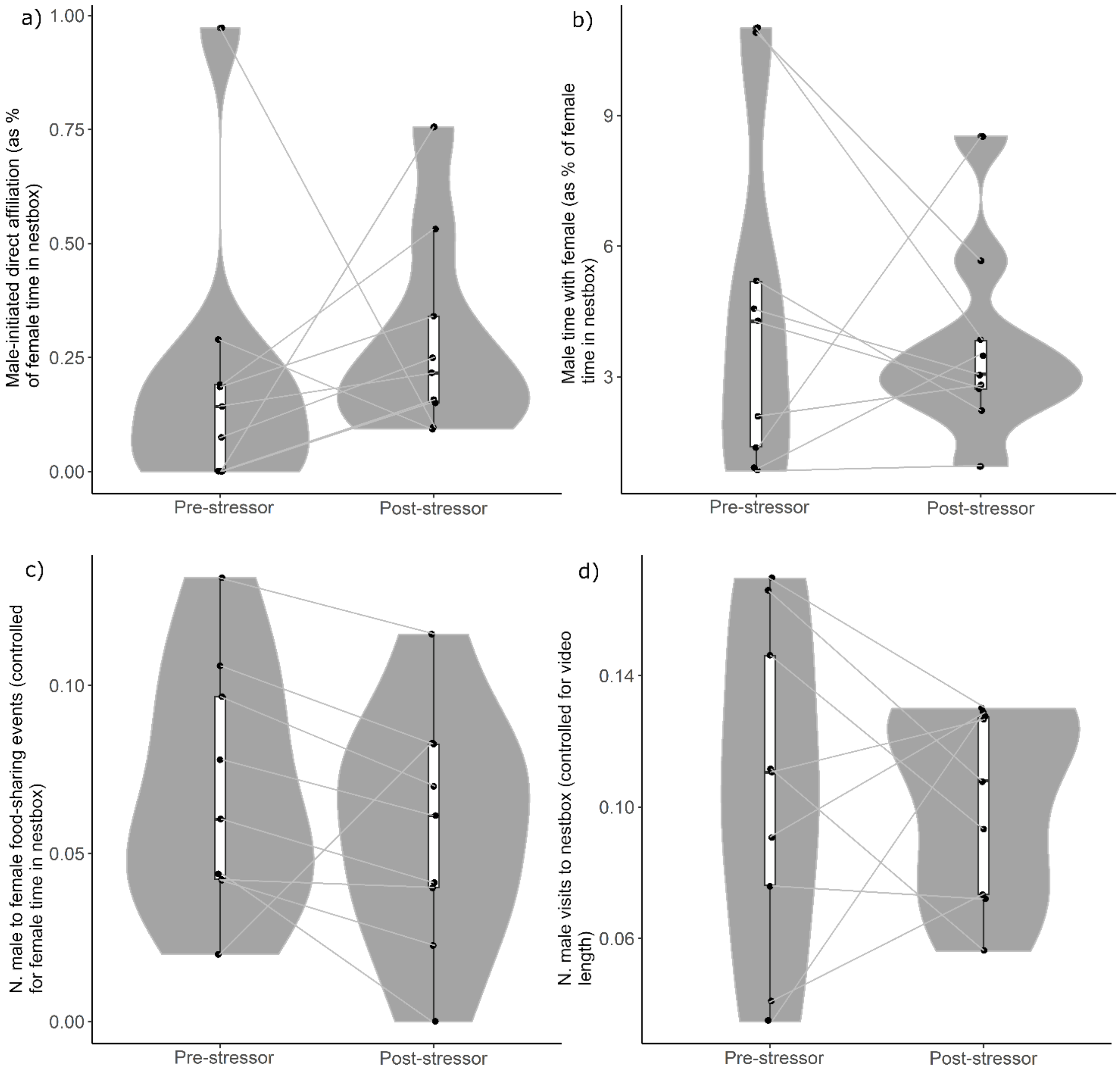


Figure S1 Behaviour changes across the pre- and post-stressor period, at the scale of the whole video. 1a shows the percentage of female time in the nest-box spent in male-initiated direct affiliation; b shows the percentage of female time in the nest-box that the male spent with the female; c shows the number of food- sharing visits made by the male, controlled for female time in the nest-box; d shows the male's visit number, controlled for video length. Grey ties connect the same individuals.

In the experimental data there were nine cases where the male returned to the nest-box before the female returned post-stressor. This subset of males responded differently to males who returned when the female was back in the box: no behaviour showed a significant pre- to post-stressor change (see Tables below). Whether males returned before or after the female did not correlate with the length of the female's absence from the nest-box ($\beta = 0.03$, SE = 0.27, $\chi^2 = 0.01$, $p = 0.91$), so this different pattern in results (relative to cases where the male returned after the female) is unlikely to be associated with the intensity of the female's stress.

Prediction 1a: There will be an increase in male-initiated direct affiliation post-stressor (whole video)

```
model101 <- glmmTMB(DIRECT_AFFILIATION_DURATION~
                TREATMENT +  VIDEO_STARTS + offset(log(F_IN_DUR)) + (1|PAIR),
                family = nbinom1, data = whole_video_male_first)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	3.553	(2.453)
Treatment (post-stressor)	-0.546	(0.900)
Video starts	-0.292	(0.234)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

One influential point was identified in this model, but diagnostic plots indicated that the model with this influential point removed did not perform well, even when the model was simplified by removing "Video_starts". We therefore used a paired t-test with standardised direct affiliation (seconds of direct affiliation/female duration in box) as the response and treatment (before- or post-stressor) as the predictor. The relationship was not significant ($t = 2.117$, df = 7, $p = 0.072$).

Prediction 1b: There will be an increase in male-initiated direct affiliation post-stressor (LPFP)

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	3.553	(2.453)
Treatment (post-stressor)	-0.546	(0.900)
Video starts	-0.292	(0.234)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1c: Male-initiated direct affiliation will negatively correlate with time since the stressor occurred (FP)

```
model107 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
                TOGETHER_LATENCY_POST_PLAYBACK + VIDEO_STARTS,
            data = LPFP_post_first,family = nbinom1)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-1.025	(3.368)
Time until pair reunite	0.002	(0.001)
Video starts	0.025	(0.312)
n.obs	9	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1d: There will be an increase in male chatter post-stressor (whole video)

```
model108 <- glmmTMB(MALE_CHATTER_DURATION  ~
                TREATMENT + VIDEO_STARTS + + offset(log(F_IN_DUR)) + (1|PAIR),
            data = whole_video_Nona_first ,family = nbinom1)
```

Model diagnostic plots showed that this model performed poorly; we therefore used a Wilcoxon test (given non-normality of differences between pairs) with standardised chatter (seconds of chatter/video length) as the response variable and treatment (before- or post-stressor) as the predictor. The relationship was non-significant (V = 6, p = 0.402).

Prediction 1e: There will be an increase in male chatter post-stressor (LPFP)

There were only 4 cases of male-initiated chatter (all in the post-stressor visit)

Prediction 1f: There will be an increase in time the male spends with the female post-stressor (whole video)

```
model111 <- glmmTMB(TIME_TOGETHER  ~
                TREATMENT + VIDEO_STARTS + + offset(log(F_IN_DUR)) + (1|PAIR),
              data = whole_video_male_first ,family = nbinom2)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-1.750 *	(0.703)
Treatment (post-stressor)	-0.211	(0.305)
Video starts	-0.126 *	(0.061)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1g: There will be an increase in time the male spends with the female post-stressor (LPFP)

```
model114 <- glmmTMB(TIME_TOGETHER  ~
                TREATMENT*DATA_TYPE + VIDEO_STARTS + (1|PAIR),
              data = LPFP_first ,family = nbinom2)
```

	Estimate (significance)	Standard Error
Intercept (mean)	4.868 ***	(1.146)
Treatment (post-stressor)	0.376	(0.408)
Video starts	-0.147	(0.096)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2a: There will be a decrease in male vigilance post-stressor (whole video)

Full model

```
model117 <- glmmTMB(MALE_PEEK_DURATION  ~
                TREATMENT + VIDEO_STARTS + offset(log(VIDEO_DURATION)) +
                (1|PAIR),
              data = whole_video_male_first ,family = nbinom2)
```

	Estimate (significance)	Standard Error
Intercept (mean)	-1.188	(1.180)
Treatment (post-stressor)	-0.064	(0.510)
Video starts	-0.268 **	(0.100)
n.obs	18	

Influential points removed

One influential point was identified in this model, but diagnostic plots indicated that the model with this influential point removed did not perform well, even when the model was simplified by removing "Video_starts". We therefore used a paired t-test with standardised vigilance (seconds of vigilance/video duration) as the response and treatment (before- or post-stressor) as the predictor. The relationship was not significant (t = 0.235, df = 7, p-value = 0.821).

Prediction 2b: There will be a decrease in male vigilance post-stressor (LPFP)

```
model120 <- glmmTMB(MALE_PEEK_DURATION  ~
               TREATMENT + VIDEO_STARTS + (1|PAIR),
             data = LPFP_first,,family = nbinom2)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	5.714 ***	(1.632)
Treatment (post-stressor)	1.099	(0.568)
Video starts	-0.373 **	(0.138)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2c: There will be a decrease in visit number post-stressor (whole video)

```
model127 <- glmmTMB(M_IN  ~
                TREATMENT + VIDEO_STARTS + offset(log(VIDEO_DURATION)) +
                (1|PAIR),
              data = whole_video_male_furst ,family = poisson)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-6.318 ***	(0.439)
Treatment (post-stressor)	-0.006	(0.196)
Video starts	-0.052	(0.039)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2d: There will be a decrease in male food-sharing post-stressor (whole video)

```
model127 <- glmmTMB(M_FS  ~
                TREATMENT + VIDEO_STARTS + offset(log(F_IN_DUR)) + (1|PAIR),
              data = whole_video_male_furst ,family = poisson)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-6.014 ***	(0.590)
Treatment (post-stressor)	-0.164	(0.265)
Video starts	-0.119 *	(0.055)
n.obs	18	

*** p < 0.001; ** p < 0.01; * p < 0.05.

SECTION 5: External data

Alongside internal nest-box footage, we filmed the nest-box exterior from the playback until the female returned (e.g. see Supplementary Video 2 for a clip of this footage). Paired with the internal nest-box footage, this was used as independent verification of nest-box identity. This footage shows that in four cases where the male entered the nest-box after the female and in four cases where the male entered first, the partners interacted outside of the box before an individual entered. Some interesting behavior was observed in this footage. For example, when females displayed hesitance to re-enter the box, males sometimes also appeared hesitant (e.g. delaying entering despite carrying nest material; inspecting the interior of the box from outside). Future studies of a similar experimental design may benefit from incorporating external video footage. In addition to filming externally, we took notes of external behaviours observed from the hide throughout the experimental period; however, this was difficult to record in a standardised way. Nonetheless, we did not observe any evidence to suggest that males increased external nest-box defense or mate-guarding post-stressor.

SECTION 6: Analysis of control data

Prediction 1a: There will be an increase in male-initiated direct affiliation post-stressor (whole video)

```
model7 <- glmmTMB(DIRECT_AFFILIATION_DURATION ~
                 TREATMENT + VIDEO_STARTS + offset(log(F_IN_DUR))
                + (1|PAIR), zi = ~1, family = nbinom1, data = control)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-13.384 **	(4.160)
Treatment (second-half)	0.145	(0.382)
Video starts	1.081 *	(0.527)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-14.452 *	(6.649)
Treatment (second-half)	0.104	(0.331)
Video starts	1.131	(0.837)
n.obs	32	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1d: males chatter more with females in the post-stressor period relative to the pre-stressor period (whole video)

```
model29 <- glmmTMB(MALE_CHATTER_DURATION ~
                 TREATMENT + VIDEO_STARTS
                + (1|PAIR), zi = ~1, family = nbinom1, data = control_Nona)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-13.868	(14.179)
Treatment (second-half)	-0.221	(0.215)

	Estimate (significance)	Standard Error
Video starts	1.913	(1.784)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	21.082	(26.322)
Treatment (second-half)	-0.616 **	(0.191)
Video starts	-2.702	(3.408)
n.obs	34	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 1f: There will be an increase in time the male spends with the female post-stressor (whole video)

```
model131 <- glmmTMB(TIME_TOGETHER ~ TREATMENT + VIDEO_STARTS +
                offset(log(F_IN_DUR)) + (1|PAIR), family = nbinom2, data = control)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-8.766	(5.915)
Treatment (second-half)	0.136	(0.111)
Video starts	0.773	(0.749)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2a: There will be a decrease in male vigilance post-stressor (whole video)

```
model55 <- glmmTMB(MALE_PEEK_DURATION ~ TREATMENT + VIDEO_STARTS +
                offset(log(VIDEO_DURATION))
              + (1|PAIR), family = nbinom1, data = control)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-12.436	(6.399)
Treatment (second-half)	0.139	(0.163)
Video starts	1.129	(0.807)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-9.180	(7.455)
Treatment (second-half)	0.157	(0.157)
Video starts	0.673	(0.948)
n.obs	32	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2c: There will be a decrease in visit number post-stressor (whole video)

```
model71 <- glmmTMB(MALE_VISIT_N  ~ TREATMENT + VIDEO_STARTS +
             offset(log(VIDEO_DURATION))
            + (1|PAIR), family = poisson, data = control)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-9.876 ***	(2.366)
Treatment (second-half)	0.033	(0.138)
Video starts	0.374	(0.298)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Prediction 2d: There will be a decrease in male food-sharing post-stressor (whole video)

```
Model80 <- glmmTMB(MALE_FS   ~ TREATMENT + VIDEO_STARTS +
              offset(log(F_IN_DUR))
             + (1|PAIR), family = poisson, data = control)
```

Full model

	Estimate (significance)	Standard Error
Intercept (mean)	-9.343 **	(3.026)
Treatment (second-half)	0.072	(0.190)
Video starts	0.230	(0.382)
n.obs	36	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Influential points removed

	Estimate (significance)	Standard Error
Intercept (mean)	-11.679 ***	(2.709)
Treatment (second-half)	-0.096	(0.206)
Video starts	0.528	(0.339)
n.obs	34	

*** p < 0.001; ** p < 0.01; * p < 0.05.

CHAPTER 4: Pair-bond strength is repeatable and related to partner responsiveness in a wild corvid

Abstract

The Social Intelligence Hypothesis (SIH) posits that the maintenance and management of social bonds generates selection on cognitive ability. While the SIH has gained much traction over the past few decades, its fundamental predictions have never been tested within a single study system. Here, we tested four key predictions of the SIH within a single wild study system, the jackdaw. We found support for three predictions: that pair-bond strength is variable between pairs, repeatable within pairs and related to socio-cognitive performance. We did not find any evidence that pair-bond strength influences reproductive success. However, we did find that pairs with stronger bonds were better able to adjust hatching synchrony to environmental conditions. Taken together, our results provide support for several key predictions of the SIH, including the first direct evidence that bond strength and cognitive performance are related in non-humans. However, more work is necessary to understand the fitness consequences of pair-bonds.

Introduction

Many societies across the animal kingdom are characterised by differentiated social bonds (Bergman & Beehner, 2015; Seyfarth & Cheney, 2012; M. J. Silk & Hodgson, 2021). Individuals who form and maintain ‘strong’ bonds, involving frequent and consistent affiliative interactions, often derive fitness benefits (Cameron et al., 2009; McFarland et al., 2017; Schülke et al., 2010; J. B. Silk et al., 2010b; M. J. Silk & Hodgson, 2021). This has major implications for our understanding of social evolution, but important questions remain. In particular, an individual’s ability to form strong bonds is assumed to be a trait on which selection can act. However, for this to be the case, the strength of social bonds must vary within a population and show consistency (repeatability) over time. To date, evidence for bond strength consistency is largely restricted to primates (Brent et al., 2013; Koski et al., 2012; Massen & Sterck, 2013; Mitani, 2009; J. B. Silk et al., 2010b, 2010a), with limited evidence in non-primate mammals (Blumstein et al., 2013; Finkenwirth & Burkart, 2017; Stanley et al., 2018) and non-mammalian taxa (Aplin et al., 2015; Brandl et al., 2021; Jacoby et al., 2014). Moreover, whether selection can act on an individual’s ability to maintain strong bonds is integral to the highly influential Social Intelligence (or Social Brain) Hypothesis (Dunbar, 1998; Humphrey, 1976), but crucial evidence is lacking.

The Social Intelligence Hypothesis (SIH) suggests that the information-processing challenges posed by navigating social relationships are a central driver of cognitive evolution (Dunbar, 1998). Evidence from both behavioural and neuroanatomical studies support the SIH (Ashton, Ridley, et al., 2018; Ashton, Thornton, et al., 2018; Dunbar, 1998; Dunbar & Bever, 1998; Fox et al., 2017; Maclean et al., 2008; Pérez-Barbería et al., 2007; Shultz & Dunbar, 2006, 2007; Street et al., 2017). However, results are often contradictory (DeCasien et al., 2017; González-forero & Gardner, 2018; MacLean et al., 2009, 2014; Powell et al., 2017; Wartel et al., 2019), and generally controversial due to methodological caveats (Healy & Rowe, 2007; Powell et al., 2017; Wartel et al., 2019). To date, multiple key predictions of the SIH have never been tested within a single study system. Specifically, we need to know whether individual ability to form strong bonds (1) varies within social groups, (2) is consistent, (3) is related to the ability to meet challenges that require information processing (i.e. are cognitive in nature: Shettleworth, 2010b), and (4) results in greater reproductive success (i.e., may be under selection).

We tested all four key predictions of the SIH in a species of corvid; large-brained birds that have emerged as a major model system in the study of animal cognition (Boucherie et al., 2019; Emery & Clayton, 2004; Güntürkün & Bugnyar, 2016; Seed et al., 2009). While the SIH was initially applied to primates (Dunbar, 1998; Humphrey, 1976), its core predictions should, in principle, apply across taxa. In primates, individuals that maintain multiple strong relationships often derive fitness benefits (Cheney et al., 2016; Ellis et al., 2019; J. B. Silk et al., 2003), while species that live in bigger groups (with more social connections) tend to have bigger brains (Barton, 1996; Dunbar, 1998; S. Shultz & Dunbar, 2007, but see DeCasien et al., 2017). In contrast, brain size in birds is associated not with the quantity of social connections (Beauchamp & Fernández-Juricic, 2004; Emery et al., 2007; Iwaniuk & Arnold, 2004; Sayol et al., 2016) but with long-term pair-bonds (Emery et al., 2007; Shultz & Dunbar, 2010). Accordingly, some authors argue that forming and maintaining a strong bond with a mated partner, and tracking and responding to that partner's behaviour, is the critical driver of cognitive evolution in birds. This is sometimes referred to as the 'Relationship Intelligence Hypothesis', or RIH (Emery et al., 2007), but its key predictions mirror those of the broader SIH.

Many bird species form long-term monogamous relationships, and there is evidence to suggest that the longer pairs have been together, or the more familiar they are, the greater their reproductive success (Culina et al., 2020; Sánchez-Macouzet et al., 2014; Van De Pol et al., 2006). However, few studies have explicitly quantified the strength of the pair-bond by measuring rates of affiliative interactions (for exceptions see Boucherie et al., 2018; Elie et al., 2011; Sparks, 1964; Spoon et al., 2004). In addition, only one study on captive cockatiels (*Nymphicus hollandicus*: Spoon et al., 2006) has linked pair-bond strength to reproductive outcomes (Spoon et al., 2006). Moreover, evidence for the consistency of pair-bond strength in birds is limited to one study on spatial proximity in wild greylag geese (*Anser anser*) (Kralj-Fišer et al., 2007). Thus, while there is support that pair-bond strength varies between pairs (prediction 1: Boucherie et al., 2018; Elie et al., 2011; Firth et al., 2018; Sparks, 1964; Spoon et al., 2004), there is very little evidence that pair-bond strength is consistent (prediction 2) and related to reproductive success (prediction 4). Furthermore, while there is some evidence to suggest that behavioural synchrony (a behaviour that requires some degree of partner attentiveness) is correlated with affiliative behaviours in cockatiels (Spoon et al., 2006), the link between bond strength and socio-cognitive performance (prediction 3) have yet to be directly tested in any non-human species.

Here, we test key predictions of the SIH in jackdaws (*Corvus monedula*), a species of corvid that forms long-term pair-bonds which tend to be both socially and genetically monogamous (Chapter 2; Gill et al., 2020; Henderson et al., 2000; Kubitza et al., 2015; Röell, 1978; Wechsler, 1989). During the breeding season, both members of the pair cooperate to build the nest (Hahn et al., 2020; Röell, 1978), and the female incubates the eggs while her partner brings her food (Röell, 1978). To test the four key predictions of the SIH, we captured video footage of jackdaws inside the nest-box across four years, during the nest-building and incubation stage of the breeding season. This video footage was coded with a detailed behavioural ethogram to allow for fine-scale quantification and analysis of pair-bond strength, thus allowing us to test prediction (1) that pair-bond strength varies between pairs, and (2) that pair bond- strength is consistent within pairs. To test whether pair-bond strength may be under selection (prediction 4), we recorded various fitness proxies, including the total number and mass of fledglings per pair per year, and cumulative fledging success of pairs over multiple years. To examine whether a strong pair-bond may facilitate plastic responsiveness to changing environmental conditions, which may present an informational (and thus, cognitive) challenge (prediction 3) and result in differential fitness outcomes (prediction 4), we also recorded hatching synchrony across years. More asynchronous hatching is thought to be more advantageous in resource-poor years because the brood is quickly reduced, thus increasing the probability that a small number of chicks survive rather than entire brood failure (Lack, 1966; Parejo et al., 2015; Wiebet et al., 1998 but see Podlas & Richner, 2013). Conversely, more synchronous hatching is thought to be advantageous in resource-rich years (Lack, 1966; Parejo et al., 2015). The synchronicity of hatching depends on the female's incubation behaviour (Wang & Beissinger, 2011), which is influenced by the male's provisioning behaviour (Henderson & Hart, 1993). Hatching synchrony may therefore be related to how well a pair are able to coordinate their behaviour in the face of environmental variability. Finally, as a more direct test of prediction (3), we experimentally tested whether individuals in strong pair-bonds are more responsive to their partner's behaviour. To do this, we used results from a 2019 playback experiment designed to test a specific facet of partner responsiveness: consolation (Chapter 3). In this experiment, we tested whether males consoled their partner by exposing incubating females to a minor stressor while their partners were absent, and comparing male-initiated partner-directed behaviour before and after the stressor. Males did respond to their partner's distress, but there was substantial inter-individual variation in response and no evidence for consolation. We therefore tested whether the *absolute* change in male behaviour, i.e. his general responsiveness to his partner (as opposed to any specific behavioural response), was associated with pair-bond strength. We expected responsiveness to increase with pair-bond strength.

Methods

Ethics This study was carried out with approval from the University of Exeter Biosciences Research Ethics Committee (eCORN002970; eCORN001858) following the ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching (Buchanan et al., 2012). All subjects were ringed and blood was collected by Cornish Jackdaw Project team members licensed by the British Trust for Ornithology and UK Home Office (project licence 30/3261).

Study population Data were collected from wild nest-box-occupying jackdaws during the 2014, 2015, 2018 and 2019 breeding seasons across three study sites in Cornwall, UK (Site X: 50°17′32″N; 5°11′96″W; Site Y: 50°11′26″N, 5°10′51″W; Site Z: 50°11′56″N, 5°10′9″W). Nest-box owners were individually ringed after being caught using baited walk-in traps or trapdoors at the nest-box. During ringing, tarsus length (an indicator of skeletal size; Rising & Somers, 1989) and age (see below) were recorded for each individual. Where individuals had multiple measurements of tarsus length across different ringing sessions, tarsus length was averaged for measurements taken when the individual was fully grown. Blood samples were also collected and used for molecular sexing as described by Griffiths et al. (1998).

Study subjects *Individual ID* Of 125 individuals (63 females, 62 males) included in this study, 17 were not ringed (5 females, 12 males). For pairs where a ringed bird (A) paired with an unringed bird (B) in year one, then in year two A paired with a ringed bird (C), and C was not ringed during or before the year one breeding season, B was presumed to be C (n = 7 instances). Given low levels of observed re-pairing (in seven years of observation, ~7% of females and ~11% of males were observed to re-pair), this is likely to be an accurate reflection of identity. Altogether, 66 unique pairs were involved in this study. Four males and three females re-paired during the course of the research.

Individual age 41 individuals were of known (exact) age, while we had ‘minimum age’ information for 70 individuals, based on morphology and previous sighting data. For pairs where both individuals were of known age (n = 14 pairs), partners were always the same age, which aligns with research suggesting that jackdaws form pair-bonds during their first year (Lorenz, 1931; Von Bayern et al., 2007). Therefore, for individuals who were of unknown age, yet had a partner of known or minimum age (n = 4 females, n = 8 males), we assumed their age to match that of their partner.

Data collection *Video data collection* In 2014, 2015, 2018 and 2019 CCTV cameras were placed inside nest-boxes during the initiation phase of nest-building. Using Digital Video Recorders, we recorded footage of pairs inside the nest-box during the nest-building stage and incubation stage. The onset of the nest-building stage was defined as when a layer of material covered the floor of the nest-box (Hahn et al., 2020), while the incubation stage was defined as 1 – 10 days after the female's fertile period (where the fertile period ends on the day before the final egg is laid; Birkhead & Møller, 1993; Gill et al., 2020).

Video data coding Following video data collection, students were trained to code the videos with a detailed behavioural ethogram (see Table S1) using either Excel (2014 and 2015) or BORIS (2018 and 2019; Friard & Gamba, 2016). Inter-rater reliability (IRR) of 36 hours of repeated coding across 12 raters revealed high IRR for 'male in', 'female in', 'male vigilance' and 'female vigilance' (see Supplementary Materials for details). Affiliative behaviours were re-defined in 2018 to make them more conservative (see Table S1). Therefore, the lead author re-coded affiliative behaviour for every section of every video where the pairs were coded as together in the nest-box (and thus where affiliation could potentially occur). The lead author was blind to the 'strength' of the pair-bond during coding, given that (i) videos were randomised in regards to coding and (ii) pair-bond strength was calculated following objective criteria (see Methods: Quantifying pair-bond strength) only after videos were coded, and it was unknown at the time of coding which behaviours would be included in the final index.

Some videos (n = 14) contained forced extra-pair copulations (FEPC), which involved the female being attacked by a non-partner male (Chapter 2). In Chapter 3, we found that following a stressful event such as FEPC, partner males changed their affiliative behaviour towards the female by decreasing affiliation with their partner and visit rate to the nest-box. This could skew pair-bond strength calculations. Therefore, to calculate pair-bond strength from these videos we only used footage from the section of video before FEPC occurred.

Partner responsiveness data In 2019, we ran a playback experiment to test whether males responded to their partner's distress (Chapter 3). To do this, we exposed incubating females (n = 27) to the sound of a foreign male at the nest-box while their partners were absent from the area (and therefore blind to the stressor). Using internal CCTV footage, we then compared the male's affiliative behaviour towards his partner in the pre- and post-stressor periods. 'Partner responsiveness' was measured as the absolute change in male-initiated direct affiliative behaviour towards his partner pre- and post-stressor (i.e., the absolute difference in affiliation during these periods). We supplemented experimental data with data from natural FEPC events where we had a measure of male behaviour towards the female both pre- and post-stressor, and the male was absent for the stressor (n = 6). See Chapter 3 for detailed methods. The pre-stressor period of these videos were included in pair-bond calculation data.

Fitness data Daily observation allowed us to monitor the date of egg-hatching in order to calculate hatching synchrony. Hatching synchrony was calculated as the date of the last hatch minus the date of the first hatch, divided by the number of eggs that hatched. Only clutches where more than one egg hatched were included in this calculation. Controlling for the number of eggs that hatched means that clutch size, infertile eggs or eggs where the embryo died during development did not skew the estimate of hatching synchrony. When the oldest chick was 25 days old, the remaining chicks in each nest-box were ringed, and final biometrics (tarsus and weight) were taken. While chicks usually fledge at around 35 days old, they can fledge as early as day 29. As nest disturbance can cause early fledging, we did not disturb chicks following ringing, and assumed that they had fledged.

Statistics All statistical analyses were undertaken in R v4.0.2 (R Core Team 2017).

Quantifying pair-bond strength Behaviours recorded from internal nest-box video were standardised by the length of the video. Video length was calculated as the second an individual returned to the nest-box following video set-up to the second the recording was terminated.

We used principal components analysis (PCA) to identify interrelationships between potential affiliative behaviours (Fraser et al., 2008; Fraser & Bugnyar, 2010b; Spoon et al., 2004) with the aim of identifying reduced dimensionality components of bondedness. All behaviours previously hypothesised to be affiliative between bird partners were selected for inclusion in the PCA. These were: ‘food-share’ (Boucherie et al., 2018; Oe, 2008; Von Bayern et al., 2007); ‘contact’ (Boucherie et al., 2018; Fraser & Bugnyar, 2010; Kubitza et al., 2015; Wechsler, 1989) ; ‘allopreen’ (Boucherie et al., 2018; Fraser & Bugnyar, 2010a; Kubitza et al., 2015; Picard et al., 2020; Spoon et al., 2004; Wechsler, 1989), ‘time together’ (Boucherie et al., 2018; Kubitza et al., 2015; Spoon et al., 2004; Wechsler, 1989) and ‘copulation’ (Spoon et al., 2004). We also included ‘male visit rate’ in the incubation stage because females must remain in the nest-box to incubate, but the rate at which the males visit their partner may vary and may be correlated with other affiliative behaviours. Finally, we included ‘chatter’ (Hahn et al., 2020), a distinctive call that partners often make when together at the nest-box (see Supplementary Video 1). Allopreening was split into ‘male-initiated’ and ‘female-initiated’ for the nest-build stage, but not at the incubation stage because almost all allopreening (94.55%) was male-initiated. See Table S1 for a definition of each behaviour.

Prior to the PCA, we tested sampling adequacy using the Kaiser–Meyer– Olkin (KMO) test (threshold 0.5) and the Bartlett sphericity test (Budaev, 2010). A PCA was then performed on the selected variables, and loadings and scree plots were examined to evaluate whether and which PC axes could be used to reflect pair-bond strength. PCA and associated tests were conducted in package *psych* (Revelle, 2018).

Statistical modelling All modelling was undertaken in glmmTMB v1.0.1 (Brooks et al., 2017) in R v4.0.2 (R Core Team, 2017). For cases where competing models were built, we compared models using Akaike’s Information Criterion (AIC). If models differed in AIC by 2 or more, the model with the lowest AIC was selected (Thomas, 2015). Otherwise, the model with the best diagnostic plots was retained (assessed using DHARMa: Hartig, 2020). If diagnostics for the full models were essentially equivalent, model diagnostics were examined for models once influential points were excluded, and the best fitting model retained. Influential points were identified as any datapoints more than four times the mean Cook’s distance (Cook, 1979). Each final model was run both with and without influential points. Models without influential points are presented in the main manuscript (and corresponding tables and figures). If results from full models and models without influential points differed, we report the results of both models. All models were tested for zero-inflation and dispersion (Hartig, 2020). No final models were over-dispersed or zero-inflated and all showed acceptable model fit.

Note that 'minimum male age', a covariate in many models, is used as a proxy for both male and female age given a high correlation between ages. When exact ages are known for both partners, they are fully correlated. When testing the relationship for all pairs with known age or minimum age (but before inputting the age of individuals with unknown age as that of their partner's) Pearson's $r = 0.76$, $p < 0.001$.

Sample sizes per model differ slightly because some pairs were missing data for minimum male age, male tarsus length and/or female tarsus length, and number of influential points per model varied.

1: Predictors of pair-bond strength We used GLMMs to test potential predictors of pair-bond strength. For every model, pair ID and site were included as random effects. The pool of covariates for possible inclusion were: male tarsus size, female tarsus size, the size difference between partners, years together, male age, female age, time the video started and days since the end of the female's fertile window. Days since the female's fertile window was included because males spend more time with their partner while the female is fertile (Gill et al., 2020). This is possibly to guard them against attempted extra-pair copulation, although rates of extra-pair paternity are very low in jackdaws (Chapter 2; Gill et al., 2020; Henderson et al., 2000; Liebers & Peter, 1998; Turjeman et al., 2021). Including this covariate therefore controlled for any changes in behaviour related to female fertility. We compared models with either absolute male and female tarsus size (an indicator of structural size: (Rising & Somers, 1989)) and size difference (male tarsus length minus female tarsus length). We hypothesised that where there is a larger disparity in size between partners, affiliation may be higher because the smaller (lower quality: Verhulst et al., 2014) individual may invest more in its partner, to reduce the probability of that partner searching for a better-quality mate (Culina et al., 2015), and that this may be more important to control for than absolute tarsus size. Next, we compared models with size difference as a linear or quadratic term, given that plots indicated a potential quadratic relationship with pair-bond strength. Finally, we compared a model including years together against a model including minimum male age. Years together and minimum male age are correlated to a degree considered to be problematic for inclusion ($r > 0.7$; Dormann et al., 2013), hence why model comparison techniques were used for variable selection. All models were run for pair-bond strength calculated in the nest-build stage and incubation stage. To improve model fit (based on diagnostics using DHARMa: Hartig, 2020), pair-bond strength was log-transformed. All models were Gamma-distributed.

2: Consistency of pair-bond strength We tested whether within-pair bond strength remained consistent over time using repeatability analysis in *rptR* (Stoffel et al., 2017). Specifically, we tested whether pair-bond strength was repeatable within-year and between-years for both the nest-building and incubation stage. We first ran repeatability models with no covariates to obtain an unadjusted estimate of repeatability of pair-bond strength. Following this, we controlled for covariates and obtained an adjusted repeatability estimate. Covariates for the within-year models were days since the female's fertile period and time the video was started. Given within-year repeatability, a mean value of pair-bond strength per year was calculated, and this was used to estimate repeatability between years. For between-year models, the age of the male and year were included as covariates. Days since the female's fertile window and video start time were not included as covariates due to the use of mean value of within-year pair-bond strength as the response variable.

In all models, pair-bond strength was log-transformed for model fit. For each model, parametric bootstrapping ($n_{boot} = 1000$) quantified uncertainty while significance testing was implemented using likelihood ratio tests (LRT) and through the permutation of residuals ($n_{perm} = 1000$) (Stoffel et al., 2017).

To validate the robustness of repeatability estimates, we also estimated repeatability using Markov Chain Monte-Carlo (MCMC) sampling, with confidence intervals of 0.95, a burn-in of 3000, 100000 iterations and a null prior (Hadfield, 2010; Stoffel et al., 2017).

3: Partner responsiveness and pair-bond strength Partner responsiveness, the dependent variable, was log-transformed for improved model fit. Pair-bond strength, data type (experimental or natural), whether the male returned before the female (Chapter 3) and minimum age of the male (which resulted in a better model fit than minimum number of years together) were included as covariates for the full model. Due to convergence issues, minimum age of the male was not included as a predictor in the model without influential points; however, the relationship between pair-bond strength and partner responsiveness was qualitatively identical for both models. The model used a Gaussian error structure with an identity link.

We followed this analysis with two further models to test whether female behaviour changed as a function of pair-bond strength, and thus whether males could simply have been responding to the differential magnitude of female behavioural change, details of which are presented in the Supplementary Material.

4: Fitness consequences of pair-bond strength We built GLMMs to test whether pair-bond strength correlated with fitness outcomes. Because pair-bond strength was not repeatable within or between years for the nest-build stage of the breeding season (see Results), we tested whether pair-bond strength measured in the incubation stage influenced reproductive outcomes. We removed any pairs (i) with more than one female laying eggs in the nest-box (n = 1 pair) (Chapter 2), and (ii) that may have been impacted by researcher activity in the incubation stage (n = 2 pairs). We also removed one abnormal pair where field observations suggested repeat brood failure (across five years) occurred due to the mother over-zealously cleaning chicks, leading to injury and death.

The fitness outcomes we tested for each pair (per year) were (4.1) number of fledglings, (4.2) total mass of fledglings, (4.3) proportion of hatched chicks that fledged and (4.4) hatching synchrony. We also tested (4.5) whether hatching synchrony related to number and mass of fledglings per pair per year for the entire population (i.e., not just for those pairs for whom we had pair-bond strength data). For each response variable, we tested competing models which included either minimum number of years together or minimum male age. Pair-bond strength (per video), male tarsus, female tarsus, lay date, year and the rate of male provisioning during female incubation (uncorrelated with affiliative behaviours yet potentially important for reproductive success) were always included as covariates. Models with minimum male age and minimum number of years the pair had been together were compared as described above. Pair ID and site were included as random effects. Error distributions were as follows: for (4.1-2, 4.4-5) zero-inflated Gamma with a log link and (4.3) Binomial with a probit link.

Because of low levels of variance in number of fledglings per year (72.87% of pairs fledged two or three offspring), we also tested (4.6) the cumulative number of fledglings per pair over five years. This required subsetting the data to pairs for whom we had five years of reproductive success data (n = 12). For this model, pair-bond strength was calculated as the mean value of pair-bond strength per pair across all available data points in the incubation stage. A Poisson error structure was used.

For all fitness outcomes, we compared models with linear and quadratic pair-bond strength, to test for directional and stabilising selection on pair-bond strength (Boogert et al., 2018). We also compared models with and without an interaction term between pair-bond strength and year. We tested this interaction because how selection acts on behaviour can vary according to environmental conditions (e.g. Dingemanse et al., 2004).

To test how reproductive success varied across years for the whole population, we also examined the relationship between year and population-wide number (4.7) and mass (4.8) of fledglings (error structure: zero-inflated Gamma with log link). Year, male and female tarsus length, minimum male age or years together, and lay date were included as predictor variables. Pair ID was included as a random effect, but site was not included due to convergence issues. These models allowed us to gain insight into whether some years appeared particularly difficult, most likely due to limited resource availability.

5: Fitness consequences of re-pairing Several studies have found that, in long-term monogamous species, individuals experience lower reproductive success when they re-pair (Sánchez-Macouzet et al., 2014; Van De Pol et al., 2006). This suggests that pair familiarity influences reproductive outcomes. To test whether individuals who re-paired due to the death or divorce of their partner experienced negative fitness consequences, we modelled whether there was a significant change in reproductive success in the year before and the year after re-pairing. We used data from our entire population, rather than only those pairs for whom we had pair-bond information. Across our entire nest-box-occupying population, we observed 16 males and 12 females re-pairing across seven years, where both the old and new pair initiated their breeding attempt in a nest-box. One female was removed from the analysis because although she re-paired, she was not observed for two years between the re-pairing event. Because some re-pairs were not independent (i.e., one male and one female split and both re-paired), we modelled male and female re-pairing separately. We tested whether number of fledglings (5.1), mass of fledglings (5.2), and proportion of hatchlings that fledged (5.3) decreased the year following re-pairing. Models 5.1 and 5.4 used a zero-inflated Gamma error structure with a log link; 5.2 used a zero-inflated Gaussian error structure; 5.3 used a Binomial error structure with a probit link. We included pair (original or re-pair), minimum age (male age for male models, female age for female models), lay date and year as predictor variables. We did not include tarsus length of the male and female because we did not have this information for all individuals and did not want to further limit sample size. All models included focal ID and site as a random effect, except male re-pairing models 5.1, 5.2 and 5.3 and female re-pairing model 5.2, where site had to be removed due to convergence issues. Lay date was also excluded as a covariate from male re-pairing model 5.2 due to convergence issues.

Results

Video data In the nest-building stage, we recorded 142.18 hours of footage across 54 videos (mean length = 2.63 ± 1.13 hours) for 39 pairs.

In the incubation stage, we recorded 362.24 hours of footage across 132 videos (mean length = 2.74 ± 1.07 hours) for 65 pairs.

Testing Prediction 1: 1.1 Quantifying pair-bond strength Correlation plots of all recorded behaviours show that (i) behaviours correlate in differing ways depending on the breeding stage (nest-building or incubation) at which they are measured and (ii) while most behaviours that are hypothesised to be affiliative significantly correlate, some do not. See Figure S1 for further detail.

PCA: nest-building stage KMO tests showed that the variables suitable for inclusion were: male chatter, male allopreening, female allopreening, contact and time together. Bartlett's test of sphericity suggested adequate sampling ($\chi^2 = 146.31$, $p < 0.001$). PC1 and PC2 (see Figure S2a) collectively explained 66.9% of the data. PC1 (explaining 43.3% of variation) was dominated by the affiliative behaviours allopreening, contact and time together, which all loaded positively (see Table S2 for loadings), so we interpreted it as a measure of pair-bond strength. PC2 (23.6%) was principally dominated by male chatter. Given that in the nest-building stage chatter significantly correlated with vigilance (see Figure S1), we interpreted PC2 as a reflection of nest-box defence.

PCA: incubation stage KMO tests showed that the variables suitable for inclusion were: male chatter, female chatter, allopreening, contact and time together. Bartlett's test of sphericity suggested adequate sampling ($\chi^2 = 436.46$, $p < 0.001$). PC1 and PC2 (see Figure S2b) collectively explained 73.8% of the data. All behaviours loaded positively onto PC1 (see Table S2) which explained 59.5% of variation in the data, and PC2 was dominated by allopreen (positive) and contact (negative). While affiliative behaviours loaded onto both PC1 and PC2, a scree plot (Figure S3) showed clear justification for using PC1 only, because it explained a large proportion of variation in the data relative to other PCs. We therefore used PC1 as a measure of pair-bond strength.

1.2 Predictors of pair-bond strength *Nest-building stage* Results for the nest-build stage differed depending on whether influential points were retained. With their inclusion, pairs were more affiliative earlier in the day; with their exclusion, pairs were more affiliative closer to the female’s fertile window. Full results are presented in the Supplementary Material, but should be interpreted with caution given their instability.

Incubation stage Pairs with older males (and thus females; see Methods) had stronger pair-bonds ($n = 101$ datapoints, $n = 49$ pairs, $\beta = 0.07$, SE = 0.03, $\chi^2 = 8.17$, 95% CI [0.02,0.13], $p < 0.01$). Year also had a significant effect ($\chi^2 = 19.68$, $p < 0.01$), with pair-bond strength lower in 2019 than in previous years (see Supplementary Materials for pairwise comparisons).

Testing prediction 2: Repeatability of pair-bond strength *Nest-building stage* Figure S4a shows the variation in nest-building stage pair-bond strength between pairs. Six pairs were recorded more than once within a year, while nine pairs were recorded in more than one year. Neither within-year pair-bond strength (R = 0.06, 95% CI[0,0.75], P_{perm} = 0.52, P_{LRT} = 0.5; adjusted R = 0.65 , 95% CI[0,0.99], P_{perm} = 0.07, P_{LRT} = 0.23) nor between-year pair-bond strength (R = 0.02, 95% CI[0,0.63], P_{perm} = 0.45, P_{LRT} = 1; adjusted R = 0.09 , 95% CI[0,0.74], P_{perm} = 0.47, P_{LRT} = 0.52) were repeatable. However, this result was not replicated when repeatability was estimated using MCMC sampling. In this analysis, both within- and between-year pair-bond strength were repeatable, albeit with wide confidence intervals (Table S3).

Incubation stage Figure S4b shows the variation in incubation stage pair-bond strength between pairs. 21 pairs were recorded more than once within a year (Figure S4c), while 34 pairs were recorded in more than one year (Figure S4d). We found repeatability for both within-year pair-bond strength (R = 0.52, 95% CI[0.13,0.77], P_{perm} = <0.01, P_{LRT} = <0.01; adjusted R = 0.65 , 95% CI[0.32,0.86], P_{perm} = 0.003, P_{LRT} = <0.01) and between-year pair-bond strength (R = 0.41, 95% CI[0.14,0.64], P_{perm} = <0.01, P_{LRT} = <0.01; adjusted R = 0.50 , 95% CI[0.28,0.72], P_{perm} = <0.01, P_{LRT} = <0.01). There was one highly affiliative pair (Figure S4c,d); results were robust to their removal when controlling for

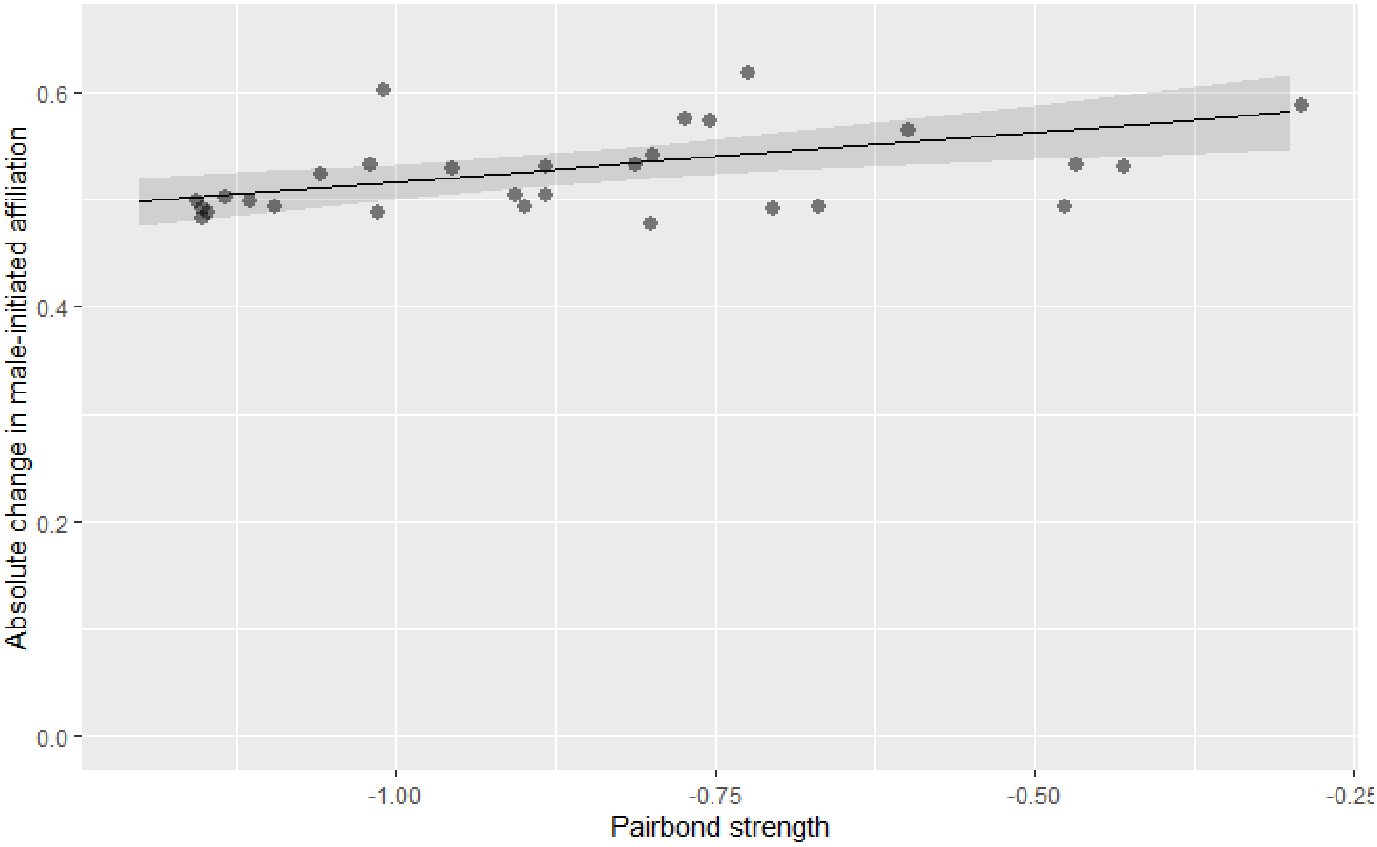


Figure 1 Incubation-stage pair-bond strength plotted against the (log-transformed) absolute change in male-initiated affiliation following partner distress. The plotted line shows the predicted values of absolute change in male-initiated affiliation following partner distress as a function of pair-bond strength.

covariates (within-year: adjusted R = 0.49 , 95% CI[0.08,0.80], $P_{perm} = 0.02$, $P_{LRT} = 0.04$; between-year: adjusted R = 0.29 , 95% CI[0.06,0.57], $P_{perm} = 0.02$, $P_{LRT} = 0.05$). Repeatability estimated from MCMC sampling supported these results (Table S3).

Figure 2 (a) the interaction between pair-bond strength and year on hatching asynchrony. Increasing levels of hatching asynchrony indicate more asynchronous hatching, whereas zero indicates synchronous hatching. (b) shows number of fledglings per pair across the entire population (not just for those pairs with pair-bond strength information) for whom model covariate information (lay date, female tarsus length, male tarsus length, minimum number of years together) was available. Asterisks show predicted mean number of fledglings per year produced by the aforementioned model.

Testing prediction 3: Partner responsiveness and pair-bond strength

There was a significant relationship between pair-bond strength and the responsiveness of the male to his partner's distress, where males in stronger pair-bonds showed a larger *absolute* change in behaviour following their partner experiencing a stressor (n pairs = 29; β = 0.09, SE = 0.03, χ^2 = 11.46, 95% CI [0.04,0.15], p < 0.01; Figure 1). This was not driven by the female's magnitude of behavioural change following the stressor: there was no significant relationship between the absolute change in female begging or chatter and pair-bond strength (begging rate: β = -0.06, SE = 0.17, X2 = 0.13, 95% CI [-0.40,0.28], p = 0.72; chatter duration: β = -0.10, SE = 0.09, X2 = 1.10, 95% CI [-0.28,0.08], p = 0.29).

Testing prediction 4: Fitness consequences of pair-bond strength

Incubation stage Pair-bond strength did not influence number of fledglings, mass of fledglings, proportion of hatched chicks that fledged or cumulative fledging (Table 1). However, the interaction between pair-bond strength and year was significantly associated with hatching synchrony (Table 1; Figure 2a).

According to population-wide models of number and mass of fledglings per year, 2018 and 2019 were poor years for jackdaw reproductive success relative to 2014 and 2015 (Figure 2b; see Supplementary Materials for further details).

Pairs with stronger bonds hatched their clutches more asynchronously in 2018 and 2019 relative to 2015, where pairs with stronger bonds hatched their clutches more synchronously (n datapoints = 90, n pairs = 46; 2018 relative to 2015: β = 0.27, SE = 0.09, 95% CI [0.08,0.45], p <0.01; 2019 relative to 2015: β = 0.24, SE = 0.12, 95% CI [0.005,0.48], p = 0.046). Pairs with stronger bonds also hatched their clutches more asynchronously in 2018 relative to 2014, where pairs with stronger bonds hatched their clutches more synchronously (n datapoints = 90, n pairs = 46; 2018 relative to 2014: β = 0.22, SE = 0.11, 95% CI [0.02,0.43], p = 0.03). Given that 2018 and 2019 were poor years relative to 2014 and 2015, this suggests that pairs with stronger bond strength were better able to adjust their hatching synchronicity to environmental conditions. However, we detected no signal that hatching synchrony interacted with year to influence reproductive success across the four years analysed (Table S4).

Testing Prediction 4: Fitness consequences of re-pairing There was no influence of re-pairing on any measure of reproductive success (Table S5).

Table 1 Results of models testing whether pairbond strength correlated with reproductive success per pair per year, or reproductive success per pair across five years for the response variable 'cumulative fledgling number'. Pair ID and site were included as random effects for all models except the cumulative fledgling model, where only site was included as a random effect. The best models (lowest AIC) with influential points removed are presented here, after testing whether (i) pairbond strength should be included as a linear or quadratic term, (ii) an interaction between pairbond strength and year should be included and (iii) whether male minimum age or minimum years together should be included. Bold indicates significant results.

Response variable	n (datapoints, pairs)	Predictor variables	β	SE	X2	2.5% CI	97.5% CI	P-value
Fledgling number	93,47	Pair-bond strength	0.003	0.019	0.032	-0.040	0.033	0.858
		Male minimum age	0.016	0.036	0.213	-0.054	0.087	0.645
		Year:2015	-0.133	0.116		-0.360	0.095	
		Year:2018	-0.296	0.167	3.976	-0.625	0.032	0.264
		Year:2019	-0.384	0.194		-0.764	-0.000	
		Food-sharing rate	**-0.036**	**0.017**	**4.65**	**-0.069**	**0.003**	**0.031**
		Male tarsus length	-0.018	0.042	0.180	-0.101	0.065	0.671
		Female tarsus length	**0.066**	**0.033**	**4.051**	**0.002**	**0.130**	**0.044**
		Lay date	-0.016	0.016	1.059	-0.048	0.015	0.303
Fledgling mass	98,49	Pair-bond strength	0.015	0.018	0.639	-0.021	0.050	0.424
		Male minimum age	-0.014	0.031	0.212	-0.074	0.046	0.645
		Year:2015	-0.075	0.108		-0.286	0.136	
		Year:2018	-0.188	0.141	5.812	-0.465	0.090	0.121
		Year:2019	-0.350	0.160		-0.663	-0.036	
		Food-sharing rate	-0.023	0.017	1.891	-0.056	0.010	0.169
		Male tarsus length	0.032	0.031	1.079	-0.028	0.092	0.299
		Female tarsus length	0.035	0.028	1.546	-0.020	0.090	0.214
		Lay date	-0.019	0.014	2.015	-0.046	0.007	0.156
Proportion hatched that fledged	94,47	Pair-bond strength	-0.006	0.035	0.035	-0.075	0.062	0.853
		Male minimum age	-0.085	0.075	1.297	-0.232	0.061	0.255
		Year:2015	-0.017	0.252		-0.512	0.478	
		Year:2018	-0.120	0.289	1.844	-0.687	0.447	0.605
		Year:2019	-0.362	0.342		-1.032	0.308	
		Food-sharing rate	-0.047	0.045	1.078	-0.135	0.042	0.299
		Male tarsus length	-0.027	0.072	0.143	-0.168	0.113	0.705
		Female tarsus length	0.087	0.054	2.609	-0.018	0.192	0.106
		Lay date	-0.035	0.031	1.284	-0.095	0.025	0.257
Hatching synchrony	90,46	Pair-bond strength	-0.158	0.099	0.468	-0.352	0.037	0.494
		Years together	0.035	0.034	1.080	-0.031	0.101	0.299
		Year:2015	-0.133	0.144		-0.414	0.149	
		Year:2018	-0.452	0.152	14.072	-0.749	-0.154	**0.003**
		Year:2019	-0.271	0.193		-0.650	0.107	
		Food-sharing rate	-0.024	0.023	1.109	-0.068	0.021	0.292
		Male tarsus length	-0.051	0.035	2.095	-0.120	0.018	0.148
		Female tarsus length	-0.046	0.028	2.737	-0.101	0.009	0.098
		Lay date	**0.037**	**0.015**	**5.868**	**0.007**	**0.066**	**0.015**

		Pair-bond strength:Year15	**-0.041**	**0.122**	**10.325**	**-0.280**	**0.198**	**0.016**
		Pair-bond strength:Year18	**0.225**	**0.106**		**0.017**	**0.432**	
		Pair-bond strength:Year19	**0.202**	**0.136**		**-0.065**	**0.468**	
Cumu. fledging	12,12	Mean pair-bond strength	0.113	0.140	0.650	-0.161	0.387	0.420
		Male tarsus length	0.122	0.107	1.290	-0.088	0.332	0.256
		Female tarsus length	0.089	0.072	1.549	-0.051	0.229	0.213

Discussion

Our combination of behavioural and life history data allowed us to test four key predictions of the Social Intelligence Hypothesis (SIH) within one study system for the first time. We found strong support for the predictions that (1) pair-bond strength varies between pairs, (2) is consistent within pairs, and (3) positively correlates with partner responsiveness. While we found that pairs with stronger bonds are better able to adjust hatching synchrony to environmental conditions, we found no evidence to support prediction (4) that having a stronger pair-bond directly translates into fitness benefits.

We found that pair-bond strength varies between and is consistent within jackdaw pairs. During the incubation stage of the breeding season, adjusted pair-bond strength repeatability was 0.65 (within-year) and 0.50 (between-year). This is higher than the average repeatability of behaviour in general (0.37) as reported in a meta-analysis of behavioural consistency (Bell et al., 2009), and similar to the repeatability of association strength with conspecifics (0.41 – 0.62) in wild great tits (*Parus major*; Aplin et al., 2015). This adds support to the finding that social phenotypes are highly repeatable in wild birds, while furthering our understanding of the consistency of pair-bond strength in particular. Previous work has found that the personality of wild great tits is related to pair-bond strength, as measured through relative association propensity (i.e., association with the partner relative to the sum of all other assocations; Firth et al., 2018). If this is true in jackdaws, then personality may be an important driver of inter-individual variation and intra-individual consistency of pair-bond strength. While bond strength was highly repeatable during the incubation stage, in the nest-building stage of the breeding season it was not significantly repeatable when analysed using a frequentist approach. An alternative analysis using a Bayesian approach, conducted to test robustness of results, showed that pair-bond strength is likely to be repeatable in this stage (Table S3). Given the inconsistency in this result, we suggest a larger sample size is required before conclusions about bond consistency during nest-building should be drawn. Nonetheless, our results provide clear support for two fundamental predictions of the SIH: that bond strength varies between and is consistent within pairs.

We also found strong support for a prediction of the SIH that had not yet been tested in non-humans, namely that bond strength correlates with cognitive performance. First, we found that jackdaw pairs with stronger bonds hatched their clutches significantly more synchronously in years with high population-level reproductive success (implying resource rich years) relative to pairs with weaker bonds. Conversely, pairs with stronger bonds hatched their clutches significantly more asynchronously in years with low population-level reproductive success (implying resource poor years) relative to pairs with weaker bonds. The precise mechanisms through which more strongly bonded pairs are better able to adjust hatching synchrony to environmental conditions is unclear, and further work must be done to understand exactly how this occurs. Hatching synchrony is, however, related to incubation initiation (Wang & Beissinger, 2011), which is itself likely to be successful only if males recognise incubation has begun and thus start food-sharing with their partner. Therefore, one explanation as to why more strongly bonded pairs are better able to adjust hatching synchrony to environmental conditions is that partners are more responsive to one another’s behaviour, and thus can better coordinate the beginning of incubation. Second, our experimental results provided further, direct support that pair-bond strength is related to a measure of cognitive performance. While males responded in substantially variable ways to their partner following the partner’s exposure to a stressor, males in stronger pair-bonds showed a larger absolute change in partner-directed behaviour, indicating that males in stronger bonds are more responsive to subtle changes in their partner’s behaviour. This is similar to results found in humans, where individuals who show better socio-cognitive performance (e.g., better recognition of and response to the emotional state of other individuals) form stronger friendships (Cutting & Dunn, 2006; Smith, 2015). Taken together, our results suggest that socio-cognitive performance and relationship strength are positively correlated in jackdaws; the first evidence of such a link in non-humans.

Despite evidence linking pair-bond strength to partner responsiveness, we found no evidence that this translates to reproductive success. Furthermore, we did not find evidence to suggest that re-pairing, and the assumed lack of familiarity with a new partner, influenced reproductive outcomes (see Supplementary Materials for further discussion). While we did find that pairs with stronger bonds were better able to adjust hatching synchrony to environmental conditions, hatching synchrony did not interact with year to influence fitness outcomes at the population level. Although this may be because hatching synchrony is unrelated to fitness outcomes in jackdaws, it is possible that the effect of hatching synchrony on reproductive success may only be detectable with more drastic environmental variation than captured in our study. It is also important to emphasise that we only measured pair-bond strength during the breeding season, and that our data covers a relatively short proportion of jackdaws' long lifespan. Further longitudinal studies of pair-bond strength and fitness outcomes would be necessary to fully interrogate the relationship between pair-bond strength and fitness. Nevertheless, the absence of a relationship between pair-bond strength and fitness found in this study is at odds with previous work showing that pair-bond strength correlates positively with fledging success in captive cockatiels (Spoon et al., 2006), and with many studies showing relationship strength to be key to reproductive outcomes in mammals (Cameron et al., 2009; McFarland et al., 2017; Schülke et al., 2010; Silk et al., 2010b; Silk & Hodgson, 2021). Moreover, this result potentially undermines a central tenet of the SIH, lending support to the idea that forming and maintaining strong relationships may not be a key driver of cognitive evolution in birds. However, before drawing this conclusion, alternative hypotheses must be addressed.

While the pair-bond is the most valuable relationship in corvid society, pairs do not exist in a social vacuum (Emery et al., 2007). Indeed, partners work together to interrupt relationship formation between potential competitors in ravens (*Corvus corax;* Massen et al., 2014), while in jackdaws and rooks

(*Corvus frugilegus*) partners aid each other in fights against third parties (Clayton & Emery, 2007), and associate with flock-members independently of one another (Boucherie et al., 2016; Kubitza et al., 2015). Moreover, jackdaws with more central social network positions have better reproductive outcomes (Kings, 2018). Given that time spent together is an important component of pair-bond strength, and that it takes time to maintain relationships beyond the pair and to monitor the wider social landscape, there is an implied trade-off between the management and maintenance of pair and non-pair relationships. Future work should interrogate the interaction between pair-bond strength, relationships beyond the pair and reproductive success. Indeed, such studies may reveal a more complex relationship between pair-bond strength and reproductive outcomes than we were able to test in this study. For example, there may be a trade-off between cultivating a strong pair-bond and maintaining an integrated position in a social network. Being strongly bonded to a partner may also limit access to valuable social and cultural information. Such trade-offs could obscure the detection of a direct relationship between pair-bond strength and reproductive success. Such trade-offs have been investigated in some species. In wild great tits, males who formed their pair-bonds early sampled fewer potential mates than males who formed pair-bonds later (Firth et al., 2018), while in humans, couples have fewer friends and invest less in friendship than single individuals (Johnson & Leslie, 1982; Kalmijn, 2003). However, there appears to be no trade-off between relationship quality and quantity in at least one species of non-human primate, the chacma baboon (*Papio ursinus*, McFarland et al., 2017). Understanding how jackdaws navigate their pair-bond within and as well as the wider social landscape will not only reveal potential trade-offs, but may also give insight into the cognitive demands of avian relationships. In particular, investigating multi-layered jackdaw sociality in detail would be foundational to testing whether maintaining the pair-bond itself is cognitively demanding, or whether cognitive challenges are posed by the navigation of the pair-bond within a wider social network. This is a key unknown in the study of avian cognitive evolution. In addition, it must be considered that while we find that socio-cognitive *performance* correlates with relationship strength, it is not clear that socio-cognitive *ability* does. Indeed, it is possible that an individual's investment in their partner (relative to non-partners) is correlated with both their socio-cognitive performance within the pair-bond and pair-bond strength, but not correlated with socio-cognitive ability. The SIH predicts cognitive ability is under selection given that an individual's ability to form strong bonds is related to socio-cognitive ability and to reproductive success. Thus, untangling relationship investment, socio-cognitive performance and socio-cognitive ability is a challenging yet vital direction for future investigations of the SIH.

Here, we tested four key predictions of the SIH within a single study system, the jackdaw, and found support for three predictions: that pair-bond strength is variable between pairs, repeatable within pairs and related to socio-cognitive performance. We did not find any direct evidence that pair-bond strength influences reproductive success, although we did find that pairs with stronger bonds were better able to adjust hatching synchrony to environmental conditions. We suggest that although pair-bond strength may not be related to fitness outcomes, it is too soon to draw a robust conclusion. Future work should investigate whether there is an adaptive benefit to adjusting hatching synchrony according to resource availability in jackdaws. Furthermore, pairs do not exist in a social vacuum (Emery et al., 2007). Individuals may therefore trade investment in the pair-bond for investment in relationships beyond the pair. Future research should interrogate the link between pair-bond strength, relationships beyond the pair and reproductive success, and test for potential trade-offs. How relationship investment, socio-cognitive performance and socio-cognitive ability interact is also a crucial consideration for future work. We suggest that intra-specific testing of the key predictions of the SIH is an important direction for a field currently dominated by broad-scale comparative studies, and that such an approach may help to clarify the debate as to whether social relationships are a key driver of cognitive evolution.

Supplementary Material

Ethogram

Table S1 The behavioural ethogram used in BORIS v7.4.6 to code video data.

Behaviour	Type	Description
ALLOPREEN	Duration	One individual preens another
CONTACT	Duration	An individual is stood or sat close enough to their partner that they would not have to move their bodies in order to make physical contact (i.e., they are within a beak's distance of one another). They are not actively engaged in any other behaviour.
CHATTER	Duration	A specific call often made between partners.
IN	Duration	An individual is in the box (time together and male visit rate are extracted from this variable).
FS	Point	Food-sharing between adults. In the incubation stage this is always from the male to the female. Coded only once per male visit to the box.
PEEK (AKA vigilance)	Duration	Individual looks out of nest-box.
LAY (AKA incubation)	Duration	Individual sitting on eggs (start behaviour when the individual 'wiggles' onto the eggs).
COPULATION	Duration	One individual attempts to or does copulate with another.

Interrater reliability (IRR; Cohen’s Kappa with a three second sliding time window) was calculated using 36 hours of repeated footage across twelve raters. IRRs were high for male ‘IN’ (0.99), female ‘IN’ (0.99), male ‘PEEK’ (0.76), female ‘PEEK’ (0.82) and moderate for ‘LAY’ (0.53) (Landis & Koch, 1977). Because affiliative behaviour definitions were re-defined in 2018 (becoming more conservative), the lead author re-coded all instances where the pair were together in a video.

Correlation between behaviours recorded within the nest-box

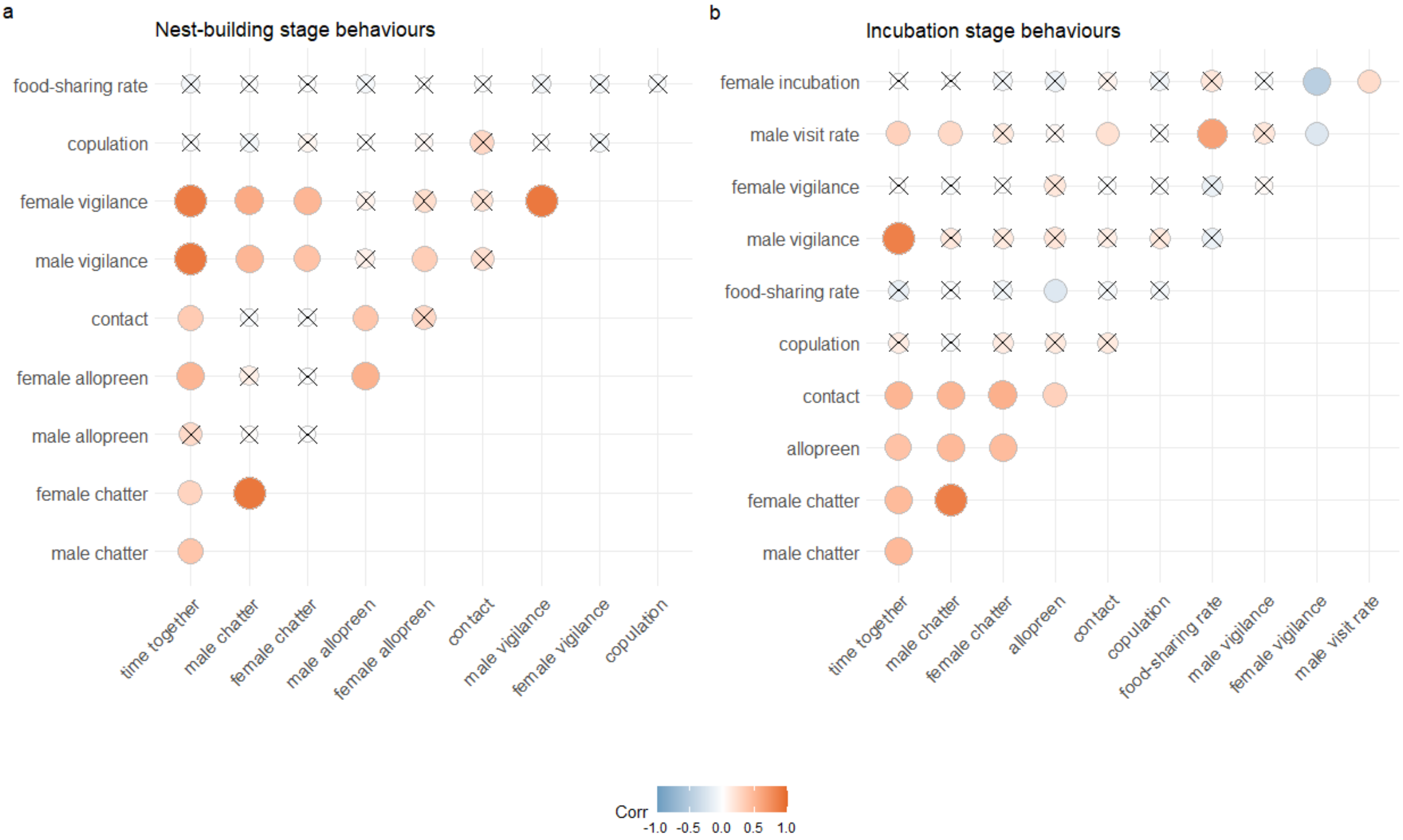


Figure S1 Correlation plots of behaviours in (a) the nest-building stage of the breeding season and (b) the incubation stage of the breeding season. Crosses indicate a non-significant (α > 0.05) pairwise correlation. Red indicates a positive correlation; blue indicates a negative correlation.

Loadings and scree plot

Table S2 PCA loadings for the nest-building and incubation stage.

Stage	Behaviours	PC1	PC2	PC3	PC4	PC5
Nest-build	Male allopreen	0.480	-0.349	0.375	-0.584	-0.409
	Female allopreen	0.537	-0.063	0.486	0.400	0.558
	Contact	0.418	-0.358	-0.754	-0.154	0.324
	Male chatter	0.209	0.783	-0.031	-0.513	0.282
	Time together	0.512	0.365	-0.233	0.461	-0.581
Incubation	Time together	0.413	-0.323	0.593	0.610	-0.018
	Male chatter	0.505	0.148	-0.442	0.186	0.702
	Female chatter	0.507	0.089	-0.462	0.133	-0.710
	Allopreen	0.377	0.707	0.484	-0.352	-0.023
	Contact	0.418	-0.605	0.072	-0.672	0.050

Figure S2 PCA biplots of (a) nest-building and (b) incubation stage behaviours.

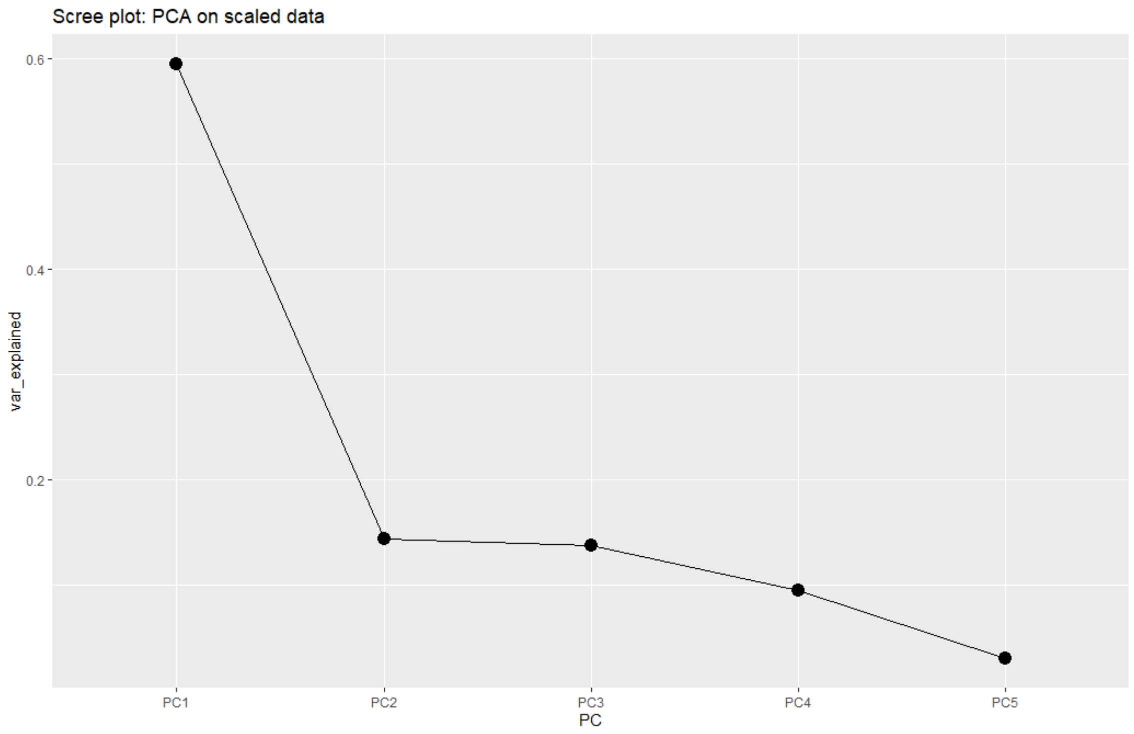


Figure S3 A scree plot showing variation explained by each Principal Component in the incubation-stage PCA. Although affiliative behaviours loaded onto both PC1 and PC2, the scree plot shows clear justification for keeping only PC1 for further analyses given the relative amount of variation explained.

Variation in pair-bond strength

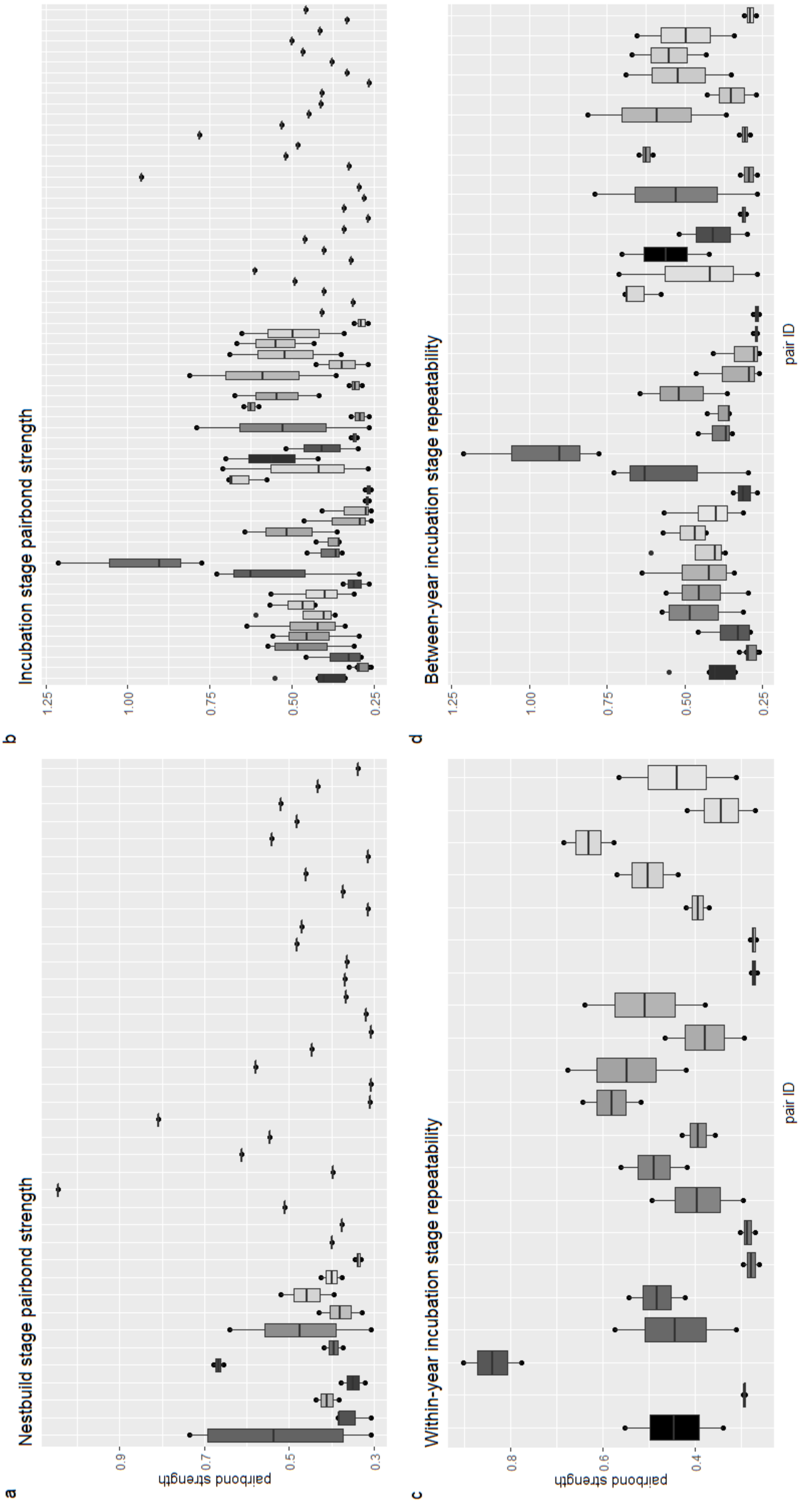


Figure S4 Nest-building (a) and incubation (b) stage pairbond strength (log-transformed) across all pairs. (c) shows (log-transformed) pairbond strength for pairs for whom we had more than one measurement of pairbond strength in the incubation stage within a year; (d) shows pairbond strength (log-transformed) for pairs for whom we had more than one measurement of pairbond strength in the incubation stage between years.

Repeatability of pair-bond strength

Table S3 Repeatability estimates using MCMC simulations in *rptR*

Model	Repeatability estimate	2.5% CI	97.5% CI
Nest-build: within-year	0.529	0.208	0.895
Nest-build: between-year	0.564	0.287	0.849
Incubation: within-year	0.486	0.296	0.735
Incubation: between-year	0.49	0.297	0.683

Hatching synchrony and fitness

Pairs with stronger bonds were better able to adjust hatching synchrony so that their broods were more synchronous in better years and more asynchronous in resource-poor years (as assumed based on population-wide reproductive success; see Results). To test whether hatching synchrony interacted with year to influence reproductive success, we ran GLMMs with (1) fledgling number, (2) fledgling mass and (3) proportion of hatchlings that fledged across the entire population (i.e., not just for those pairs for whom we had pair-bond strength data; n = 191 datapoints and 112 pairs). We included male and female tarsus length, minimum male age/years together (see Methods), hatching synchrony and year as covariates. Site and pair ID were included as random effects. However, models (1) and (2) would not converge with the inclusion of site, so it was removed.

For each response variable, we compared model performance with and without an interaction term between year and hatching synchrony. All models without the interaction term were better than models with its inclusion (where a ‘better’ model has an AIC more than or equal to two less than the competing model; Table S4). This suggests that hatching synchrony did not interact with year to influence reproductive outcome, at least across the four years of our study.

Table S4 AIC values for models including the interaction between year and hatching synchrony (Year*hatching synchrony incl.) and excluding it. Models excluding the interaction term always performed better.

Response variable	Year*hatching synchrony incl.	Year*hatching synchrony excl.
N. fledglings	480.386	475.25
Mass fledglings	2430.691	2425.709
Proportion hatchlings that fledge	512.65	507.648

Re-pairing

Table S5 Results of models testing whether individuals who re-pair (due to partner death or divorce) experience a fitness cost of changing partners in the year of re-pairing relative to their final breeding attempt with their original partner. Focal ID was included as a random effect in each model. Site was included in all models with a female focal, but resulted in convergence issues for male models and thus was excluded.

Re-paired focal	Response variable	n(datapoints, individuals)	Predictor variables	β	SE	X^2	2.5% CI	97.5% CI	P-value
Male	**Fledgling number**	32,16	Re-pair	0.188	0.101	3.454	-0.010	0.387	0.063
			Male min age	0.097	0.067	2.101	-0.034	0.229	0.147
			Lay date	0.006	0.015	0.193	-0.022	0.035	0.661
			Year	-0.074	0.045	2.702	-0.163	0.014	0.100
	Fledgling mass	34,17	Re-pair	42.280	49.410	0.732	-54.560	139.112	0.392
			Male min age	39.460	26.190	2.270	-11.874	90.791	0.132
			Year	-33.230	20.110	2.731	-72.638	6.181	0.098
	Proportion hatched that fledged	34,17	Re-pair	0.123	0.245	0.254	-0.356	0.603	0.614
			Male min age	0.067	0.098	0.463	-0.126	0.260	0.496
			Lay date	-0.008	0.032	0.056	-0.071	0.055	0.813
			Year	-0.076	0.076	0.988	-0.225	0.074	0.320
Female	**Fledgling number**	22,11	Re-pair	-0.203	0.204	0.991	-0.602	0.197	0.319
			Female min age	0.045	0.079	0.333	-0.109	0.199	0.564
			Lay date	-0.040	0.023	2.966	-0.085	0.005	0.085
			Year	0.133	0.114	1.354	-0.091	0.356	0.245
	Fledgling mass	24,12	Re-pair	-0.155	0.174	0.792	-0.495	0.186	0.373
			Female min age	0.037	0.067	0.302	-0.095	0.168	0.583
			Lay date	-0.034	0.020	2.967	-0.073	0.005	0.085

		Year	0.080	0.099	0.661	-0.113	0.273	0.416
Proportion hatched that fledged	24,12	Re-pair	-0.504	0.420	1.444	-1.327	0.318	0.230
		Female min age	0.138	0.142	0.947	-0.140	0.417	0.330
		Lay date	**-0.107**	**0.048**	**5.025**	**-0.200**	**-0.013**	**0.025**
		Year	0.290	0.223	1.689	-0.147	0.727	0.194

Supplementary Methods

Partner responsiveness: female behaviour After the analysis of the change in male behaviour following partner distress, we ran two further models to test whether female behaviour changed as a function of pair-bond strength, and thus whether males could simply have been responding to the differential magnitude of female behavioural change. Post-stressor, females did not change their incubation, vigilance or self-preening behaviour in the male's first visit to the nest-box, relative to his last visit pre-stressor (Chapter 3). However, females did show a slight change in chatter and begging. We therefore tested whether the absolute change in female chatter and begging was significantly correlated with pair-bond strength. Both begging and chatter were log-transformed to improve model fit.

Supplementary Results

Predictors of pair-bond strength: Nest-build stage Significant predictors of pair-bond strength differed depending on whether influential points were included or not. With their inclusion, the only predictor of pair-bond strength at the nest-build stage was time the video was started, where pairs were more affiliative earlier in the day (n = 45 datapoints, n = 33 pairs, β = -0.09, SE = 0.04, X^2 = 5.30, 95% CI [-0.16,-0.01], p = 0.02). With influential points excluded, only the number of days until the female's fertile window was significant, where the closer the female was to being fertile, the more affiliative the pair (n = 42 datapoints, n = 32 pairs, β = 0.01, SE = 0.01, X^2 = 5.22, 95% CI [0.001,0.02], p = 0.02). Given the instability of these results, they should be interpreted with caution.

Predictors of pair-bond strength: Incubation stage In the incubation stage, pairs in 2019 has significantly lower pair-bond strength than in 2014, 2015 and 2019 (2019 relative to 2014: $\beta = -0.30$, SE = 0.15, 95% CI [-0.58,-0.01], $p = 0.04$; 2019 relative to 2015: $\beta = -0.39$, SE = 0.12, 95% CI [-0.62,-0.16], $p < 0.01$; 2019 relative to 2018: $\beta = -0.39$, SE = 0.09, 95% CI [-0.55,-0.2], $p < 0.01$). Results were robust to the inclusion/exclusion of influential points.

Pairwise comparisons of reproductive success across years: Pairs in 2018 and 2019 fledged significantly fewer chicks than 2015 (n datapoints = 180; n pairs = 106; 2018 relative to 2015: $\beta = -0.14$, SE = 0.07, 95% CI [-0.27,-0.01], $p = 0.04$; 2019 relative to 2015: $\beta = -0.28$, SE = 0.07, 95% CI [-0.41,-0.14], $p < 0.01$), and in 2019 pairs also fledged significantly fewer chicks than in 2014 and 2018 (n datapoints = 180, n pairs = 106; 2019 relative to 2014: $\beta = -0.32$, SE = 0.10, 95% CI [-0.53,-0.12], $p <0.01$; 2019 relative to 2018: $\beta = -0.14$, SE = 0.06, 95% CI [-0.25,-0.02], $p = 0.02$). Similarly, in 2019 the cumulative mass of fledglings per pair was significantly lower than in 2014, 2015 and 2018 (n datapoints = 180, n pairs = 105; 2019 relative to 2014: $\beta = -0.35$, SE = 0.10, 95% CI [-0.55,-0.16], $p <0.01$; 2019 relative to 2015: $\beta = -0.29$, SE = 0.07, 95% CI [-0.43,-0.15], $p <0.01$; 2019 relative to 2018: $\beta = -0.17$, SE = 0.06, 95% CI [-0.29,-0.06], $p <0.01$). In 2018, cumulative mass of chicks per pair was also lower than in 2014 (n datapoints = 180, n pairs = 105: $\beta = -0.19$, SE = 0.10, 95% CI [-0.39,-0.003], $p = 0.046$). Together these results show that 2019 was the hardest year for jackdaws in terms of reproductive success, followed by 2018. 2014 and 2015 were comparatively good years.

Supplementary Discussion

Re-pairing Previous work has found that the more familiar a pair are with one another (i.e., the longer they have known each other), the greater their reproductive success (Culina et al., 2020; Sánchez-Macouzet et al., 2014; Van De Pol et al., 2006). To understand whether pair familiarity influenced fitness outcomes in jackdaws, we tested whether re-pairing led to decreased reproductive success in the first year of a new partnership, relative to the final year of the former partnership. However, at odds with previous literature, we did not find an effect of pair familiarity on reproductive success (Table S5). Previous work on re-pairing rooks found that newly formed pairs were immediately as strong as the old pair-bond (Boucherie et al., 2018). This suggests that in corvids, pair-bond strength may not be a function of familiarity. Indeed, our analyses show that years together is not a predictor of pair-bond strength in jackdaws, again indicating that partner familiarity may not be related to pair-bond strength. However, although familiarity may be unimportant for reproductive success in jackdaw pair-bonds, it is also possible that this analysis did not fully capture pair familiarity. Because jackdaw pairs spend a large proportion of time with one another throughout the year (Kubitza et al., 2015; Röell, 1978; Wechsler, 1989), familiarity with new partners can be built up over the course of a year, thus meaning that 'newly' paired birds in our analysis were actually already familiar with one another by the time we recorded them in the breeding season. We must also consider that some individuals in our analysis may have divorced their first partner, rather than re-pairing due to partner death. Birds who re-pair following divorce often 'trade-up', and have improved fitness consequences with their new partner (Culina et al., 2015). Such a pattern could conceivably have created noise in our analysis, and divorce or re-pairing due to partner death would be important variables to control for in future analyses.

CHAPTER 5: Further methodological caveats of comparative studies of brain size

Abstract

There are multiple hypotheses for the evolution of cognition. The most prominent hypotheses are the Social Intelligence Hypothesis (SIH) and the Ecological Intelligence Hypothesis (EIH), which are often pitted against one another. These hypotheses tend to be tested using broad-scale comparative studies of brain size, where brain size is used as a proxy of cognitive ability, and various social and/or ecological variables are included as predictors. Here, we test how methodologically robust such analyses are. First, we quantify how much variation there is in brain and body size measurements across >900 species of bird, to understand how much variation in estimates there is between datasets. Following this, we subset our data to the Corvides infraorder and ask whether variable classification, source and inclusion influences model results. We demonstrate that there is substantial variation in brain and body size estimates across datasets, indicating that conclusions drawn from comparative brain size models are likely to differ depending on data source. We also show that model results change substantially depending on variable classification, source and inclusion. Indeed, we could have found support for either the SIH or EIH as a result of how we analysed our data. These results chime with recent concerns that comparative brain size studies do not give robust results. We add our voices to a growing community of researchers to suggest that we move on from such studies, and focus instead on understanding cognitive evolution at the intra-specific, behavioural scale.

Introduction

The principal drivers of cognitive evolution have been debated for decades (Barton, 1996; Dunbar, 1992; Dunbar & Shultz, 2017; Holekamp, 2007; Humphrey, 1976; Jolly, 1966; Rosati, 2017). Researchers often fall into two broad camps, focusing primarily on either social or ecological factors. Briefly, the Social Intelligence Hypothesis (SIH) posits that cognitive evolution is principally driven by the informational challenges of navigating a dynamic social environment, such as the need to track, anticipate and respond to the behaviour of social partners and monitor the relationships of others (Dunbar, 1998;

Humphrey, 1976; Jolly, 1966). In contrast, the Ecological Intelligence Hypothesis (EIH) emphasises informational challenges posed by ecological variables, such as variable food sources and climatic conditions (Barton, 1996; Clutton-Brock & Harvey, 1980; Harvey & Krebs, 1990). A large body of research has investigated social and ecological correlates of cognitive performance and neuroanatomy (e.g., Ashton et al., 2018; DeCasien et al., 2017; Dunbar, 1992; MacLean et al., 2009; Pérez-Barbería et al., 2007; Sayol et al., 2016; Shultz & Dunbar, 2007; Street et al., 2017; van Woerden et al., 2010; West, 2014), but results are often inconsistent and contradictory (Healy & Rowe, 2007; Logan et al., 2018; Powell et al., 2017; Wartel et al., 2019). For instance, Dunbar (1992), found that primate social group size positively correlated with a measure of brain size, which is commonly used as a proxy for cognitive ability. In contrast, DeCasien et al. (2017), found that diet is an important driver of primate brain size but social group size is not. Wartel et al. (2019), on the other hand, found that either diet or social group size could predict primate brain size, depending on specific methodological choices. While most research interrogating the SIH and EIH has focused on primates, birds have emerged as a major model system in cognitive evolution over the last 20 years (Güntürkün & Bugnyar, 2016; Iwaniuk & Arnold, 2004; Overington et al., 2009; Sayol et al., 2016; Seed et al., 2009). Some species of bird show convergent cognitive performance to primates (Güntürkün & Bugnyar, 2016; Seed et al., 2009), yet have divergent neuroanatomy (Güntürkün & Bugnyar, 2016) and differing constraints on brain size, such as those imposed by long-range migration (Vincze, 2016). Here we interrogate the potential pitfalls that arise in the comparative study of cognitive evolution in birds. Moreover, we highlight potential pitfalls of current methodologies that have not previously been investigated in any taxa, to date.

The relationship between brain size and cognitive ability is largely unknown and highly contentious (Chittka & Niven, 2009; Healy & Rowe, 2007; Logan et al., 2018); nevertheless, most studies investigating comparative cognitive evolution use some measure of brain size as a proxy of cognitive ability (Wartel et al., 2019). Most comparative studies of brain size use a single measurement of brain size per species, taken either from one “type specimen” individual or averaged across a number of individuals. The degree of intra-specific variation in brain size and its influences on the results of comparative analyses are therefore poorly understood. There is, however, evidence that the dataset used can result in substantially different conclusions for primate brain size studies (Wartel et al., 2019), suggesting that intraspecific variation and/or measurement technique may change results. For example, while social variables were not significantly correlated with primate brain size using one brain size dataset (DeCasien et al., 2017), an identical model using a different dataset did find a significant positive correlation (Wartel et al., 2019). Among birds, one comparative study of brain size evolution found that results did not differ when subsetting a combined dataset of brain mass (converted to volume) and endocranial volume to a single dataset on endocranial volume only (Overington et al., 2009). However, an explicit test of whether bird brain size dataset influences results has yet to be undertaken. Moreover, to control for the allometric relationship between brain and body size, most studies of brain size control for body size ('relative brain size') (Logan et al., 2018) but the impacts of variation in body size estimates from different datasets have not been investigated. Here, we quantify intraspecific variation in brain and body size measurements across bird species, to understand whether the magnitude of variation may influence the results of comparative brain size analyses. We then explicitly test whether results are robust when using different bird brain size datasets.

Beyond the influence of brain and body size estimates on model results, variable classification, source and combination are important considerations in models of brain size evolution. It is common practice in models to include covariates associated with the hypothesis of interest (broadly, the SIH or EIH), and either omit (e.g., Emery et al., 2007) or include less detailed (e.g. Sayol et al., 2016) variables associated with the competing hypothesis. However, the source and combination of variables are known to have a substantial influence on results in primate brain size models (Powell et al., 2017; Wartel et al., 2019). Moreover, the classification of variables can be a somewhat subjective decision made by authors. For example, in birds, if some populations of a species are migratory, but most are resident (e.g., as in jackdaws, *Corvus monedula*: Madge & de Juana, 2020), should they be classified as migrants, residents, or a more detailed covariate, in which case over-parameterisation may become a concern? How classification decisions influence model results is as yet unquantified, but such decisions have the potential to influence results.

Here, we use multiple datasets of brain and body size to quantify variation in estimates for >900 species of bird. Following this, we interrogate whether conclusions drawn from models testing alternative hypotheses for brain size evolution differ dependent on (1) the dataset of whole brain size and body size used and (2) the classification, source and combination of variables included. To do this, we collated detailed social and ecological variables for species in the Corvides infraorder; a relatively well-studied group of birds with well-resolved taxonomy (Jønsson et al., 2016) and large variation in brain size (Iwaniuk & Arnold, 2004; Sayol et al., 2016). Together, these investigations allow us to (i) identify novel pitfalls in the study of comparative cognition, and (ii) highlight parallel pitfalls to those previously identified in the field of primate comparative cognition (Powell et al., 2017; Wartel et al., 2019).

Methods

Quantifying intraspecific variation in brain and body size Whole brain volumes across bird species were collated from five published datasets (García-Peña et al., 2013; Iwaniuk et al., 2004; Iwaniuk & Arnold, 2004; Iwaniuk & Nelson, 2003; Sayol et al., 2016), all of which measured brain volume using either the endocranial volume technique (see Iwaniuk & Nelson, 2002 for details), brain mass converted to volume (Iwaniuk & Nelson, 2002) , or both. García-Peña et al., 2013, measured and analysed male and female brain sizes separately, and used one estimate of brain size per sex. All other studies used sex-averaged brain size and had one datapoint per species. We therefore also averaged male and female brain sizes from García-Peña et al., 2013. Datasets are non-independent, with some measurements shared between them; thus, quantification of variability between datasets is likely to be an underestimate.

Body sizes were collated from ten published datasets (Corfield et al., 2013; Fristoe et al., 2017; Garamszegi et al., 2002; Iwaniuk et al., 2004, 2005; Iwaniuk & Arnold, 2004; Lendvai et al., 2013; Minias & Podlaszczuk, 2017; Sayol et al., 2016; Sol et al., 2010). All body sizes were measured in grams. Again, datasets are not independent, with some using overlapping sources, and each dataset had one datapoint per species.

Altogether we collated information on 2399 bird species. Of these, 954 species had brain measurements in more than one dataset. 1546 species had body mass in more than one dataset.

Corvides: Analysis 1 – dataset dependency To test whether model results remained consistent across datasets, we collated whole brain and body size data from Iwaniuk & Arnold, 2004 (dataset 1) and Sayol et al., 2016 (dataset 2). For dataset 1, two methods were used to obtain brain size estimates: whole brain mass (converted into volume) and endocranial volume (measured using lead shot; see Iwaniuk & Nelson, 2002, for details). Where possible, the authors collected body mass data from the same specimens whose brain measurements were used. Where this was not possible, body mass measurements were collected from published literature. For dataset 2, endocranial volume was measured using the same method as used in dataset 1, but brain size and body mass were always measured from the same specimen. Altogether, we collated information for 40 species with brain size and body mass measurements from both datasets. We analysed the data as described in the section *Methods: Statistical Modelling,* using variables collected as described in *Methods: Variables.*

Corvides: Analysis 2 – variable inclusion, classification and source We tested whether including detailed ecological *and* social covariates, relative to including ecological *or* social covariates, qualitatively changed conclusions of models. To do this, we extracted/collated detailed ecological and social variables that have previously been shown to have a significant relationship with brain size (see *Methods: Variables*). In addition to constructing models with differing sets of predictor variables, we examined how sensitive model results were to choices regarding the categorisation of variables (see *Methods: Variables: Re-classification*). We also tested whether collecting variables from differing sources changed model results (see *Methods: Variables: Environmental variation*).

Variables We extracted/collated the following detailed ecological and social variables that have previously been shown to have a significant relationship with brain size.

1. Ecological variables We included species movement, environmental variability and diet. Species that migrate are thought to have smaller brains than resident species (Pravosudov et al., 2007; Shultz & Dunbar, 2010; Sol et al., 2010; Vincze, 2016). This is hypothesised to be because the energetic cost of the brain constrains selection on increased brain size in migrating species, who have large energetic demands during migration (Pravosudov et al., 2007; Sol et al., 2010; Vincze, 2016). Meanwhile, species that live in fluctuating environments (Fristoe et al., 2017; Sayol et al., 2016; Schuck-paim et al., 2008) and species with broader diets (Sayol et al., 2016) are thought to have bigger brains than those in more stable environments or with specialist diets. This is potentially because species that encounter more uncertainty must process more information in order to respond appropriately (Dall et al., 2005; Sayol et al., 2016; Schmidt et al., 2010), and therefore require more 'processing power' (i.e., bigger brains).

1.1 Movement We coded species movement using four categories: resident, partial migrant, migrant or nomadic. Previous studies including migration as a covariate tend to include migration as a binary variable (resident or migratory: Fristoe et al., 2017; Shultz & Dunbar, 2010). However, some species (e.g., the jackdaw, *Corvus monedula*) are only migratory in certain regions. Such species were therefore coded as partial migrants. Meanwhile, other species move often but do not undertake migrations; these were classified as nomadic. For the first analysis (comparing datasets), some categories had small sample sizes (nomadic = 3 species; migratory = 2 species). We therefore simplified this variable into resident and non-resident for *Corvides: analysis 1*, while we used variables of full complexity for *Corvides: analysis 2* (although see *Methods: Variables: Re-classification*).

1.2 Environmental variability We collected environmental variability from two sources. The first measure of environmental variability was 'temperature variation', as reported in Fristoe et al., 2017, where higher values indicate more variability. The second was a measure of environmental variability calculated by Sayol et al., 2016. Briefly, Sayol et al. included multiple environmental variables in a phylogenetic principal component analysis. The resultant phylogenetic principal component 1 (PPC1) captured seasonal variation, duration of snow cover and among-year variation, with higher values indicating higher variation, longer snow-cover and larger among-year variation. PPC1 can therefore be interpreted as an axis describing general environmental variation, with higher values at higher latitudes. Meanwhile, phylogenetic principal component 2 (PPC2) captured variation in vegetation at lower latitudes (e.g., drought events). Although Fristoe et al.'s 'temperature variation' is not as comprehensive as Sayol et al.'s 'PPCs', this data was available for more species in our sample. Temperature variation and PPCs were never used in the same models; instead, they were interpreted as two independent sources of 'environmental variation', which we used to quantify whether differing variable source may influence results.

1.3 Diet breadth We used diet breadth as reported in Sayol et al., 2016, who used Rao's quadratic entropy (de Cáceres et al., 2011) with diet frequency for seven diet types. Higher values indicate a broader diet.

2. Social variables We used two social variables in our models, both of which have been suggested to be involved in brain size evolution: group structure and cooperative breeding. While long-term monogamy has been shown to positively correlate with brain size (Emery et al., 2007; Shultz & Dunbar, 2010), almost all species in our sample form long-term monogamous pair bonds (see Supplementary Data), so there was not enough variation for this variable to be included.

2.1 Foraging group structure Foraging group structure has previously been shown to correlate with relative brain size (Emery et al., 2007; Shultz & Dunbar, 2010). Specifically, species that forage in pairs or bonded groups have been shown to have larger brains than those that forage in large aggregations (Shultz & Dunbar, 2010), and species that live in small groups have bigger brains than those that live in large aggregations (Emery et al., 2007). This is argued to be because the *quality* rather than *quantity* of social bonds is a key driver of cognitive evolution in birds (Emery et al., 2007; Shultz & Dunbar, 2010). However, in other studies foraging group structure appears to be unimportant (Sayol et al., 2016). A common problem with the inclusion of social variables in comparative studies is that they may not capture the underlying informational demands which, according to the SIH, drive cognitive evolution (Boucherie et al., 2019; Dunbar, 1998; Lukas & Clutton-Brock, 2018). We therefore expanded on previous categorisations of foraging group structure by trying to capture variables thought to be associated with information-processing. Specifically, species were coded as foraging solitarily, in pairs, in small groups (<30 individuals), in aggregations (>30 individuals), or as nested versions of these variables (e.g., forages in pairs nested within larger groups). If a species is known to forage in different social contexts but not necessarily in a nested fashion, we categorised these species using the largest group size commonly recorded (e.g., if the species forages in pairs and in small groups, but not necessarily in a nested manner, we recorded this as small group foraging). Following predictions of the SIH, we expected that solitary foragers would have the smallest brains, given the relatively limited demands for processing social information. Moreover, we expected that species foraging in nested groups would generally have larger brains, given the informational demands of managing relationships within a multi-layered context (e.g., managing the pair-bond relationship within a wider social group).

2.2 Cooperative breeding The role of cooperative breeding in cognitive evolution is contentious. Some authors argue that cooperative breeding entails substantial cognitive demands because individuals need to cooperate and coordinate with multiple others to raise offspring (Burkart et al., 2009; Burkart & van Schaik, 2009; Hrdy, 2009). Conversely, others suggest that the typically high levels of relatedness and shared interests within cooperatively breeding groups may in fact reduce cognitive demands relative to independent breeding (Lukas & Clutton-Brock, 2018; Thornton et al., 2016; Thornton & McAuliffe, 2015). Relevant empirical evidence remains limited and controversial. For instance, Burkart & van Schaik, 2009, suggest that cooperatively breeding monkeys show elevated socio-cognitive performance, but these species also have particularly small brains (Thornton & McAuliffe, 2015), and rank poorly in meta-analyses of cognitive performance across primates (Deaner et al., 2006). Among birds, the only comparative study to date found no relationship between cooperative breeding and brain size (Iwaniuk & Arnold, 2004), but this study did not include variables since shown to be significantly related to brain size, such as diet and environmental variation (Sayol et al., 2016). We therefore included cooperative breeding as a binary variable in our analyses. We note that species such as American crows (*Corvus brachyrhynchos*) and carrion crows (*Corvus corone*) are facultative cooperative breeders, but as there were few species in our sample that could be defined as such, we classified all facultative cooperative breeders as cooperative (but see *Methods: Variables: Re-classification)*.

Re-classification Some classifications are ambiguous and multiple different classifications can be justified. We therefore tested whether re-classifying variables changed model results. We re-classified one ecological and one social variable. “Partial migrants”, where at least one population of a species migrates but often most populations are resident, were re-classified as residents. Facultative or suspected cooperative breeders were re-classified as non-cooperative, rather than cooperative breeders.

Statistical modelling All statistical analyses were undertaken in R v4.0.2 (R Core Team 2017). We used a phylogenetic generalized least squares (PGLS) modelling framework (Freckleton et al., 2002) in the package *caper* (Orme, 2018), which controls for non-independence of datapoints due to relatedness and is the most commonly used technique in the comparative brain size literature (Fristoe et al., 2017; Fristoe & Botero, 2019; Sayol et al., 2016; Shultz & Dunbar, 2010; Sol et al., 2010; Vincze, 2016). We constructed a consensus tree by downloading 1000 equally plausible phylogenetic trees for the species in our sample from www.BirdTree.org (Sayol et al., 2016). We used the Hackett rather than Ericson backbone because it is the most recently constructed; however, differences between backbones are small and they tend to produce consistent results (Rubolini et al., 2015). Using TreeAnnotator in BEAST v1.10.4, a maximum clade credibility consensus tree was built from these equally plausible trees. This tree was then used to control for phylogenetic non-independence in the following PGLS models. Lambda was estimated using Maximum Likelihood. Model diagnostics and variance inflation factor (VIF) were checked to ensure assumptions were met and variables were not unacceptably collinear, respectively.

Corvides - analysis 1 To understand whether dataset had an influence on results, we subsetted our data to the 40 species with brain and body size data from both Iwaniuk & Arnold, 2004 (Dataset 1), and Sayol et al., 2016 (Dataset 2). We built four models using this data. First, we built model (1) using Iwaniuk & Arnold’s brain size data and their corresponding body size data (collected from both the specimens used for brain size measurement and from the literature). Second, we built model (2) using Sayol et al.'s brain and corresponding body size data (collected only from the same specimens that brain size was measured from). Next, we built models (3) using Iwaniuk & Arnold's brain size data and (4) using Sayol's brain size data, with *averaged* body mass estimates from across the two datasets. In all models, all ecological (movement, diet breadth, environmental variation) and social variables (social foraging, cooperative breeding) were used, with Fristoe et al.'s measure of environmental variability (rather than Sayol et al.'s) to maximise sample size. To test whether different datasets resulted in differing conclusions, we compared the output of models (1) and (2). We then compared the output of models (3) and (4), with averaged body size data, to test whether differences in brain size data alone, as opposed to the combination of different brain and body size estimates, qualitatively altered results.

Corvides - analysis 2 Here, we tested how variable classification (partial migrant/resident; cooperative breeder/non-cooperative breeder), source (environmental variability: temperature variation or PPC), and combination changed conclusions. See Table 1 for a summary of model formulations. We used Sayol et al.'s (2016) brain size data only, which consisted of 59 Corvides species. We chose to use this dataset because only one method was used to measure brain size, and all body mass data came from the same specimens that brain volume was taken from. It is therefore likely to be the most precise data currently available. Using this data, we built three models: an SIH model (brain size in response to body size, cooperative breeding and group structure), an EIH model (brain size in response to body size, migration, environmental variability and diet breadth), and a 'combined' model with all covariates included. For every model including environmental variation, we built two models: one with Sayol et al.'s measure of variability, and one with Fristoe et al.'s measure (a larger sample size; see Table 1). Furthermore, we tested each model with the initial and reclassified variables. All brain and body size measurements were log-transformed.

Results

Quantifying intraspecific variation in brain and body size Figure 1a visualises variation in (log-transformed) brain size estimates across datasets. There was considerable variation in brain size estimate within species. Of 954

Table 1 Different models testing a specific hypothesis (SIH/EIH) to test how variable inclusion changes conclusions drawn from results.

Hypothesis	**Species number**	**Model predictors**	**Conclusions**
Ecological Intelligence Hypothesis (EIH)	59	body size + diet breadth + environmental variation (temperature variation) + **movement**	1. Species movement correlates with brain size (residents have bigger brains than nomads) 2. Environmental variation is not a driver of brain size evolution
Social Intelligence Hypothesis (SIH)	59	body size + cooperative breeding + **social foraging**	1. Group structure during foraging is a driver of brain size evolution (non-nested small groups = bigger brains than foraging in pairs)
Combined (EIH + SIH)	59	body size + diet breadth + environmental variation (temperature variation) + **movement** + cooperative breeding + **social foraging**	1. Species movement correlates with brain size (residents have bigger brains than nomads) 2. Group structure during foraging is a driver of brain size evolution (non-nested small groups *and solitary foraging* = bigger brains than foraging in pairs)
EIH (PPC subset)	46	body size + diet breadth + environmental variation (**PPC1** + **PPC2)** + movement	1. Species movement *is not* correlated with brain size 2. Environmental variation *is* a driver of brain size evolution: more variation in snow cover and vegetative cover drives bigger brains
		body size + diet breadth +	1. Species movement *is not* correlated with brain size 2. Environmental variation *is* a driver of brain size evolution: more variation in

Combined (EIH + SIH; PPC subset)	46	environmental variation (**PPC1** + PPC2) + movement + cooperative breeding + **social foraging**	snow cover drives bigger brains but *not* vegetative cover 3. Group structure during foraging is a driver of brain size evolution (*nested* small groups *and solitary foraging* = bigger brains than foraging in pairs)

species in more than one dataset, 20% had variation equating to at least a 5% difference in brain mass and 7% of species showed a difference in brain mass of at least 10% between estimates. The maximum difference in mass was 40.52%, while the mean was 1.92±3.93%.

Figure 1b visualises variation in (log-transformed) body size estimates across datasets. Of 1546 species in more than one dataset, over one in five had variation between estimates equating to at least a 20% change in body mass. Meanwhile, 10% of species had variation equating to at least a 30% change in body mass. The average difference in estimate was 15.68±33.86%.

Corvides - ***analysis 1 (Table 2; Figure 2a)*** Datasets 1 and 2 consisted of the same 40 species with known social and ecological variables, and identical model formulations were used for both datasets. Results varied depending on whether we used brain and body size from dataset 1 or dataset 2, although effect sizes and confidence intervals were similar between models (Figure 2a). Model (1) (dataset 1) showed that species that forage in non-nested small groups tend to have larger brains than those that forage in pairs, but no other social or ecological variables were significant. Model (2) (dataset 2) also showed that species that forage in non-nested small groups tend to have larger brains than species that forage in pairs, but also found that solitary foragers have bigger brains than those that forage in pairs and that resident species have larger brains than migrants.

For models (3) (dataset 1) and (4) (dataset 2), we averaged body size across the datasets 1 and 2, but kept dataset-specific brain size as a response variable. Results for model (4) remained the same. For model (3), results differed: while species that forage in non-nested small groups still have bigger brains than species that forage in pairs, resident species also have significantly bigger brains than migrants.

See Table 2 for a summary of model results.

Corvides - ***analysis 2 (Table 3; Figure 2b)*** The full dataset used for this analysis included 59 species, where all species had known social and ecological variables (excluding PPCs). The PPC subset contained 46 species. Conclusions on the principal drivers of brain size evolution in Corvides differed dependent on modelling

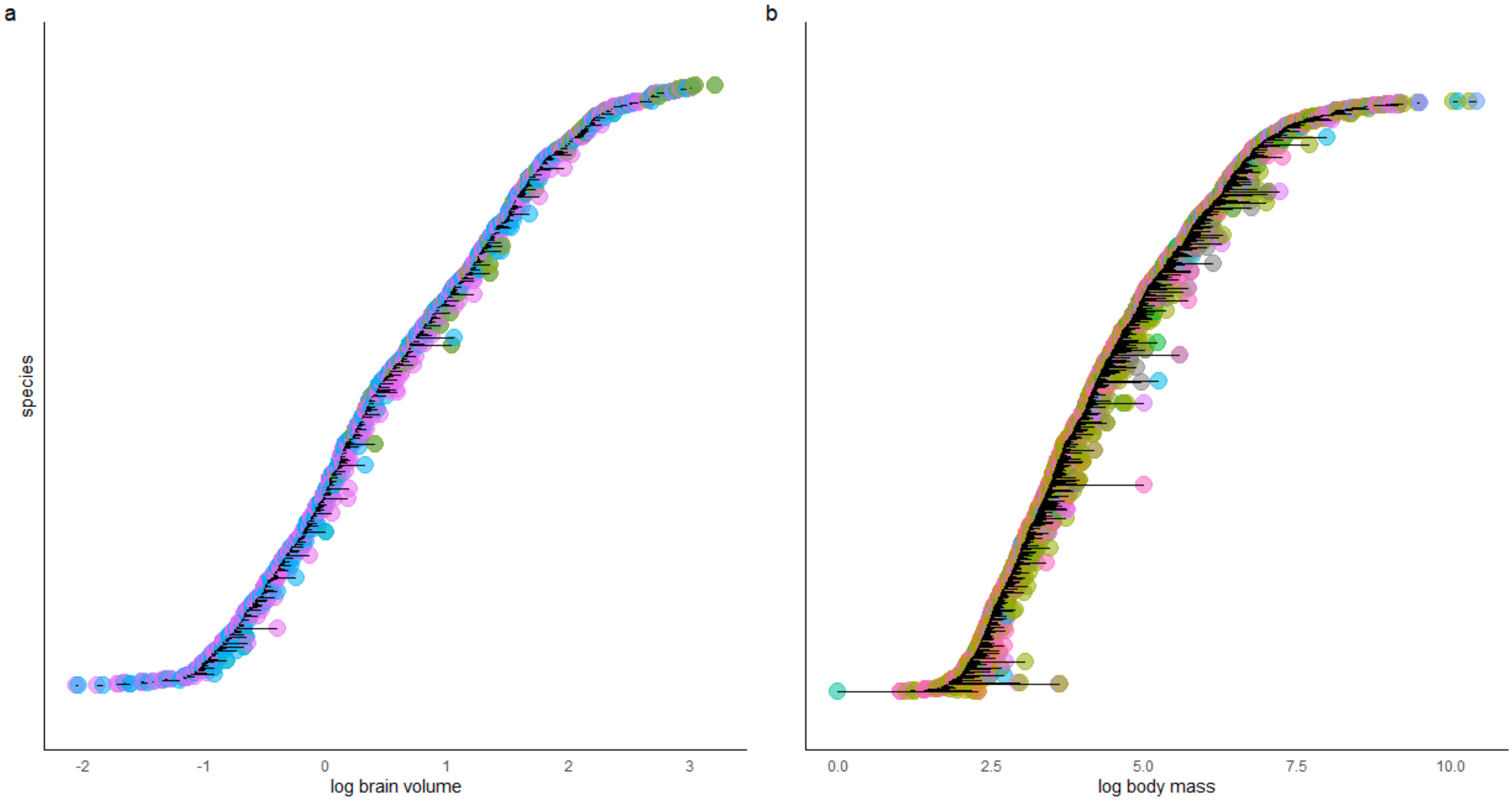


Figure 1 Log brain volume and log body mass, respectively, plotted for each species with more than one datapoint across datasets (N = 954 species for a; 1546 species for b). Points are ordered by minimum estimate, and lines connect points for the same species. Colour corresponds to dataset (for a, n = 5; for b, n = 10).

Table 2 Phylogenetic Generalised Least Squares model results, comparing species-matched models using two different brain size datasets. All significant pairwise contrasts for categorical variables are shown.

	Predictors	**λ**	**Estimate**	**Standard error**	**T-value**	**P-value**
Model 1 (Dataset 1)	**Body size**	**0.599**	**0.654**	**0.041**	**16.041**	**<0.001**
	Diet breadth		0.312	0.530	0.590	0.560
	Temperature variation		0.077	0.052	1.480	0.149
	Movement (resident)		0.120	0.061	1.954	0.060
	Cooperative breeding (binary)		-0.052	0.064	-0.805	0.427
	Social foraging (nested pairs)		0.160	0.094	1.706	0.098
	Social foraging (solitary)		0.189	0.100	1.892	0.068
	Social foraging (non-nested small groups)		**0.212**	**0.100**	**2.126**	**0.042**
	Social foraging (nested small groups)		0.098	0.138	0.710	0.483
Model 2 (Dataset 2)	**Body size**	**0.721**	**0.629**	**0.041**	**15.437**	**<0.001**
	Diet breadth		0.391	0.522	0.749	0.46
	Temperature variation		0.042	0.05	0.828	0.414
	Movement (resident)		***0.132***	***0.058***	***2.269***	***0.031***
	Cooperative breeding (binary)		-0.057	0.062	-0.91	0.37
	Social foraging (nested pairs)		0.185	0.094	1.965	0.059
	Social foraging (solitary)		***0.22***	***0.096***	***2.305***	***0.028***
	Social foraging (non-nested small groups)		**0.224**	**0.099**	**2.253**	**0.032**
	Social foraging (nested small groups)		0.152	0.134	1.136	0.265
Model 3 (Dataset 1, mean body size)	**Body size**	**0.672**	**0.642**	**0.042**	**15.291**	**<0.001**
	Diet breadth		0.247	0.538	0.460	0.649
	Temperature variation		0.079	0.052	1.500	0.144
	Movement (resident)		***0.127***	***0.061***	***2.084***	***0.046***
	Cooperative breeding (binary)		-0.064	0.065	-0.988	0.331
	Social foraging (nested pairs)		0.172	0.096	1.784	0.085
	Social foraging (solitary)		0.188	0.100	1.882	0.070
	Social foraging (non-nested small groups)		**0.228**	**0.102**	**2.231**	**0.033**
	Social foraging (nested small groups)		0.134	0.139	0.962	0.344
	Body size	**0.662**	**0.64**	**0.04**	**16.24**	**<0.001**
	Diet breadth		0.44	0.51	0.863	0.395
	Temperature variation		0.039	0.05	0.775	0.444

Model 4 (Dataset 2, mean body size)	***Movement (resident)***		***0.124***	***0.058***	***2.138***	***0.041***
	Cooperative breeding (binary)		-0.045	0.061	-0.732	0.47
	Social foraging (nested pairs)		0.175	0.091	1.918	0.065
	Social foraging (solitary)		***0.223***	***0.095***	***2.355***	***0.025***
	Social foraging (non-nested small groups)		**0.208**	**0.097**	**2.15**	**0.04**
	Social foraging (nested small groups)		0.118	0.132	0.893	0.379

approach (see Table 1 for a general summary, and Table 3 and Figure 2b for model results). For example, the EIH model with temperature variation as a proxy of environmental variation, rather than PPCs, suggests that environmental variation is not correlated with brain size, but that species movement is. When PPCs are used instead, environmental variation (both PPC1 and PPC2) but not species movement is significant. In the combined model with temperature variation included, species movement was again correlated with brain size, while PPC2 was no longer significant in the combined model with PPCs.

Table 3 Phylogenetic Generalised Least Squares model results, comparing different model formulations. All significant pairwise contrasts for categorical variables are shown.

Model type	Predictors	λ	Estimate	Standard error	T-value	P-value
EIH	**Body size**	**0.625**	**0.663**	**0.03**	**22.434**	**<0.001**
	Diet breadth		0.184	0.377	0.487	0.629
	Temperature variation		-0.019	0.039	-0.479	0.634
	Movement (partial)		0.133	0.086	1.542	0.129
	Movement (resident)		**0.186**	**0.083**	**2.227**	**0.030**
	Movement (migrant)		0.185	0.115	1.611	0.113
SIH	**Body size**	**0.698**	**0.668**	**0.029**	**22.685**	**<0.001**
	Cooperative breeding (binary)		-0.072	0.053	-1.358	0.180
	Social foraging (nested pairs)		0.079	0.061	1.292	0.202
	Social foraging (solitary)		0.111	0.075	1.473	0.147
	Social foraging (non-nested small groups)		**0.153**	**0.068**	**2.238**	**0.030**
	Social foraging (nested small groups)		0.062	0.111	0.560	0.578

Combined (EIH + SIH)	**Body size**	**0.662**	**0.658**	**0.031**	**21.565**	**<0.001**
	Diet breadth		0.161	0.389	0.414	0.680
	Temperature variation		0.000	0.038	-0.001	0.999
	Movement (partial)		0.070	0.863	0.812	0.421
	Movement (resident)		**0.173**	**0.082**	**2.0967**	**0.041**
	Movement (migrant)		-0.018	0.135	-0.132	0.896
	Cooperative breeding (binary)		-0.079	0.055	-1.430	0.159
	Social foraging (nested pairs)		0.120	0.061	1.966	0.055
	Social foraging (solitary)		**0.226**	**0.099**	**2.291**	**0.027**
	Social foraging (non-nested small groups)		**0.165**	**0.070**	**2.363**	**0.022**
	Social foraging (nested small groups)		0.112	0.114	0.981	0.332
EIH (PPC subset)	**Body size**	**0.948**	**0.605**	**0.034**	**17.952**	**<0.001**
	Diet breadth		0.392	0.398	0.984	0.331
	PPC1		**0.013**	**0.006**	**2.207**	**0.033**
	PPC2		**0.031**	**0.015**	**2.072**	**0.045**
	Movement (partial)		0.054	0.061	0.887	0.381
	Movement (resident)		-0.006	0.085	-0.075	0.940
Combined (EIH + SIH; PPC subset)	**Body size**	**0.913**	**0.600**	**0.033**	**18.269**	**<0.001**
	Diet breadth		0.451	0.396	1.139	0.263
	PPC1		**0.012**	**0.006**	**2.075**	**0.046**
	PPC2		0.028	0.015	1.903	0.066
	Movement (partial)		0.013	0.064	0.209	0.836
	Movement (resident)		0.089	0.092	0.964	0.342
	Cooperative breeding (binary)		-0.079	0.056	-1.409	0.168
	Social foraging (nested pairs)		0.107	0.068	1.567	0.126
	Social foraging (solitary)		**0.357**	**0.156**	**2.283**	**0.029**
	Social foraging (non-nested small groups)		0.124	0.070	1.770	0.086
	Social foraging (nested small groups)		**0.267**	**0.121**	**2.202**	**0.035**

In the SIH model and combined models, cooperative breeding was consistently non-significant. Social foraging was consistently significant when included as a predictor, but the significant pairwise contrasts changed across models, leading to inconsistent conclusions about which type of social foraging is correlated with bigger brains. For example, the SIH model suggests that non-nested small group foragers have bigger brains than species that forage in pairs, the combined full model suggests that solitary and non-nested small group foragers have bigger brains than species that forage in pairs, and the combined PPC model suggests that solitary and *nested* small group foragers have bigger brains than species that forage in pairs.

Reclassification of ambiguous variables Changing facultative or suspected cooperative breeders from their initial categorisation of cooperative breeders to non-cooperative breeders did not qualitatively change SIH model results. However, changing partial migrants to residents did change EIH model results. For the EIH model with PPCs (see Table 1), PPC1 and PPC2 went from significantly influencing brain size (PPC1: $\beta = 0.01$, $SE = 0.01$; $p = 0.03$; PPC2: $\beta = 0.03$, $SE = 0.02$; $p = 0.04$) to having no significant effect (PPC1: $\beta = 0.01$, $SE = 0.01$; $p = 0.055$; PPC2: $\beta = 0.03$, $SE = 0.01$; $p = 0.07$).

In combined models, where both changed variables were included, the combined model without PPCs showed solitary foraging (relative to foraging in pairs) changing from a significant ($\beta = 0.23$, $SE = 0.10$; $p = 0.03$) to no significant effect on brain size ($\beta = 0.16$, $SE = 0.10$; $p = 0.11$). In the combined model with PPCs, PPC2 changed from no significant ($\beta = 0.03$, $SE = 0.01$; $p = 0.07$) to a significant effect on brain size ($\beta = 0.03$, $SE = 0.01$; $p = 0.047$).

Discussion

In agreement with a growing body of literature (Healy & Rowe, 2007; Logan et al., 2018; Powell et al., 2017; Wartel et al., 2019), our analyses raise concerns that comparative brain size studies are not methodologically robust. We show that there is considerable variation in bird brain and body size estimates across datasets (Figure 1), most likely due to intraspecific variation in brain and body size. This has the potential to substantially influence results. Models using brain size estimates from different datasets give differing results (Figure 2a) and results change again when dataset-specific body size is used (Figure 2a). The classification, source and combination of social and ecological variables used also substantially changes results (Figure 2b). Indeed, we could have found support for either the social or ecological intelligence hypothesis, depending on how we classified and sourced variables, and which we chose to include. Our results chime with and add to concerns raised in the primate brain size literature (Powell et al., 2017; Wartel et al., 2019) that current methods in the comparative study of brain size evolution are not methodologically sound.

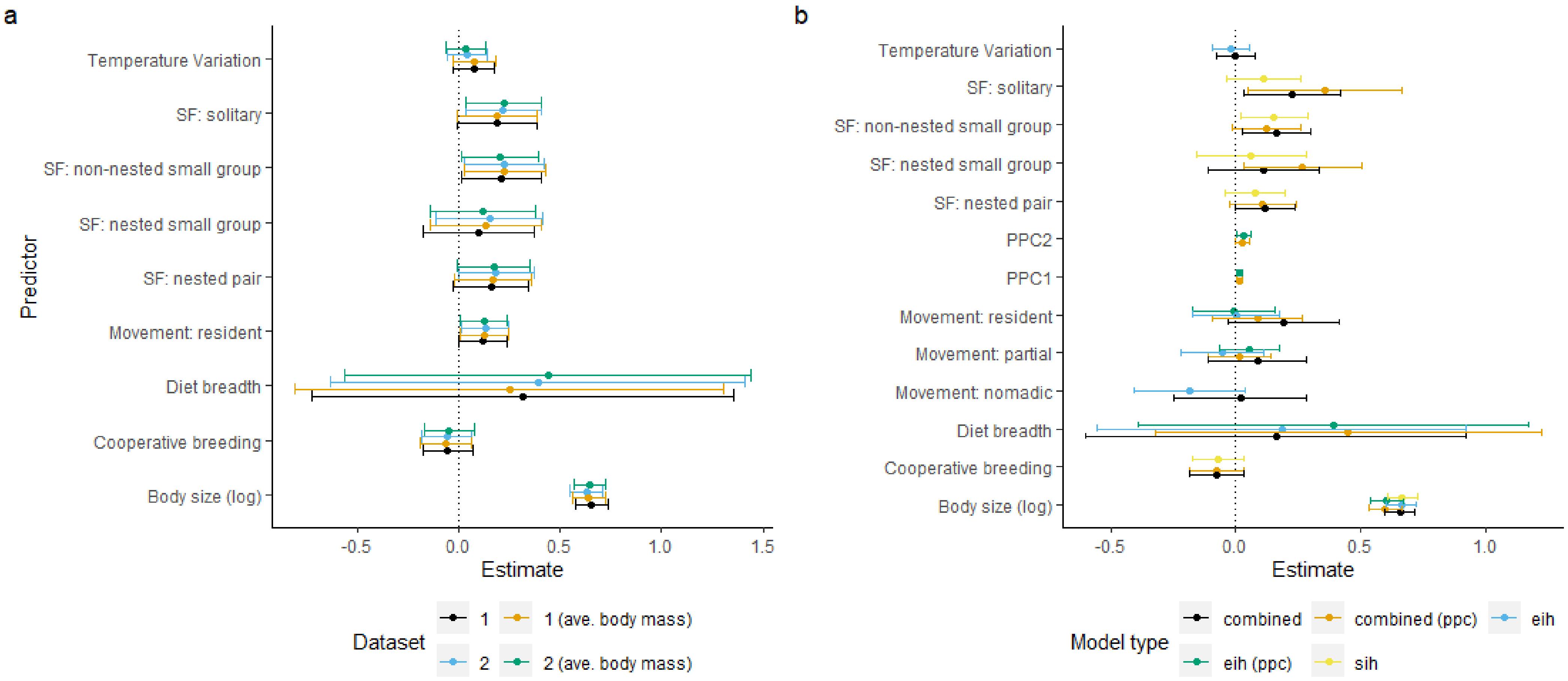


Figure 2 Estimates and confidence intervals for a) analysis 1 (comparing brain size datasets in Corvides) and b) analysis 2 (comparing models with differing variable sources and combinations in Corvides). SF is an abbreviation of Social Foraging.

Comparative brain size studies typically use brain size averaged from multiple specimens of a single species to obtain one brain size estimate per species. Often, important yet key information such as the number, sex and source population of the specimens is not reported despite their potential influence on the final estimate. Despite all brain sizes being estimated either from endocranial volume, or brain mass converted to volume (which results in a strong positive correlation with volume measurements (Iwaniuk & Nelson, 2002)), we found substantial variation in brain size estimates. One in five of 954 species showed at least a 5% change in brain volume estimate between datasets, equating to, on average, an overlap with two other species in our sample. Body size estimates rarely have any associated information reported, such as source or number of specimens the estimate is derived from, yet have the potential to influence results to the same magnitude as brain size estimates when included as a control variable. We also found substantial variation in body size across datasets: one in ten of 1546 species showed at least a 30% change in body size between species, again corresponding to an average overlap with two other species. Variation in brain and body size was not driven by any dataset in particular (Figure 1), suggesting that this was not the result of a specific methodological approach, but rather the result of natural intraspecific variation in brain and body size, and the use of different specimens by different studies. This result clearly highlights the caveats of using one estimate of brain and body size per species when specimen sample sizes are limited and associated information is not accounted for. Conclusions are likely to be dependent on the specific specimen individual(s) from which brain and body size measures are taken.

Following the quantification of variation in brain and body size estimates across bird species, we tested whether using different datasets changed results for a subset of species (those in the infraorder, Corvides). Many studies combine brain size measurements by converting between mass and volume. Here, we show that despite a high correlation between measurements, a dataset combining brain volume and (converted) mass (Iwaniuk & Arnold, 2004) gives significantly different results to a dataset containing brain volume only (Sayol et al., 2016), even when body mass estimates are identical between models. Given high correlations between measurement techniques (Iwaniuk & Nelson,

2002) and no pattern of particular datasets driving variation in results (Figure 1), it is unlikely that such discrepancies are due to differences between measurement types. A more likely explanation is that variation is driven by the use of different specimens between datasets. Using dataset-specific body mass further changed results, indicating that variation in both brain and body size estimates can have a substantial influence on results. However, it is also important to note that effect sizes and confidence intervals were very similar across differing datasets (Figure 2a). Relying on p-values when effect and sample sizes are small is likely to result in inconsistent results even when datasets are similar. However, our findings are consistent with analyses of dataset sensitivity in the primate brain size literature (Wartel et al., 2019). Furthermore, notwithstanding the reduction in sample size, there are important advantages to modelling only closely related species. Specifically, modelling brain size for closely related species reduces interference introduced by studying evolutionarily divergent species (Logan et al., 2018). As a specific example, the intercept and slope of the allometric relationship between brain and body size varies substantially among taxa (Ksepka et al., 2020; Smaers et al., 2021). Studying a smaller group of closely related species means that this source of variation is somewhat mitigated. Nonetheless, even when we limit the variation by focusing on a particular group, discrepancies across datasets can generate different conclusions.

As well as important effects of the dataset used, we also found that variable classification, source and inclusion substantially influenced results. Some classifications of variables are ambiguous; for instance, species with both cooperatively and non-cooperatively breeding populations could be classified as either. We therefore changed categorical variables that could justifiably be re-classified, and tested how this influenced results. Re-classifying suspected/facultative cooperative breeders as non-cooperative breeders did not change SIH model results; however, re-classifying partial migrants (i.e., where at least one population of a species migrate) as residents substantially changed EIH model results. While two measures of environmental variation (PPC1 and PPC2) were significantly associated with bigger brains before re-classification, there was no significant effect following re-classification. In the combined models, where both re-classified variables were included, results changed yet again. Thus, how variables are classified has the potential to substantially changes results. Similarly, variable source significantly influenced model results. Using temperature variation from Fristoe et al. (2017) as a proxy of environmental variation resulted in no support for environmental uncertainty driving the evolution of bigger brains. Meanwhile, using more detailed measures of environmental variation from Sayol et al. (2016) resulted in strong support (although this support was absent when "partial migrants" were re-classified). Note, however, that models with temperature variation rather than PPC had a larger sample size, so this result may be somewhat driven by sample size. Nevertheless, these results parallel those reported in the primate brain size literature (Powell et al., 2017), where using differing variable sources resulted in differing results even when sample sizes were matched. Similar to Wartel et al.'s (2019) analyses of primate brain size data, we found that variable combination also substantially changed results. For example, depending on which covariates were included in models, we could have concluded that species that forage in *non-nested* small groups tend to have larger brains than those that forage in pairs (SIH model only), that species that forage in *non-nested* small groups *and* forage solitarily tend to have larger brains than those than forage in pairs (SIH + EIH model), or that species that forage in *nested* small groups and solitarily tend to have bigger brains than those that forage in pairs (SIH + EIH with a different and more detailed measure of environmental variability). Taken together, our investigation into variable classification, source and combination shows that the SIH or EIH could be supported or refuted depending on somewhat arbitrary decisions made by researchers. This throws previous claims of support for the SIH (e.g., Dunbar, 1992; Shultz & Dunbar, 2010) or the EIH (e.g., DeCasien et al., 2017) into question. In addition to issues with methodological approach, we argue that framing the SIH and EIH as dichotomous hypotheses is not logically sound, given that (i) the hypothesised underlying driver of cognitive evolution, informational uncertainty, is shared for both, and (ii) social and ecological variables are not independent, i.e., social species solve ecological problems in a social context, and sociality itself may evolve in response to ecological variables (Ashton, Thornton, et al., 2018; Jetz & Rubenstein, 2011). We therefore suggest not only that our methodological approach to studying comparative brain size evolution needs to change, but also the conceptual framework itself. This is discussed further in Chapter 6.

In this study, we show that there is substantial variation in both bird brain and body size estimates across datasets, and suggest that this variation has the potential to alter the results of comparative analyses substantially. Furthermore, we show that the classification, source and combination of variables included in models can change results so that either the SIH or EIH could be supported or refuted, which parallels and adds to concerns of model instability in the primate literature (Powell et al., 2017; Wartel et al., 2019). When considering the accumulating literature on issues associated with comparative studies of brain size evolution (here; Healy & Rowe, 2007; Logan et al., 2018; Powell et al., 2017; Wartel et al., 2019), we add our voices to a growing number in the field suggesting that we move away from such methods, and towards a more robust approach that researchers may wish to consider instead of large-scale comparative brain size models. For instance, intra-specific tests of how specific variables relate to cognitive performance. In addition, we recommend a shift away from treating the SIH and EIH as dichotomous hypotheses, and working instead to understand the sources of informational uncertainty and how these relate to cognitive ability.

CHAPTER 6: General Discussion

Why there is so much variation in cognitive ability has been a major source of scientific debate for decades. One hypothesis for why cognition evolves is the Relationship Intelligence Hypothesis (RIH). The RIH was formulated in response to an apparent relationship between long-term pair-bonds and big brains in birds. The RIH proposes that the management and maintenance of such pair-bonds is cognitively demanding, and better socio-cognitive ability results in higher quality (“stronger”) pair-bonds (Emery et al. 2007). The RIH also posits that stronger pair-bonds result in higher fitness, and thus that the management and maintenance of pair-bonds is a driver of (socio-)cognitive evolution (Emery et al., 2007).

The RIH was formulated more than a decade ago, and has gained some traction in the field of cognitive evolution. For instance, the original paper has been cited >350 times. Despite this, key predictions of the RIH have never been tested, and multiple predictions have never been tested in a single study system. In the previous chapters, I used the jackdaw as a wild study system to test key predictions of the RIH at the intra-specific scale (Chapters 2 – 4). Subsequently, I examined whether the current methodology employed in broad-scale comparative brain size studies, the results of which are the basis for the formulation of the RIH, are robust (Chapter 5). In this final chapter, I discuss my findings within the context of the RIH, while critically evaluating the RIH in light of contemporary empirical evidence and theoretical work. Specifically, I ask:

1. Is the RIH built on strong empirical foundations?
2. Is there evidence to support the predictions that pair-bond strength is variable between pairs, consistent within pairs and heritable?
3. Is pair-bond strength positively related to fitness outcomes?
4. Is pair-bond management and maintenance socio-cognitively demanding
    4.1. within the pair-bond?
    4.2. within the wider social landscape?
5. Is the RIH empirically and theoretically supported, and what are future directions in the field?

Finally, I address the issue of anthropomorphism and anthropocentrism in the field of cognitive evolution, and suggest ways we can attempt overcome our inherent biases.

The RIH: a critical evaluation

1. *Shaky foundations?*

In their 2007 paper introducing the RIH, Emery et al. argued that relative brain size in birds tends to be largest for both long-term monogamous species and cooperative breeders (Emery et al., 2007). It was this finding that led to the proposal that cooperation between long-term partners, including in cooperatively breeding societies, may generate selection on cognitive ability. Despite using more than 400 species of birds, with widely variable phylogenetic relatedness, Emery et al. did not conduct a phylogenetically controlled analysis. However, a previous study had conducted a phylogenetically controlled comparative study interrogating the relationship between brain size and cooperative breeding in the parvorder Corvida (Iwaniuk & Arnold, 2004), and no relationship was found. Moreover, subsequent to Emery et al.'s initial analysis, several studies implemented phylogenetically controlled analyses to examine the relationship between bird brain size and various social and ecological covariates, with contradictory results. For instance, one study found no relationship between mating system and brain size in birds (Sayol et al., 2016), while another found mating system to be an important predictor (Shultz & Dunbar, 2010). In Chapter 5, I show that results of comparative brain size studies are not robust, and it is therefore not surprising that the field is dominated by inconsistent results. Indeed, results from Chapter 5 chime with and extend concerns raised in other studies (Logan et al., 2018; Powell et al., 2017; Wartel et al., 2019) that the current approach to comparative brain size studies is methodologically flawed. Beyond methodology, a conceptual issue with such studies is that we do not actually know whether brain size approximates cognitive ability (Healy & Rowe, 2007); thus, even if results were consistent, whether we could infer anything about cognitive evolution is unclear. Taken together, this body of work demonstrates not only why we should move away from comparative studies of brain size to examine cognitive evolution, but also that we should interpret results of previous such studies with extreme caution. Indeed, an intraspecific approach to studying cognitive evolution, focussed on animal behaviour rather than broad-scale proxies of cognitive ability, is likely to be a more fruitful way forward.

In addition to the comparison of relative brain size across hundreds of bird species, Emery et al. (2007) also examined the relationship between brain and body size for a select few species. They selected 18 species with long-term monogamous pair-bonds to demonstrate that even within these species, there is considerable variation in relative brain size. The taxa selected were species of geese, albatrosses, corvids and parrots. By plotting brain size against body size for these species, Emery et al. showed that corvids and parrots tend to have bigger brains for their body size than the other species. They went on to suggest that this difference in relationship is because pair-bonds in corvids and parrots are more complex than those observed in albatrosses and geese. Specifically, they argue that pair-bonds in corvids and parrots are complex because they are characterised by high levels of affiliation, maintaining the bond year-round, and the need to cooperate and coordinate to raise altricial offspring. They argue that these factors generate selection for cognitive ability, and that such selection pressure is not present in species with more “simplistic” pair-bonds. Here, simplistic refers to pair-bonds that are characterised by little affiliation (e.g., only mutual proximity), and no need for high levels of cooperation to successfully raise offspring. There are three key points to consider in response to these assertions.

First, this is an arguably weak analytical approach to the question. The sample size is small, “*the choice of these species is rather arbitrary*” (Emery et al. 2007, p. 494), and no phylogenetic control is undertaken despite a mixture of closely and distantly related species being included. Second, alternative explanations for the observed relationships are not adequately addressed. Although Emery et al. (2007) acknowledge that differential foraging ecology is a potential alternative explanation for the variation in brain size, the idea is not discussed in detail. They merely suggest that the large range size of geese (who migrate) and albatrosses (who undergo vast journeys to find food) would be more likely to result in the opposite of the observed relationship. Another alternative explanation, which is not discussed, is that long journeys may impose energetic constraints on the evolution of brain size. Indeed, there is some evidence that migratory species have smaller brains (Vincze, 2016). It is therefore possible that the relationship observed by Emery et al. (2007) could in fact be explained by differences in ecological constraint, rather than in pair-bonding behaviour. Third, Emery et al. make a rather sweeping generalisation regarding the "simplistic" pair-bonding behaviour of geese and albatrosses. As addressed in detail by Scheiber et al. (2008), geese can show what Emery et al. regard as more complex behaviours. For instance, greylag geese (*Anser anser*) partners coordinate with one another by synchronising behaviours (Nedelcu & Hirschenhauser, 2010), and actively support one another in agonistic interactions (Scheiber et al., 2005). In albatrosses, little research has been done to understand the characteristics of the pair-bond. However, a recent study suggests that in black-browed albatross (*Thalassarche melanophris*), pair-bonds are characterised by affiliative behaviour over and above mutual proximity (Gillies et al., 2021). Specifically, partners allopreen one another, and preliminary data suggest that allopreening rate may be associated with coordination of biparental care (Gillies et al., 2021). Together, these arguments somewhat undermine the point that corvid and parrot relationships are qualitatively different to, and more cognitively demanding than, relationships observed in smaller-brained species.

Returning then to the case of the RIH, the relationships that underpin the hypothesis – the correlation between long-term monogamous pair-bonds and brain size, and the different grades of brain to body size allometry between a subset of taxa - are questionable, as are the conclusions drawn from them. It is therefore clear that the original formulation of the RIH rests on rather shaky empirical ground. Nevertheless, the hypothesis has not in any sense been disproved, merely shown to be built on flimsy foundations. It is thus important to interrogate contemporary empirical and theoretical support at the intra-specific scale, especially because since the RIH was originally proposed, a large body of relevant research has been undertaken.

2. *Is pair-bond strength variable between pairs, consistent within pairs, and heritable?*

The RIH posits that better socio-cognitive ability is under selection because it allows individuals to better manage and maintain their pair-bonds, which in turn increases fitness consequences. For this to occur, an individual's ability to form strong bonds must (i) vary between individuals, (ii) be consistent and (iii) be heritable. If these predictions are not met, then selection cannot act on socio-cognitive ability through its influence on pair-bond strength, and thus this cannot be the mechanism through which cognition evolves. However, measuring an individual's *ability* to form strong bonds directly is not possible. An approximation of this is to measure the strength of the pair-bond that the individual forms. However, it must be noted that added complexities are introduced here because the trait becomes a measure of dyadic behaviour, dependent on the phenotypes of two individuals. Nonetheless, we would expect that an individual's ability to form strong pair-bonds, and the strength of the pair-bonds they form, to correlate. Thus, testing whether pair-bond strength (i) varies between pairs, (ii) is consistent and (iii) is heritable should allow for a better understanding of whether an individual's ability to form strong bonds meets any of these predictions.

In Chapter 4, I tested whether pair-bond strength varies between pairs and is consistent within pairs in wild jackdaws. Variation in pair-bond strength has been quantified using an array of behavioural measures only in captive cockatiels (*Nymphicus hollandicus:* Spoon et al., 2004, 2006, 2007), captive rooks (*Corvus frugilegus:* Boucherie et al., 2018), captive zebra finches (*Taeniopygia guttata*; Elie et al., 2011) and wild red avadavats (*Amandava amandava*; Sparks, 1964). Although captive studies are of course valuable, rates of affiliative behaviours have been shown to vary between wild and captive settings (Brummer et al., 2010; Inoue & Shimada, 2020; Pacheco & Madden, 2021). This is most likely to be because of differing time budgets (e.g., captive animals need not invest much time in food-searching and so have more time for socialising) (Inoue & Shimada, 2020), as well as living within spatially and socially restricted environments when captive (Brummer et al., 2010). Thus, studying pair-bond dynamics in the wild is vital to understand natural levels of variation. It is therefore important to note that studies on both captive (Boucherie et al., 2018; Elie et al., 2011; Spoon et al., 2004, 2006, 2007) and wild (Chapter 4; Sparks, 1964) birds show substantial variation in pair-bond strength. Thus, there is strong support for the RIH's prediction that pair-bond strength varies between pairs.

We also tested the prediction that pair-bond strength is consistent within pairs. I found that it is indeed significantly repeatable in wild jackdaws. The repeatability of jackdaw pair-bond strength was slightly higher than that reported for wild greylag geese (Kralj-Fišer et al., 2007), where bond strength was measured using a single dimension - proximity of partners; similar to the repeatability of social group size and social association strength in wild great tits (*Parus major*) (Aplin et al., 2015); and higher than the average repeatability of behaviour in general (Bell et al., 2009). This was true for both within- and between-year measures, but only during a specific stage of the breeding season (the incubation stage). More data is necessary to interrogate the consistency of pair-bond strength at other stages. Although consistency of pair-bond strength in particular has only been examined in jackdaws and greylag geese, both of which have long-term pair-bonds, there is some evidence that pair-bond strength may also be consistent in species with shorter-term bonds, such as great tits. Great tits pair with one partner throughout the breeding season, and may sometimes pair with the same partner again in the following season if both individuals survive (Gosler et al., 2020). However, great tits are short-lived birds, with only around half of adults surviving the winter season each year (Gosler et al., 2020). They therefore tend not to maintain long-term pair-bonds. In wild great tits, personality is repeatable and associated with relative association strength (based on co-occurrence at feeders) between partners (Firth et al., 2018), where bolder individuals associate relatively more with their partner compared to non-partners. This suggests that pair-bond strength (or at least one facet of pair-bond strength) is likely to be repeatable in species with short-term pair-bonds, too. Thus, current data suggests that pair-bond strength is repeatable in at least some species with long-term pair-bonds, but that repeatable pair-bond strength may not be exclusive to species with long-term partnerships.

For pair-bond strength to be a substrate of selection, it must be heritable. Whether an individual’s ability to form strong bonds is heritable has not been investigated in birds. In mammals, humans (*Homo sapiens*) (Fowler et al., 2009), marmots (*Marmota flaviventris*) (Lea et al., 2010) and rhesus macaques (*Macaca mulatta*) (Brent et al., 2013) have been shown to have heritable social network metrics, indicating that there is a genetic basis to an individual’s ability to form bonds. However, there are criticisms of the techniques used in these studies. Specifically, indirect genetic effects, which are the influence of other individuals' genotypes on the phenotype of the focal individual, were not quantified, yet they are likely to influence heritability estimates to some degree because social behaviour inevitably involves the interaction of two or more genotypes (Wilson et al., 2009). A recent study, however, quantified heritability of social network position in drosophila (*Drosophila melanogaster*) controlling for the genotypes of social partners, and found that social network position is heritable (Wice & Saltz, 2021). Whether an individual's ability to form a strong bond (as opposed to social network position) is heritable has only been investigated in prairie voles (*Microtus ochrogaster*), a species of rodent. In this species, males often form monogamous pair-bonds with females (Vogel et al., 2018). However, while pair-bonding is known to have a genetic basis in these voles (Sadino & Donaldson, 2018), a male's preference for spending time with his partner rather than another female showed very low heritability (Vogel et al., 2018). While this suggests that pair-bonding behaviour may not heritable, it could also be the case that the one specific measure of pair-bonding used did not fully capture pair-bonding behaviour. Indeed, prosocial behaviours, which are a key component of bond strength, seem to be highly heritable in humans (Knafo et al., 2011), which would suggest that bonding behaviour is heritable. Taken together, it is clear that more work is necessary to understand whether an individual's ability to form strong social bonds is heritable, but current evidence tentatively supports that it may be.

If an individual's ability to form bonds is heritable, this could potentially arise from non-cognitive factors; for example, endocrine profiles (Burkett et al., 2016). If the RIH holds, the ability to form strong bonds should (at least in part) be mediated by cognitive processes, and these cognitive processes should also be heritable. There are few estimates of heritability for socio-cognitive abilities. In humans, empathetic concern (Melchers et al., 2016; Warrier et al., 2018) and facial recognition (Wilmer et al., 2010; Zhu et al., 2010) have both been shown to be moderately heritable, while studies in non-human species have demonstrated that some cognitive abilities, such as reversal learning (Sorato et al., 2018) and inhibitory control (Gnanadesikan et al., 2020; Langley et al., 2020) are heritable. However, little is known about socio-cognitive heritability in non-humans. This is a key area of future research.

3. *Do stronger pair-bonds lead to better fitness outcomes?*

For the RIH to hold, stronger pair-bonds must result in better fitness outcomes. Although a previous study has found pair-bond strength to positively correlate with reproductive success (Spoon et al., 2006), this study was undertaken on a captive population. Testing fitness consequences in the wild is crucial, given that without examining a behaviour in its ecological context, it is difficult to know in which contexts it would naturally evolve (Boesch, 2020; Cauchoix et al., 2020; Pritchard et al., 2016; Thornton & Lukas, 2012). In Chapter 4 I investigated whether pair-bond strength influences fitness in the wild. I did not find a linear or quadratic relationship between pair-bond strength and reproductive success in wild jackdaws, suggesting that pair-bond strength was not under either directional or stabilising selection at the time I conducted our study. It is important to note, though, that jackdaws are long-lived, and have relatively little variation in reproductive success within a year (Chapter 4). It is thus not completely clear whether there could be an influence of pair-bond strength on lifetime reproductive success, which is ultimately what matters in terms of fitness. I did, however, find that pairs with stronger bonds were better able to adjust hatching synchrony to environmental conditions. While this did not have any detectable effect on reproductive success, it is possible that within the scope of our study, environmental conditions did not vary enough for any such effect to be evident. On the other hand, pair-bond strength may not be important for fitness outcomes in jackdaws, which would undermine a central tenet of the RIH.

At a broader scale than fine-scale pair-bond strength, several studies find an effect of partner choice, partner familiarity and pair-bond duration on fitness outcomes, which are all likely to be important facets of pair-bond strength. For example, zebra finches (*Taeniopygia guttata*) and convict cichlids (*Amatitlania siquia*) have higher reproductive success when they are able to choose their partner (Ihle et al., 2015; Laubu et al., 2019); great tits who meet earlier in the pre-breeding season, and thus are more familiar before breeding, produce more fledglings (Culina et al., 2020), and blue-footed boobies (*Sula nebouxii*) fledge more offspring when they have been together for longer (Sánchez-Macouzet et al., 2014). In Chapter 4, I did not find an effect of re-pairing on jackdaw fitness outcomes, indicating that the length of the pair-bond is not an important predictor of fitness in this species. A potential explanation for this is that jackdaws stay with their partner year-round, so may become familiar with their partner extremely quickly, relative to species that spend time with their partner only during the breeding season. Taken together, current evidence suggests that pair-bond dynamics may have fitness consequences in multiple species, which means that a core prediction of the RIH is plausible. However, of the two studies explicitly testing how pair-bond strength influences fitness consequences (Spoon et al., 2006; Chapter 4), only one study, on captive cockatiels, finds an effect (Spoon et al., 2006). I suggest that further work is necessary to understand the link between pair-bond strength and fitness outcomes. Moreover, focussing on wild species in particular is vital for our understanding of if and when strong pair-bonds offer an adaptive advantage (Pritchard et al., 2016; Thornton & Lukas, 2012). Longitudinal studies would be of particular value, given that long-term monogamous species tend to be long-lived, and understanding lifetime rather than annual reproductive success would be key to quantifying actual fitness outcomes.

4. Is pair-bond management and maintenance socio-cognitively demanding?

Perhaps the most central tenet of the RIH is that the management and maintenance of a pair-bond is cognitively demanding. In particular, the management and maintenance of long-term, year-round pair-bonds is posited to be more demanding than managing and maintaining shorter-term relationships, or relationships where individuals do not stay with one another year-round. Here, I evaluate current empirical and theoretical support for this prediction.

4.1 Socio-cognitive demands of managing the pair-bond

In their formulation of the RIH, Emery et al (2007) suggest that for individuals in a pair-bond, *“keeping track of the accumulating, subtle behavioural characteristics of a bonded partner over the course of a relationship requires a … form of relationship intelligence, which enables them to accurately read the social signals of their partner [and] respond appropriately to them”*. Such responsiveness requires socio-cognitive ability because it necessitates that an individual attends to the actions of another, and uses this information to inform their own behaviour (Wascher et al., 2018).

An example of partner responsiveness, as proposed by Emery et al. (2007), is post-conflict third-party affiliation. This is where, following a conflict, an observer of the conflict (or “bystander”) offers affiliation to another in distress (Pérez-Manrique & Gomila, 2018). Third-party affiliation is common between individuals in valuable relationships, and has been identified in a range of taxa including great apes (Cordoni et al., 2006; de Waal & van Roosmalen, 1979; Palagi & Norscia, 2013), monkeys (Call et al., 2002; Palagi et al., 2014), canids (Quervel-Chaumette et al., 2016) and corvids (Fraser & Bugnyar, 2010a; Logan, Emery, et al., 2013; Logan, Ostojić, et al., 2013; Sima et al., 2018). Whether and how post-conflict third-party affiliation is relevant to pair-bonds in particular has been interrogated in three species of corvid: jackdaws, rooks and jays (Logan, Emery, et al., 2013). Jackdaws and rooks form long-term pair-bonds and stay with each other year-round, while jays (*Garrulus glandarius*) tend to stay with their partner only during the breeding season. Following conflict, captive jackdaws and rooks engage in third-party affiliation with their partners more than with other individuals (Logan, Emery, et al., 2013), indicating that third-party post-conflict affiliation may be relevant to pair-bond management and maintenance. This is supported by the fact that jays, who do not need to manage and maintain the pair-bond outside of the breeding season, engage in post-conflict third-party affiliation with partners and non-partners indiscriminately (Logan, Emery, et al., 2013). The socio-cognitive demands of post-conflict third-party affiliation are unclear, given that the proximate and ultimate underpinnings of the behaviour can differ substantially depending on context and species. For instance, affiliation may be solicited or unsolicited by the receiver of affiliation (e.g.(Fraser & Bugnyar, 2010a)). Arguably, the socio-cognitive demands of solicited third-party affiliation are relatively low, given that an individual merely has to respond to a salient signal, parallel to, for example, a parent’s response to begging offspring. Unsolicited third-party affiliation, however, has the potential to be more socio-cognitively challenging.

Unsolicited third-party affiliation is often interpreted by researchers as an example of consolation (de Waal, 2010; Pérez-Manrique & Gomila, 2018). Consolation is said to occur when an individual directs unsolicited affiliation towards a distressed individual in order to alleviate their negative emotional state (Burkett et al., 2016), and is thought to be motivated by empathetic or sympathetic concern (Pérez-Manrique & Gomila, 2018). Empathy is regarded as an “advanced” cognitive ability, given that it is thought to require theory of mind (de Waal, 2010), i.e. the ability to predict another individual’s behaviour by inferring their unobservable mental state (*sensu* Premack and Woodruff 1978). The neuronal basis of empathy has been well-studied in rodents, non-human primates and humans (Paradiso et al., 2021). It is specifically associated with the amygdala and the anterior cingulate cortex, which is also implicated in other supposedly “complex” cognitive abilities, such as inhibitory control (Paradiso et al., 2021). The oxytocin-pathway, which is thought to mediate pair-bonding in mammals (Young & Wang, 2004), is also thought to be heavily involved in empathy (Burkett, et al., 2016; Li et al., 2019). Thus, empathy and its associated behaviours, such as consolation, are strong candidates for socio-cognitive behaviour that may be involved in the maintenance and management of social bonding.

While consolation has been reported in a wide range of species, in Chapter 3 I discuss how almost all studies of consolation suffer from one or more potential caveats that preclude the conclusion that the behaviour is definitely consolation. I designed an experiment that controlled for common confounds, and tested for consolation between wild jackdaw partners. I found that while male jackdaws did respond to their partner’s distress, they did not console them. Instead, they generally decreased affiliation and visit rate to the female, indicating that they avoided the area where the stressor occurred. Adaptively, this makes intuitive sense: if your partner has experienced danger, it is sensible to attempt to avoid that danger yourself. Given that this study controlled for key and common confounds in the field, these results raise an important question as to whether previous studies that report consolation actually do find the behaviour, rather than a proximately similar but ultimately different behaviour. I also suggest in Chapter 3 that to expect consolation between non-human partners is a somewhat anthropocentric prediction based on human expectations of relationship norms.

Arguably, a more neutral prediction than “partners with a stronger bond should offer more consolation to one another” is that partners with a stronger bond should show a stronger response (i.e., a larger change in behaviour) to a partner’s behavioural changes. This is because pairs that are more strongly bonded should be more attentive to subtle changes in partner behaviour, which should generate a larger change in behaviour. In Chapter 4, I show that this is true. Male jackdaws in stronger pair-bonds are more responsive to subtle changes in their partner's behaviour following partner distress. The cognitive basis of this responsiveness is unclear, but is consistent with the idea that, as Emery et al. suggest, individuals in a long-term pair-bond do track the subtleties of their partner's behaviour and use this to inform their own behaviour (Emery et al., 2007). Although likely to be less cognitively "sophisticated" than empathy, this still implies a socio-cognitive basis to pair-bond management and maintenance. What is not clear, and must be tested in the future, is whether pair-bond strength and partner responsiveness are correlated only for species in long-term pair-bonds where partners stay with one another throughout the year. If it is unique to these species, this would support the RIH's prediction that long-term pair-bonds generate selection for cognitive abilities to a degree that other bond types do not. However, it is possible that such a pattern is common to animals that form close social bonds, regardless of the bond type, and this cannot be ruled out until we have more data.

Other forms of partner responsiveness have also been investigated. Behavioural coordination is a specific form of partner responsiveness that requires individuals to attend to another's behaviour so that they can coordinate their response. Behavioural synchronisation, a form of coordination where individuals match their behaviour to that of another, has been researched extensively across taxa (Duranton & Gaunet, 2016). In particular, it has been interrogated in-depth in humans, where it facilitates economic cooperation (Lang et al., 2017) and endorphin release (Cohen et al., 2010), and is correlated with how much individuals "like" one other (Paxton & Dale, 2013). Behavioural synchronisation has also been found to be important in the relationships of non-humans. For instance, bottlenose dolphins (*Tursiops aduncus*) who are more affiliative also show more synchronised surfacing (Sakai et al., 2010), and allied males synchronise their vocalisations when coercing females (Moore et al., 2020). In orangutans *(Pongo pygmaeus*), facial expressions become synchronised during play (Davila Ross et al., 2008). Behavioural synchronisation may also be an important facet of pair-bonding in birds. For instance, for cockatiel pairs, behavioural synchronisation correlates with some affiliative behaviours, including partner proximity and responsiveness to a partner's solicitation of allopreening (Spoon et al., 2006). During behavioural synchronisation, individuals must attend to the behaviour of others in order to synchronise with them, and this is likely to have some sort of socio-cognitive underpinning (Wascher et al., 2018). Indeed, during behavioural synchronisation in humans, "inter-brain" synchronisation occurs, where electrical frequencies implicated in socio-cognition (alpha-mu bands) act symmetrically between individuals (Dumas et al., 2010). Again, however, the sophistication of cognition required here is unclear. Evidence from flocking and shoaling species would suggest that behavioural synchrony can be achieved through simple rules (Ballerini et al., 2008; Herbert-Read et al., 2011; Kotrschal et al., 2018). Moreover, a study in guppies (*Poecilia reticulata*) showed that behavioural synchronisation during shoaling was not related to brain size or cognitive abilities such as learning and memory (Kotrschal et al., 2018). Current evidence therefore suggests that behavioural synchrony may involve cognitive architecture implicated in socio-cognitive processing, but does not necessarily require sophisticated cognitive ability; indeed, static rules that require little information-processing may underpin the behaviour.

In birds, perhaps one of the most salient examples of behavioural coordination is bi-parental care. ~90% of bird species, with both short and long-term pair-bonds, biparentally care for their offspring (Kendeigh, 1952). Recently, much work has been done to understand the dynamics of bi-parental coordination (Savage et al., 2020). Mathematical modelling shows that, when the fitness interests of parents are not completely aligned (for instance because of extra-pair copulation), individuals should monitor their partner's investment in rearing offspring, e.g. by monitoring their provisioning behaviour (Johnstone & Savage, 2019). This is so that individuals can decide whether it is advantageous for them to continue provisioning the brood (Johnstone & Savage, 2019). This results in apparent coordination between partners, where they take turns to provision young (Johnstone & Savage, 2019). Indeed, there is empirical support for such a pattern in some species (Ihle et al., 2019; Johnstone et al., 2014; Savage et al., 2017). Whether responding to partner provisioning behaviour is particularly cognitively demanding is, however, unclear. Indeed, previous work has shown that more complex coordination rules between partners, where each individual has a different role, does not require cooperative cognition (Jelbert et al., 2015). How provisioning coordination relates to pair-bond management and maintenance is also unknown. However, importantly for understanding partner coordination in the context of the RIH, mathematical models predict that pairs may *not* need to coordinate (Johnstone & Savage, 2019). Specifically, when parents “cost-share”, such that the cost of raising offspring incurred by one parent decreases the fitness of the other (i.e., where the reproductive success of one parents relies on the other staying alive), individuals need not monitor and respond to their partner’s contribution because there is no risk of their partner exploiting them (Johnstone & Savage, 2019). In other words, the partner’s behaviour is predictable, and it is thus unnecessary for them to be monitored. Therefore, partners may not coordinate provisioning when fitness interests are fully aligned (Johnstone & Savage, 2019). Counterintuitively to the predictions of the RIH, this implies that partners with completely aligned interests may coordinate *less* than species where pairs have conflicting interests.

The RIH predicts that the management and maintenance of long-term, year-round pair-bonds generates more socio-cognitive challenges than other bond types. In this section, I show that there are threads of empirical support for this prediction. For instance, jackdaws and rooks offer third-party affiliation to their partners, and cockatiel pairs engage in behavioural synchrony. While it is unclear whether non-humans engage in empathetic behaviours such as consolation (Chapter 3), I do show in Chapter 4 that jackdaws with stronger bonds are more able to track the fine-scale behaviours of their partners and use subtle behavioural changes to inform their own actions. This is consistent with the key prediction of the RIH that socio-cognitive ability and pair-bond strength are positively correlated. However, how cognitively demanding these behaviours are is not at all clear. Furthermore, whether long-term, year-round pair-bonds require better socio-cognitive ability to be maintained and managed relative to other bond types is not known, yet is crucial for testing the RIH. Moreover, given that long-term, year-round pair-bonds are often characterised by highly aligned fitness interests between partners (e.g. (Gill et al., 2020; Henderson et al., 2000)) and that theory predicts that in at least one scenario, partners with aligned fitness interests need not coordinate (Johnstone &

Savage, 2019), long-term year-round pair-bonds may actually have *relaxed* selection on socio-cognitive ability, at least in some contexts.

4.2 Socio-cognitive demands of managing the pair-bond within a wider social context

The RIH focusses on the cognitive challenges of navigating the pair-bond; however, as Emery et al. acknowledge, pairs do not exist in a social vacuum (Emery et al., 2007). In this section, I evaluate whether theoretical and empirical work support the prediction that (i) the aligned interests of long-term, year-round pair-bonds are cognitively demanding and (ii) long-term, year-round pair-bonds are likely to be cognitively demanding when embedded in a wider social landscape.

Theory suggests that the less predictable an environment (so long as it is not totally unpredictable), the more information-processing required by an individual in order to reduce the uncertainty of that environment (McNamara & Dall, 2010). Reducing uncertainty allows individuals to make better decisions about how to act (Dall & Cuthill, 2016). Thus, in unpredictable environments, better information-processing ability (i.e., better cognitive ability) should generally lead to better decision-making, and thus better fitness. Conversely, successfully navigating an entirely predictable environment should not be cognitively demanding. This concept is perhaps best explained using a “brain-less” example. For instance, microbes that pass through the mammalian gastro-intestinal tract appear to “predict” the journey from mouth to stomach by preparing, at the molecular level, for their new environment before they reach it. Of course, they are not planning for the future. Instead, they have evolved a rule where, if a particular environment is detected, a biomolecular chain reaction is instigated to prepare the cell for what, over evolutionary time, has proved extremely likely to be the following environment (Freddolino & Tavazoie, 2012). Thus, if an environment is extremely predictable, cognition is not required. Indeed, the previously discussed theoretical prediction that partners with aligned interests need not monitor one another’s provisioning is another example of this (Johnstone & Savage, 2019). This highlights the relative simplicity of predictability.

Empirically, the theory that less predictable environments generate more selection on cognitive ability receives some support. For example, in birds, the invasion of novel environments correlates with innovation rate (Sol et al., 2005), and inhabiting a variable and/or harsh habitat correlates with problem-solving and speed of reversal learning (Kozlovsky et al., 2015; T. C. Roth et al., 2010; Tebbich & Teschke, 2014). Meanwhile, pheasants (*Phasianus colchicus*) raised in a spatially unpredictable environment show increased inhibitory control (van Horik et al., 2019). In a species of primate, the mouse lemur (*Microcebus murinus*), increased problem-solving ability positively correlates with the magnitude of increase in body condition after the harsh dry season, suggesting that the individuals who are most efficient at problem-solving are also most efficient at foraging in a harsh environment (Huebner et al., 2018). Primates that live in fission-fusion societies, reflecting a constantly shifting and thus less predictable social landscape, show increased inhibitory control relative to primates that live in more stable social groups (Amici et al., 2008). At the intraspecific scale, Australian magpies (*Cracticus tibicen dorsalis*) show increased cognitive performance when raised in larger groups (Ashton, Ridley, et al., 2018). Larger groups do not necessarily imply a more cognitively demanding landscape, but when relatedness is low the environment is more likely to be less predictable because of more conflicts of interest (Lukas & Clutton-Brock, 2018). Indeed, Lukas and Clutton-Brock found that in groups with lower levels of relatedness more complex societies emerged, with, for example, increased coalition formation (Lukas & Clutton-Brock, 2018). Meanwhile, in humans, navigating social interactions characterised by conflicting interests increases neuronal activity relative to scenarios where there is little conflicting interest (Emonds et al., 2012). Thus, navigating social landscapes where individuals have conflicting interests (and thus, where there is less predictability) requires more information-processing and is more cognitively demanding.

The RIH focusses on the cognitive demands of managing and maintaining a long-term pair-bond. Arguably, a long-term pair-bond introduces predictability into an individual's social environment, where the behaviour of the partner is likely to be predictable given (i) aligned interests and (ii) familiarity. Given strong theoretical and empirical support that less predictable environments require more cognitive ability, the RIH thus seems to be at odds with current ideas about why cognitive ability evolves.

While the RIH focusses mainly on the interactions between partners (Emery et al., 2007), in species with long-term pair-bonds, pairs are often embedded within a wider social landscape of non-kin individuals. For instance, rook and jackdaw pairs tend to breed colonially, forage in small flocks, and roost in large flocks comprising of hundreds or even thousands of individuals (Clayton & Emery, 2007; Röell, 1978). In jackdaws and rooks, pair-bonded individuals form non-pair relationships independently of their partner (Boucherie et al., 2016; Kubitza et al., 2015). Given that maintaining and managing such bonds takes time, and that, at least in jackdaws, time together is a key facet of pair-bond strength (Chapter 4), there is an implied trade-off between investment in the pair-bond and investment in other relationships. There is also likely to be a trade-off in terms of social and cultural information available to individuals: strongly bonded pairs may be less likely to learn important information from the wider social group, and this may potentially decrease fitness outcomes. Moreover, relationships beyond the pair have been shown to be crucial to fitness. In jackdaws, individuals that are more central in their social network have higher reproductive success (Kings, 2018). However, a strong pair-bond may also be important for fitness outcomes (Spoon et al., 2006). Thus, how the trade-off in investment between pair and non-pair relationships is navigated (for instance, spending time with a partner rather than forming other bonds that may give access to valuable social learning opportunities) is likely to impact fitness outcomes. Navigating this trade-off may potentially introduce cognitive challenges that would not be generated if pairs existed in a social vacuum, given that the introduction of non-partner individuals (who do not have fully aligned fitness interests) presumably makes for a much less predictable social landscape.

Living in a group also introduces the opportunity for extra-pair copulation. While Emery et al. do mention extra-pair copulation in their formulation of the RIH, it is mostly to say that species with long-term partnerships, where pairs remain together throughout the year, tend to be genetically monogamous (Emery et al., 2007). As discussed in Chapter 2, only a very few species have been shown to be genetically monogamous, and even then, this is likely to be population-specific (Birkhead & Moller, 1998; Gray, 1996; Hatchwell, 1988; Westneat, 1992). Indeed, in jackdaws, which have previously been reported to be socially and genetically monogamous, there is evidence of low levels of extra-pair fertilisation in some populations (Gill et al. 2020; Liebers and Peter 1998; Turjeman, Chen, and Nathan 2021; Chapter 2). In our population, some of these extra-pair offspring were the result of males mating with more than one female, who then laid her eggs in the male's nest-box. This clearly indicates a conflict of interest between partners, given that the resident female is then forced to incubate and provision offspring that are not hers. Following from Johnstone and Savage's (2019) mathematical model, such a conflict of interest may result in selection for the monitoring of partner contributions (Johnstone & Savage, 2019). Therefore, this conflict of interest could potentially increase the information-processing required by partners.

In addition to "follower" relationships where extra-pair females laid eggs fertilised by a paired male in his nest-box, I found that pair-bonded male jackdaws engage in FEPCs, which may (rarely) result in extra-pair fertilisation. Why males engage in FEPC is unclear, and potential hypotheses are discussed in detail in Chapter 2. There is currently not enough evidence to either support or refute the "reproductive suppression hypothesis", which suggests that males may benefit from engaging in FEPC because it reduces the fitness of the target female, thus reducing competition for the male and his partner's offspring. In this case, despite the costs of FEPC to the male and to his partner, FEPC may be beneficial to both members of the pair and therefore represent no conflict of interest. However, a more likely explanation is that FEPC persists because it once commonly resulted in siring success for the male. If this is the case, then this behaviour does represent a conflict of interest between the pair, because of the minimal benefit and multiple potential costs incurred by the male's partner. Costs to her are likely to be non-negligible. For instance, I found that males with laying or incubating partners would monitor neighbouring nest-boxes until the resident male was absent, and then force copulations with the resident female. Given that incubating females must be provisioned by their partner during incubation (Röell, 1978), males who use their time to monitor neighbours and engage in FEPCs, rather than forage, are not acting in the interest of their partner. Moreover, FEPC introduces significant risk of injury to the males engaging in the behaviour. Since a pair must work together to raise offspring (Henderson & Hart, 1993), and an injured partner is likely to provide poorer care (G. F. Wagner et al., 2019), this again introduces a potential negative impact of FEPC on the female whose partner engages in FEPC. While FEPC has not been explored in detail in many species, it has been anecdotally reported from a wide range of species, most of which form socially monogamous pair-bonds. Again, such a conflict of interest between partners is likely to introduce uncertainty into an individual's environment, which in turn is likely to generate selection for information-processing (cognitive) ability.

The conflicts of interest that are introduced when examining the pair relationship within the context of wider society are little discussed in the original formulation of the RIH. However, they are discussed in detail by some researchers interrogating the evolution of cognition in birds. In particular, there is much work examining the wider social bonds of ravens (*Corvus corax*), who form long-term pair-bonds that stay with each other year-round, but also form fission-fusion flocks throughout the non-breeding season (Boucherie et al., 2019). It has been argued that the informational demands, and thus cognitive challenges, which arise from navigating such a multi-level society drive cognitive evolution (Boucherie et al., 2019). For instance, remembering many individuals within the flock, monitoring partner interactions with others and tracking third party relationships are all considered to be cognitively demanding behaviours associated with the navigation of multi-level society (Boucherie et al., 2019). Indeed, as discussed throughout this section, the multi-level society that long-term pair-bonded bird species usually inhabit is a key facet of life as a pair; thus, pair-bonds should not be discussed as if they exist in isolation. The informational demands, and thus cognitive challenges, faced by pairs are inherently tied to both their bond and their wider social landscape.

5. *The RIH: conclusions and future directions*

Here, I have interrogated support for the RIH in light of contemporary empirical and theoretical work. I find that the RIH was built on weak foundations (Chapter 5), and empirical support is mixed. For instance, I investigated key predictions of the RIH within one wild study system, the jackdaw, and found that some predictions were met and some were not. I found strong support that pair-bond strength is variable between pairs, repeatable within pairs and related to a measure of socio-cognitive performance; however, I did not find convincing evidence that pair-bond strength influenced fitness (Chapter 4). I also did not find any evidence that empathetic concern is used as a strategy through which individuals manage and maintain their pair-bond (Chapter 3). I also found that, at odds with general assumptions of the RIH, partners are likely to have to navigate conflicts of interest (Chapter 2) as a result of interactions with non-pair individuals.

Beyond our intraspecific study there is piecemeal support for the RIH, scattered across behavioural and neurobiological studies from a wide variety of taxa. For instance, one other species of bird, the greylag goose, shows repeatability of an aspect of pair-bond strength (Kralj-Fišer et al., 2007); a link between pair-bond strength and reproductive success has been identified in one species, the cockatiel (Spoon et al., 2006); and a link between different but related aspects of the pair-bond (e.g., length or partner familiarity) has been found in a range of species (e.g.(Culina et al., 2020; Sánchez-Macouzet et al., 2014)). Various socio-cognitive behaviours relevant to social bonding have been quantified at the behavioural and neural level, but pair-bond strength has only been directly linked to a measure of socio-cognitive ability in jackdaws. Moreover, many of the studies discussed in this chapter have been conducted on captive species. While still valuable, it must be considered that bonding behaviour between individuals can differ substantially between captive and wild settings (Inoue & Shimada, 2020; Pacheco & Madden, 2021), and captive individuals are subject to a much less challenging and dangerous environment. Therefore, to what degree these studies can tell us about the evolutionary basis and ecological relevance of pair-bonding behaviour and its associated cognitive challenges is limited (Pritchard et al., 2016; Thornton et al., 2014).

While there are threads of empirical support for the RIH, which suggest that pair-bond strength may be a substrate of selection and may be socio-cognitively demanding, I propose that there are two key flaws with this hypothesis. The first is a flaw that could potentially be rectified with more data. It is that we currently have absolutely no evidence to suggest that species with long-term, year-round partnerships actually do have better socio-cognitive performance relative to species with other bond types, yet this is fundamental to understanding if long-term pair-bonds do drive cognitive evolution. I therefore propose that an important area of future work is to expand our investigations of the cognitive basis of pair-bonding beyond species with long-term pair-bonds. Specifically, explicit comparisons of cognitive performance between (ideally closely related) species with long- versus short-term pair-bonds would be extremely valuable. Currently, most investigations are restricted to species that we already know to (i) show sophisticated cognitive performance and (ii) have long-term pair-bonds. Without establishing a full picture of pair-bonding dynamics across species, we risk drawing false inferences about the cognitive demands of pair-bonding due to our own anthropomorphic biases of which species should be capable of behaviours that we think are cognitively complex (Shettleworth, 2010a). Importantly, I also recommend an intraspecific approach to testing predictions of the RIH, in-line with other researchers in the field of cognitive evolution (Ashton, Thornton, et al., 2018; Logan et al., 2018) and the analyses presented in Chapters 2 - 4. By testing predictions of the RIH within species, we will gain a finer-grained understanding of the cognitive demands and fitness consequences of pair-bonds, thus elucidating fine-scale patterns that may not be evident at a broader scale**.** Following intra-specific studies of the cognitive demands of pair-bonding across a range of species, comparative work can be done to understand whether and how patterns vary between species with different bond types.

The second major flaw with the RIH is a theoretical one. Current theory suggests that cognition should evolve when information-processing can create a more predictable environment for an individual, thus meaning an individual can make better decisions (Dall & Cuthill, 2016; McNamara & Dall, 2010). This means that the less predictable an environment, the more selection there should be on cognitive ability. Long-term, year-round pair-bonds, when viewed in isolation (as they primarily are in the original formulation of the RIH) are likely to create a more predictable social environment for partners. I suggest in section 4.2. that perhaps year-round long-term pair-bonds generate informational challenges because individuals must navigate *both* the pair-bond and the wider social landscape. Indeed, even in jackdaws, where pairs have almost completely interdependent fitness, partners appear to have some conflicting interests when the pair-bond is examined in the context of wider society. While this extension of the RIH may prove fruitful, I wish to stress that it is “patching a hole” in a hypothesis that otherwise does not align well with theoretical expectations.

We would like to posit that a useful way to move forward as a field may be to contextualise cognitive evolution within information theory and use a formal theory-driven approach to generate predictions. Currently, there are multiple hypotheses for the evolution of cognition (e.g., the Social Intelligence Hypothesis, the Relationship Intelligence Hypothesis, the Ecological Intelligence Hypothesis), and they are often presented as separate and competing hypotheses (e.g.(DeCasien et al., 2017; Rosati, 2017)). However, underpinning them all is the uniting theory that cognition is an adaptation to allow for more information-processing. Framing our investigations of cognitive evolution in terms of the information-processing demands faced by individuals, and using this to generate predictions, will allow for a more cohesive and tractable study of cognitive evolution than verbal arguments alone (McNamara & Leimar, 2020; Smaldino, 2020), and will avoid unnecessary competition between falsely dichotomous explanations. It will also allow for the generation of predictions that are grounded in evolutionary theory. This will help to mitigate the influence of our inherent anthropocentric biases in formulating predictions in the field, an issue addressed in detail in the next section.

Anthropocentrism in cognitive research: how to move forward?

During the development of this thesis, two very different approaches to research became apparent: the anthropocentric versus the theory-driven approach. In Chapter 3, I reviewed research into empathetic concern in non-humans. Initially, empathy was investigated in a rather anthropocentric manner, whereby close relatives of humans were tested for traits that we recognise in ourselves (e.g.(de Waal & van Roosmalen, 1979; Povinelli et al., 1992)). Following this, the Valuable Relationship Hypothesis was used as a framework to explain why one might expect empathy in non-humans (de Waal & Aureli, 1997). Thus, the theory followed the research, rather than informing it, and the research itself was originally conducted based on rather anthropocentric perspectives. On the flip side of this approach is the study of Alternative Reproductive Tactics (ARTs), as briefly reviewed in the introduction to Chapter 2. Although forced copulations are relevant to human behaviour, the study of ARTs has, from its beginning, been grounded in evolutionary theory (e.g.(Dominey, 1984; Thornhill, 1984)), even when used to explore the evolutionary roots of forced copulation in humans (Thornhill & Palmer, 2000; Thornhill & Wilmsen Thornhill, 1983). Indeed, Thornhill and Palmer (very controversially - see response to criticisms here (Palmer & Thornhill, 2003)) tested specific predictions generated from research in the field of sexual conflict in order to understand whether and how human rape fits into our knowledge of the evolution of forced copulations in other animals (Thornhill & Palmer, 2000). The approach to studying empathy and the approach to studying ARTs, both of which are relevant to human behaviour, therefore differs substantially: while one is historically rooted in an anthropocentric exploration of non-human behaviour, the other has always been rooted in an evolutionary view of behaviour (human and non-human). While this is perhaps a drastic example, such a differential approach to the study of behaviour (anthropocentric versus non-anthropocentric) is pervasive.

Why does it matter that researchers take different approaches to the study of behaviour? The answer is that there are considerable differences in inference depending on approach. For example, as discussed by Vasconselos et al. (2012), two studies investigating a similar behaviour came to vastly different conclusions as to *what* the behaviour showed (Vasconcelos et al., 2012). Both studies investigated rescue behaviour: one in rats (Ben-Ami Bartal et al., 2011) and one in ants (*Cataglyphis cursor* (Nowbahari et al., 2009)). In these studies, an individual was restrained, and experimenters tested whether conspecifics worked to release them. In both species, conspecifics released trapped individuals in very similar ways. For rats, this was interpreted as a prosocial behaviour rooted in empathetic concern. For ants, it was interpreted as a programmed response of colony members to a specific scenario, selected for due to indirect fitness benefits. Critically, empathy (or indeed any psychological mechanism) was not invoked.

Another example of the danger of differing approaches comes from the study of consolation. Burkett et al. tested for consolation in prairie voles (Burkett et al., 2016). They found evidence in-line with the conclusion that prairie voles console distressed partners. However, they did not control for the possibility that the distressed partner solicited affiliation through ultra-sonic calls, and thus it cannot be concluded that the behaviour was consolation, as opposed to solicited affiliation. This is highlighted in a published comment on the article (Pérez-

Manrique & Gomila, 2016), and it is a fair and important criticism. However, claims of consolation in primates seem not to be held to such a high standard. For instance, in a review of empathy, Pérez-Manrique and Gomila (2018) discuss a study of bonobos (*Pan paniscus*) as finding “unsolicited” affiliation (Palagi et al., 2004). However, this study does not describe how solicitation was measured, merely that affiliation was considered unsolicited if it was not “invited”. Given that no description is given as to what “invited” involves, this raises questions as to whether a careful analysis of potentially subtle soliciting behaviour (e.g., quiet calls or subtle gestures) was conducted. This is not to criticise Palagi et al.’s study; merely to highlight that Palagi et al.’s study and Burkett et al.’s studies seem to have been subjected to differing levels of criticism.

From the above examples, it could be perceived that we are more willing to ascribe empathetic behaviours to species more similar to ourselves (i.e., bonobos more readily than rodents, but rodents more readily than ants). However, experiments should always be held to the same standard, regardless of species, and while shared ancestry should not be discounted, neither should the possibility of deep evolutionary roots of behaviour, convergent evolution across disparate species, nor the possibility that something really might be a uniquely human trait.

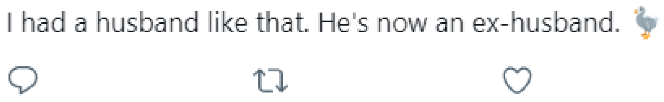
It seems that a large influence on the approach to testing cognitive ability may be the value system of the researcher. Frans de Waal’s popular science article “What I learned from tickling apes”, an endorsement of anthropomorphism of closely related species, suggests that “anthrodenialism” (where researchers have a baseline assumption that non-human animals, even our close relatives, do not possess human traits) is related to the human need to distance ourselves from animals following our switch from hunting to farming (de Waal, 2016). He suggests that “*to justify how [we] treated other species, [we] had to play down their intelligence and deny them a soul*”. Thus, it appears that de Waal’s argument in favour of anthropomorphism is partly motivated by a moral stance that this helps humans treat animals ethically, and that this is something we should strive to do. With this latter point, most researchers in animal behaviour would probably passionately agree. However, whether animals should need to have human-like qualities in order to deserve our respect is most likely a point of contention: depending on philosophical outlook, some researchers probably strongly agree and others strongly oppose this stance. Perhaps such differences in value systems partially underlies the disparities we observe in approaches to testing animal cognition.

What, then, is the way forward? I propose that formal theoretical approaches, such as the approach proposed at the end of Section 5, will allow us not only to unite falsely dichotomous hypotheses regarding the evolution of cognition, but also to mitigate the influence of our inherent anthropocentric biases. In conjunction with this approach, we should work to design experimental and observational studies that mitigate our anthropocentric and anthropomorphic biases by ruling out alternative explanations for behaviours (Barrett et al., 2007; Shettleworth, 2010a). We must also ensure that the same rigour in study design is applied across species, independently of their relation to us, and, importantly, that the conclusions we reach are not taxa dependent. If empathy is invoked to explain a behaviour in a rodent, we must ask: would we also consider this to be empathy in an insect? If not, we must address why, and ask whether our biases are influencing our conclusions. To adopt such an approach is not to be "anthrodenialist" (de Waal, 2016), or "killjoy" (Shettleworth, 2010a) and is certainly not to undermine advances in animal welfare. It is to support rigorous and robust science to comprehend the minds of animals, in order to understand what an animal mind *is*, not what we think it should be, or want it to be, according to anthropocentric ideals.

APPENDIX

Chapter 3: Wild jackdaws respond to their partner’s distress, but not with consolation

After publication this article was shared widely on Twitter, with some interesting responses. Our favourite was:



I had a husband like that. He's now an ex-husband. 🦤

Luckily, female jackdaws do not seem to mind quite so much.

BIBLIOGRAPHY

Abbot, P., & Chapman, T. (2017). Sociality in aphids and thrips. In *Comparative social evolution* (pp. 124–153).

Adler, M. (2010). Sexual conflict in waterfowl: Why do females resist extrapair copulations? *Behavioral Ecology*, *21*(1), 182–192. https://doi.org/10.1093/beheco/arp160

Adriaense, J. E. C., Koski, S. E., Huber, L., & Lamm, C. (2020). Challenges in the comparative study of empathy and related phenomena in animals. *Neuroscience and Biobehavioral Reviews*, *112*(April 2019), 62–82. https://doi.org/10.1016/j.neubiorev.2020.01.021

Adriaense, J. E. C., Martin, J. S., Schiestl, M., Lamm, C., & Bugnyar, T. (2019). Negative emotional contagion and cognitive bias in common ravens (Corvus corax). *Proceedings of the National Academy of Sciences of the United States of America*, *166*(23), 11547–11552. https://doi.org/10.1073/pnas.1817066116

Amici, F., Aureli, F., Call, J., & Street, B. (2008). *Fission-Fusion Dynamics , Behavioral Flexibility , and Inhibitory Control in Primates*. 1415–1419. https://doi.org/10.1016/j.cub.2008.08.020

Anderson, C., & McShea, D. W. (2001). Individual versus social complexity, with particular reference to ant colonies. *Biological Reviews of the Cambridge Philosophical Society*, *76*(2), 211–237. https://doi.org/10.1017/S1464793101005656

Aplin, L. M., Firth, J. A., Farine, D. R., Voelkl, B., Crates, R. A., Culina, A., Garroway, C. J., Hinde, C. A., Kidd, L. R., Psorakis, I., Milligan, N. D., Radersma, R., Verhelst, B. L., & Sheldon, B. C. (2015). Consistent individual differences in the social phenotypes of wild great tits, Parus major. *Animal Behaviour*, *108*, 117–127. https://doi.org/10.1016/j.anbehav.2015.07.016

Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature*, *554*(7692), 364–367. https://doi.org/10.1038/nature25503

Ashton, B. J., Thornton, A., & Ridley, A. R. (2018). An intraspecific appraisal of the social intelligence hypothesis. *Philosophical Transactions of the Royal Society B*, *373*, 20170288. https://doi.org/http://dx.doi.org/10.1098/rstb.2017.0288

Audacity Team. (2019). *Audacity(R): Free Audio Editor and Recorder*. https://audacityteam.org/

Austin, V. I., Awelbergen, J., Maisey, A. C., Lindsay, M. G., & Dalziell, A. H. (2019). Destruction of a conspecific nest by a female Superb Lyrebird: Evidence for reproductive suppression in a bird with female-only parental care. *Behaviour*, *156*(15), 1459–1469. https://doi.org/10.1163/1568539X-00003574

Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Filho, W. J., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, *513*(5), 532–541. https://doi.org/10.1002/cne.21974

Baker, R. R., & Bellis, M. A. (1993). Human sperm competition: Ejaculate manipulation by females and a function for the female orgasm. *Animal Behaviour*, *46*(5), 887–909.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., & Zdravkovic, V. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1232–1237. https://doi.org/10.1073/pnas.0711437105

Barrett, L. (2017). The (r)evolution of primate cognition: Does the social intelligence hypothesis lead us around in anthropocentric circles? In Julian Kiverstein (Ed.), *The Routledge Handbook of Philosophy of the Social Mind* (pp. 19–34). https://doi.org/10.4324/9781315530178

Barrett, L., Henzi, P., & Kendall, D. (2007). Social brains, simple minds: Does social complexity really require cognitive complexity? *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *362*(1480), 561–575. https://doi.org/10.1098/rstb.2006.1995

Barton, R. A. (1996). Neocortex size and behavioural ecology in primates. *Proceedings of the Royal Society B: Biological Sciences*, *263*(1367), 173–177. https://doi.org/10.1098/rspb.1996.0028

Beauchamp, G., & Fernández-Juricic, E. (2004). Is there a relationship between forebrain size and group size in birds? *Evolutionary Ecology Research*, *6*(6), 833–842.

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*(4), 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Ben-Ami Bartal, I., Decety, J., & Mason, P. (2011). Empathy and Pro-Social Behavior in Rats. *Science*, *334*, 1427–1430.

Bergman, T. J., & Beehner, J. C. (2015). Measuring social complexity. *Animal Behaviour*, *103*, 203–209. https://doi.org/10.1016/j.anbehav.2015.02.018

Birkhead, T. R. (1979). Mate guarding in the magpie Pica pica. *Animal Behaviour*, *27*, 866–874.

Birkhead, T. R., & Moller, A. P. (Eds.). (1998). *Sperm competition*. New York: Academic.

Birkhead, T. R., & Møller, A. P. (1993). Why do male birds stop copulating while their partners are still fertile? In *Animal Behaviour* (Vol. 45, Issue 1, pp. 105–118). https://doi.org/10.1006/anbe.1993.1010

Black, J. M. (2001). *Fitness consequences of long-term pair bonds in barnacle geese : monogamy in the extreme*. *12*(5), 640–645.

Blumstein, D. T., Petelle, M. B., & Wey, T. W. (2013). Defensive and social aggression: Repeatable but independent. *Behavioral Ecology*, *24*(2), 457–461. https://doi.org/10.1093/beheco/ars183

Boesch, C. (2020). The Human Challenge in Understanding Animal Cognition. In L. Johnson, A. Fenton, & A. Shriver (Eds.), *Neuroethics and Nonhuman Animals* (pp. 33–52). Springer. https://doi.org/10.1007/978-3-030-31011-0

Boogert, N. J., Madden, J. R., Morand-Ferron, J., & Thornton, A. (2018). Measuring and understanding individual differences in cognition. *Philosophical Transactions of the Royal Society B*, *373*, 20170280. https://doi.org/10.1093/oxfordhb/9780198570097.013.0009

Boucherie, P. H., Loretto, M. C., Massen, J. J. M., & Bugnyar, T. (2019). What constitutes “social complexity” and “social intelligence” in birds? Lessons from ravens. *Behavioral Ecology and Sociobiology*, *73*(1). https://doi.org/10.1007/s00265-018-2607-2

Boucherie, P. H., Mariette, M. M., Bret, C., & Dufour, V. (2016). Bonding beyond the pair in a monogamous bird: Impact on social structure in adult rooks (Corvus frugilegus). *Behaviour*, *153*(8), 897–925. https://doi.org/10.1163/1568539X-00003372

Boucherie, P. H., Poulin, N., & Dufour, V. (2018). Not Much Ado about Something: Behavioural Mechanisms of Pair Bond Separation and Formation in Long-Term Pairing Rooks. *Ecoscience*, *25*(1), 71–83. https://doi.org/10.1080/11956860.2017.1414671

Brandl, H. B., Griffith, S. C., Farine, D. R., & Schuett, W. (2021). Wild zebra finches that nest synchronously have long-term stable social ties. *Journal of Animal Ecology*, *90*(1), 76–86. https://doi.org/10.1111/1365-2656.13082

Braun, A., & Bugnyar, T. (2012). Social bonds and rank acquisition in raven nonbreeder aggregations. *Animal Behaviour*, *84*(6), 1507–1515. https://doi.org/10.1016/j.anbehav.2012.09.024

Brekke, P., Cassey, P., Ariani, C., & Ewen, J. G. (2013). *Evolution of extreme-mating behaviour : patterns of extrapair paternity in a species with forced extrapair copulation*. 963–972. https://doi.org/10.1007/s00265-013-1522-9

Brekke, P., Wang, J., Bennett, P. M., Cassey, P., Dawson, D. A., Horsburgh, G. J., & Ewen, J. G. (2012). Postcopulatory mechanisms of inbreeding avoidance in the island endemic hihi (Notiomystis cincta). *Behavioral Ecology*, *23*(2), 278–284. https://doi.org/10.1093/beheco/arr183

Brent, L. J. N., Heilbronner, S. R., Horvath, J. E., Gonzalez-Martinez, J., Ruiz-Lambides, A., Robinson, A. G., Skene, J. H. P., & Platt, M. L. (2013). Genetic origins of social networks in rhesus macaques. *Scientific Reports*, *3*(1), 1042. https://doi.org/10.1038/srep01042

Briskie, J. V., & Montgomerie, R. (1997). *Intromittent Organ of Birds*. *28*(1), 73–86.

Briskie, J. V., & Montgomerie, R. (2001). Efficient copulation and the evolutionary loss of the avian intromittent organ. *Journal of Avian Biology*, *32*, 184–187.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400.

Brouwer, L., & Griffith, S. C. (2019). Extra-pair paternity in birds. *Molecular Ecology*, *28*(22), 4864–4882. https://doi.org/10.1111/mec.15259

Brummer, S. P., Gese, E. M., & Shivik, J. A. (2010). The effect of enclosure type on the behavior and heart rate of captive coyotes. *Applied Animal Behaviour Science*, *125*(3–4), 171–180. https://doi.org/10.1016/j.applanim.2010.04.012

Buchanan, K., Burt de Perera, T., Carere, C., Carter, T., Hailey, A., Hubrecht, R., Jennings, D., Metcalfe, N., Pitcher, T., Peron, F., Sneddon, L., Sherwin, C., Talling, J., Thomas, R., & Thompson, M. (2012). Guidelines for the treatment of animals in behavioural research and teaching. *Animal Behaviour*, *83*(1), 301–309.

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*(5), 472–480. https://doi.org/10.1111/j.1439-0310.2010.01758.x

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, *18*(5), 175–186. https://doi.org/10.1002/evan.20222

Burkart, J. M., & van Schaik, C. P. (2009). Cognitive consequences of cooperative breeding in primates? *Animal Cognition*, *13*(1), 1–19. https://doi.org/10.1007/s10071-009-0263-7

Burkett, J. P., Andari, E., Johnson, Z. V., Curry, D. C., de Waal, F. B. M., Young, L. J., Waal, F. B. M. De, & Young, L. J. (2016). Oxytocin-dependent consolation behavior in rodents. *Science*, *351*(6271), 375–378. https://doi.org/10.1126/science.aac4785

Byrne, R. W., & Whiten, A. (1988). *Machiavellian Intelligence.* Clarendon Press.

Call, J., Aureli, F., & de Waal, F. B. M. (2002). Postconflict third-party affiliation in stumptailed macaques. *Animal Behaviour*, *63*(2), 209–216. https://doi.org/10.1006/anbe.2001.1908

Cameron, E. Z., Setsaas, T. H., & Linklater, W. L. (2009). Social bonds between unrelated females increase reproductive success in feral horses. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13850–13853. https://doi.org/10.1073/pnas.0900639106

Carere, C., & Van Oers, K. (2004). Shy and bold great tits (Parus major): Body temperature and breath rate in response to handling stress. *Physiology and Behavior*, *82*(5), 905–912. https://doi.org/10.1016/j.physbeh.2004.07.009

Cauchoix, M., Chaine, A. S., & Barragan-Jason, G. (2020). Cognition in Context: Plasticity in Cognitive Performance in Response to Ongoing Environmental Variables. *Frontiers in Ecology and Evolution*, *8*(April), 106. https://doi.org/10.3389/fevo.2020.00106

Chance, M. R. A., & Mead, A. P. (1953). Social behaviour and primate evolution. *Symp. Soc. Exp. Biol.*, *7*, 395–439.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *17*, i884–i890.

Cheney, D. L., Silk, J. B., & Seyfarth, R. M. (2016). Network connections, dyadic bonds and fitness in wild female baboons. *Royal Society Open Science*, *3*(7). https://doi.org/10.1098/rsos.160255

Chittka, L., & Niven, J. (2009). Are Bigger Brains Better ? Review. *Current Biology*, *19*(21), R995–R1008. https://doi.org/10.1016/j.cub.2009.08.023

Clayton, N. S., & Emery, N. J. (2007). The social life of corvids. *Current Biology*, *17*(16), 652–656.

Clutton-Brock, T. H., & Parker, G. A. (1995). Sexual coercion in animal societies. *Animal Behaviour*, *49*(5), 1345–1365.

https://doi.org/10.1006/anbe.1995.0166

Clutton-Brock, T. H., & Harvey, P. H. (1980). Primates, brains and ecology. *Journal of Zoology*, *190*(3), 309–323. https://doi.org/10.1111/j.1469-7998.1980.tb01430.x

Cohen, E. E. A., Ejsmond-Frey, R., Knight, N., & Dunbar, R. I. M. (2010). Rowers' high: Behavioural synchrony is correlated with elevated pain thresholds. *Biology Letters*, *6*(1), 106–108. https://doi.org/10.1098/rsbl.2009.0670

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, *74*(365), 169–174. https://doi.org/10.1080/01621459.1979.10481634

Cordoni, G., Palagi, E., & Tarli, S. B. (2006). Reconciliation and consolation in captive western gorillas. *International Journal of Primatology*, *27*(5), 1365–1382. https://doi.org/10.1007/s10764-006-9078-4

Corfield, J. R., Birkhead, T. R., Spottiswoode, C. N., Iwaniuk, A. N., Boogert, N. J., Gutiérrez-Ibáñez, C., Overington, S. E., Wylie, D. R., & Lefebvre, L. (2013). Brain size and morphology of the brood-parasitic and cerophagous honeyguides (Aves: Piciformes). *Brain, Behavior and Evolution*, *81*(3), 170–186. https://doi.org/10.1159/000348834

Cramp, S., & Perrins, C. (Eds.). (1994). *Handbook of the birds of Europe, the Middle East and North Africa*. Oxford University Press.

Culina, A., Firth, J. A., & Hinde, C. A. (2020). Familiarity breeds success: Pairs that meet earlier experience increased breeding performance in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1941). https://doi.org/10.1098/rspb.2020.1554rspb20201554

Culina, A., Radersma, R., & Sheldon, B. C. (2015). Trading up: The fitness consequences of divorce in monogamous birds. *Biological Reviews*, *90*(4), 1015–1034. https://doi.org/10.1111/brv.12143

Dall, S. R. ., & Cuthill, I. C. (2016). The Information Costs of Generalism. *Oikos*, *80*(1), 197–202.

Dall, S. R. X., Giraldeau, L., Olsson, O., Mcnamara, J. M., & Stephens, D. W.

(2005). Information and its use by animals in evolutionary ecology. *Trends in Ecology and Evolution*, *20*(4). https://doi.org/10.1016/j.tree.2005.01.010

Das, M., Penke, Z., & Van Hooff, J. A. R. A. M. (1998). Postconflict affiliation and stress-related behavior of long-tailed macaque aggressors. *International Journal of Primatology*, *19*(1), 53–71. https://doi.org/10.1023/A:1020354826422

Davila Ross, M., Menzler, S., & Zimmermann, E. (2008). Rapid facial mimicry in orangutan play. *Biology Letters*, *4*(1), 27–30. https://doi.org/10.1098/rsbl.2007.0535

Davis, E. S. (2002). Male reproductive tactics in the mallard, Anas platyrhynchos: Social and hormonal mechanisms. *Behavioral Ecology and Sociobiology*, *52*(3), 224–231. https://doi.org/10.1007/s00265-002-0513-z

de Cáceres, M., Sol, D., Lapiedra, O., & Legendre, P. (2011). A framework for estimating niche metrics using the resemblance between qualitative resources. *Oikos*, *120*(9), 1341–1350. https://doi.org/10.1111/j.1600-0706.2011.19679.x

de Waal, F. B. M. (2010). The ‘Russian doll’ model of empathy and imitation. In S. Braten (Ed.), *On Being Moved, From Mirror Neurons to Empathy* (pp. 49–69). John Benjamins Publishing Company. https://doi.org/10.1007/978-90-481-3316-1

de Waal, F. B. M. (2016). What I learned from tickling apes. *New York Times*.

de Waal, F. B. M., & Aureli., F. (1996). Consolation, reconciliation, and a possible cognitive difference between macaques and chimpanzees. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching Into Thought: The Minds of the Great Apes* (pp. 80 –110). Cambridge University Press.

de Waal, F. B. M., & Aureli, F. (1997). Conflict resolution and distress alleviation in monkeys and apes. In I. I. Lederhendler, C. S. Carter, & B. Kirkpatrick (Eds.), *The integrative neurobiology of affiliation* (pp. 317–328). The MIT Press.

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/nrn.2017.72

de Waal, F. B. M., & van Roosmalen, A. (1979). Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology*, *5*(1), 55–66. https://doi.org/10.1007/BF00302695

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do Some Taxa Have Better Domain-General Cognition than others? A Meta-Analysis of Nonhuman Primate Studies. *Evolutionary Psychology*, *4*(1), 147470490600400. https://doi.org/10.1177/147470490600400114

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution*, *1*(5), 0112. https://doi.org/10.1038/s41559-017-0112

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2017). Skin temperature changes in wild chimpanzees upon hearing vocalizations of conspecifics. *Royal Society Open Science*, *4*(1), 160816. https://doi.org/10.1098/rsos.160816

Dicke, U., & Roth, G. (2016). Neuronal factors determining high intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1685). https://doi.org/10.1098/rstb.2015.0180

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 847–852. https://doi.org/10.1098/rspb.2004.2680

Dominey, W. J. (1984). Alternative mating tactics and evolutionarily stable strategies. *Integrative and Comparative Biology*, *24*(2), 385–396. https://doi.org/10.1093/icb/24.2.385

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE*, *5*(8). https://doi.org/10.1371/journal.pone.0012166

Dunbar, R. I. M. (1992). *Neocortex size as a constraint size in primates on group ecologically*. *March 1989*.

Dunbar, R. I. M. (1995). Neocortex size and group size in primates: A test of the hypothesis. In *Journal of Human Evolution* (Vol. 28, Issue 3, pp. 287–296). https://doi.org/10.1006/jhev.1995.1021

Dunbar, R. I. M. (1998). The Social Brain Hypothesis. *Evolutionary Anthropology*, 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.3.CO;2-P

Dunbar, R. I. M., & Bever, J. (1998). Neocortex size predicts group size in carnivores and some insectivores. *Ethology*, *104*(8), 695–708. https://doi.org/10.1111/j.1439-0310.1998.tb00103.x

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. In *Science* (Vol. 317, Issue 5843, pp. 1344–1347). https://doi.org/10.1126/science.1145463

Dunbar, R. I. M., & Shultz, S. (2017). Why are there so many explanations for primate brain evolution? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1727), 20160244. https://doi.org/10.1098/rstb.2016.0244

Duranton, C., & Gaunet, F. (2016). Behavioural synchronization from an ethological perspective: Overview of its adaptive value. *Adaptive Behavior*, *24*(3), 181–191. https://doi.org/10.1177/1059712316644966

Eberhard, W. G. (1982). Beetle horn dimorphism: making the best of a bad lot. *American Naturalist*, *119*, 420–426. https://doi.org/10.1525/jung.1.1986.6.2.38

Elie, J. E., Mathevon, N., & Vignal, C. (2011). Same-sex pair-bonds are equivalent to male-female bonds in a life-long socially monogamous songbird. *Behavioral Ecology and Sociobiology*, *65*(12), 2197–2208. https://doi.org/10.1007/s00265-011-1228-9

Ellis, S., Snyder-Mackler, N., Ruiz-Lambides, A., Platt, M. L., & Brent, L. J. N. (2019). Deconstructing sociality: The types of social connections that predict longevity in a group-living primate. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917). https://doi.org/10.1098/rspb.2019.1991

Emery, N. J. (2004). Are corvids 'feathered apes'? Cognitive evolution in crows, rooks and jackdaws. In S. Watanabe (Ed.), *Comparative Analysis of Minds* (pp. 181–213). Keio University Press. https://doi.org/10.1098/rstb.2006.1991

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. In *Science* (Vol. 306, Issue 5703, pp. 1903–1907). https://doi.org/10.1126/science.1098410

Emery, N. J., Seed, A., Bayern, A. M. P. Von, & Clayton, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society B*, *362*(1480), 489–505. https://doi.org/10.1098/rstb.2006.1991

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2012). The cognitive demands on cooperation in social dilemmas: An fMRI study. *Social Neuroscience*, *7*(5), 494–509. https://doi.org/10.1080/17470919.2012.655426

Ewen, J. G., Armstrong, D. P., & Lambert, D. M. (1999). Floater males gain reproductive success through extrapair fertilizations in the stitchbird. *Animal Behaviour*, *58*(2), 321–328. https://doi.org/10.1006/anbe.1999.1150

Finkenwirth, C., & Burkart, J. M. (2017). Long-term-stability of relationship structure in family groups of common marmosets, and its link to proactive prosociality. *Physiology and Behavior*, *173*, 79–86. https://doi.org/10.1016/j.physbeh.2017.01.032

Firth, J. A., Cole, E. F., Ioannou, C. C., Quinn, J. L., Aplin, L. M., Culina, A., McMahon, K., & Sheldon, B. C. (2018). Personality shapes pair bonding in a wild bird social system. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-018-0670-8

Fitch, M. A., & Shugart, G. W. (1984). Requirements for a mixed reproductive strategy in avian species. *American Naturalist*, *124*(1), 116–126. https://doi.org/10.1086/284255

Fowler, J. H., Dawes, C. T., & Christakis, N. A. (2009). Model of genetic variation in human social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 1720–1724. https://doi.org/10.1073/pnas.0806746106

Fox, K. C. R., Muthukrishna, M., & Shultz, S. (2017). The social and cultural roots of whale and dolphin brains. *Nature Ecology and Evolution*, *1*(11), 1699–1705. https://doi.org/10.1038/s41559-017-0336-y

Fraser, O. N., & Bugnyar, T. (2010a). Do ravens show consolation? Responses to distressed others. *PLoS ONE*, *5*(5), e10605. https://doi.org/10.1371/journal.pone.0010605

Fraser, O. N., & Bugnyar, T. (2010b). The quality of social relationships in ravens. *Animal Behaviour*, *79*(4), 927–933. https://doi.org/10.1016/j.anbehav.2010.01.008

Fraser, O. N., Schino, G., & Aureli, F. (2008). Components of relationship quality in chimpanzees. *Ethology*, *114*(9), 834–843. https://doi.org/10.1111/j.1439-0310.2008.01527.x

Freddolino, P. L., & Tavazoie, S. (2012). Beyond homeostasis: A predictive-dynamic framework for understanding cellular behavior. *Annual Review of Cell and Developmental Biology*, *28*, 363–384. https://doi.org/10.1146/annurev-cellbio-092910-154129

Freeberg, T. M., Dunbar, R. I. M., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1785–1801. https://doi.org/10.1098/rstb.2011.0213

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1324–1330. https://doi.org/10.1111/2041-210X

Fristoe, T. S., & Botero, C. A. (2019). Alternative ecological strategies lead to avian brain size bimodality in variable habitats. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11757-x

Fristoe, T. S., Iwaniuk, A. N., & Botero, C. A. (2017). Big brains stabilize populations and facilitate colonization of variable habitats in birds. *Nature Ecology and Evolution*, *1*(11), 1706–1715. https://doi.org/10.1038/s41559-017-0316-2

Funahashi, S. (2001). Neuronal mechanisms of executive control by the prefrontal cortex. *Neuroscience Research*, *39*(2), 147–165. https://doi.org/10.1016/S0168-0102(00)00224-8

Galdikas, B. M. F. (1985). Subadult male orangutan sociality and reproductive behavior at Tanjung Puting. *American Journal of Primatology*, *8*(2), 87–99. https://doi.org/10.1002/ajp.1350080202

Garamszegi, L. Z., Møller, A. P., & Erritzøe, J. (2002). Coevolving avian eye size and brain size in relation to prey capture and nocturnality. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1494), 961–967. https://doi.org/10.1098/rspb.2002.1967

García-Peña, G. E., Sol, D., Iwaniuk, A. N., & Székely, T. (2013). Sexual selection on brain size in shorebirds (Charadriiformes). *Journal of Evolutionary Biology*, *26*(4), 878–888. https://doi.org/10.1111/jeb.12104

Gilbert, L., Burke, T., & Krupa, A. (1998). No evidence for extra-pair paternity in the western gull. *Molecular Ecology*, *7*(11), 1549–1552. https://doi.org/10.1046/j.1365-294x.1998.00488.x

Gill, L. F., van Schaik, J., von Bayern, A. M. P., & Gahr, M. L. (2020). Genetic monogamy despite frequent extrapair copulations in “strictly monogamous” wild jackdaws. *Behavioral Ecology*, *31*, 247–260. https://doi.org/10.1093/beheco/arz185

Gillies, N., Guilford, T., & Catry, P. (2021). Allopreening in the Black-browed Albatross (Thalassarche melanophris): an exploration of patterns and possible functions. *Ibis*. https://doi.org/10.1111/ibi.12960

Gladstone, D. E. (1979). Promiscuity in monogamous colonial birds. *The American Naturalist*, *114*(4), 545–557.

Gnanadesikan, G. E., Hare, B., Snyder-Mackler, N., & MacLean, E. L. (2020). Estimating the heritability of cognitive traits across dog breeds reveals highly heritable inhibitory control and communication factors. *Animal Cognition*, *23*(5), 953–964. https://doi.org/10.1007/s10071-020-01400-4

González-forero, M., & Gardner, A. (2018). Inference of ecological and social drivers of human brain-size evolution. *Nature*.

Gosler, A., Clement, P., & Christie, D. A. (2020). Great Tit (Parus major). In S. . Billerman, B. . Keeney, P. . Rodewald, & T. . Schulenberg (Eds.), *Birds of the World* (Version 1.). Cornell Lab of Ornithology. https://doi.org/10.2173/bow.gretit1.01

Gowaty, P. A., & Buschhaus, N. (1998). Ultimate causation of aggressive and forced copulation in birds: Female resistance, the CODE hypothesis, and social monogamy. *American Zoologist*, *38*(1), 207–225. https://doi.org/10.1093/icb/38.1.207

Gray, E. M. (1996). Female control of offspring paternity in a western population of red-winged blackbirds (Agelaius phoeniceus). *Behavioral Ecology and Sociobiology*, *38*(4), 267–278. https://doi.org/10.1007/s002650050242

Greggor, A. L., McIvor, G. E., Clayton, N. S., & Thornton, A. (2016). Contagious risk taking: Social information and context influence wild jackdaws' responses to novelty and risk. *Scientific Reports*, *6*, 27764. https://doi.org/10.1038/srep27764

Griffith, S. C., Holleley, C. E., Mariette, M. M., Pryke, S. R., & Svedin, N. (2010). Low level of extrapair parentage in wild zebra finches. *Animal Behaviour*, *79*(2), 261–264. https://doi.org/10.1016/j.anbehav.2009.11.031

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. G. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7*, 1071–1075.

Gross, M. R. (1996). Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology & Evolution*, *11*, 92–98.

Güntürkün, O., & Bugnyar, T. (2016). Cognition without Cortex. *Trends in Cognitive Sciences*, *20*(4), 291–303. https://doi.org/10.1016/j.tics.2016.02.001

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22. https://doi.org/10.18637/jss.v033.i02

Hahn, L. G., Hooper, R., McIvor, G. E., & Thornton, A. (2020). Cooperative nest building in wild jackdaw pairs. *BioRxiv*.

Hartig, F. (2020). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.3.2.0.* https://cran.r-project.org/package=DHARMa

Harvey, P. H., & Krebs, J. R. (1990). Comparing brains. In *Science* (Vol. 249, Issue 4965, pp. 140–146). https://doi.org/10.1126/science.2196673

Hatchwell, B. J. (1988). Intraspecific Variation in Extra-Pair Copulation and Mate Defence in Common Guillemots Uria aalge. *Behaviour*, *107*(3), 157–185.

Healy, S. D., & Rowe, C. (2007). A critique of comparative studies of brain size. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 274, Issue 1609, pp. 453–464). https://doi.org/10.1098/rspb.2006.3748

Heathcote, R. J. P., Darden, S. K., Franks, D. W., Ramnarine, I. W., & Croft, D. P. (2017). Fear of predation drives stable and differentiated social relationships in guppies. *Scientific Reports*, *7*(February), 1–10. https://doi.org/10.1038/srep41679

Heinsohn, R. G. (1988). Inter-group ovicide and nest destruction in cooperatively breeding white-winged choughs. *Animal Behaviour*, *36*(6), 1856–1858. https://doi.org/10.1016/S0003-3472(88)80134-9

Helfenstein, F., Tirard, C., Danchin, E., & Wagner, R. H. (2004). Low frequency of extra-pair paternity and high frequency of adoption in black-legged kittiwakes. *The Condor*, 115–120.

Henderson, I. G., & Hart, P. J. B. (1993). Provisoning, Parental Investment and Reproductive Success in Jackdaws Corvus monedula. *Ornis Scandinavica1*, *24*(2), 142–148.

Henderson, I. G., Hart, P. J. B., & Burke, T. (2000). Strict monogamy in a semi-colonial passerine: the Jackdaw (Corvus monedula). *Journal of Avian*

Biology, *31*(2), 177–182. https://doi.org/10.1034/j.1600-048X.2000.310209.x

Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J. T., & Ward, A. J. W. (2011). Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(46), 18726–18731. https://doi.org/10.1073/pnas.1109355108

Herculano-Houzel, S. (2011). Scaling of brain metabolism with a fixed energy budget per neuron: Implications for neuronal activity, plasticity and evolution. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0017514

Heyes, C. (2012). Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Hogg, J. T., & Forbes, S. H. (1997). Mating in bighorn sheep: Frequent male reproduction via a high-risk “unconventional” tactic. *Behavioral Ecology and Sociobiology*, *41*(1), 33–48. https://doi.org/10.1007/s002650050361

Holekamp, K. E. (2007). Questioning the social intelligence hypothesis. *Trends in Cognitive Sciences*, *11*(2), 65–69. https://doi.org/10.1016/j.tics.2006.11.003

Hrdy, S. B. (2009). *Mother Nature: A History of Mothers, Infants, and Natural Selection*. Harvard University Press.

Huebner, F., Fichtel, C., & Kappeler, P. M. (2018). Linking cognition with fitness in a wild primate: Fitness correlates of problem-solving performance and spatial learning ability. *Philosophical Transactions of the Royal Society B-Biological Sciences*. https://doi.org/10.1098/rstb.2017.0295

Humphrey, N. K. (1976). The social function of intellect. *Growing Points in Ethology*, *37*(5), 303–317. https://doi.org/10.2307/375925

Ihle, M., Kempenaers, B., & Forstmeier, W. (2015). Fitness Benefits of Mate Choice for Compatibility in a Socially Monogamous Species. *PLoS Biology*, *13*(9), 1–21. https://doi.org/10.1371/journal.pbio.1002248

Ihle, M., Pick, J. L., Winney, I. S., Nakagawa, S., & Burke, T. (2019). Measuring up to reality: Null models and analysis simulations to study parental coordination over provisioning offspring. *Frontiers in Ecology and Evolution*, *7*(APR), 1–15. https://doi.org/10.3389/fevo.2019.00142

Ikkatai, Y., & Watanabe, S. (2015). Eye surface temperature detects stress response in budgerigars (Melopsittacus undulatus). *NeuroReport*, *26*(11), 642–646. https://doi.org/10.1097/WNR.0000000000000403

Inoue, N., & Shimada, M. (2020). Comparisons of activity budgets, interactions, and social structures in captive and wild chimpanzees (Pan troglodytes). *Animals*, *10*(6), 1–16. https://doi.org/10.3390/ani10061063

Iwaniuk, A. N., & Arnold, K. E. (2004). Is cooperative breeding associated with bigger brains? A comparative test in the Corvida (Passeriformes). *Ethology*, *110*(3), 203–220. https://doi.org/10.1111/j.1439-0310.2003.00957.x

Iwaniuk, A. N., Dean, K. M., & Nelson, J. E. (2004). A mosaic pattern characterizes the evolution of the avian brain. *Proceedings of the Royal Society B: Biological Sciences*, *271*(SUPPL. 4), 148–151. https://doi.org/10.1098/rsbl.2003.0127

Iwaniuk, A. N., Dean, K. M., & Nelson, J. E. (2005). Interspecific allometry of the brain and brain regions in parrots (Psittaciformes): Comparisons with other birds and primates. *Brain, Behavior and Evolution*, *65*(1), 40–59. https://doi.org/10.1159/000081110

Iwaniuk, A. N., & Nelson, J. E. (2002). Can endocranial volume be used as an estimate of brain size in birds? *Canadian Journal of Zoology*, *80*(1), 16–23. https://doi.org/10.1139/z01-204

Iwaniuk, A. N., & Nelson, J. E. (2003). Developmental differences are correlated with relative brain size in birds: a comparative analysis. *Canadian Journal of Zoology*, *81*(12), 1913–1928. https://doi.org/10.1139/z03-190

Jacoby, D. M. P., Fear, L. N., Sims, D. W., & Croft, D. P. (2014). Shark personalities? Repeatability of social network traits in a widely distributed predatory fish. *Behavioral Ecology and Sociobiology*, *68*(12), 1995–2003. https://doi.org/10.1007/s00265-014-1805-9

Jelbert, S. A., Singh, P. J., Gray, R. D., Taylor, A. H., & Marshall, J. A. R. (2015). New caledonian crows rapidly solve a collaborative problem without cooperative cognition. *PLoS ONE*, *10*(8), 1–17. https://doi.org/10.1371/journal.pone.0133253

Jerem, P., Jenni-Eiermann, S., Herborn, K., McKeegan, D., McCafferty, D. J., & Nager, R. G. (2018). Eye region surface temperature reflects both energy reserves and circulating glucocorticoids in a wild bird. *Scientific Reports*, *8*(1), 1907. https://doi.org/10.1038/s41598-018-20240-4

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

Johnson, M. P., & Leslie, L. (1982). Couple Involvement and Network Structure : A Test of the Dyadic Withdrawal Hypothesis Author. *Social Psychology Quarterly*, *45*(1), 34–43.

Johnstone, R. A., Manica, A., Fayet, A. L., Stoddard, M. C., Rodriguez-Gironés, M. A., & Hinde, C. A. (2014). Reciprocity and conditional cooperation between great tit parents. *Behavioral Ecology*, *25*(1), 216–222. https://doi.org/10.1093/beheco/art109

Johnstone, R. A., & Savage, J. L. (2019). Conditional Cooperation and Turn-Taking in Parental Care. *Frontiers in Ecology and Evolution*, *7*(October), 1–11. https://doi.org/10.3389/fevo.2019.00335

Jolly, A. (1966). Lemur social behavior and primate intelligence. *Science*, *153*(3735), 501–506. https://doi.org/10.1126/science.153.3735.501

Kalmijn, M. (2003). Shared friendship networks and the life course: An analysis of survey data on married and cohabiting couples. *Social Networks*, *25*(3), 231–249. https://doi.org/10.1016/S0378-8733(03)00010-8

Kano, F., Hirata, S., Deschner, T., Behringer, V., & Call, J. (2016). Nasal temperature drop in response to a playback of conspecific fights in chimpanzees: A thermo-imaging study. *Physiology and Behavior*, *155*, 83–94. https://doi.org/10.1016/j.physbeh.2015.11.029

Kendeigh, S. C. (1952). Parental Care and Its Evolution in Birds. *Illinois Biological Monographs*, *22*, 1–3.

Kilham, L. (1984). Intra- and extrapair copulatory behavior of American Crows.

The Wilson Bulletin, *96*(4), 716–717.

Kings, M. (2018). Foraging tactics and social networks in wild jackdaws (Doctoral Dissertation). *Unpublished*.

Knafo, A., Israel, S., & Ebstein, R. P. (2011). Heritability of children’s prosocial behavior and differential susceptibility to parenting by variation in the dopamine receptor D4 gene. *Development and Psychopathology*, *23*(1), 53–67. https://doi.org/10.1017/S0954579410000647

Koski, S. E., de Vries, H., van de Kraats, A., & Sterck, E. H. M. (2012). Stability and Change of Social Relationship Quality in Captive Chimpanzees (Pan troglodytes). *International Journal of Primatology*, *33*(4), 905–921. https://doi.org/10.1007/s10764-012-9623-2

Koski, S. E., & Sterck, E. H. M. (2009). Post-conflict third-party affiliation in chimpanzees: What’s in it for the third party? *American Journal of Primatology*, *71*(5), 409–418. https://doi.org/10.1002/ajp.20668

Kotrschal, A., Szorkovszky, A., Romenskyy, M., Perna, A., Buechel, S. D., Zeng, H. L., Pelckmans, K., Sumpter, D., & Kolm, N. (2018). Brain size does not impact shoaling dynamics in unfamiliar groups of guppies (Poecilia reticulata). *Behavioural Processes*, *147*(September 2017), 13–20. https://doi.org/10.1016/j.beproc.2017.12.006

Kozlovsky, D. Y., Branch, C. L., & Pravosudov, V. V. (2015). Elevation-related differences in parental risk-taking behavior are associated with cognitive variation in mountain chickadees. *Ethology*, *121*(4), 383–394. https://doi.org/10.1111/eth.12350

Kralj-Fišer, S., Scheiber, I. B. R., Blejec, A., Moestl, E., & Kotrschal, K. (2007). Individualities in a flock of free-roaming greylag geese: Behavioral and physiological consistency over time and across situations. *Hormones and Behavior*, *51*(2), 239–248. https://doi.org/10.1016/j.yhbeh.2006.10.006

Kramer, G. (1932). Beobachtungen und Fragen zur Biologie des Kolkraben (Corvus c. corax). *Journal Fuer Ornithologie*, *80*, 329–342.

Ksepka, D. T., Balanoff, A. M., Smith, N. A., Bever, G. S., Bhullar, B. A. S., Bourdon, E., Braun, E. L., Burleigh, J. G., Clarke, J. A., Colbert, M. W., Corfield, J. R., Degrange, F. J., De Pietri, V. L., Early, C. M., Field, D. J., Gignac, P. M., Gold, M. E. L., Kimball, R. T., Kawabe, S., … Smaers, J. B. (2020). Tempo and Pattern of Avian Brain Size Evolution. *Current Biology*, *30*(11), 2026-2036.e3. https://doi.org/10.1016/j.cub.2020.03.060

Kubitza, R. J., Bugnyar, T., & Schwab, C. (2015). Pair bond characteristics and maintenance in free-flying jackdaws Corvus monedula: Effects of social context and season. *Journal of Avian Biology*, *46*(2), 206–215. https://doi.org/10.1111/jav.00508

Kummer, H. (1978). On the value of social relationships to nonhuman primates: A heuristic scheme. *Social Science Information*, *17*(4–5), 687–705. https://doi.org/10.1177/053901847801700418

Kustra, M. C., & Alonzo, S. H. (2020). Sperm and alternative reproductive tactics: A review of existing theory and empirical data: Sperm and alternative tactics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1813). https://doi.org/10.1098/rstb.2020.0075rstb20200075

Lack, D. (1966). *Population studies of birds.* Oxford University Press.

Lack, D. (1968). Ecological adaptations for breeding in birds. *The Auk*, *86*(4), 774–777.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lang, M., Bahna, V., Shaver, J. H., Reddish, P., & Xygalatas, D. (2017). Sync to link: Endorphin-mediated synchrony effects on cooperation. *Biological Psychology*, *127*(June), 191–197. https://doi.org/10.1016/j.biopsycho.2017.06.001

Langley, E. J. G., Adams, G., Beardsworth, C. E., Dawson, D. A., Laker, P. R., Van Horik, J. O., Whiteside, M. A., Wilson, A. J., & Madden, J. R. (2020). Heritability and correlations among learning and inhibitory control traits. *Behavioral Ecology*, *31*(3), 798–806. https://doi.org/10.1093/beheco/araa029

Laubu, C., Louâpre, P., & Dechaume-Moncharmont, F. X. (2019). Pair-bonding influences affective state in a monogamous fish species. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904). https://doi.org/10.1098/rspb.2019.0760

Lea, A. J., Blumstein, D. T., Wey, T. W., & Martin, J. G. A. (2010). Heritable victimization and the benefits of agonistic relationships. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21587–21592. https://doi.org/10.1073/pnas.1009882107

Lee, V. E., McIvor, G. E., & Thornton, A. (2019). Testing relationship recognition in wild jackdaws (Corvus monedula). *Scientific Reports*, *9*(1), 6710. https://doi.org/10.1038/s41598-019-43247-x

Lee, V. E., Régli, N., McIvor, G. E., & Thornton, A. (2019). Social learning about dangerous people by wild jackdaws. *Royal Society Open Science*, *6*(9), 191031. https://doi.org/10.1098/rsos.191031

Lendvai, Á. Z., Bókony, V., Angelier, F., Chastel, O., & Sol, D. (2013). Do smart birds stress less? An interspecific relationship between brain size and corticosterone levels. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1770). https://doi.org/10.1098/rspb.2013.1734

Levine, B. A., Douglas, M. R., Adams, A. A. Y., Lardner, B., Reed, R. N., Savidge, J. A., & Douglas, M. E. (2019). *Genomic pedigree reconstruction identifies predictors of mating and reproductive success in an invasive vertebrate*. *August*, 11863–11877. https://doi.org/10.1002/ece3.5694

Li, L. F., Yuan, W., He, Z. X., Wang, L. M., Jing, X. Y., Zhang, J., Yang, Y., Guo, Q. Q., Zhang, X. N., Cai, W. Q., Hou, W. J., Jia, R., & Tai, F. D. (2019). Involvement of oxytocin and GABA in consolation behavior elicited by socially defeated individuals in mandarin voles. *Psychoneuroendocrinology*, *103*(December 2018), 14–24. https://doi.org/10.1016/j.psyneuen.2018.12.238

Liebers, D., & Peter, H.-U. (1998). Intraspecific interactions in jackdaws Corvus monedula: A field study combined with parentage analysis. *Ardea*, *86*(2), 221–235.

Ling, H., Mclvor, G. E., van der Vaart, K., Vaughan, R. T., Thornton, A., &

Ouellette, N. T. (2019). Costs and benefits of social relationships in the collective motion of bird flocks. *Nature Ecology and Evolution*, *3*(6), 943–948. https://doi.org/10.1038/s41559-019-0891-5

Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., Jelbert, S., Lukas, D., Mares, R., Navarrete, A. F., Shigeno, S., & Montgomery, S. H. (2018). Beyond brain size: Uncovering the neural correlates of behavioral and cognitive specialization. *Comparative Cognition and Behavior Reviews*, *13*, 55–90. https://doi.org/10.3819/CCBR.2018.130008

Logan, C. J., Emery, N. J., & Clayton, N. S. (2013). Alternative behavioral measures of postconflict affiliation. *Behavioral Ecology*, *24*(1), 98–112. https://doi.org/10.1093/beheco/ars140

Logan, C. J., Ostojić, L., Clayton, N. S., Ostoji, L., & Clayton, N. S. (2013). Rook, But Not Jackdaw, Post-Conflict Third-Party Affiliation Reduces Aggression for Aggressors. *Ethology*, *119*(5), 427–435. https://doi.org/10.1111/eth.12078

Lorenz, K. (1931). Beitrage zur Ethologie sozialer Corviden. *Journal Fuer Ornithologie*, *79*, 67–127.

Low, M., Castro, I., & Berggren, Å. (2005). Cloacal erection promotes vent apposition during forced copulation in the New Zealand stitchbird (hihi): Implications for copulation efficiency in other species. *Behavioral Ecology and Sociobiology*, *58*(3), 247–255. https://doi.org/10.1007/s00265-005-0935-5

Lukas, D., & Clutton-Brock, T. (2018). Social complexity and kinship in animal societies. *Ecology Letters*, *21*(8), 1129–1134. https://doi.org/10.1111/ele.13079

MacLean, E. L., Barrickman, N. L., Johnson, E. M., & Wall, C. E. (2009). Sociality, ecology, and relative brain size in lemurs. *Journal of Human Evolution*, *56*(5), 471–478. https://doi.org/10.1016/j.jhevol.2008.12.005

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140–E2148. https://doi.org/10.1073/pnas.1323533111

Maclean, E. L., Tin, D. U. S., Ritt, J. M. E. R., & Non, E. M. B. (2008). *Social complexity predicts transitive reasoning in prosimian primates*. 479–486. https://doi.org/10.1016/j.anbehav.2008.01.025

Madge, S., & de Juana, E. (2020). Eurasian Jackdaw (Corvus monedula). In J. del Hoyo, A. Elliott, J. Sargatal, D. E. Christie, & E. de Juana (Eds.), *Birds of the World*. Cornell Lab of Ornithology.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12.

Massen, J. J. M., & Sterck, E. H. M. (2013). Stability and Durability of Intra- and Intersex Social Bonds of Captive Rhesus Macaques (Macaca mulatta). *International Journal of Primatology*, *34*(4), 770–791. https://doi.org/10.1007/s10764-013-9695-7

Massen, J. J. M., Szipl, G., Spreafico, M., & Bugnyar, T. (2014). Ravens intervene in others’ bonding attempts. *Current Biology*, *24*(22), 2733–2736. https://doi.org/10.1016/j.cub.2014.09.073

Mayer, C., & Pasinelli, G. (2013). New support for an old hypothesis: Density affects extra-pair paternity. *Ecology and Evolution*, *3*(3), 694–705. https://doi.org/10.1002/ece3.489

Mcauliffe, K., & Thornton, A. (2015). The psychology of cooperation in animals: An ecological approach. *Journal of Zoology*, *295*(1), 23–35. https://doi.org/10.1111/jzo.12204

McFarland, R., Murphy, D., Lusseau, D., Henzi, S. P., Parker, J. L., Pollet, T. V., & Barrett, L. (2017). The ‘strength of weak ties’ among female baboons: fitness-related benefits of social bonds. *Animal Behaviour*, *126*, 101–106. https://doi.org/10.1016/j.anbehav.2017.02.002

McKinney, F., Cheng, K. M., & Bruggers, D. J. (1984). Sperm Competition in Apparently Monogamous Birds. In R. L. Smith (Ed.), *Sperm Competition and the Evolution of Animal Mating Systems* (pp. 523–545). Academic Press, INC.

McKinney, F., & Evarts, S. (1998). Sexual Coercion in Waterfowl and Other Birds. *Ornithological Monographs, No. 49, Avian Reproductive Tactics: Female and Male Perspectives*, *49*, 163–195.

McNamara, J. M. (2013). Towards a richer evolutionary game theory. *Journal of the Royal Society Interface*, *10*(88), 20130544. https://doi.org/10.1098/rsif.2013.0544

McNamara, J. M., & Dall, S. R. X. (2010). Information is a fitness enhancing resource. *Oikos*, *119*(2), 231–236. https://doi.org/10.1111/j.1600-0706.2009.17509.x

McNamara, J. M., & Leimar, O. (2020). *Game Theory in Biology: Concepts and Frontiers.* Oxford University Press.

Melchers, M., Montag, C., Reuter, M., Spinath, F. M., & Hahn, E. (2016). How heritable is empathy? Differential effects of measurement and subcomponents. *Motivation and Emotion*, *40*(5), 720–730. https://doi.org/10.1007/s11031-016-9573-7

Minias, P., & Podlaszczuk, P. (2017). Longevity is associated with relative brain size in birds. *Ecology and Evolution*, *7*(10), 3558–3566. https://doi.org/10.1002/ece3.2961

Mitani, J. C. (2009). Male chimpanzees form enduring and equitable social bonds. *Animal Behaviour*, *77*(3), 633–640. https://doi.org/10.1016/j.anbehav.2008.11.021

Montgomerie, R., & Briskie, J. V. (2007). Anatomy and evolution of copulatory structures. In B. G. M. Jamieson (Ed.), *Reproductive biology and phylogeny of birds Part A: Phylogeny, Morphology, Hormones and Fertilization Science Publishers* (pp. 115–148).

Montgomery, S. H., Currie, A., Lukas, D., Boogert, N., Buskell, A., Cross, F. R., Jelbert, S., Avin, S., Mares, R., Navarrete, A. F., Shigeno, S., & Logan, C. J. (2018). Ingredients for understanding brain and behavioral evolution: Ecology, phylogeny, and mechanism. *Comparative Cognition and Behavior Reviews*, *13*, 99–104. https://doi.org/10.3819/CCBR.2018.130011

Moore, B. L., Connor, R. C., Allen, S. J., Krützen, M., & King, S. L. (2020).

Acoustic coordination by allied male dolphins in a cooperative context. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1924). https://doi.org/10.1098/rspb.2019.2944

Mumme, R. L., Koenig, W. D., & Pitelka, F. A. (1983). Reproductive competition in the communal acorn woodpecker: sisters destroy each other's eggs. *Nature*, *306*(5943), 583–584. https://doi.org/10.1038/306583a0

Naves, L. C., Cam, E., & Monnat, J. Y. (2007). Pair duration, breeding success and divorce in a long-lived seabird: benefits of mate familiarity? *Animal Behaviour*, *73*(3), 433–444. https://doi.org/10.1016/j.anbehav.2006.10.004

Nedelcu, I. T., & Hirschenhauser, K. (2010). Maintenance of the monogamous pair bond. *The Social Life of Greylag Geese: Patterns, Mechanisms and Evolutionary Function in an Avian Model System*, 65–87. https://doi.org/10.1017/CBO9781139049955.007

Nowbahari, E., Scohier, A., Durand, J. L., & Hollis, K. L. (2009). Ants, Cataglyphis cursor, use precisely directed rescue behavior to free entrapped relatives. *PLoS ONE*, *4*(8), 4–7. https://doi.org/10.1371/journal.pone.0006573

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences*, *113*(26), 7255–7260. https://doi.org/10.1073/pnas.1517131113

Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*, *78*(4), 1001–1010. https://doi.org/10.1016/j.anbehav.2009.06.033

Pacheco, X. P., & Madden, J. R. (2021). Does the social network structure of wild animal populations differ from that of animals in captivity? *Behavioural Processes*, *190*(April 2020), 104446. https://doi.org/10.1016/j.beproc.2021.104446

Palagi, E., & Cordoni, G. (2009). Postconflict third-party affiliation in Canis lupus: do wolves share similarities with the great apes? *Animal Behaviour*, *78*(4), 979–986. https://doi.org/10.1016/j.anbehav.2009.07.017

Palagi, E., Dall’Olio, S., Demuru, E., & Stanyon, R. (2014). Exploring the evolutionary foundations of empathy: Consolation in monkeys. *Evolution and Human Behavior*, *35*(4), 341–349. https://doi.org/10.1016/j.evolhumbehav.2014.04.002

Palagi, E., & Norscia, I. (2013). Bonobos protect and console friends and kin. *PLoS ONE*, *8*(11), e79290. https://doi.org/10.1371/journal.pone.0079290

Palagi, E., Paoli, T., & Tarli, S. B. (2004). Reconciliation and Consolation in Captive Bonobos (Pan paniscus). *American Journal of Primatology*, *62*(1), 15–30. https://doi.org/10.1002/ajp.20000

Palmer, C. T., & Thornhill, R. (2003). Straw men and fairy tales: Evaluating reactions to A Natural History of Rape. *Journal of Sex Research*, *40*(3), 249–255. https://doi.org/10.1080/00224490309552189

Paradiso, E., Gazzola, V., & Keysers, C. (2021). Neural mechanisms necessary for empathy-related phenomena across species. *Current Opinion in Neurobiology*, *68*, 107–115. https://doi.org/10.1016/j.conb.2021.02.005

Parejo, D., Avilés, J. M., & Expósito, M. (2015). Hatching Asynchrony and Spring Climatic Conditions in the European Roller Hatching Asynchrony and Spring Climatic Conditions in the European Roller. *Evolutionary Biology*, *42*(4), 443–451. https://doi.org/10.1007/s11692-015-9337-4

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. In *Methods in Ecology and Evolution* (Vol. 8, Issue 10). https://doi.org/10.1111/2041-210X.12775

Paxton, A., & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, *45*(2), 329–343. https://doi.org/10.3758/s13428-012-0249-2

Penn, D. C. (2011). How folk psychology ruined comparative psychology: And how scrub jays can save it. *Animal Thinking: Contemporary Issues in Comparative Cognition*, *8*, 253–265.

Pérez-Barbería, F. J., Shultz, S., & Dunbar, R. I. M. (2007). Evidence for coevolution of sociality and relative brain size in three orders of mammals. *Evolution*, *61*(12), 2811–2821. https://doi.org/10.1111/j.1558-

5646.2007.00229.x

Pérez-Manrique, A., & Gomila, A. (2016). RE: Comments on the article: “Oxytocin-dependent consolation behavior in rodents”. *Science*, *351*(6271). https://science.sciencemag.org/content/351/6271/375/tab-e-letters

Pérez-Manrique, A., & Gomila, A. (2018). The comparative study of empathy: sympathetic concern and empathic perspective-taking in non-human animals. *Biological Reviews*, *93*(1), 248–269. https://doi.org/10.1111/brv.12342

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Picard, A. M., Mundry, R., Auersperg, A. M., Boeving, E. R., Boucherie, P. H., Bugnyar, T., Dufour, V., Emery, N. J., Federspiel, I. G., Gajdon, G. K., Matjaž, J. G., Lisa, H., Eithne, H., Lambert, M. L., Massen, J. J. M., Rodrigues, M. A., Schiestl, M., Schwing, R., Szabo, B., … Slocombe, K. E. (2020). *Why preen others ? Predictors of allopreening in parrots and corvids and comparisons to grooming in great apes*. *May 2019*, 207–228. https://doi.org/10.1111/eth.12999

Podlas, K. A., & Richner, H. (2013). The adaptive function of hatching asynchrony: An experimental study in great tits. *Animal Behaviour*, *86*(3), 567–576. https://doi.org/10.1016/j.anbehav.2013.06.012

Povinelli, D. J., Nelson, K. E., & Boysen, S. T. (1992). Comprehension of role reversal in chimpanzees: evidence of empathy? *Animal Behaviour*, *43*(4), 633–640. https://doi.org/10.1016/S0003-3472(05)81022-X

Powell, L. E., Isler, K., Barton, R. A., Powell, L. E., & Barton, R. A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proc Royal Soc*, 1–8.

Pravosudov, V. V., Sanford, K., & Hahn, T. P. (2007). On the evolution of brain size in relation to migratory behaviour in birds. *Animal Behaviour*, *73*(3), 535–539. https://doi.org/10.1016/j.anbehav.2006.10.005

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioural and Brain Sciences*, 515–526. http://www.informaworld.com/index/713824049.pdf

Pritchard, D. J., Hurly, T. A., Tello-Ramos, M. C., & Healy, S. D. (2016). Why study cognition in the wild (and how to test it)? *Journal of the Experimental Analysis of Behavior*, *105*(1), 41–55. https://doi.org/10.1002/jeab.195

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., Bakker, P. I. W. De, Daly, M. J., & Sham, P. C. (2007). PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(September), 559–575. https://doi.org/10.1086/519795

Quervel-Chaumette, M., Faerber, V., Faragó, T., Marshall-Pescini, S., & Range, F. (2016). Investigating empathy-like responding to conspecifics' distress in pet dogs. *PLoS ONE*, *11*(4), e0152920. https://doi.org/10.1371/journal.pone.0152920

R Core Team. (2017). *R: A language and environment for statistical computing.* (3.4.0). R Foundation for Statistical Computing. https://www.r-project.org/

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research. R package version, 1(10).*

Rios-Cardenas, O., Bono, L., & Morris, M. R. (2018). Frequency-dependent selection and fluctuations around an equilibrium for alternative reproductive tactics in a swordtail. *Animal Behaviour*, *140*, 19–28. https://doi.org/10.1016/j.anbehav.2018.03.018

Rising, J. D., & Somers, K. M. (1989). The Measurement of Overall Body Size in Birds. *The Auk*, *106*(4), 666–674.

Rochette, N. C., Rivera-Colón, A. ., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Röell, A. (1978). Social behaviour of the jackdaw, Corvus monedula, in relation to its niche. *Behaviour*, *64*(1–2), 1–124.

https://doi.org/10.1163/156853978X00459

Rosati, A. G. (2017). Foraging Cognition: Reviving the Ecological Intelligence Hypothesis. *Trends in Cognitive Sciences*, *21*(9), 691–702. https://doi.org/10.1016/j.tics.2017.05.011

Roskaft, E. (1983). Male promiscuity and female adultery by the rook Corvus frugilegus. *Ornis Scandinavica*, *14*(3), 175–179. https://doi.org/10.2307/3676150

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, *9*(5), 250–257. https://doi.org/10.1016/j.tics.2005.03.005

Roth, T. C., LaDage, L. D., & Pravosudov, V. V. (2010). Learning capabilities enhanced in harsh environments: A common garden approach. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3187–3193. https://doi.org/10.1098/rspb.2010.0630

Sadino, J. M., & Donaldson, Z. R. (2018). Prairie Voles as a Model for Understanding the Genetic and Epigenetic Regulation of Attachment Behaviors. *ACS Chemical Neuroscience*, *9*(8), 1939–1950. https://doi.org/10.1021/acschemneuro.7b00475

Sakai, M., Morisaka, T., Kogi, K., Hishii, T., & Kohshima, S. (2010). Fine-scale analysis of synchronous breathing in wild Indo-Pacific bottlenose dolphins (Tursiops aduncus). *Behavioural Processes*, *83*(1), 48–53. https://doi.org/10.1016/j.beproc.2009.10.001

Sánchez-Macouzet, O., Rodríguez, C., & Drummond, H. (2014). Better stay together: Pair bond duration increases individual fitness independent of age-related variation. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1786). https://doi.org/10.1098/rspb.2013.2843

Savage, J. L., Browning, L. E., Manica, A., Russell, A. F., & Johnstone, R. A. (2017). Turn-taking in cooperative offspring provisioning: by-product of foraging behaviour or active response rule? *Behavioral Ecology and Sociobiology*, *71*(162).

Savage, J. L., Hinde, C. A., & Johnstone, R. A. (2020). Editorial: Cooperation and Coordination in the Family. *Frontiers in Ecology and Evolution*, *8*. https://doi.org/10.3389/fevo.2020.614863

Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13971

Scheiber, I. B. R., Weiß, B. M., Frigerio, D., & Kotrschal, K. (2005). Active and passive social support in families of greylag geese (Anser anser). *Behaviour*, *142*(11–12), 1535–1557. https://doi.org/10.1163/156853905774831873

Scheiber, I. B. R., Weiβ, B. M., Hirschenhauser, K., Wascher, C. A. F., Nedelcu, I. T., & Kotrschal, K. (2008). Does “Relationship Intelligence” Make Big Brains in Birds? *The Open Biology Journal*, *1*(1), 6–8. https://doi.org/10.2174/1874196700801010006

Schino, G., & Marini, C. (2012). Self-protective function of post-conflict bystander affiliation in mandrills. *PLoS ONE*, *7*(6), 2–7. https://doi.org/10.1371/journal.pone.0038936

Schmidt, K. A., Dall, S. R. X., & van Gils, J. A. (2010). The ecology of information: An overview on the ecological significance of making informed decisions. *Oikos*, *119*(2), 304–316. https://doi.org/10.1111/j.1600-0706.2009.17573.x

Schuck-paim, C., Alonso, W. J., & Ottoni, B. (2008). Cognition in an Ever-Changing World : Climatic Variability Is Associated with Brain Size in Neotropical Parrots. *Brain, Behavior and Evolution*, *71*, 200–215. https://doi.org/10.1159/000119710

Schülke, O., Bhagavatula, J., Vigilant, L., & Ostner, J. (2010). Social bonds enhance reproductive success in male macaques. *Current Biology*, *20*(24), 2207–2210. https://doi.org/10.1016/j.cub.2010.10.058

Schürmann, C. L., & van Hooff, J. A. R. A. M. (1986). Reproductive strategies of the orang-utan: New data and a reconsideration of existing sociosexual models. *International Journal of Primatology*, *7*(3), 265–287. https://doi.org/10.1007/BF02736392

Seed, A., Clayton, N. S., & Emery, N. J. (2007). Postconflict Third-Party Affiliation in Rooks, Corvus frugilegus. *Current Biology*, *17*(2), 152–158. https://doi.org/10.1016/j.cub.2006.11.025

Seed, A., Emery, N. J., & Clayton, N. S. (2009). Intelligence in corvids and apes: A case of convergent evolution? *Ethology*, *115*(5), 401–420. https://doi.org/10.1111/j.1439-0310.2009.01644.x

Seyfarth, R. M., & Cheney, D. L. (2012). Social Relationships, Social Cognition, and the Evolution of Mind in Primates. *Handbook of Psychology, Second Edition*, 574–594. https://doi.org/10.1002/9781118133880.hop203021

Sheldon, B. C. (1994). Male phenotype, fertility, and the pursuit of extra-pair copulations by female birds. *Proceedings of the Royal Society B: Biological Sciences*, *257*(1348), 25–30. https://doi.org/10.1098/rspb.1994.0089

Shettleworth, S. J. (2010a). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481. https://doi.org/10.1016/j.tics.2010.07.002

Shettleworth, S. J. (2010b). *Cognition, evolution, and behavior* (Second). Oxford University Press.

Shultz, S., & Dunbar, R. I. M. (2006). Both social and ecological factors predict ungulate brain size. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 207–215. https://doi.org/10.1098/rspb.2005.3283

Shultz, S., & Dunbar, R. I. M. (2007). The evolution of the social brain: anthropoid primates contrast with other vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1624), 2429–2436. https://doi.org/10.1098/rspb.2007.0693

Shultz, S., & Dunbar, R. I. M. (2010). Social bonds in birds are associated with brain size and contingent on the correlated evolution of life-history and increased parental investment. *Biological Journal of the Linnean Society*, *100*(1), 111–123. https://doi.org/10.1111/j.1095-8312.2010.01427.x

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social Bonds of Female Baboons Enhance Infant Survival. *Science*, *302*(5648), 1231–1234. https://doi.org/10.1126/science.1088580

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2010a). Female chacma baboons form strong, equitable, and enduring social bonds. *Behavioral Ecology and Sociobiology*, *64*(11), 1733–1747. https://doi.org/10.1007/s00265-010-0986-0

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2010b). Strong and consistent social bonds enhance the longevity of female baboons. *Current Biology*, *20*(15), 1359–1361. https://doi.org/10.1016/j.cub.2010.05.067

Silk, M. J., & Hodgson, D. J. (2021). Differentiated Social Relationships and the Pace-of-Life-History. *Trends in Ecology & Evolution*, 1–9. https://doi.org/10.1016/j.tree.2021.02.007

Sima, M. J., Matzinger, T., Bugnyar, T., & Pika, S. (2018). Reconciliation and third-party affiliation in carrion crows. *Ethology*, *124*(1), 33–44. https://doi.org/10.1111/eth.12699

Sima, M. J., Pika, S., & Bugnyar, T. (2016). Experimental Manipulation of Food Accessibility Affects Conflict Management Behaviour in Ravens. *Ethology*, *122*(2), 114–126. https://doi.org/10.1111/eth.12451

Smaers, J. B., Rothman, R. S., Hudson, D. R., Balanoff, A. ., Beatty, B., Dechmann, D. K. ., de Vries, D., Dunn, J. C., Fleagle, J. G., Gilbert, C. C., Goswami, A., Iwaniuk, A. N., Jungers, W. L., Kerney, M., Ksepka, D. T., & Safi, K. (2021). The evolution of mammalian brain size. *Science Advances*, *7*. https://doi.org/10.4324/9780203470640_chapter_7

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*, 207–218.

Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences*, *102*(15), 5460–5465. https://doi.org/10.1073/pnas.0408145102

Sol, D., Garcia, N., Iwaniuk, A., Davis, K., Meade, A., Boyle, W. A., & Székely, T. (2010). Evolutionary divergence in brain size between migratory and resident birds. *PLoS ONE*, *5*(3), 1–8. https://doi.org/10.1371/journal.pone.0009617

Sorato, E., Zidar, J., Garnham, L., Wilson, A., & Løvlie, H. (2018). Heritabilities and co-variation among cognitive traits in red junglefowl. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756). https://doi.org/10.1098/rstb.2017.0285

Sorenson, L. G. (1994). Forced extra-pair copulation and mate guarding in the white-cheeked pintail: Timing and trade-offs in an asynchronously breeding duck. *Animal Behaviour*, *48*(3), 519–533. https://doi.org/10.1006/anbe.1994.1273

Sparks, J. H. (1964). Flock structure of the red avadavat with particular reference to clumping and allopreening. *Animal Behaviour*, *12*(1), 125–136. https://doi.org/10.1016/0003-3472(64)90113-7

Spoon, T. R., Millam, J. R., & Owings, D. H. (2004). Variation in the stability of cockatiel (Nymphicus hollandicus) pair relationships: The roles of males, females, and mate compatibility. *Behaviour*, *141*(10), 1211–1234. https://doi.org/10.1163/1568539042729711

Spoon, T. R., Millam, J. R., & Owings, D. H. (2006). The importance of mate behavioural compatibility in parenting and reproductive success by cockatiels, Nymphicus hollandicus. *Animal Behaviour*, *71*(2), 315–326. https://doi.org/10.1016/j.anbehav.2005.03.034

Spoon, T. R., Millam, J. R., & Owings, D. H. (2007). Behavioural compatibility, extrapair copulation and mate switching in a socially monogamous parrot. *Animal Behaviour*, *73*(5), 815–824. https://doi.org/10.1016/j.anbehav.2006.10.010

Stanley, C. R., Mettke-Hofmann, C., Hager, R., & Shultz, S. (2018). Social stability in semiferal ponies: networks show interannual stability alongside seasonal flexibility. *Animal Behaviour*, *136*, 175–184. https://doi.org/10.1016/j.anbehav.2017.04.013

Sterling, P., & Laughlin, S. (2015). *Principles of neural design.* MIT Press.

Stockley, P. (1997). Sexual conflict resulting from adaptations to sperm competition. *Trends in Ecology and Evolution*, *12*(4), 154–159. https://doi.org/10.1016/S0169-5347(97)01000-8

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *8*(11), 1639–1644. https://doi.org/10.1111/2041-210X.12797

Street, S. E., Navarrete, A. F., Reader, S. M., & Laland, K. N. (2017). Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences*, *114*(30), 7908–7914. https://doi.org/10.1073/pnas.1620734114

Tebbich, S., & Teschke, I. (2014). Coping with uncertainty: Woodpecker finches (Cactospiza pallida) from an unpredictable habitat are more flexible than birds from a stable habitat. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091718

Thomas, R. J. (2015). *Data analysis with R statistical software: a guidebook for scientists*. Eco-explore.

Thornhill, R. (1981). Panorpa (Mecoptera: Panorpidae) Scorpionflies: Systems for Understanding Resource-Defense Polygyny and Alternative Male Reproductive Efforts. *Ann. Rev. Ecol. Syst.*, *12*, 355–386. https://doi.org/10.1007/978-3-319-47829-6_990-1

Thornhill, R. (1984). Alternative female choice tactics in the scorpionfly Hylobittacus apicalis(mecoptera) and their implications. *Integrative and Comparative Biology*, *24*(2), 367–383. https://doi.org/10.1093/icb/24.2.367

Thornhill, R., & Palmer, C. T. (2000). *A natural history of rape.*

Thornhill, R., & Wilmsen Thornhill, N. (1983). Human rape: An evolutionary analysis. *Ethology and Sociobiology*, *4*(3), 137–173. https://doi.org/10.1016/0162-3095(83)90027-4

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology*, *25*(6), 1299–1301. https://doi.org/10.1093/beheco/aru095

Thornton, A., & Lukas, D. (2012). Individual variation in cognitive performance:

Developmental and evolutionary perspectives. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 367, Issue 1603, pp. 2773–2783). https://doi.org/10.1098/rstb.2012.0214

Thornton, A., & McAuliffe, K. (2015). *Cognitive consequences of cooperative breeding ? A critical appraisal*. *295*, 12–22. https://doi.org/10.1111/jzo.12198

Thornton, A., Mcauliffe, K., Dall, S. R. X., Fernandez-Duque, E., Garber, P. A., & Young, A. J. (2016). Fundamental problems with the cooperative breeding hypothesis. A reply to Burkart & van Schaik. *Journal of Zoology*, *299*(2), 84–88. https://doi.org/10.1111/jzo.12351

Townsend, A. K. (2009). Extra-pair copulations predict extra-pair fertilizations in the American Crow. *Condor*, *111*(2), 387–392. https://doi.org/10.1525/cond.2009.090010

Townsend, A. K., Clark, A. B., Mcgowan, K. J., & Lovette, I. J. (2009). Reproductive partitioning and the assumptions of reproductive skew models in the cooperatively breeding American crow. *Animal Behaviour*, *77*(2), 503–512. https://doi.org/10.1016/j.anbehav.2008.10.030

Trail, P. W., Strahl, S. D., & Brown, J. L. (1981). Infanticide in Relation To Individual and Flock Histories in a Communally Breeding Bird, the Mexica Jay (Aphelocoma ultramarina). *The American Naturalist*, *118*(I), 72–82.

Turjeman, S., Chen, R., & Nathan, R. (2021). Absence of strict monogamy in the Eurasian jackdaw, Coloeus monedula. *Israel Journal of Ecology and Evolution*, *49*(5659), 1–5. https://doi.org/10.1163/22244662-bja10012

Utami, S. S., Goossens, B., Bruford, M. W., De Ruiter, J. R., & Van Hooff, J. A. R. A. M. (2002). Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology*, *13*(5), 643–652. https://doi.org/10.1093/beheco/13.5.643

Van De Pol, M., Heg, D., Bruinzeel, L. W., Kuijper, B., & Verhulst, S. (2006). Experimental evidence for a causal effect of pair-bond duration on reproductive performance in oystercatchers (Haematopus ostralegus). *Behavioral Ecology*, *17*(6), 982–991. https://doi.org/10.1093/beheco/arl036 van Horik, J. O., Beardsworth, C. E., Laker, P. R., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2019). Unpredictable environments enhance inhibitory control in pheasants. *Animal Cognition*, *22*(6), 1105–1114. https://doi.org/10.1007/s10071-019-01302-0

van Horik, J. O., Clayton, N. S., & Emery, N. J. (2012). Convergent Evolution of Cognition in Corvids, Apes and Other Animals. *The Oxford Handbook of Comparative Evolutionary Psychology*, *January*. https://doi.org/10.1093/oxfordhb/9780199738182.013.0005

van Woerden, J. T., van Schaik, C. P., & Isler, K. (2010). Effects of Seasonality on Brain Size Evolution: Evidence from Strepsirrhine Primates. *The American Naturalist*, *176*(6), 758–767. https://doi.org/10.1086/657045

Vasconcelos, M., Hollis, K., Nowbahari, E., & Kacelnik, A. (2012). Pro-sociality without empathy. *Biology Letters*, *8*(6), 910–912. https://doi.org/10.1098/rsbl.2012.0554

Venier, L. A., Dunn, P. O., Lifjeld, J. T., & Robertson, R. J. (1993). Behavioural patterns of extra-pair copulation in tree swallows. *Animal Behaviour*, *45*(2), 412–415. https://doi.org/10.1006/anbe.1993.1050

Verbeek, N. A. M. (1972). The exploitation system of the yellow-billed magpie. *Univ. Calif. Publ. Zool.*, *76*, 1–58. https://doi.org/10.2307/4512011

Verbeek, N. A. M., & Butler, R. W. (1999). Northwestern Crow (Corvus caurinus), no. 407. In A. Poole & F. Gill (Eds.), *The birds of North America*.

Verhulst, S., Geerdink, M., Salomons, H. M., Boonekamp, J. J., & B, P. R. S. (2014). *Social life histories : jackdaw dominance increases with age , terminally declines and shortens lifespan Social life histories : jackdaw dominance increases with age , terminally declines and shortens lifespan*. *August*.

Vincze, O. (2016). Light enough to travel or wise enough to stay? Brain size evolution and migratory behavior in birds. *Evolution; International Journal of Organic Evolution*, *70*(9), 2123–2133. https://doi.org/10.1111/evo.13012

Vogel, A. R., Patisaul, H. B., Arambula, S. E., Tiezzi, F., & McGraw, L. A. (2018). Individual Variation in Social Behaviours of Male Lab-reared Prairie voles (Microtus ochrogaster) is Non-heritable and Weakly Associated with V1aR Density. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-19737-9

Von Bayern, A. M. P., De Kort, S. R., Clayton, N. S., & Emery, N. J. (2007). The role of food- and object-sharing in the development of social bonds in juvenile jackdaws (Corvus monedula). *Behaviour*, *144*(6), 711–733. https://doi.org/10.1163/156853907781347826

Wagner, G. F., Mourocq, E., & Griesser, M. (2019). Distribution of Experimentally Increased Costs of Parental Care Among Family Members Depends on Duration of Offspring Care in Biparental Birds. *Frontiers in Ecology and Evolution*, *7*(November), 1–10. https://doi.org/10.3389/fevo.2019.00366

Wagner, R. H., Helfenstein, F., & Danchin, E. (2004). Female choice of young sperm in a genetically monogamous bird. *Proceedings of the Royal Society B: Biological Sciences*, *271*(SUPPL. 4), 134–137. https://doi.org/10.1098/rsbl.2003.0142

Wang, J. M., & Beissinger, S. R. (2011). Partial incubation in birds: Its occurrence, function, and quantification. *Auk*, *128*(3), 454–466. https://doi.org/10.1525/auk.2011.10208

Warrier, V., Grasby, K. L., Uzefovsky, F., Toro, R., Smith, P., Chakrabarti, B., Khadake, J., Mawbey-Adamson, E., Litterman, N., Hottenga, J. J., Lubke, G., Boomsma, D. I., Martin, N. G., Hatemi, P. K., Medland, S. E., Hinds, D. A., Bourgeron, T., & Baron-Cohen, S. (2018). Genome-wide meta-analysis of cognitive empathy: Heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Molecular Psychiatry*, *23*(6), 1402–1409. https://doi.org/10.1038/mp.2017.122

Wartel, A., Lindenfors, P., & Lind, J. (2019). Whatever you want: Inconsistent results are the rule, not the exception, in the study of primate brain evolution. *PLoS ONE*, *14*(7), 1–15. https://doi.org/10.1371/journal.pone.0218655

Wascher, C. A. F., Fraser, O. N., & Kotrschal, K. (2010). Heart rate during conflicts predicts post-conflict stress-related behavior in greylag geese.

PLoS ONE, *5*(12). https://doi.org/10.1371/journal.pone.0015751

Wascher, C. A. F., Kulahci, I. G., Langley, E. J. G., & Shaw, R. C. (2018). How does cognition shape social relationships? *Philosophical Transactions of the Royal Society B*, *This issue*, 15–18. https://doi.org/10.1098/rstb.2017.0293

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*, 337–340.

Wechsler, B. (1989). Measuring Pair Relationships in Jackdaws. *Ethology*, *80*(1–4), 307–317. https://doi.org/10.1111/j.1439-0310.1989.tb00749.x

Weiss, M. N., Franks, D. W., Giles, D. A., Youngstrom, S., Wasser, S. K., Balcomb, K. C., Ellifrit, D. K., Domenici, P., Cant, M. A., Ellis, S., Nielsen, M. L. K., Grimes, C., & Croft, D. P. (2021). Age and sex influence social interactions, but not associations, within a killer whale pod. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1953), 20210617. https://doi.org/10.1098/rspb.2021.0617

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K., Knief, U., Heim, W., Peona, V., Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M., & Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, 1–11. https://doi.org/10.1038/s41467-020-17195-4

West, R. (2014). The evolution of large brain size in birds is related to social , not genetic , monogamy social , not genetic , monogamy. *Biological Journal of the Linnean Society*, *111*, 668–678. https://doi.org/10.1111/bij.12193

Westneat, D. F. (1992). Do Female Red-winged Blackbirds Engage in a Mixed Mating Strategy? *Ethology*, *92*(1), 7–28. https://doi.org/10.1111/j.1439-0310.1992.tb00946.x

Westneat, D. F., & Stewart, I. R. . (2003). Extra-pair paternity in birds: Causes, correlates and conflict. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 365–396. https://doi.org/10.1146/132439

White, J. G., Southgate, E., Thomson, J. N., & Brenner, F. R. S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans.

Philosophical Transactions of the Royal Society of London. B, Biological Sciences, *314*(1165), 1–340. https://doi.org/10.1098/rstb.1986.0056

Wice, E. W., & Saltz, J. B. (2021). Selection on heritable social network positions is context-dependent in Drosophila melanogaster. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-23672-1

Wiebet, K. L., Korpimaki, E., & Wiehn, J. (1998). *Hatching Asynchrony in Eurasian Kestrels in Relation to the Abundance and Predictability of Cyclic Prey Author (s): Karen L . Wiebe , Erkki Korpimaki and Jurgen Wiehn Source : Journal of Animal Ecology , Nov ., 1998 , Vol . 67 , No . 6 (Nov ., 1998),* . *67*(6), 908–917.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Wilson, A. J., Gelin, U., Perron, M. C., & Réale, D. (2009). Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1656), 533–541. https://doi.org/10.1098/rspb.2008.1193

Wittenberg, J. (1968). Freilanduntersuchungen zur Brutb Rabenkrahe (Corvus c. corone). *Zool. Jahrb. Abt. Syst.*, 97–232.

Wittig, R. M., & Boesch, C. (2010). Receiving post-conflict affiliation from the enemy's friend reconciles former opponents. *PLoS ONE*, *5*(11), e13995. https://doi.org/10.1371/journal.pone.0013995

Wu, T. D., & Nacu, S. (2010). *Fast and SNP-tolerant detection of complex variants and splicing in short reads*. *26*(7), 873–881. https://doi.org/10.1093/bioinformatics/btq057

Young, L. J., & Wang, Z. (2004). The neurobiology of pair bonding. *Nature Neuroscience*, *7*(10), 1048–1054. https://doi.org/10.1038/nn1327

Zduniak, P., Kosicki, J. Z., & Yosef, R. (2016). Sexual aggression by intruders in hooded crow Corvus cornix. *Acta Ethologica*, *19*(1), 91–94.

https://doi.org/10.1007/s10211-015-0222-z

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the Specific Cognitive Ability of Face Perception. *Current Biology*, *20*(2), 137–142. https://doi.org/10.1016/j.cub.2009.11.067