# Cognition gone wild: A test of the social intelligence hypothesis in wild birds

Kelsey B. McCune

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2018

Reading Committee:
Renee L. Ha, Chair
Michael D. Beecher
Jessica A. Sommerville

Program authorized to offer degree:
Psychology

University of Washington

## Abstract

# Cognition gone wild: A test of the social intelligence hypothesis in wild birds

Kelsey B. McCune

Chair of Supervisory Committee:

Renee L. Ha

Department of Psychology

Complex sociality evolved in many species across the animal kingdom, including humans. Species in complex social systems are likely adapted to not just the physical environmental niche (like climate and foraging resources), but also to social environmental pressures. Therefore one consequence of complex sociality might be the evolution and development of complex cognitive abilities to enable successful navigation of social interactions for increase survival and fitness within the social group. The social intelligence hypothesis (SIH) posits that species exhibiting dynamic and larger group sizes will also evolve increased cognitive abilities. However, previous research testing this hypothesis relied on broad proxy measures for both sociality and cognition to compare species across taxa. Studies that do experimentally assess cognition are conducted in tightly controlled, artificial

lab environments. As a result, there is still much conflicting evidence for the effect of sociality on the evolution of cognition. In the research presented here we used social network analysis to directly quantify sociality of two congeneric jay species, the social Mexican Jay (MEJA) and the asocial California Scrub-Jay (CASJ). We conducted cognitive assessments on individuals from both species in the wild, and experimentally tested whether wild subjects and subjects tested in captivity performed similarly. We also quantified whether sociality had a divergent effect on the behavioral response to a threatening stimuli in two contexts (boldness). We found that naïve individuals from the social species did not imitate novel foraging behavior of knowledgeable conspecifics and did not significantly outperform CASJ. Instead, social facilitation significantly predicted the behavioral responses of MEJA social groups to the novel foraging task and the two threat contexts. In contrast, CASJ have higher quality social relationships with their mate, but showed no indication of a social effect on behavioral responses to the foraging or threat tasks. Therefore we found limited evidence for the SIH in this system. However we believe this research is an important step in the comparative cognition field because we quantified cognitive and behavioral performance on an ecologically relevant task in the social and physical environment in which selection is occurring. Future research should expand these methods to other congeneric species pairs, and test additional socio-cognitive mechanisms like transitive inference. In this way we can better understand the effect of sociality on brain evolution and development.

## Table of Contents

## Introduction

Species across the animal kingdom exhibit incredibly diverse and plastic social systems. There are various proximate (i.e. hormonal or genetic) and ultimate (i.e. predation pressure or patchy resources) explanations for the occurrence of sociality (Kelley et al. 2011; Napper and Hatchwell 2016; Tanner and Jackson 2012). However, we still know very little about the individual-level evolutionary consequences of social life.

The high quantity and quality of social relationships exhibited by human groups throughout our evolutionary history is thought to have led to the evolution of brain structure and function. The social intelligence hypothesis (hereafter SIH), also called the social brain hypothesis, posits that the unique difficulty of predicting and responding to behavior of conspecific group-mates necessitated the evolution of complex cognitive abilities (Humphrey 1976; Seyfarth and Cheney 2015; Taborsky and Oliveira 2012). Socio-cognitive mechanisms increase the ability of individuals to navigate potentially dangerous social interactions. For example, transitive inference would allow an individual to know the dominance rank of a stranger relative to self by observing group-mate interactions (Bond et al. 2003). Alternatively, individuals can increase survival and fitness by attending to and replicating the behavior of group-mates towards novel stimuli (social learning; Hoppitt and Laland 2013).

Many studies have used broad cross-taxon comparisons to search for evidence to support the SIH. Frequently, this involves proxy measures for social complexity and cognitive abilities that may be impossible to directly compare across species (Bergman and Beehner 2015). As a result, many contradictory lines of evidence complicate the field. To understand the environmental factors relating to cognitive evolution it is necessary to design more detailed, ecologically valid experiments.

The research presented here improves on previous SIH experiments in several ways. First, our study system is comprised of two congeneric species that have very different social environments, but live in similar physical environments. Consequently, any differences in cognitive abilities can be attributed to divergent sociality rather than genetic, morphological, or ecological differences (Bergman and Beehner 2015; Shettleworth 2010). In our first experiment, we used social network analysis to directly assess quality and quantity of individual social associations (Croft et al. 2008). We compared dyadic social relationship quality of our two species to test the assumption that increases social system complexity defined at the group-level result in a different degree of evolutionary pressures on cognition.

While most researchers testing the SIH use brain size as a proxy measure of cognitive ability (i.e. Dunbar 1998; Dunbar and Bever 1998), there are a minority of studies that quantify cognitive performance with experimental tasks (Balda and Kamil 2002; Templeton et al. 1999). Nearly all of these assessments occur in the laboratory with captive animal subjects, but assume that results are generalizable to wild conspecifics. In our second experiment we tested this assumption by comparing performance of captive and wild individuals on the same novel foraging task.

Third, we conducted a test of the SIH by quantifying the socio-cognitive ability social learning using a complex foraging task demonstrated by trained conspecifics. We compared performance at the task of naïve subjects in the wild from the social and the asocial species. Our task structure also allowed us to discern what types of social, or asocial information naïve individuals use when deciding where and how to interact with the task (Laland 2004; Zentall 2012).

Lastly, one benefit of sociality is increased predator vigilance and deterrence (Kelley et al. 2011), and appropriate behavioral responses to threat can be learned (Cornell et al. 2011; Griffin 2008). Therefore, a potential consequence of evolution and development in a complex social group is behavioral coordination in time and space. To test for this, we quantified the occurrence and assortment patterns of repeatable boldness behaviors towards two different threatening stimuli within and across territories.

We believe this research program will push the boundaries of comparative cognition research by demonstrating that data to test evolutionary hypotheses can be collected on wild animals. *In situ* experiments occur in the location where natural selection has taken place, and so results will be more ecologically valid and generalizable (Coussi-Korbel and Fragaszy 1995; Shettleworth 2010; Thornton and Lukas 2012). Despite a decreased ability to tightly control experiments on wild animal groups, it is vital to observe natural behaviors of animals in their social environment to better understand consequences of sociality.

## Literature cited

Balda RP, Kamil AC. 2002. Spatial and social cognition in corvids: an evolutionary approach. In: *The Cognitive Animal*. 129–133.

Bergman TJ, Beehner JC. 2015. Measuring social complexity. Anim. Behav. 103:203–209; doi:10.1016/j.anbehav.2015.02.018.

Bond AB, Kamil AC, Balda RP. 2003. Social complexity and transitive inference in corvids. Anim. Behav. 65:479–487; doi:10.1006/anbe.2003.2101.

Cornell HN, Marzluff JM, Pecoraro S. 2011. Social learning spreads knowledge about dangerous humans among American crows. Proc. Biol. Sci. / R. Soc. 1–11; doi:10.1098/rspb.2011.0957.

Coussi-Korbel S, Fragaszy DM. 1995. On the relation between social dynamics and social learning. Anim. Behav. 50:1441–1453; doi:10.1016/0003-3472(95)80001-8.

Croft DP, James R, Krause J. 2008. Exploring Animal Social Networks. Princeton, NJ: Princeton University Press.

Dunbar RIM. 1998. The social brain hypothesis. Evol. Anthropol. Issues, News, Rev. 6:178–190; doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

Dunbar RIM, Bever J. 1998. Neocortex size predicts group size in carnivores and some insectivores. Ethology 104:695–708; doi:10.1111/j.1439-0310.1998.tb00103.x.

Griffin AS. 2008. Socially acquired predator avoidance: is it just classical conditioning? Brain Res. Bull. 76:264–71; doi:10.1016/j.brainresbull.2008.02.005.

Hoppitt W, Laland KN. 2013. Social learning: An introduction to Mechanism, Methods, and Models. Princeton University Press, Princeton NJ.

Humphrey NK. 1976. The social function of intellect. In: *Growing Points in Ethology*. 303–317.

Kelley JL, Morrell LJ, Inskip C, Krause J, Croft DP. 2011. Predation Risk Shapes Social Networks in Fission-Fusion Populations. PLoS One 6:e24280; doi:10.1371/journal.pone.0024280.

Laland KN. 2004. Social learning strategies. Anim. Learn. Behav. 32:4–14; doi:10.3758/BF03196002.

Napper CJ, Hatchwell BJ. 2016. Social dynamics in nonbreeding flocks of a cooperatively breeding bird: causes and consequences of kin associations. Anim. Behav. 122:23–35; doi:10.1016/j.anbehav.2016.09.008.

Seyfarth RM, Cheney DL. 2015. Social cognition. Anim. Behav. 103:191–202; doi:10.1016/j.anbehav.2015.01.030.

Shettleworth S. 2010. Cognition, evolution and behavior. Oxford University Press, Oxford UK.

Taborsky B, Oliveira RF. 2012. Social competence: An evolutionary approach. Trends Ecol. Evol. 27:679–688; doi:10.1016/j.tree.2012.09.003.

Tanner CJ, Jackson AL. 2012. Social structure emerges via the interaction between local ecology and individual behaviour. J. Anim. Ecol. 81:260–267; doi:10.1111/j.1365-2656.2011.01879.x.

Templeton JJ, Kamil AC, Balda RP. 1999. Sociality and social learning in two species of corvids: The pinyon jay (Gymnorhinus cyanocephalus) and the Clark's nutcracker (Nucifraga columbiana). J. Comp. Psychol. 113:450–455; doi:10.1037/0735-7036.113.4.450.

Thornton A, Lukas D. 2012. Individual variation in cognitive performance: Developmental and evolutionary perspectives. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:2773–83; doi:10.1098/rstb.2012.0214.

Zentall TR. 2012. Perspectives on observational learning in animals. J. Comp. Psychol. 126:114–28; doi:10.1037/a0025381.

# Defining social complexity at the individual level

Kelsey B. McCune[1*], Renee R. Ha[1], Piotr Jablonski[2,3], Sang-im Lee[2,4], and T. Brandt Ryder[5]

1. The University of Washington Psychology Department, Seattle WA, USA
2. Laboratory of Behavioral Ecology and Evolution, School of Biological Sciences, Seoul National University, Seoul, South Korea
3. Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
4. Daegu-Gyeongbuk Institute of Science and Technology School of Undergraduate Studies, Daegu, South Korea
5. Migratory Bird Center, Smithsonian Conservation Biology Institute, Washington, DC USA
* Corresponding author: kmccune@uw.edu

## Abstract

Current interest in the evolutionary origins of cognition has driven researchers to seek a predictive relationship between social behavior and the evolution of complex cognitive abilities. In social groups with characteristics such as dominance hierarchies, cooperative behaviors and long-term interactions, social competence should select for increased socio-cognitive abilities like social learning, transitive inference, and theory of mind. Previous work has assumed that broad group-level measures of social complexity, like group size accurately capture the degree of social pressure underlying the evolution of cognitive and behavioral adaptations to social life. Group-level traits result from emergent properties of individual-level dyadic interactions, yet variation in dyadic relationship quality is rarely considered. Here we used social network analysis to compare two species differing in social system to test whether complex social group structure predicts increased frequency and duration of individual-level social interactions. In contrast to assumptions of previous research using group size as a proxy for social complexity, we found social Mexican Jays had lower quality dyadic interactions than asocial California Scrub-Jay mate pairs. As a result, we show that it is dangerous to assume that broad group-level measures accurately capture the features of individual social interactions that might select for cognitive and behavioral adaptations.

## Introduction

Sociality is nearly ubiquitous across vertebrate and invertebrate taxa, but varies widely both within and among species. Group size, group stability, and the type of breeding system are used to broadly define complexity of social groups. Larger groups, with traits such as reproductive skew or fission-fusion movement patterns are assumed to reflect increased social environmental pressure on the evolution of cognition and behavior over "simpler" social systems like long-term monogamy (Dunbar 1998; Humphrey 1976; Jolly 1966). However, the features of complex groups arise from the emergent properties of individual-level dyadic interactions.

Considerable research has found that individuals within social groups interact non-randomly (Krause et al. 2015; Seyfarth 1977), resulting in groups with dominance hierarchies, long-term interactions between over-lapping generations, and occasionally behavioral coordination in time and space for cooperation (Ryder et al. 2011). Individuals in a group can gain survival and fitness benefits such as social foraging, cooperative breeding and predator vigilance, but there are also costs like increased competition for resources and mating opportunities (Dunbar 2018; Lima et al. 1999; Sewall 2015; Silk 2007; Wittig et al. 2008). Therefore, successfully navigating dyad-level social interactions can positively affect survival and/or fitness for socially competent individuals (Gersick et al. 2012; Taborsky and Oliveira 2012; White et al. 2010).

In order to successfully navigate dynamic social interactions, socially competent individuals likely pay increased attention to behavior of conspecifics, have larger memory capacity for previous social interactions, or abilities such as social learning, transitive inference, and aspects of perspective taking (Dally et al. 2005; Emery et al. 2007; Range et al. 2009; Scheid et al. 2007; Seyfarth and Cheney 2015). Therefore, in addition to the proximate fitness consequences of group living, variation in group composition and function could shape the evolution of cognitive and behavioral mechanisms (Humphrey 1976; Jolly 1966). The theoretical link between social complexity and evolution of cognition has led many researchers to search for synergies between social complexity, social competence, and cognitive abilities in animal models (Dunbar 1998). Current popular proxy measures of social complexity focus on group-level traits and fail to capture variation in individual-level social interactions (Bergman and Beehner 2015), even though selection for increased capacity for some cognitive abilities likely occurs on the level of individual dyadic interactions (Duckworth 2009; Taborsky and Oliveira 2012). To date, however, few studies have measured variation at the group and individual levels to test the assumption that group-level measures accurately characterize social complexity and the underlying selection dynamics (Whitehead 2008).

Group size as a measure of complex sociality is problematic because it is overly general, and spatially and temporally subjective (Bergman and Beehner

2015). For some of the largest animal groups (i.e. wildebeest, colonial seabirds) no consistent individual interactions occur. Instead, large aggregations more likely result from ecological factors such as predator avoidance and patchy resources (Beauchamp and Fernández-Juricic 2004; Shultz and Dunbar 2006). Furthermore, group size for many species depends on season or age (Alcock 2009). Species like the American Crow have long-term pair bonds and associate in small groups or pairs during the breeding season, but aggregate in huge numbers during the non-breeding season (Verbeek and Caffrey 2002). Similarly, juvenile Common Ravens aggregate in groups for several years before they sexually mature, then pair bond for life and become aggressively territorial and asocial (Boarman and Heinrich 1999). Therefore, it is unclear whether cognitive and behavioral adaptations for social competence in a large, dynamic social group would evolve in these temporally variable systems. Lastly, group size as a definition of social complexity overlooks the intricacy of dyad-level interactions within small groups or pairs.

Previous research has found evidence that social competence in these "simpler" social systems relates to increased cognitive abilities like directed social attention (Scheid et al. 2007), memory of previous interactions (Dally et al. 2006), social learning ability (Templeton et al. 1999), and components of perspective taking (Ostojić et al. 2013). For example, pair-bonded species like Graylag Geese exhibit long-term social bonds and non-random coordination in time and space for territory defense or predator vigilance (Emery et al. 2007). In many cases, cognitive performance of species living in small groups or pairs is equivalent to that from species classically considered socially complex (Emery 2004). As such, to clarify the importance of social environmental pressures on the evolution of cognition and behavior, it is necessary to redefine social “complexity”.

Social network analysis (SNA) explicitly quantifies the dyad-level social properties leading to emergent group-level characteristics. Unlike the broad measure of group size, group stability and breeding system to characterize complex groups, SNA uses observations of repeated social interactions for bottom-up quantification of social behavior at the individual, dyadic and group-level. Therefore, this approach can highlight variation in social competence at multiple scales (i.e. between individuals, groups, and species) to provide more detailed information about the social pressures that could underlie evolved adaptations like social cognition. SNA uses weighted association indices to quantify the intensity of dyadic relationships, like preferred social partners or behavioral coordination, that result in long-term and frequent social interactions. These spatial and temporal associations between individuals are quantified with weighted edges (or links, ties) between individuals that are larger than edge weights between dyads that rarely associate. Therefore, edge weights describe the relationships between individuals, whereas variation in social competence of an individual within social networks can be defined from network graphs with metrics such as degree (number of associates) and strength (sum of connected edge weight values). Consequently, SNA allows us to make specific predictions about the degree of synchrony between group-level and individual-level social complexity.

Frequent, long-term social interactions (i.e. pair-bonded species like ravens) will result in larger edge weight values, but not necessarily large degree values, whereas individuals in large aggregations (i.e. wildebeest) will show the opposite trend and have large degree values, but small edge weights. We expect systems that are classically considered as socially complex (e.g. long-term interactions, dominance hierarchies, reproductive skew, or behavioral coordination in large groups) would exhibit both large dyadic edge weights and large degree values (which equates to individuals with large strength values).

Although SNA is frequently used to describe social structure of populations in isolation, only very recently have researchers begun to use this approach to test evolutionary hypotheses (Belton et al. 2018; Dunbar 2018; Ilany and Akçay 2016). Yet species comparisons are necessary to understand the biological relevance of relatively larger or smaller network metrics values. Here we used a comparative approach to test the assumption that complexity at the group-level necessitates a high degree of social competence and drives the number and frequency of social interactions at the individual-level. We used SNA to compare social associations of a species that exhibits larger group sizes, long-term stability of group composition and a dynamic breeding system, to a pair-bonded solitary species. In accord with the differences in species group size, we expect higher degree values for individuals in the “complex” social group species. If broad group-level measures of size, stability and breeding system accurately predict individual-level social competence, then we also expect higher edge weight values between individuals in the complex social group species. An alternative hypothesis is that long-term monogamy necessitates equally high levels of social attention, behavioral coordination, and perspective taking (i.e. Eurasian Jays; Ostojić et al. 2013). If this occurs, then we predict a difference in degree values, but no difference in dyadic edge weight values between the two types of social groups. This study will test the common assumption that complex social groups defined using broad definitions are also most likely to contain individuals with high social competence and complex dyadic relationships that could lead to the evolution of cognitive and behavioral adaptations.

## Methods

*Study system:* We studied two species in the *Aphelocoma* genus at extreme ends of the sociality spectrum: the highly social Mexican Jay (*A. wollweberi*; hereafter MEJA), and the asocial California-Scrub Jay (*A. californica*; hereafter CASJ). CASJ live in long-term, solitary monogamous pairs that defend a territory year-round against all other jays, including fledglings, which are chased from the territory before the next breeding attempt begins (Curry et al. 2002). While CASJ only have one constant social partner, most mate for life which could require frequent affiliative interactions and social coordination to maintain the pair bond (Emery et al. 2007). In contrast, MEJA live in flocks of 5-25 individuals on a year-round territory that they defend against other groups. Within the flock, several pairs form during the breeding season to build nests and lay eggs. Pairs are not constant across breeding seasons, or genetically monogamous within a breeding season (Brown 1994). Once nestlings hatch, all jays in the flock cooperate to feed and protect the young, regardless of relatedness (Brown and Brown 1981). Young delay dispersal for several years or remain in their natal flock for life. Membership in flocks is very stable, but low levels of emigration and immigration prevent inbreeding depression (Brown and Brown 1981). As such, MEJA flocks show group-level social complexity in that there are stable, large groups, with a breeding system that involves cooperation and reproductive skew (Brown et al. 1997).

*Data collection*: From April 2014 – September 2016 we tracked 49 CASJ from 20 territories in Willamette Mission State Park in Keizer, OR (supplementary Fig. 1). We also tracked 73 MEJA from 7 flocks around the Southwestern Research Station in Portal, AZ from May – November 2015 (supplementary Fig. 2). The majority of jays were color-banded for individual identification, but 7 unbanded jays (3 MEJA, 4 CASJ) were included in analyses because their space use was sufficiently distinct to allow for identification. We conducted up to 6 behavioral observations on all banded jays of both species. The average (± SE) number of observations for MEJA was 4.9 ± 0.07 observations, and for CASJ it was 5.1 ± 0.09 observations. Average duration of MEJA behavioral observations was 11.3 ± 0.3 minutes, and average observation time for CASJ was 12.3 ± 0.3 minutes. In total, we spent 85.3 hours observing jays.

During focal observations, we defined associations using spatial proximity (i.e., any individual within 10m of the focal jay). We also recorded approximate distance of all associates and duration of associations. We excluded all agonistic interactions from these data (displacements, pecks, chases). To prevent temporal autocorrelation of social behavior, we did not conduct an observation on any jays that were seen in the same day associating with a previous focal individual. Additionally, each jay's observations were separated by at least two weeks (Whitehead 2008). We attempted to follow every banded bird in both breeding and non-breeding seasons, but variation in behavior precluded complete replication of observations for all birds. To ensure equivalent sampling effort across all individuals, we systematically searched for the individuals who had not been seen in more than two weeks.

In our system, where there are very stable groups/pairs, we assumed that dyads that are never seen within 10m have less secure relationships than those that associate more closely (Fraser and Bugnyar 2010). Therefore an association index that includes the degree of spatial proximity is appropriate for more fine-scaled discrimination of preferred, stable relationships. To weight the association between each dyad we used an association index that included the distance two jays were seen associating, and the duration at that distance (Blaszczyk 2018). For each pair, we calculated our association index for a given dyad ($\mathrm{AI}_{\mathrm{AB}}$) as:

$$\mathrm{AI}_{\mathrm{AB}} = \frac{t_{AB} \; x \; \frac{1}{d_{AB}}}{(T_A + T_B)}$$

Where $t_{\mathrm{AB}}$ is the sum duration in seconds across observations periods that jay A and jay B spent within 10m, $d_{\mathrm{AB}}$ is the closest average distance in meters that jay A and jay B were observed across observation periods (average of closest distance in observation 1 plus closest distance in observation 2, etc.). $\mathrm{T}_{\mathrm{A}}$ represents the summed total time during all of its focal observations that jay A was in the observer's view, and $\mathrm{T}_{\mathrm{B}}$ represents jay B's total time in view during observations (Blaszczyk 2018). Jays that spent a larger proportion of time in close proximity have higher association index values and are more likely to be involved in an interaction that requires social competence such as directed social attention or behavioral coordination (Fraser and Bugnyar 2010). This differs from common association indices (Cairns and Schwager 1987) that only quantify association frequency, but yields a more detailed picture of individual-level dyadic social relationships.

*Data analysis*: We used the R package *igraph,* version 1.1.2 (Csardi and Nepusz 2006), to create undirected, weighted social network graphs from our association data, and to calculate network metrics. To quantify social complexity we used *edge weight* (the value of the association index) to describe the dyadic relationship. We used social network metrics *degree* (the number of jays observed within 10m of the focal jay), and *strength* (sum of edge weights between focal jay and associating jays) to describe individual social propensity. It is likely impossible in the CASJ social system to have degree values as high as that seen in MEJA. Additionally, larger degree values will artificially increase strength values, even if dyadic relationship quality is low with all associating jays. Therefore to ensure a more straightforward comparison of species with very different network sizes, we scaled degree and strength values by dividing the observed value by the maximum value for its species (Croft et al. 2008), but we included the non-scaled results in the supplementary material. We used linear mixed models (LMM) to test for the effect of species on three dependent variables: edge weight, strength and degree. We log transformed all variables to meet normality assumptions, and verified with visual inspections of residuals, as well as Shapiro-Wilk (Shapiro and Wilk 1965) and Kolmogorov-Smirnov (Lilliefors 1967) tests of normality. For scaled strength and degree we included a random effect of territory of the focal jay to account for differences in group mean network metric due to assortativity by gregariousness and variability in group size. Edge weights of associations between jays from different territories were significantly smaller than edge weights between jays from the same territory for both species (*t*-value = -2.69, *p* = 0.01). We calculated the intra-class correlation coefficient of within/between flock edge status, and found that it accounts for 15% of the variance in edge weight. Therefore, we also included a random effect of within/between territory in our analysis of edge weights to account for these distinct mean values.

Network data violate the assumption of independence required for most statistical tests (Whitehead 2008). Additionally, many animals reside in groups or appear in spatial proximity because of ecological constraints rather than social preference. To determine significance of species differences in our dependent variables, we generated 10,000 datasets in which the species label of nodes (individuals) were permuted. In these random networks, individual-level sociality is decoupled from group-level complexity (species). We re-ran the same LMM on each of the 10,000 datasets to create a null beta coefficient distribution. We compared our observed coefficient values describing the species effect in our empirical data to the distribution of coefficients resulting from randomized data (Croft et al. 2008; Whitehead 2008). We considered empirical coefficient values significant if they were more extreme than 97.5% of the distribution of coefficients resulting from randomized data.

## Results

Over the whole study period we observed 389 unique dyadic association events in the MEJA population and 109 dyadic associations between CASJ individuals. MEJA were less likely to associate with members of other territories. Only 2% of MEJA ties (4 out of 214 ties) occurred between members of different flocks (Fig. 1a), whereas 43% of CASJ ties (16 out of 37) occurred between non-mate pairs (Fig. 1b) in contexts that were not obviously agonistic.

As expected, the number of associates (degree) of MEJA was significantly greater than that of CASJ (Fig. 2a). MEJA degree values were normally distributed in the population with a mean ± SE of 5.9 ± 0.3, whereas CASJ degree values were left-skewed and the average was 1.5 ± 0.1. The effect of species was significantly greater than the coefficients from the null distribution that used node-label permutations to decouple scaled degree values from species identity (Fig. 3a; MEJA $\beta_1 = 0.45$, $p = 0$).

Contrary to our prediction that group-level complexity relates to greater quality in individual relationships, MEJA log-transformed edge weight values were smaller than those between CASJ (Fig. 2b; MEJA mean ± SE: 0.67 ± 0.3; CASJ: 0.72 ± 0.2). The observed coefficient for the species effect on edge weight fell outside of the null distribution of coefficients resulting from the permuted data (Fig. 3b; MEJA $\beta_1 = -1.03$, $p = 0.0001$), indicating a significant difference.

Strength (i.e., the sum of an individual's connected edge weights) was significantly different between species (Fig. 2c); the observed coefficient for the species effect also fell outside of the null distribution of coefficients from permuted data (Fig. 3c). However, in contrast to our predictions, MEJA had significantly smaller scaled strength values than CASJ (MEJA mean ± SE: 0.06 ± 0.02; CASJ: 0.36 ± 0.4; MEJA $\beta_1 = -0.76$, $p = 0.0004$).

## Discussion

Large group sizes, variation in group stability, and multifaceted breeding systems, are often used to identify social complexity. Much of the interest in defining whether species have complex social systems comes from hypotheses predicting that social complexity drives the evolution of complex cognitive abilities (i.e. Dunbar 1998; Humphrey 1976; Jolly 1966). MEJA are a communal, cooperatively breeding species. Therefore, these hypotheses would predict that MEJA individuals require a high amount of social competence and associated cognitive abilities to navigate the dominance hierarchy, and coordinate behaviors in time and space for cooperative breeding. Conversely, we show that these broad group-level traits to define social complexity are not predictive of increased individual-level social relationship quality measured from observations of repeated social associations.

Instead of variation in group-size as the mechanism for evolution of all complex social cognitive abilities, it's likely that cognitive traits evolve in a modular fashion depending on the pressures from the local social environment. As a result, we would expect various behaviors to be differentially expressed depending on the type of social competence needed. We found support for the hypothesis that the "simpler" social system of the life-long pair bond in CASJ requires high levels of social attention and behavioral coordination because dyadic edge weights (i.e. more frequent and close associations) and individual strength values were greater than those seen between MEJA. Since most CASJ have just one social connection (their mate), they can dedicate more energy to attending to the behavior of that partner. Given this social environment, it's likely that CASJ might socially learn a task (or acquire a pathogen, for a non-cognitive example) much more quickly from their social mate than a MEJA would from a given group-mate (Hoppitt and Laland 2013). Furthermore, these data add to previous research that found captive CASJ and another asocial monogamous species, the Eurasian Jay, exhibit cognitively advanced perspective-taking behaviors rarely seen in other taxa (Dally et al. 2005; Ostojić et al. 2013). Jays of both species were found to observe and remember past behavioral interactions and visual perspectives of a conspecific, demonstrating that repeated dyadic interactions might be a key mechanism facilitating social competence in these taxa.

Not surprisingly, we found higher scaled degree values in group-living MEJA than in solitary, monogamous CASJ. MEJA may not spend as much time with any one associate because of the heterogeneous social environment. There may be a general, evolved trade-off between a few high quality relationships and many lower quality relationships (Taborsky and Oliveira 2012). Previous research using SNA within populations of great tits (Aplin et al. 2013) and three-spined stickleback (Pike et al. 2008) found a similar trend where bold, exploratory individuals have many low quality relationships, but shy slow-exploring individuals have a few high quality relationships. While CASJ social competence requires repeated interactions with one conspecific, MEJA experience larger group sizes, dominance hierarchies and overlapping generations. Social competence in this system may lead to fewer interactions with more conspecifics and could require increased memory capacity or cognitive load for individual identification and abilities like transitive inference to track the dominance hierarchy (Seyfarth and Cheney 2015). Since MEJA do not form long-term monogamous mate-pairs, there may be no need for directed social attention towards the mate. In contrast, there might be selection against especially strong dyadic relationships in MEJA because individual fitness and group functioning requires labile responses to social information from any flock member (Brosnan et al. 2010). All MEJA flock members are vigilant for predators, communicate a threat or new food source, and help to feed and protect the young after hatching (Brown 1994). Similarly, diffuse social attention permits more opportunities for scrounging, an important individual benefit of social life (Giraldeau and Caraco 2000; McCormack et al. 2007). If there were preferred strong relationships, then flock-level coordination for cooperative breeding, foraging, and predator defense could decrease (Brosnan et al. 2010; Couzin 2009; Webster and Ward 2011). Future comparative research should investigate more closely the link between social competence and cooperation at both the dyadic and group levels.

The current field of SNA has rarely been used in a comparative framework. This is largely due to a few quantitative challenges. Association data are inherently non-independent, which can artificially result in significant p-values from parametric statistical models if not addressed (Croft et al. 2011). Consequently, we assessed whether our species differences were significant by comparing observed coefficients describing the species effect to coefficients obtained from null models of systematically randomized data. Secondly, many network metrics change with sample size or number of edges (Croft et al. 2008). Although the number of jays observed in each species was different, the edge density (proportion of actual associations out of total possible associations among all individuals in the population) was very similar (MEJA density = 0.08; CASJ density = 0.03), and so likely had little effect on our findings. The observed difference in density between our species is much smaller than what has previously been compared in other taxa (Faust 2006). As a result, we believe the findings presented here are ecologically valid.

We have shown that it is dangerous to assume that individuals in groups exhibiting broadly defined characteristics of social complexity are also associating more frequently and closely. These results have important implications for research on the evolution of sociality, complex cognitive abilities and social competence. Researchers comparing species, especially to make inferences about the evolution of cognitive abilities, should measure sociality at the individual *and* group-level. Our system compared two species with very different group sizes, but it would be interesting to quantify dyadic relationships in species with a range of intermediate group sizes to test if there is an upper limit on number of associates with which one can maintain a high quality relationship (Sutcliffe et al. 2012). Furthermore, future research could compare species that live in groups of similar sizes, but that vary in characteristics such as dominance hierarchies, reproductive skew, or cooperative behaviors to better understand when social complexity necessitates high levels of social competence.

## Figures

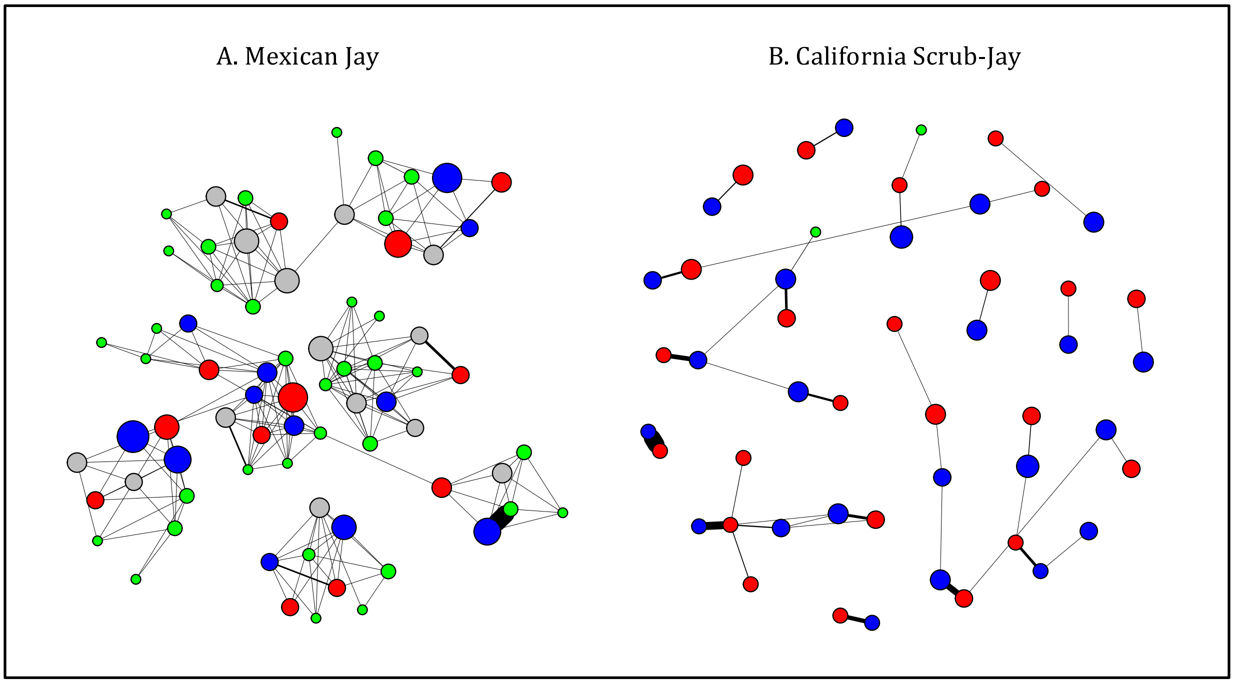

Figure 1: Non-agonistic social network graphs for A) Mexican Jays (MEJA), and B) California Scrub-Jays (CASJ). Size of node corresponds to jay age, where the smallest nodes are age 0 (a hatch year bird). Color of the node indicates the breeding status of the jay where red is a breeding female, blue is a breeding male, green is a non-sexually mature juvenile, and grey is a jay of unknown status (only MEJA). The edges between nodes are weighted by our association index, and thicker lines represent dyads that associated more closely and frequently.

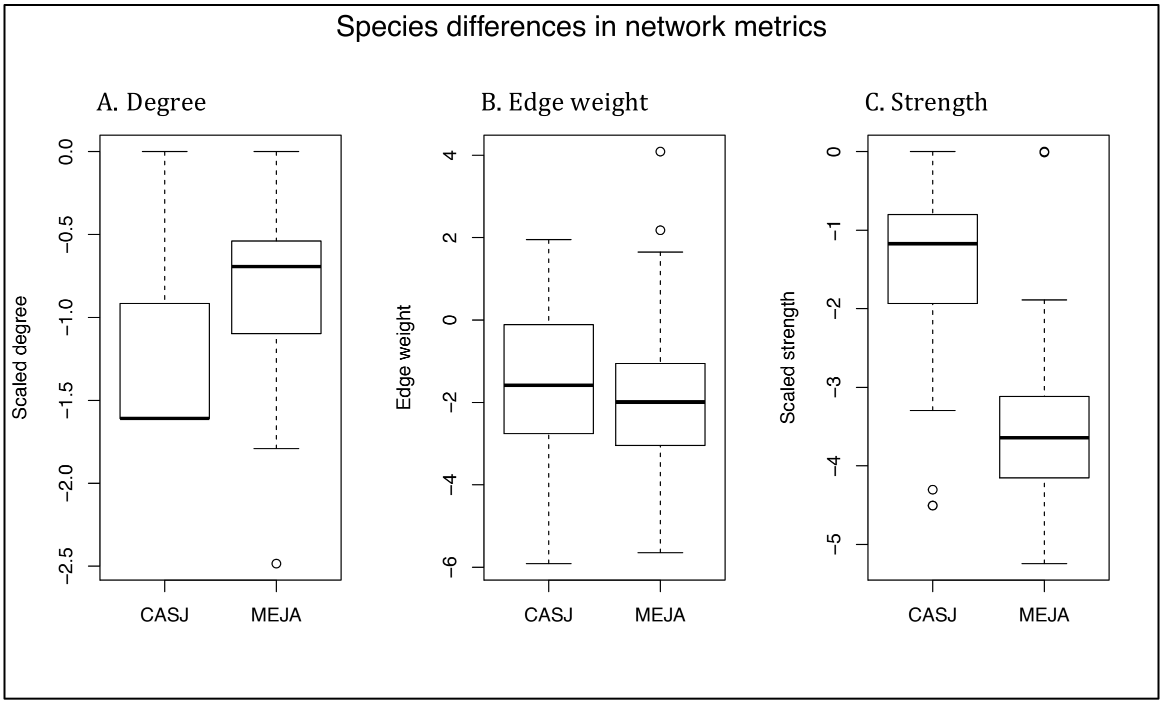

Figure 2: Boxplots showing the difference in network metrics between California Scrub-Jays (CASJ) and Mexican Jays (MEJA). All y-values were log-transformed for readability and normality. Note that there appears to be no difference in edge weights between species. However, when we compare our observed effect to that expected from random networks we find a statistically significant species difference.

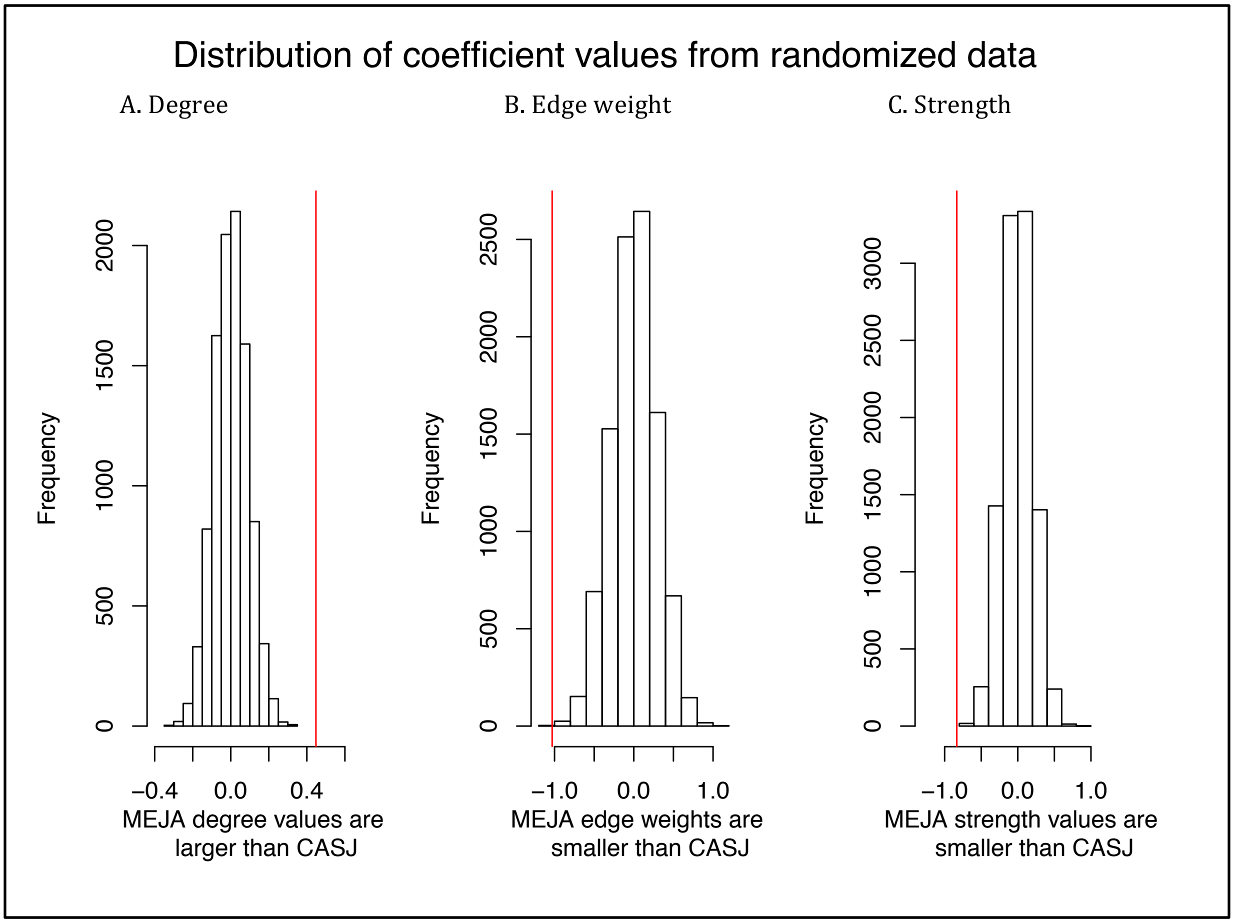

Figure 3: To assess significance of the difference between Mexican Jay (MEJA) and California Scrub-Jay (CASJ) network metrics, we created 10,000 datasets where the species label for each node was randomized. These histograms show the coefficient describing the effect of MEJA network metrics in relation to CASJ for the randomized data. The vertical red line indicates our observed coefficient value.

## Supplementary material

We classified individual MEJA by several attributes that have previously been found to relate to social network position (Krause et al. 2015): Dominance rank, breeding status, and age.

*A. Dominance rank:* We conducted scramble competition assays where jays competed with group-mates for access to peanuts from a container that could be monopolized. We calculated a dominance rank as the number of displacements an individual initiated compared to its flock-mates. Dominance was not correlated with age.

*B. Breeding status:* MEJA are not sexually dimorphic, only a few adults in the flock breed each season, and all group members feed nestlings once they hatch. Therefore breeding status can only be assigned based on observation of nest building, copulation, female incubation, or a male mate-feeding an incubating female. Jays two years and younger (juveniles) did not attempt to breed but are identifiable by white coloration on the beak. We labeled all other individuals as unknown breeding status.

*C. Age:* We looked at the effect of age as a continuous numeric variable as well as a dichotomized factor variable. We created the factor age variable by dividing MEJA into juveniles (2 years old and younger), and adults (jays 3 years and older).

### 1. MEJA degree and strength

We ran linear regression models for log-transformed strength values, and Poisson regression for degree values as the dependent variables. We incorporated the above attributes as fixed effects. MEJA network metric values were unrelated to individual attributes (Table 1). Given the correlations in other taxa between network metrics and attributes, it's possible that missing attribute values in our data are obscuring significant relationships (Table 2).

Results are similarly non-significant when we compared the observed coefficients from the models in Table 1 to the coefficients resulting from identical models run on 10,000 random datasets. We created random datasets by permuting observed social network ties within flocks and recalculating strength and degree.

## 2. Assortativity of MEJA network ties

Two hypotheses explain the evolution of cooperation in groups, with different predictions about the type of social interactions that would occur (reviewed in Brown 1987). The "skill development" hypothesis posits that non-breeding adults may forgo dispersal and personal fitness by remaining in their natal group to gain skills that will help them live longer and become more successful breeders in the future. Alternatively, non-breeding adults could "pay-to-stay" on the territory so that they may eventually rise in rank to the breeding position.

We tested these hypotheses by comparing the edge weight (relationship quality) between jays in relation to the dyadic similarity or dissimilarity of the attributes we measured. If the skill-development hypothesis is supported in MEJA, we expect non-breeding jays will have no preferential associations with jays of certain breeding status or dominance rank because they can learn skills from group members with high or low dominance, breeder or non-breeder status. Moreover, we expect to see a positive relationship between edge weight and difference in age, or higher weights between adult-juvenile dyads than juvenile-juvenile dyads. Young jays likely cannot learn skills from the youngest jays, and so will associate more with older jays. In contrast, if the pay-to-stay hypothesis is supported, we will see the edge weights between dyads of disassortative breeding status, dominance, and age will be significantly higher. To prevent eviction from the group, non-breeding jays will show frequent associations with dominant individuals and breeders because there would be coordination of cooperative activities such as feeding nestlings and fledglings, and defending the territory.

To test if non-breeders have preferential ties to breeders we excluded dyads where both jays were of unknown breeding status, and dyads comprised of both breeding individuals. We used linear regression of log-transformed edge weight values to determine the coefficient comparing non-breeder associations with breeders to non-breeder associations with other non-breeders. Similarly, we modeled whether juvenile class jays (ages 2 and below) have greater edge weight values when associating with adults (ages 3 and above) than when associating with other juveniles, and adult-adult dyads were excluded. Additionally, we looked for a general trend of increased edge weight value with increased difference in numeric age of the dyad members. Finally, we tested if jays associate most with those of higher dominance status by using a fixed effect of a dominance difference score for each dyad. We compared our observed coefficients to the distribution of coefficients generated from 10,000 permuted datasets.

We did not find evidence that non-breeding jays are associating specifically with breeding jays rather than other non-breeders, or that juvenile class jays associate more with adult jays than with other juveniles (Table 3). Additionally, we found no relationship between edge weight values and the dominance difference score, or age difference score for the dyad. Although it is possible that our null results indicate some support for the skill-development hypothesis, further evidence is needed to definitely claim social association trends support either of the hypotheses.

### 3. Consistency in MEJA network position over time and after a disturbance

It's possible that stable social groups will fragment after a disturbance or loss of a particularly central individual (Flack et al. 2006). Additionally, in groups where there is reproductive skew and only a few dominant individuals breed, there may be changes to the social network across the breeding season as sexual selection, courtship, and competition occur. To see if individuals differed in their place within the network over time, I divided the data for each individual before and after the captive period for their flock. I only took jays into captivity after the breeding season, so this corresponds to differences in network position during breeding and non-breeding. I had 5 flocks where I held two jays per flock in captivity for 3 weeks (manipulated/Experimental treatment), and 2 flocks that were undisturbed (unmanipulated/Control treatment) for the entire season. I split the data of each individual from the two unmanipulated groups, GA and HI, at the same time point as the periods of captivity for the manipulated groups (see Table 4).

I re-calculated the association index (edge weights) and social network metrics Degree (number of associating jays), Strength (sum of the edge weight values of all associated jays), and Betweenness (the number of ties/paths/edges connecting other jays that pass through the focal jay). However, this resulted in a sample size decreased by half, *and* some jays disappeared during the post-captivity period or where not banded in the pre-captivity period, so cannot be included in the comparison. Our interpretation of whether there are any changes in network structure pre- and post-captivity could be affected by number of observations, and due to a time limited field season the observations are unequally distributed with more conducted pre- than post-captivity. Therefore it is possible I did not detect enough associations post-captivity to calculate reliable and robust network metrics.

To see if network metrics are susceptible to sample size I calculated the network metrics for each individual after including additional focal observations. I did find that network metrics change as the number of observations increase (Fig. 4). Degree values from 2 and 3 observations are significantly different than values from 4, 5, and 6 observations. For strength, values from 5 and 6 observations are significantly less than values from only 1 or 2 observations. However there was no difference in betweenness across observation number. This indicates that measures of degree and strength are likely not reliable from the divided pre- and post-captivity data.

To compare network metrics from experimental and control groups, I subtracted the pre-captivity metric from the post-captivity metric to get a difference score where positive means the network metric got larger and negative means it got smaller in the non-breeding season/post-captivity period. Experimental and control groups do not differ in the amount of metric change pre- and post-captivity, except strength increased in the experimental groups post-captivity compared to control group strength which didn't change (Fig. 5). However, this is attributable to Flock differences and the significance disappears with inclusion of a Flock random effect.

In only the manipulated groups, I looked at the correlation between network metric rank order within the flock before and after the captive period of the removed jays, and the jays that remained in the flocks to see if network position *within the flock* was affected by the removal. Removing the jays into captivity resulted in a change in their network position when they were released back into their groups - there was not a correlation in pre- and post-captivity metric rank order (Table 5). Jays that remained in the group mostly retained their network position, as indicated by the significant correlations for Degree and Betweenness rank order scores. Strength does differ, however the sample size of captive jays is very small (n = 7), and the correlation coefficient for strength is high (Table 5). So, it is unclear if there are enough data to detect all associations in the post-captivity period to yield reliable network metrics. Overall, most removed jays became less central after captivity compared to pre-captivity, and the two jays that became more central were among the least dominant in their flocks pre-captivity (Fig. 6).

## Supplemental tables and figures

Table 1: Results from regression models of network metrics on attribute variables. SE stands for standard error. We log-transformed Strength to use linear regression, and used Poisson regression for the dependent variable Degree.

| **Strength Models** | **Estimate** | **SE** | ***p*-value** |
|---|---|---|---|
| Strength ~ Dominance rank | 0.44 | 0.51 | 0.40 |
| Strength ~ Breeding status (F) | 0.90 | 0.27 | 0.47 |
| Breeding status (M) | -0.60 | 0.34 | 0.13 |
| Breeding status (J) | -0.22 | 0.32 | 0.49 |
| Breeding status (UNK) | -0.34 | 0.37 | 0.34 |
| Strength ~ Age (factor) | 0.12 | 0.25 | 0.64 |
| Strength ~ Age (numeric) | -0.01 | 0.05 | 0.83 |
| **Degree Models** | | | |
| Degree ~ Dominance rank | -0.07 | 0.21 | 0.72 |
| Degree ~ Breeding status (F) | 1.79 | 0.12 | 0.26 |
| Breeding status (M) | -0.22 | 0.18 | 0.22 |
| Breeding status (J) | -0.03 | 0.14 | 0.83 |
| Breeding status (UNK) | 0.11 | 0.16 | 0.47 |
| Degree ~ Age (factor) | 0.008 | 0.11 | 0.94 |
| Degree ~ Age (numeric) | 0.003 | 0.02 | 0.90 |

Table 2: The raw number, and the percent of the total attribute values that are unknown or missing in the Mexican Jay population (n = 73).

| **Attribute** | **Number unknown** | **Percent** |
|---|---|---|
| Age | 0 | 0 |
| Dominance rank | 22 | 30 |
| Breeding status | 14 | 19 |

Table 3: Output of models testing for assortativity of ties by attributes. The dependent variable was the log-transformed edge weight value of a dyad. The p-value of permutation tests is calculated as the number of coefficient values from the permuted data that are more extreme than the observed coefficient, divided by the number of datasets.

| **Model parameter** | **Coefficient** | ***p*-value** |
|---|---|---|
| Non-breeder assortativity | -0.16 | 0.36 |
| Age class assortativity | -0.70 | 0.19 |
| Numeric age assortativity | -0.01 | 0.41 |
| Dominance assortativity | -0.84 | 0.08 |

Table 4: Dates and duration of captivity for experimental flocks. We split Control flocks HI and GA at identical dates to create pre- and post-captivity categories for these unmanipulated flocks.

| **Date** | **Flocks in Captivity** |
|---|---|
| 19 July - 10 Aug | TA |
| 11 Aug - 1 Sept | UC, CO, (HI split here) |
| 1 Sept - 22 Sept | KI, XM, (GA split here) |

Table 5: Correlation coefficients and significance values of individual network metrics pre- and post- the captive period. We analyzed jays taken into captivity (Captive), and jays left in the flock (Non-Captive) separately.

| **Captive** | *r* | *p* |
|---|---|---|
| Degree | 0.25 | 0.59 |
| Strength | 0.63 | 0.13 |
| Betweenness | -0.05 | 0.90 |
| **Non-Captive** | *r* | *p* |
| Degree | 0.45 | 0.001 |
| Strength | 0.22 | 0.13 |
| Betweenness | 0.40 | 0.01 |

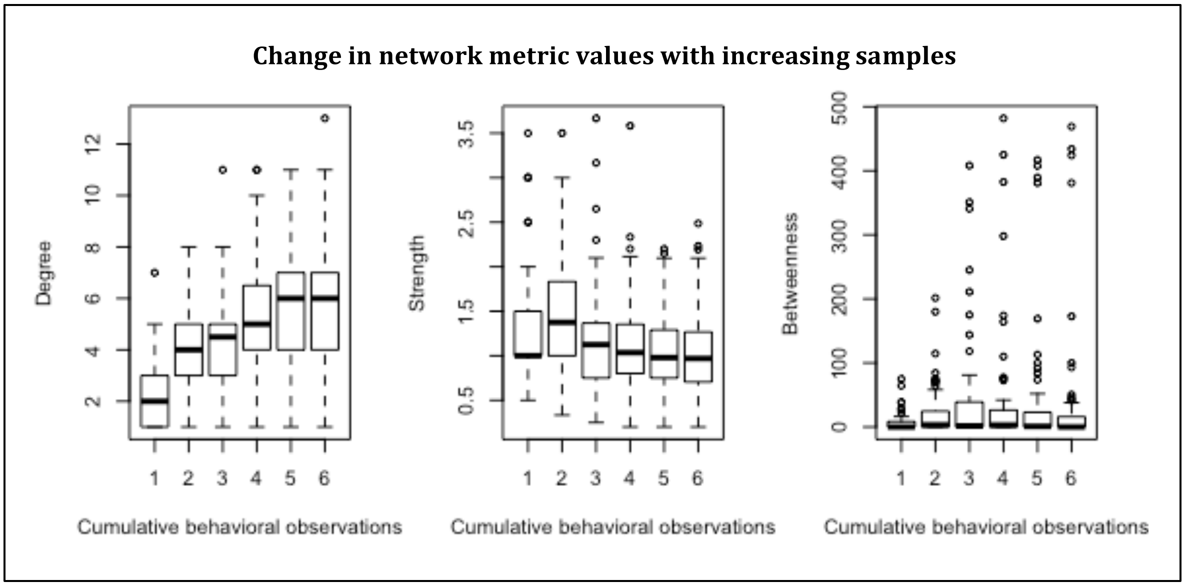
Change in network metric values with increasing samples

Figure 4: Boxplots show the change in network metrics Degree, Strength and Betweenness of individual jays as more focal observation data are added.

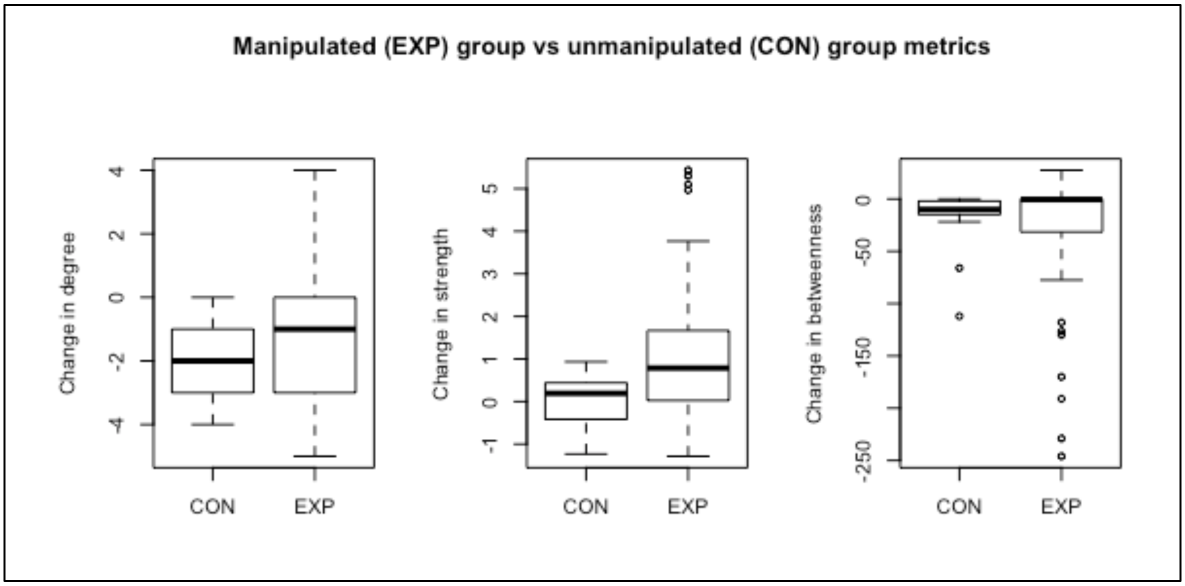
Manipulated (EXP) group vs unmanipulated (CON) group metrics

Figure 5: Boxplots show the differences in network metrics Degree, Strength, and Betweenness of groups in each of the treatment conditions.

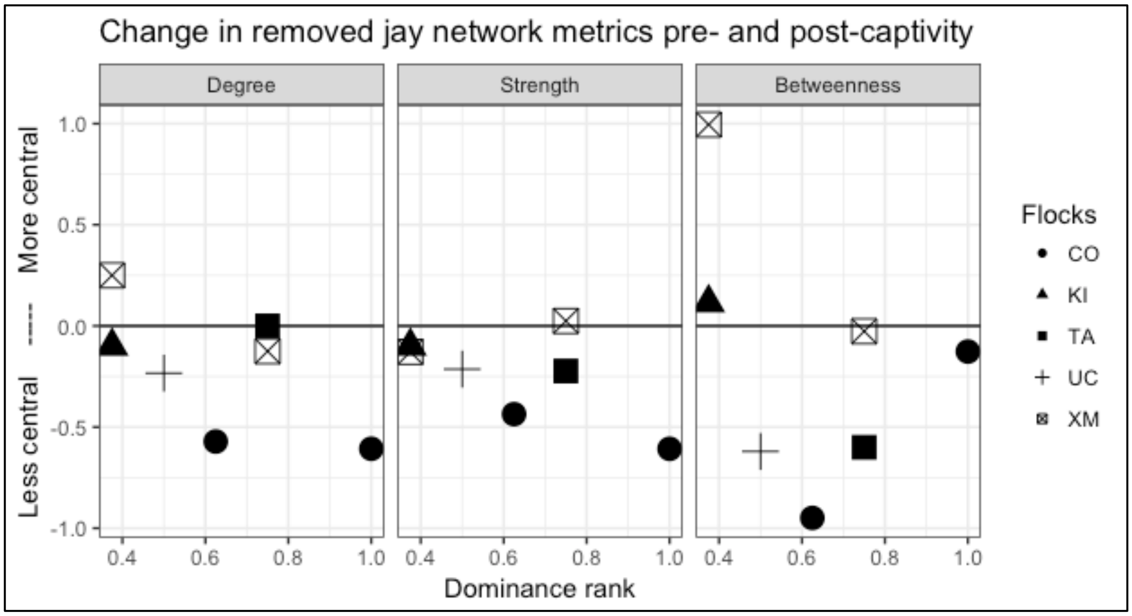

Figure 6: Change in network metric rank-order position of jays that were taken into captivity. Most jays became less central (below the horizontal line), but the two that became more central are also the least dominant in their group.

## Literature cited

Alcock J. 2009. Animal Behavior. Sunderland, MA: Sinauer Associates, Inc.

Aplin LM, Farine DR, Morand-Ferron J, Cole EF, Cockburn A, Sheldon BC. 2013. Individual personalities predict social behaviour in wild networks of great tits (Parus major). Ecol. Lett. 16:1365–72; doi:10.1111/ele.12181.

Beauchamp G, Fernández-Juricic E. 2004. Is there a relationship between forebrain size and group size in birds? Evol. Ecol. Res. 6: 833–842.

Belton LE, Cameron EZ, Dalerum F. 2018. Social networks of spotted hyaenas in areas of contrasting human activity and infrastructure. Anim. Behav. 135:13–23; doi:10.1016/j.anbehav.2017.10.027.

Bergman TJ, Beehner JC. 2015. Measuring social complexity. Anim. Behav. 103:203–209; doi:10.1016/j.anbehav.2015.02.018.

Blaszczyk MB. 2018. Consistency in social network position over changing environments in a seasonally breeding primate. Behav. Ecol. Sociobiol. 72:1–13; doi:10.1007/s00265-017-2425-y.

Boarman WI, Heinrich B. 1999. Common Raven (Corvus corax), version 2.0. In: *The Birds of North America* (P.G. Rodewald, ed.) The Birds of North America, Inc.: Philadelphia, PA.

Brosnan SF, Salwiczek L, Bshary R. 2010. The interplay of cognition and cooperation. Philos. Trans. R. Soc. B Biol. Sci. 365:2699–2710; doi:10.1098/rstb.2010.0154.

Brown JL. 1994. Mexican Jay (Aphelocoma ultramarina). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Brown JL, Brown ER. 1981. Extended family system in a communal bird. Science (80-. ). 211: 959–960.

Brown JL, Brown ER, Sedransk J, Ritter S. 1997. Dominance, age, and reproductive success in a complex society: A long-term study of the Mexican Jay. Auk 114: 279–286.

Couzin ID. 2009. Collective cognition in animal groups. Trends Cogn. Sci. 13:36–43; doi:10.1016/j.tics.2008.10.002.

Croft DP, James R, Krause J. 2008. Exploring Animal Social Networks. Princeton, NJ: Princeton University Press.

Croft DP, Madden JR, Franks DW, James R. 2011. Hypothesis testing in animal social networks. Trends Ecol. Evol. 26:502–7; doi:10.1016/j.tree.2011.05.012.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695.

Curry RL, Peterson AT, Langen TA. 2002. Western Scrub Jay (Aphelocoma californica). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Dally JM, Emery NJ, Clayton NS. 2005. Cache protection strategies by western scrub-jays, Aphelocoma californica: Implications for social cognition. Anim. Behav. 70:1251–1263; doi:10.1016/j.anbehav.2005.02.009.

Dally JM, Emery NJ, Clayton NS. 2006. Food-caching Western scrub-jays keep track of who was watching when. Science 312: 1662–1665.

Duckworth RA. 2009. The role of behavior in evolution: A search for mechanism.

Evol. Ecol. 23:513–531; doi:10.1007/s10682-008-9252-6.

Dunbar RIM. 2018. Social structure as a strategy to mitigate the costs of group living: a comparison of gelada and guereza monkeys. Anim. Behav. 136:53–64; doi:10.1016/j.anbehav.2017.12.005.

Dunbar RIM. 1998. The social brain hypothesis. Evol. Anthropol. Issues, News, Rev. 6:178–190; doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

Emery NJ. 2004. Are Corvids "Feathered Apes"? In: *Comparative Analysis of Mind*.

Emery NJ, Seed AM, von Bayern AMP, Clayton NS. 2007. Cognitive adaptations of social bonding in birds. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 362:489–505; doi:10.1098/rstb.2006.1991.

Fraser ON, Bugnyar T. 2010. The quality of social relationships in ravens. Anim. Behav. 79:927–933; doi:10.1016/j.anbehav.2010.01.008.

Gersick AS, Snyder-Mackler N, White DJ. 2012. Ontogeny of social skills: Social complexity improves mating and competitive strategies in male brown-headed cowbirds. Anim. Behav. 83:1171–1177; doi:10.1016/j.anbehav.2012.02.005.

Giraldeau LA, Caraco T. 2000. Social Foraging Theory. Princeton, NJ: Princeton University Press.

Hoppitt W, Laland KN. 2013. Social Learning: An Introduction to Mechanisms, Methods, and Models. Princeton, NJ: Princeton University Press.

Humphrey NK. 1976. The social function of intellect. In: *Growing Points in Ethology*. 303–317.

Ilany A, Akçay E. 2016. Social inheritance can explain the structure of animal social networks. Nat. Commun. 7; doi:10.1038/ncomms12084.

Jolly A. 1966. Lemur social behavior and primate intelligence. Science (80-. ). 153: 501–6.

Lilliefors HW. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 62:399–402; doi:10.1111/j.1467-9574.1971.tb00143.x.

Lima SL, Zollner PA, Bednekoff PA. 1999. Predation, scramble competition, and the vigilance group size effect in dark-eyed juncos (Junco hyemalis). Behav. Ecol. Sociobiol. 46:110–116; doi:10.1007/s002650050599.

McCormack JE, Jablonski PG, Brown JL. 2007. Producer-Scrounger Roles and Joining Based on Dominance in a Free-Living Group of Mexican Jays ( Aphelocoma ultramarina ). Behaviour 144: 967–982.

Ostojić L, Shaw RC, Cheke LG, Clayton NS. 2013. Evidence suggesting that desire-state attribution may govern food sharing in Eurasian jays. Proc. Natl. Acad. Sci. U. S. A. 110:4123–8; doi:10.1073/pnas.1209926110.

Pike TW, Samanta M, Lindstrom J, Royle NJ. 2008. Behavioural phenotype affects social interactions in an animal network. Proc. R. Soc. B Biol. Sci. 275:2515–2520; doi:10.1098/rspb.2008.0744.

Range F, Horn L, Bugnyar T, Gajdon GK, Huber L. 2009. Social attention in keas, dogs, and human children. Anim. Cogn. 12:181–92; doi:10.1007/s10071-008-0181-0.

Ryder TB, Blake JG, Parker PG, Loiselle BA. 2011. The composition, stability, and kinship of reproductive coalitions in a lekking bird. Behav. Ecol. 22:282–290; doi:10.1093/beheco/arq213.

Scheid C, Range F, Bugnyar T. 2007. When, what, and whom to watch? Quantifying attention in ravens (Corvus corax) and jackdaws (Corvus monedula). J. Comp. Psychol. 121:380–6; doi:10.1037/0735-7036.121.4.380.

Sewall KB. 2015. Social Complexity as a Driver of Communication and Cognition. Integr. Comp. Biol. 55:384–395; doi:10.1093/icb/icv064.

Seyfarth R. 1977. A model of social grooming among female monkeys. J. Theor. Biol. 65:671–698; doi:http://dx.doi.org/10.1016/0022-5193(77)90015-7.

Seyfarth RM, Cheney DL. 2015. Social cognition. Anim. Behav. 103:191–202; doi:10.1016/j.anbehav.2015.01.030.

Shapiro SS, Wilk MB. 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52:591; doi:10.2307/2333709.

Shultz S, Dunbar RIM. 2006. Both social and ecological factors predict ungulate brain size. Proc. Biol. Sci. / R. Soc. 273:207–215; doi:10.1098/rspb.2005.3283.

Silk JB. 2007. Social components of fitness in primate groups. Science 317:1347–1351; doi:10.1126/science.1140734.

Sutcliffe A, Dunbar RIM, Binder J, Arrow H. 2012. Relationships and the social brain: Integrating psychological and evolutionary perspectives. Br. J. Psychol. 103:149–168; doi:10.1111/j.2044-8295.2011.02074.x.

Taborsky B, Oliveira RF. 2012. Social competence: An evolutionary approach. Trends Ecol. Evol. 27:679–688; doi:10.1016/j.tree.2012.09.003.

Templeton JJ, Kamil AC, Balda RP. 1999. Sociality and social learning in two species of corvids: The pinyon jay (Gymnorhinus cyanocephalus) and the Clark's nutcracker (Nucifraga columbiana). J. Comp. Psychol. 113:450–455; doi:10.1037/0735-7036.113.4.450.

Verbeek NA, Caffrey C. 2002. American Crow (Corvus brachyrhynchos), version 2.0. In: *The Birds of North America* (P.G. Rodewald, ed). The Birds of North America, Inc.: Philadelphia, PA.

Webster MM, Ward AJW. 2011. Personality and social context. Biol. Rev. 86:759–773; doi:10.1111/j.1469-185X.2010.00169.x.

White DJ, Gersick AS, Freed-Brown G, Snyder-Mackler N. 2010. The ontogeny of social skills: Experimental increases in social complexity enhance reproductive success in adult cowbirds. Anim. Behav. 79:385–390; doi:10.1016/j.anbehav.2009.11.014.

Whitehead H. 2008. Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis. Chicago, IL: The University of Chicago Press.

Wittig RM, Crockford C, Lehmann J, Whitten PL, Seyfarth RM, Cheney DL. 2008. Focused grooming networks and stress alleviation in wild female baboons. Horm. Behav. 54:170–177; doi:10.1016/j.yhbeh.2008.02.009.FOCUSED.

# The effect of temporary captivity on learning performance in jays

Kelsey B. McCune[1*], Piotr Jablonski[2,3], Sang-im Lee[2,4], and Renee R. Ha[1]

1. The University of Washington Psychology Department, Seattle WA, USA
2. Laboratory of Behavioral Ecology and Evolution, School of Biological Sciences, Seoul National University, Seoul, South Korea
3. Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
4. Daegu-Gyeongbuk Institute of Science and Technology School of Undergraduate Studies, Daegu, South Korea
* Corresponding author: kmccune@uw.edu

## Abstract

The recent surge in research seeking to quantify animal cognitive abilities across taxa has revealed surprising similarities and differences among species. Frequently, experimental designs for quantifying animal cognition are adapted from human developmental cognition literature, and conducted on lab-reared subjects. While results from these types of controlled studies have some merit for understanding how human cognitive development directly compares to that of lab animals, the ecological validity and generalizability are severely limited. These studies assume that performance on lab-based assessments accurately represents abilities of wild conspecifics, however this assumption is rarely tested. In this experiment, we used the same apparatus to experimentally quantify performance of wild and captive subjects on a simple individual learning task. We found that performance is not equivalent, and wild subjects were faster learners. These results are contrary to other studies directly comparing wild and captive performance, which have found that captive subjects perform better. However, together these results indicate that we cannot assume captive animal cognitive performance is

equal to wild conspecifics. In order to understand environmental factors contributing to the evolution of cognition, or the importance of cognition to behavioral ecology and conservation management strategies, it is necessary to quantify ecologically relevant abilities on animals in their natural environment.

## Introduction

Research over the past 40 years has sought to quantify animal learning and cognitive abilities in order to understand the evolution of complex cognition. However, recent commentary (i.e. Kendal et al. 2010; Shettleworth 2010) in comparative cognition encourages scientists to improve the ecological validity of their experimental designs for several reasons. First, animal taxa evolve in environments with very distinct perceptual and cognitive challenges to those of humans. Yet many of the questions and research designs to assess animal cognition have been borrowed from the human developmental cognition field rather than created with the perspective of the specific animal species in mind. For example, assessments of Piaget's levels of object permanence (Piaget 1954) have been translated nearly identically to animal subjects in several taxa such as primates (de Blois and Novak 1994), birds (Pepperberg et al. 1997), and dogs (Gagnon and Doré 1994). Applying assessments created for humans to animals can illuminate areas of surprising similarity in cognitive ability (Wood et al. 1980), or even abilities in which animals excel over humans (Barth and Call 2006). However failure of animal subjects at these assessments is difficult to interpret because it could imply either 1) that the experimental paradigm was not ecologically relevant or behaviorally appropriate to the animal subject, or 2) the cognitive ability does not exist in that species.

The second way in which animal cognition studies could still fail to produce ecologically valid results is that the majority of experiments are conducted with captive subjects. Cognitive assessments on animals in captivity have the advantage of being tightly controlled so that there are few alternative explanations to confound the observed results. Assessments in the wild often cannot control variation in performance due to food motivation, predator vigilance, or social factors such as competition (Shettleworth 2010). However, because of the tight controls and the complex or contrived nature of some lab-based experiments, it is difficult to apply those findings to wild conspecifics, or even replicate results in other captive animals (Shettleworth 2010). For example, in search for evidence of theory-of-mind abilities in animals, primate researchers have developed increasingly elaborate experiments for captive primates, some of which involve the human experimenter wearing a bucket over the head (e.g. Call and Tomasello 2008). Results from these studies indicate primates have many of the same theory-of-mind abilities as humans, but could not be replicated in other labs (Penn and Povinelli 2007; Povinelli and Vonk 2004). This is likely due to the individual experience of those specific captive primates with the researchers, experimental equipment, and carry over effects of previous training.

Relatively few studies have used the same assessment to quantify cognitive performance of lab-based and wild, free-living subjects (Benson-Amram et al. 2013; Bouchard et al. 2007; Gajdon et al. 2006; Morand-Ferron et al. 2011; Webster and

Lefebvre 2001). Of these, only one statistically compared performance of subjects in each condition (Benson-Amram et al. 2013), and found that hyenas born in a research facility out-performed wild hyenas on a puzzle box task, likely because of background experience with novel man-made objects. Subjects that are born in captivity, or have lived the majority of their life confined, experience completely different ontogenetic environments and learning opportunities (also see van de Waal and Bshary 2011). These subjects will approach assessments with background experience that does not exist in their wild counterparts, leading to results that cannot be generalized to the species as a whole.

Researchers may still be able to take advantage of the strict control of assessments in the lab environment, while avoiding confounds of differences in previous experience, by taking wild individuals temporarily into captivity (Morand-Ferron et al. 2016). Research on great tits and blue tits first tested solving ability of birds temporarily held in captivity, then measured performance in the wild after release (Morand-Ferron et al. 2011). Results showed that performance of previously captive individuals was similar to that of wild conspecifics that had never seen the task. However, the task in the wild was slightly different from the one given to subjects in captivity, making it difficult to directly compare performance and assess the effect of temporary captivity. To our knowledge, no other previous research has directly compared performance of subjects temporarily held in captivity to performance of free-living conspecifics on the same task.

Here we conducted simple learning assessments on Mexican Jays (*Aphelocoma wollweberi*; hereafter "jays") in the wild, and on subjects held in captivity for less than 3 weeks. Assessments were conducted simultaneously, in the same local environment, so that there would be no seasonal effects on motivation or performance. We evaluated the support for three competing hypotheses: 1) testing subjects temporarily in captivity, with equivalent background experiences to wild conspecifics, will result in no differences in performance (e.g., Morand-Ferron et al. 2011); 2) the temporary captive environment, where we control food motivation and limit distractions from predators or competitors, will result in increased performance of captive individuals relative to wild conspecifics; and 3) the experience of capture and removal into a novel space for temporary captivity itself could negatively affect performance of captive jays relative to free-flying conspecifics. These results will inform the validity of lab-based assessments of cognition and whether it is appropriate to assume performance of captive individuals is equivalent to likely performance of wild conspecifics.

## Methods

*Experimental setup:* We used a simple puzzle box task to quantify jay individual learning ability (Fig. 1). This species lives in stable flocks of 5 – 25 jays on a year-round territory. Within the flock there are overlapping generations, a dominance hierarchy, and communal cooperative breeding (Brown 1994). From May – September 2015 we conducted learning trials in 7 flocks around the Southwestern Research Station, near Portal, AZ. All individuals were color-banded for individual identification, and trained to come to a whistle for food, which ensured timely participation in the task.

Our puzzle box was created out of a log, to mimic natural extractive foraging for grubs or cache-recovery behaviors (Brown 1994). We created four food-holding compartments in the log, covered by transparent doors that open in different ways (Fig. 1). This apparatus allows us to test how quickly individuals can learn through trial-and-error about the affordances of the puzzle box doors. The four diverse doors engaged individuals that varied in their initial motivation to interact with different opening methods at distinct locations on the puzzle box, and allowed us to determine whether learning performance was repeatable across options (Griffin et al. 2015). On the territory of each flock, we initially allowed the group access to the open puzzle box filled with peanuts until they habituated to this new food source. We progressed with the learning trials only after all jays were eating comfortably from all compartments in the puzzle box.

Two jays from each of 5 out of the 7 flocks were chosen for captivity based on a relatively high position in the dominance hierarchy (see supplementary material). To decrease stress from social isolation while in captivity, we held two jays per flock in an aviary. We trapped the pair of jays from a particular flock in the same day, and transported them immediately to large aviaries on the research station campus. Aviaries were constructed of sturdy metal grating, with a natural dirt floor and ground vegetation, and averaged 8 feet wide x 10 feet long x 12 feet high. An enclosed, connected room allowed visual and physical access to the aviaries through a one-way window. We outfitted aviaries with a table near the window, many branches with foliage including acorns (when in season), and a tarp attached to the ceiling and part of the walls to give jays the option for shelter. Jays were fed a maintenance diet of moistened cat food, fruit chunks, mixed seed, live arthropods when possible, and peanuts. They also had constant access to a wide shallow dish of water for drinking and bathing. We did not start trials until jays habituated to captivity. We assessed habituation as eating comfortably (no jumpiness) from food dishes and the open puzzle box on the table near the window, normal caching and preening behaviors within the aviary, and foraging on the foliage with acorns. All jays began exhibiting these behaviors indicative of habituation within 48 hours of their introduction into the aviaries. We used meal worms (*Tenebrio mollitor*) and wax worms (*Galleria mellonella*) as the reward inside the puzzle box, because jays cached peanuts around the aviary and were not motivated to interact with the task for peanuts. We conducted trials on jays in captivity each morning, before 9am, and jays were given their maintenance diet immediately after. Captive jays were released back to their flock after 3 weeks in the aviaries.

*Data collection:* We conducted learning trials identically for 10 captive and 7 wild jays. Immediately prior to each trial attempt, we put out the puzzle box with one food item in each compartment and the doors open. If subjects came within 10 minutes and ate comfortably from the puzzle box, we closed the doors and began the trial. Wild, free-flying jay subjects were chosen pseudo-randomly at the start of the first trial as the very first wild jay from each of the 7 flocks that came to contact the puzzle box after the doors had been closed. All other jays in that flock were discouraged from coming near by blocking access to the box, walking towards them or tossing sticks in their direction until they moved away. Wild jays were tested during the 3 weeks that their group-mates were in captivity.

Once the trial began, we recorded the duration the subject spent within 2m of the puzzle box, if it touched a door with foot or beak (attempt), if it successfully opened a door and retrieved the food item (solve), and latency to each solve. Video clips of jays interacting with the puzzle box are available in the online supplementary material. We manually closed doors and refilled compartments after a jay successfully opened it. A subject was allowed to solve a given door 3 times, after which that door was emptied and remained open for the rest of the trials. In this way, we assumed that the opening event occurred because the jay learned the affordance of that door and replicated the behavior, rather than accidentally opening it. Additionally, now the subject had to attempt on other doors to continue getting food rewards, so we could measure repeatability of learning performance. Each trial was up to 2 hours long. If jays lost interest and did not return within 2m of the puzzle box for 30 minutes, the trial was ended early. We conducted a maximum of 6 trials on each subject, though some jays had fewer trials because they had already opened all 4 doors on the puzzle box 3 times (for a total of 12 solves). All methods were first approved by the University of Washington Institutional Animal Care and Use Committee (protocol #4064-03), the Arizona Game and Fish Department (permit #SP697293), and U.S. Fish and Wildlife (permit #MB51894B-0).

*Statistical analyses:* All analyses were conducted in RStudio (RStudio Team 2015). We used Poisson regression and the glm function in the stats base package of RStudio to analyze the total number of successful openings for jays in each treatment (captive vs. wild). We included an offset for time within 2m to account for variable motivation and distractions that might decrease attendance to the task. To further clarify mechanisms contributing to performance, we conducted an additional Poisson regression to quantify the total number of attempts on puzzle box doors as a function of treatment, including a time offset. Our hypotheses predict not only a difference in number of solves, but also a difference in the rate of solving performance. Therefore we used the *coxme* package (Therneau 2018) and cox proportional hazard survival models with ID as a random effect and treatment as the stratification variable, to compare the rate of learning of the two treatment groups. We ran three models: the latency to first solve, latency to last solve, and given a jay solved a door one time, the time it took after the first solve for the jay to solve a door two more times (time between first and last solves). Finally, we quantified the repeatability of solving success (Griffin et al. 2015) as the ability of subjects to consistently switch to new opening methods for the other doors, using the *rptR* package (Nakagawa and Schielzeth 2010), including treatment as a fixed effect and jay ID as a random effect.

## Results

Jays from both treatments learned the affordances of the doors. With only one exception, jays that solved any door at least once were able to successfully solve that door 2 more times during the 6 learning trials. Over half (4/7) of the wild subjects solved all of the doors; three subjects did so in the span of one trial. In contrast, only 3 out of the 10 captive jays solved all of the doors and none were able to do so in fewer than 3 trials. After accounting for treatment condition, we found that door solving of individual jays was significantly repeatable (R = 0.8 ± 0.15, $p$ = 0.002), indicating that our task assesses an inherent individual trait.

Despite the fact that captive birds spent significantly more time within 2m of the puzzle box (Wild $\beta = 0.93$, $p < 0.01$), they were worse at successfully opening doors. Wild subjects undergoing trials in their natural environment were more persistent because they made more attempts per total time within 2m on puzzle box doors ($\beta = 1.32$, $p < 0.01$; raw mean ± SE: Wild = 162.9 ± 49.6, Captive = 110.1 ± 28.8; Fig. 2a), and had more successful openings per time than captive conspecifics (Wild $\beta = 1.24$, $p < 0.01$; Wild = 7.4 ± 2.0, Captive = 5.4 ± 1.7; Fig. 2b).

These results are even clearer in the cox proportional hazard models (Fig. 3). The hazard ratio for the first solve and last solve were significantly greater in wild than captive jays (Table 1; $p = 0.045$ and $p = 0.047$ respectively). There was no difference between treatment groups in the time it took jays to solve a door for the third time after solving a first time (Table 1; $p = 0.23$).

## Discussion

Despite the popularity of cognitive assessments of captive animal subjects, very little previous research has directly compared performance of captive and wild animals. Yet captive animal studies assume performance represents abilities of wild conspecifics. This is problematic because captive subjects, especially those born or long-lived in captivity, likely have very different previous experiences than wild conspecifics that govern engagement with, and performance on, cognitive assessments. Other than to understand how best to enrich zoo or lab animal life, it is important that we understand actual cognitive abilities exhibited by wild animals for application to conservation management strategies (Greggor et al. 2014, 2016), and research on the factors affecting brain evolution (Bingman 1992).

Of studies that have tested subjects in both captivity and the natural environment on identical tasks, all have found that captive animals out-perform wild conspecifics in some way (Benson-Amram et al. 2013; Bouchard et al. 2007; Gajdon et al. 2006; Webster and Lefebvre 2001; except see Morand-Ferron et al. 2011 which found no difference on slightly different tasks). In contrast, we found the opposite: wild, free-living jays outperformed captive jays. Even though captive subjects spent more time within 2m of the puzzle box, and were therefore motivated to engage with the task, they made fewer attempts and had fewer successes per time than wild conspecifics. These results support our third hypothesis that learning performance while in temporary captivity could be confounded by neophobia, even after a habituation period. It is likely that captive jays performed fewer attempts and solves, and longer latency to solve because of divided attention. Jays may be trying to remain vigilant to possible danger in the new environment, while also trying to learn the affordances of the puzzle box doors to achieve the food reward (Morand-Ferron and Quinn 2011; Overington et al. 2009). Conversely, it is interesting that wild jays performed so well given the possibility of environmental and social distractions, and our lack of ability to control food motivation, which together resulted in some trials ending before the 2hr limit.

We found significant individual differences in solving ability for both captive and wild jays; solving performance across door types was highly repeatable. This generally indicates that some jays were "solvers", and others were not. Inspection of our survival plots indicates that if a jay has not solved a door by the third trial (after 240 min), it is unlikely it will ever be a solver. There are a few possible reasons for this trend. First, previous research in this population has found short-term consistency in producer-scrounger roles (McCormack et al. 2007). Role is tied to dominance rank where scrounging tactic is most likely to be used when the difference in dominance between producer and scrounger is large. However, the probability of being a solver was not statistically related to dominance (see supplementary material). Another explanation for the repeatability of solving performance is that the bolder jays are most interactive with the puzzle box (Carere and Locurto 2011). Previous research in this population did find consistent individual differences in boldness towards a novel object (McCune et al. 2018), but boldness scores of our subjects were not significantly related to success on this task (see supplementary material). Future research should evaluate more explicitly, with larger sample sizes, the impact of these factors on variation in cognitive performance.

To further knowledge on the evolution of cognition and to understand the source of species differences, it is necessary to increase ecological validity of experimental designs. For one, it is important to design tasks that are ecologically relevant and require appropriate behavioral and perceptual abilities. Additionally, there are many developmental or situational confounds that explain performance on lab assessments that may not be relevant in the natural environment, especially for studies of animals born or long-lived in captivity. For example, Amram-Benson et al. (2013) tested possible explanations for differences in learning performance and concluded that captive animals have more experience receiving rewards after engaging in object-oriented inspection behaviors of novel, man-made objects. Furthermore, frequent experience with novel objects during development might increase object-oriented regions in the brain, resulting in captive animals with cognitive abilities that it is not possible for wild conspecifics to possess (Tomasello and Call 2004; see van de Waal and Bshary 2010 for a similar comparison of wild rural verets to wild urban vervets). Our results on subjects temporarily held in captivity, combined with those from studies on animals born or long-lived in captivity, show that in both of these situations it is dangerous to generalize performance in captivity to that of wild conspecifics.

## Tables and figures

Table 1: Output from the three Cox models

| Model | Hazard ratio | *z* | *p* |
|---|---|---|---|
| First solve | 7.17 | 2.01 | **0.045** |
| Last solve | 13.66 | 1.99 | **0.047** |
| Time between | 1.96 | 1.2 | 0.23 |

Hazard ratio refers to the likelihood of solving in the wild condition relative to the captive condition. All three models included the random effect of jay ID.

Figure 1: We created our puzzle box from a log by carving out 4 compartments and covering them with clear doors that open in different ways. The top door opens up like a hatch. The left door opens out like a car door. The front door pulls out like a drawer, and the door on the right side of the log has a hinge at the top and so pushes in (in the right photo it is pulled up for the open, habituation phase).

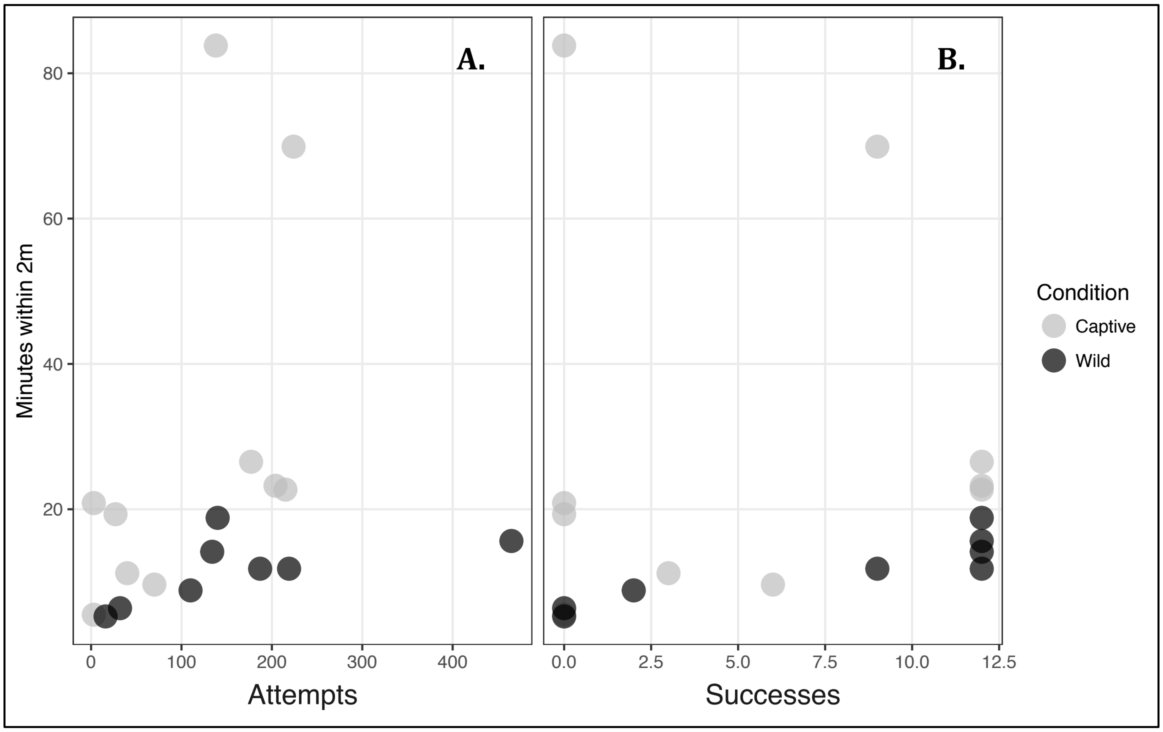

Figure 2: Wild MEJA subjects (dark grey dots) made more attempts (A) and had more successes (B) per time than subjects tested in captivity (light grey dots).

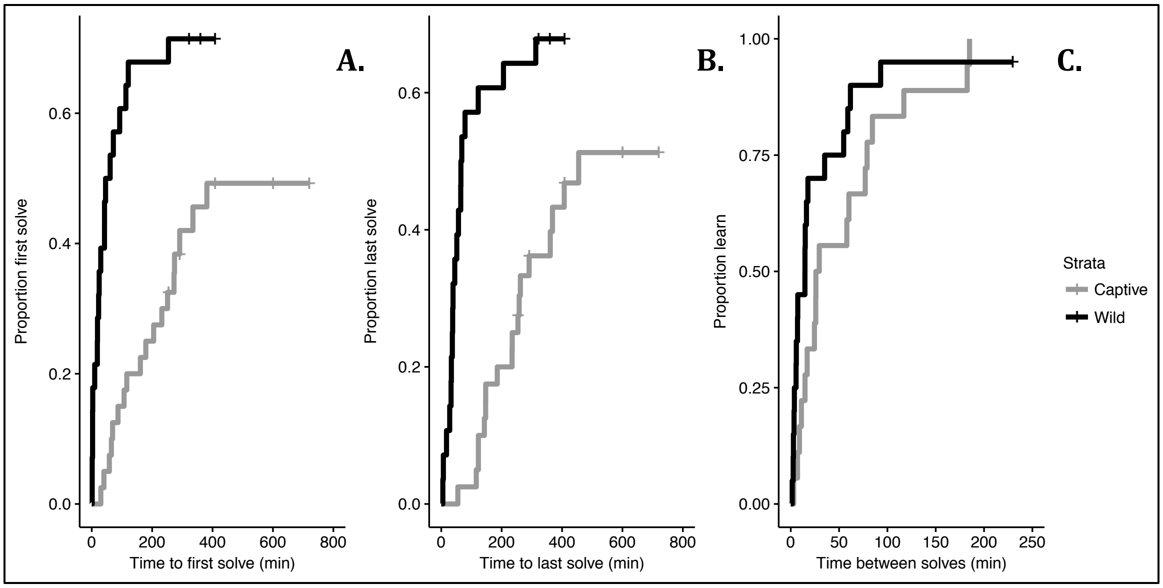

Figure 3: Survival plots showing the rate which jays in each treatment solve a door for the first time (A), or the last time (B), as well as the rate of learning illustrated by the time between first and last solves (C). Captive jay performance is shown in grey, wild jay performance is in black. The time to solve of wild jays was shorter than that for captive jays because wild jays were more likely to solve all doors sooner, or lose motivation to engage with the task.

## Supplementary material

As part of another experiment, we collected data on dominance and boldness of individuals in this population of jays. To assess dominance, we conducted scramble competition assays in the territories of each flock. We put peanuts in a monopolizable container, and observed for displacement behavior (Barkan et al. 1986). We calculated an individual's dominance rank as the proportion of initiated displacements out of total observed displacements for the flock.

Methodological details of our boldness assay are described elsewhere (McCune et al. 2018; Ch. 5 this document), so we will give a brief overview here. We quantified boldness using closest approach to a novel object surrounded by peanuts. Peanuts varied in distance from the novel object, so bolder jays took more peanuts from closer to the object during trials. We found novel object boldness to be repeatable after 4 – 11 weeks, indicating it is a valid measure of a personality trait.

We categorized jays as "solvers" if they were able to open at least one door three times. We analyzed the probability of being a solver as a function of closest approach boldness score or dominance rank using logistic regression. Dominance rank of jays was not statistically related to the probability of being a solver ($\beta$ = -4.91, $p$ = 0.09). Similarly, probability of being a solver was also not significantly related to boldness score ($\beta$ = 1.13, $p$ = 0.46).

## Literature cited

Barkan CPL, Craig JL, Strahl SD, Stewart AM, Brown JL. 1986. Social dominance in communal Mexican jays Aphelocoma ultramarina. Anim. Behav. 34:175–187; doi:10.1016/0003-3472(86)90021-7.

Barth J, Call J. 2006. Tracking the displacement of objects: A series of tasks with great apes (Pan troglodytes, Pan paniscus, Gorilla gorilla, and Pongo pygmaeus) and young children (Homo sapiens). J. Exp. Psychol. Anim. Behav. Process. 32:239–252; doi:10.1037/0097-7403.32.3.239.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 67(1):1-48.

Benson-Amram S, Weldele ML, Holekamp KE. 2013. A comparison of innovative problem-solving abilities between wild and captive spotted hyenas, Crocuta crocuta. Anim. Behav. 85:349–356; doi:10.1016/j.anbehav.2012.11.003.

Bouchard J, Goodyer W, Lefebvre L. 2007. Social learning and innovation are positively correlated in pigeons (Columba livia). Anim. Cogn. 10:259–66; doi:10.1007/s10071-006-0064-1.

Brown JL. 1994. Mexican Jay (Aphelocoma ultramarina). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Call J, Tomasello M. 2008. Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn. Sci. 12:187–192; doi:10.1016/j.tics.2008.02.010.

Carere C, Locurto C. 2011. Interaction between animal personality and animal cognition. Curr. Zool. 57: 491–499.

de Blois ST, Novak MA. 1994. Object permanence in rhesus monkeys (Macaca mulatta). J. Comp. Psychol. 108:318–327; doi:10.1037/0735-7036.108.4.318.

Gagnon S, Doré FY. 1994. Cross-sectional study of object permanence in domestic puppies (*Canis familiaris*). J. Comp. Psychol. 108:220–232; doi:10.1037/0735-7036.108.3.220.

Gajdon GK, Fijn N, Huber L. 2006. Limited spread of innovation in a wild parrot, the kea (Nestor notabilis). Anim. Cogn. 9:173–81; doi:10.1007/s10071-006-0018-7.

Greggor AL, Berger-Tal O, Blumstein DT, Angeloni LM, Bessa-Gomes C, Blackwell BF, et al. 2016. Research Priorities from Animal Behaviour for Maximising Conservation Progress. Trends Ecol. Evol. xx:1–12; doi:10.1016/j.tree.2016.09.001.

Greggor AL, Clayton NS, Phalan B, Thornton A. 2014. Comparative cognition for conservationists. Trends Ecol. Evol. 29:489–495; doi:10.1016/j.tree.2014.06.004.

Griffin AS, Guillette LM, Healy SD. 2015. Cognition and personality: An analysis of an emerging field. Trends Ecol. Evol. 30:207–214; doi:10.1016/j.tree.2015.01.012.

Healy SD, Bacon IE, Haggis O, Harris a. P, Kelley L a. 2009. Explanations for variation in cognitive ability: Behavioural ecology meets comparative cognition. Behav. Processes 80:288–294; doi:10.1016/j.beproc.2008.10.002.

Kendal RL, Galef BG, van Schaik CP. 2010. Social learning research outside the laboratory: How and why? Learn. Behav. 38:187–94; doi:10.3758/LB.38.3.187.

McCormack JE, Jablonski PG, Brown JL. 2007. Producer-Scrounger Roles and Joining Based on Dominance in a Free-Living Group of Mexican Jays ( Aphelocoma ultramarina ). Behaviour 144: 967–982.

McCune KB, Jablonski P, Lee S, Ha RR. 2018. Evidence for personality conformity, not social niche specialization in social jays. Behavioral Ecology. *In press.*

Morand-Ferron J, Cole EF, Quinn JL. 2016. Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. Biol. Rev. 91:367–389; doi:10.1111/brv.12174.

Morand-Ferron J, Cole EF, Rawles JEC, Quinn JL. 2011. Who are the innovators? A field experiment with 2 passerine species. Behav. Ecol. 22:1241–1248; doi:10.1093/beheco/arr120.

Morand-Ferron J, Quinn JL. 2011. Larger groups of passerines are more efficient problem solvers in the wild. Proc. Natl. Acad. Sci. 108:15898–15903; doi:10.1073/pnas.1111560108.

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biol. Rev. 85:935–956; doi:10.1111/j.1469-185X.2010.00141.x.

Overington SE, Cauchard L, Morand-Ferron J, Lefebvre L. 2009. Innovation in groups: does the proximity of others facilitate or inhibit performance? Behaviour 146:1543–1564; doi:10.1163/156853909X450131.

Penn DC, Povinelli DJ. 2007. On the lack of evidence that non-human animals possess anything remotely resembling a “theory of mind”. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 362:731–44; doi:10.1098/rstb.2006.2023.

Pepperberg IM, Willner MR, Gravitz LB. 1997. Development of Piagetian object permanence in a grey parrot (Psittacus erithacus). J. Comp. Psychol. 111: 63–75.

Povinelli DJ, Vonk J. 2004. We don’t need a microscope to explore the chimpanzee’s mind. Mind Lang. 19: 1–28.

Rstudio Team. 2015. RStudio: Integrated development for R. RStudio, Inc. Boston, MA.

Shettleworth SJ. 2010. Cognition, Evolution, and Behavior. Oxford University Press, Oxford, UK.

Therneau TM. 2018. coxme: Mixed effects Cox models. R package version 2.2-7.

Tomasello M, Call J. 2004. The role of humans in the cognitive development of apes revisited. Anim. Cogn. 7:213–215; doi:10.1007/s10071-004-0227-x.

van de Waal E, Bshary R. 2011. Contact with human facilities appears to enhance technical skills in wild vervet monkeys (Chlorocebus aethiops). Folia Primatol. 81: 282–291.

Webster S, Lefebvre L. 2001. Problem solving and neophobia in a columbiform–passeriform assemblage in Barbados. Anim. Behav. 62:23–32; doi:10.1006/anbe.2000.1725.

Wood S, Moriarty KM, Gardner BT, Gardner RA. 1980. Object permanence in child and chimpanzee. Anim. Learn. Behav. 8:3–9; doi:10.3758/BF03209723.

## Testing the social intelligence hypothesis in two congeneric jay species

Kelsey B. McCune[1*], Piotr Jablonski[2,3], Sang-im Lee[2,4], and Renee R. Ha[1]

1. The University of Washington Psychology Department, Seattle WA, USA
2. Laboratory of Behavioral Ecology and Evolution, School of Biological Sciences, Seoul National University, Seoul, South Korea
3. Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
4. Daegu-Gyeongbuk Institute of Science and Technology School of Undergraduate Studies, Daegu, South Korea
* Corresponding author: kmccune@uw.edu

### Abstract

Social cognitive abilities such as transitive inference, theory of mind, or social learning can improve survival and fitness of individuals that live in social groups. The social intelligence hypothesis (SIH) posits that life in complex social groups necessitates the evolution of these complex cognitive abilities, and might explain the extreme encephalization observed in humans. Previous research using broad, cross-taxon comparisons has failed to find consistent support for this hypothesis. However, these broad comparisons suffer from several confounds. First, they have used proxy measures of social and cognitive complexity such as group size and brain size. While this allows many more species to be quickly compared, it is unlikely that measures of group size and brain size are analogous measures across species. Additionally, very little comparative cognition research evaluates abilities of animal subjects in their natural environments. As a result, assessments of animals in the lab may lead to contrived performance that is not generalizable to the species as a whole. Here we conduct a rigorous test of the SIH by comparing social learning performance of wild subjects from two congeneric species divergent in social system. We used a novel foraging task with trained conspecific demonstrators to test the prediction that, compared to the asocial species, the social species would succeed more quickly at the complex task, and use social learning mechanisms rather than asocial learning. Contrary to predictions of the SIH, we found no difference in performance of naïve individuals in the social or asocial species. Additionally, we found little support for social learning mechanisms governing individual behavioral interactions with the task. Instead, behavioral interactions of naïve individuals are governed by personal information gained from trial-and-error learning; but social facilitation likely is important to motivate engagement with the task in general. These results demonstrate that a social system exhibiting communal, cooperative breeding is not sufficiently complex to select for precise social learning abilities. However the social environment does increase engagement with the task, and likely generalizes to social group engagement with natural foraging opportunities. Lastly, detecting an effect of social facilitation on social group performance would be improbable in a lab-based assay, demonstrating that it is important and interesting to test the SIH in wild individuals.

## Introduction

The social intelligence hypothesis (SIH) was developed to explain the larger brain sizes and complex cognitive abilities of humans and great apes relative to other mammals (Humphrey 1976; Jolly 1966). Optimal survival and fitness for an individual in a social group depends on successfully navigating social interactions with conspecifics. Correctly responding to the unpredictable behavior of conspecifics is more cognitively demanding than responding to the predictable occurrence of non-animate aspects of the environment like home range location and seasonally available food (Byrne and Bates 2007). Therefore, the SIH and related hypotheses predict individuals in large social groups will possess larger brains to support the need for increased memory of previous group-mate interactions, flexible decision-making, and reasoning abilities beyond the simple automatic behavioral responses to incoming stimuli (Dunbar 1998; Seyfarth and Cheney 2015).

Subsequent research on diverse taxa has yielded conflicting support for this hypothesis. For example, researchers found evidence that relative brain size was highly correlated with group size in primates (Dunbar 1998), cetaceans (Marino 1996), carnivores and insectivores (Dunbar and Bever 1998). Yet, additional analyses of brain size and group size in primates (Powell et al. 2017) and carnivores (Gittleman 1986) failed to find significant correlations. Instead, researchers have found correlations between brain size and non-social traits such as tool use and individual innovation (Lefebvre et al. 2004; Reader and Laland 2002). Finally, no relationship between sociality and brain size was found in birds (Beauchamp and Fernandez-Juricic 2004), or ungulates (Shultz and Dunbar 2006).

There are several potential reasons for such conflicting results. Previous research has relied on broad comparisons across the animal kingdom and various proxy measures for quantifying social and cognitive complexity. While this is important for sample sizes sufficient to support evolutionary hypotheses, results may be misleading because proxies such as group size or brain size may not mean the same thing within or across taxa.

Although in some species group size is a reliable indicator of pressures in the social environment that may necessitate complex cognitive abilities, in many other cases group size is overly general and temporally or spatially subjective. For some of the largest animal groups (i.e. wildebeest, colonial seabirds) there may be no need to evolve complex socio-cognitive abilities because no consistent individual interactions occur and individuals do not predicate their behavior on the behavior of others (Bergman and Beehner 2015). Furthermore, group size for many species depends on season or age (Alcock 2009). For example, during the breeding season the American Crow associates in small groups or pairs, but during the non-breeding season individuals aggregate in huge numbers (Verbeek and Caffrey 2002). Similarly, juvenile Common Ravens aggregate in non-breeding groups for several years before they sexually mature, then pair bond for life and become aggressively territorial and asocial (Boarman and Heinrich 1999). Given the dynamic nature of species sociality, it is important to explicitly hypothesize which types of social interactions should select for specific cognitive abilities (Bergman and Beehner 2015).

Brain size as a proxy for complex cognition is also problematic. The evolution and development of brain size and morphology depends on many non-social factors. For example, life history traits guiding perceptual biases (reliant on vision versus olfaction; Healy and Rowe 2007) and ecological factors (Marino 2002) likely exert strong evolutionary pressure on morphology and size of specific brain areas. Therefore, it is unclear why whole brain volume should scale with sociality, and it becomes impossible to compare brain size of land mammals to that in birds (where brain size is constrained by the need to fly), or even land mammals to marine mammals (where buoyancy creates an evolutionary release on brain size). Using relative brain part size ratios may help correct for allometric correlations with body size (Reader et al. 2005), but it still assumes function and structure of brain anatomy that are not yet verified and hard to generalize across taxa, or even across different methods to measure brain size (Healy and Rowe 2007). For instance, while the forebrain is perceived as the brain area most involved in non-reflexive, complex cognitive behaviors in mammals, the corresponding area in birds (Jarvis et al. 2013; Kirsch et al. 2008) is much smaller, but is more densely packed with neurons (Olkowicz et al. 2016). Similarly, in many cetacean species brain-to-body size ratios are similar to that seen in primates (Marino 1996), but the ratio of forebrain to the rest of the brain is considerably smaller, and enlargement occurred in the temporal and parietal lobes instead (Marino 2002). To avoid confounds associated with brain size proxies for cognitive ability, results of controlled cognitive assessments are a better way to categorize the actual extent of cognitive abilities across taxa.

Social learning is a socio-cognitive skill that is likely to be important in many species. By learning from group-mates or conspecifics, individuals will be optimally adapted to their local environment (Reader and Biro 2010), and responsive to environmental change (Whitehead 2010). Observational learning of a skill can be more efficient, and safer than individual learning through trial-and-error. For example, social learning of defenses against predators (Griffin 2008; Mathis et al. 1996) or nest parasites (Davies and Welbergen 2009) can directly increase survival and fitness. Similarly, though complex behaviors such as tool use are often partially innate, efficient practice of the behavior requires social learning (Kenward et al. 2006; Mann et al. 2012; Musgrave et al. 2016).

Social transmission of information can occur through various mechanisms. Some species (including humans; Horner and Whiten 2005) are able to exactly replicate the behaviors of a conspecific to gain a reward ("imitation"). In other taxa, naïve individuals attend to the location that a conspecific was rewarded, but not the specific behavior used to attain the reward ("local enhancement"). Additionally, it is possible that naïve individuals attend to the specific object a knowledgeable conspecific manipulated to gain the reward, but not the behavior ("stimulus enhancement"). Disentangling these mechanisms can provide more detailed knowledge of the fitness benefits of sociality, and the specific features of conspecific behaviors that individuals use to inform foraging decisions (Laland 2004; Zentall 2012).

Many studies that experimentally quantify social learning mechanisms in animals are performed in lab environments where ecological validity is low. Results from these studies are difficult to replicate and generalize to wild populations because of the tightly controlled artificial conditions (Shettleworth 2010). Developmental, social, and motivational characteristics of subjects in the lab are unlikely to approximate those in wild conspecifics (Benson-Amram et al. 2013; Reader and Biro 2010). While it may be more difficult to control some confounds during cognitive assessments conducted on wild individuals, experimentation is taking place in the location in which selection is occurring (Kendal et al. 2010), and so results will be ecologically valid and have greater interpretability (Coussi-Korbel and Fragaszy 1995; Thornton and Lukas 2012).

To date, no previous studies have directly compared performance on socio-cognitive tasks in wild, congeneric species (Reader and Biro 2010; Slagsvold and Wiebe 2011; Whiten and Mesoudi 2008). In order to provide a rigorous test of the SIH, it is necessary to compare closely related species whose life history traits vary only in degree of social behavior (Byrne and Bates 2007). Here we quantify whether social learning of a novel foraging behavior occurs in wild subjects from a social and an asocial species. California Scrub-Jays (hereafter CASJ; *Aphelocoma californica*) exhibit a pair-bonded monogamous breeding system and vigorously defend their territory from other jays, including young of the previous year (Curry et al. 2017). In contrast, Mexican Jays (hereafter MEJA; *A. wollweberi*) are communal, cooperative breeders and live in stable groups of 5-25 jays (Brown 1994). Young delay dispersal for several years or remain in the natal group for life. Therefore, MEJA have a stable social environment in which they can observe and learn from relatives and group-mates (Coussi-Korbel and Fragaszy 1995). Additionally, a dominance hierarchy governs access to food and reproductive resources (Brown et al. 1997; McCormack et al. 2007), so successfully navigating social interactions within the group likely requires increased social attention, individual recognition, and the capacity to maintain memories of, and learn from previous interactions (Byrne 1997).

Even though social behavior is very distinct, these two species occupy similar ecological niches. Both are mast-dependent species that prefer dry, open, scrub-oak and pine habitats (Rice et al. 2003). As such, differences in brain evolution or development most likely result from the divergent pressures of the social environment. If the SIH is supported, we predict 1) a significant species difference such that naïve MEJA will show a faster rate of first interaction with the social learning task after observation of a knowledgeable conspecific demonstrator than naïve CASJ, and 2) a significant treatment effect where social groups with a trained demonstrator will show a faster rate of first interaction than those in which there is no demonstrator. Although we compared only two species, we believe these results will be an important step towards a better understanding of the evolution of cognition in natural systems.

## Methods

*Study system:* Subjects for this experiment included individually color-banded MEJA from 5 flocks around the Southwestern Research Station in Portal, AZ and CASJ from 16 territories in Willamette Mission State Park, near Keizer, OR. All jays were trained to come for peanuts when we whistled, which facilitated timely participation in the social learning task. We conducted social learning trials on 49 total MEJA and 26 CASJ subjects. MEJA live in natural stable groups (flocks), whereas CASJ social groups were created from a minimum of 3 neighboring mate-pairs surrounding central, open picnic areas. For several weeks before the trials began, we put peanuts on the central tables to check that no CASJ from surrounding territories defended this “neutral zone”. Even though CASJ only maintain one enduring social bond (their mate), it is not unusual for dispersing hatch year and adult non-neighbor CASJ to interact at large natural food sources (i.e. oak trees in late summer; Pesendorfer and Koenig 2017). As such, we believe trials conducted in neutral zones do not represent a contrived experience for CASJ subjects.

*Experimental apparatus:* Our social learning apparatus consisted of a puzzle box created out of a log to mimic natural extractive foraging for grubs or cache-recovery behaviors (Brown 1994). We created four food-holding compartments in the log, covered by transparent doors that open in different ways (Fig. 1a), labeled as A, B, C, and D doors. First, in the center of each group location we allowed the group access to the open puzzle box filled with peanuts until they habituated to this new food source. Secondly, all jays were given previous experience interacting with the closed puzzle box and opening the simple doors (McCune unpub. data); we included the amount of this previous experience as a covariate in our final models. The diverse door options engaged individuals that varied in their initial motivation to interact with different opening methods at distinct locations on the puzzle box, and allowed us to determine if stimulus enhancement is an important mechanism governing jay social learning. Subsequent to habituation to the simple puzzle box doors, we added locks to doors A, B, and C. We left D door unlocked, but filled with a less preferred food (seeds rather than peanuts), to encourage naïve jay attendance at the puzzle boxes to increase opportunities to observe group-mate interactions. The lock for A and B doors similarly consisted of a stick that blocked the opening of the doors, but required different behaviors to solve. The lock on C door was a hook-and-eye type lock (Fig. 1b).

*Demonstrator training:* Only 2 select jays within a group were trained as demonstrators to open a lock on the puzzle box. For each species we trained demonstrators from 4 of the social groups, whereas in the 5th group (control) we sham-trained two individuals so that they had equal amounts of experience with the locked puzzle box, but no shaping to open the locks. Demonstrators were all adults, and only trained on either A or B door lock, such that we had groups in 3 treatments: control, trained on lock A, trained on lock B (Table 1).

Due to the differences in species social systems, we used slightly different methods to isolate demonstrator jays for training. In CASJ, the aggressive territoriality created a situation where we could train a pair as demonstrators in the wild, in the middle of their territory. However, for MEJA it was impossible to separate two individuals from group-mates for training in the wild. Therefore we took two jays from each of the experimental flocks into captivity for training (details of captive environment in Ch. 3 of this document). All other training methods are identical for the two species. We first ensured that demonstrators could efficiently open the simple unlocked door that we planned to train them on (either A or B door). If not, we used a shaping procedure in which the door was gradually closed until the jays learned to manipulate it to gain access to the food compartment. All other doors were left open and empty to discourage interaction at these loci. Next, we habituated jays to the puzzle box with the focal door closed, and the lock on but not engaged. We then moved the stick so that it was in the way of the door, but not engaged, until jays began to manipulate the stick to access the door. Subsequently, we moved the stick so that it was incrementally more fully in the locked position. We progressed with the learning trials in each group only after demonstrators were efficiently opening only the trained option, and all naïve jays were eating comfortably from compartments in the puzzle box.

*Data collection:* For all groups, we conducted between 9 and 16 trials that were a maximum of 2 hours in length (Table 1). Trials were conducted in the mornings before noon, and each trial occurred at least 24 hours after the previous trial. To begin a trial, we deployed two open, unlocked puzzle boxes filled with peanuts, approximately 2m apart, in the center of each group area. It was important to deploy two puzzle boxes during a trial to quantify the occurrence of local enhancement (Zentall 2012), and to prevent monopolization of the puzzle box by dominant individuals. We determined that jays were motivated to engage with the task if members of the social group arrived and ate from the open puzzle boxes comfortably within 10 minutes of putting them out and whistling. If no jays arrived, the trial attempt was aborted. Otherwise we closed and locked all doors and began the trial. We noted identity of all jays within 10m of the puzzle box, and noted interaction behaviors using the ethogram in Table 2. As such, we quantified: who interacted, at which of the two puzzle boxes, with which door on the puzzle box, the behavior at the door, whether the interaction was successful (obtained a peanut) or unsuccessful, and which other jays were present within 10m to potentially observe the interacting jay.

Since demonstrators were trained to open doors by interacting with door *locks*, we divided all interactions at the doors into two categories. The first category describes interactions at a door that mimic those of the demonstrator in that they are directed at the lock on the door. The second category describes interactions that mimic previous experience with the unlocked doors in that the interactions are exclusively directed at the door, not the lock. Therefore, there were 7 total interaction "Options": A1, A2, B1, B2, C1, C2, and D. We monitored social group motivation throughout the trial, and we ended the trial early if for 30 consecutive minutes no jays other than the demonstrator came within 2m of the puzzle boxes. All trials were filmed, and behaviors coded later by individuals naïve to treatment condition and hypotheses. All methods were approved by the University of Washington Institutional Animal Care and Use Committee (protocol #4064-03), and the appropriate permitting agencies (Bird Banding Lab #22802, Oregon Department of Fish and Wildlife #015-16, Oregon State Parks #007-14, US Fish and Wildlife Service #MB51894B-0, Arizona Game and Fish Department #SP697293).

*Data analysis:* We used non-parametric analyses (Mann-Whitney U Test) to compare engagement with the experimental task of each species and treatment type. We conducted cox proportional hazard analyses (*coxme* package; Therneau 2018) in the R statistical environment (RStudio Team 2015) to examine the influence of social information on the rate of first interaction with each puzzle box option of naïve jays (Logan et al. 2015). We tested the cox proportional hazard assumption that the baseline hazard rate of our covariates for all subjects is constant over time (Grambsch and Therneau 1994) and found that the "option" variable violated this assumption ($p$ = 0.02) and so we stratified data by option in each cox model. We first used the cox models to compare species, and treatment conditions. We included individual level covariates of dominance rank (methodological details of dominance assays in Ch. 1 of this document; Barkan et al.

1986) and amount of previous experience with the puzzle box for both species, as well as age (adult or juvenile) for MEJA only because all but 4 CASJ were adults. We included random effects of individual ID to account for interactions by the same jay at multiple options, and a group ID random effect to account for interactions by multiple jays in the same social group.

Secondly, to test for an effect of social information use in each species we included covariates to describe the social learning mechanisms in addition to the covariates and random effects from the first set of models. For each first interaction of a naïve jay at a given option, we evaluated the support for a mechanism of imitative learning as the number of interactions the focal jay had previously observed at the same puzzle box and same option. We quantified the effect of stimulus enhancement as the number of interactions the focal jay had previously observed at the same option, but on either puzzle box. We assessed the use of a local enhancement learning mechanism as the number of interactions a focal jay previously observed at the specific puzzle box, regardless of which option jays directed interactions to. Lastly, we needed to account for the possibility that within the social learning trials, an individual's own prior successful interactions at the puzzle box alter the rate of first interaction with different options (asocial enhancement). Therefore, we summed the number of times a given focal jay attained a peanut from different options before its first interaction with a focal option.

## Results

Jays from both species had equal opportunity to interact with, or observe conspecific interactions at the puzzle box. Demonstrators from both species performed similarly, there was no significant difference in the number of times that a MEJA or CASJ demonstrator opened the door that it was trained on ($W = 16$, $p = 0.53$). There were no significant species differences in the number of trials ($W = 13$, $p = 1$), total trial time ($W = 9$, $p = 0.55$), or number of jays in each group ($W = 7$, $p = 0.29$). Similarly, control and experimental treatment groups did not differ in trial time ($W = 2$, $p = 0.18$), number of trials ($W = 2$, $p = 0.11$), or number of jays per group ($W = 6$, $p = 0.69$). However, the demonstrator from one experimental MEJA group never performed the lock-opening behavior. When we consider this group as part of the control treatment we find jays in experimental treatments experienced more trials ($W = 0$, $p = 0.01$) and longer trial times ($W = 0$, $p = 0.02$) than jays in the control treatment. All subsequent analyses consider this group as part of the control treatment.

Naïve MEJA in groups with demonstrators trained on either A or B door locks also performed similarly. Experimental groups trained on lock A and groups trained on lock B did not differ in rate of first interaction with each option on the puzzle box (Hazard ratio of B trained groups relative to A = 1.33, $z = 0.45$, $p = 0.65$), but both had a significantly faster first rate of interaction than did MEJA from control groups (Fig. 2; HR = 8.55, $z = 6.96$, $p < 0.001$). Therefore, we combined the two types of experimental groups for the second set of analyses examining support for various social learning mechanisms. In CASJ there was no difference in the rate of first interaction between the two types of experimental groups (HR of B trained groups relative to A = 1.59, *z* = 1.50, *p* = 0.13), or for experimental groups relative to control groups (HR = 1.28, *z* = 0.59, *p* = 0.55), so all CASJ data are combined in analyses testing the social learning mechanisms. There was no difference in performance between species in the experimental groups (HR of MEJA relative to CASJ = 2.22, *z* = 1.10, *p* = 0.27) or control groups (HR = 0.84, *z* = -0.35, *p* = 0.72).

Species differed in the learning mechanisms that significantly predicted rate of first interaction (Table 3). For MEJA, prior observations of a group-mate succeeding at a given option actually *reduced* the rate of first interaction with that option (imitation; HR = 0.83, *z* = -2.68, *p* = 0.01). Additionally, there was support for a significant negative effect of both stimulus enhancement (HR = 0.88, *z* = -2.09, *p* = 0.04) and local enhancement (HR = 0.88, *z* = -2.43, *p* = 0.02). Previous success at other puzzle box options significantly reduced the rate of first interaction at a given option (asocial enhancement; HR = 0.97, *z* = -3.55, *p* = 0.0004). Finally, rate of first interaction was positively related to juvenile age class (HR = 4.38, *z* = 2.07, *p* = 0.04), and the amount of previous experience with the puzzle box (HR = 1.01, *z* = 3.43, *p* = 0.001), but not related to dominance rank (Table 3; HR = 0.80, *z* = -0.22, *p* = 0.83). In contrast, for CASJ only dominance rank was significantly related to the rate of first interaction (HR = 0.85, *z* = -1.96, *p* = 0.05), where lower ranking jays had an increased rate of first interaction.

## Discussion

This research is the first to compare an asocial and social species on a socio-cognitive task in the wild. Although we only compared two species, we believe these results provide a rigorous test of the SIH because we compared congeneric species with life history characteristics that differ only in degree of sociality. Therefore any differences in cognitive performance are likely attributable to evolution and development in social groups.

In accord with the SIH, the performance of the asocial species showed no evidence for an effect of social information use. CASJ experimental groups seeded with a demonstrator trained on this complex task did not differ in performance from control groups without demonstrators (Fig. 2). When we examined the rate of first interaction of each naïve CASJ on every puzzle box option we found no significant effect of previously observing conspecifics successfully opening puzzle box doors. Previous research on captive CASJ found individuals do attend to the foraging behavior of conspecifics, can remember who cached food, and the spatial location where it was hidden (Dally et al. 2006). Given our results, it's possible that the behavior of hand-reared, captive CASJ that live in a constant social environment is not representative of wild jay behavior. Alternatively, wild CASJ may not attend to *novel* behaviors of conspecifics, or remember novel behaviors that led to a conspecific attaining a reward.

Contrary to predictions of the SIH we found no significant difference in the rate of first interaction of naïve CASJ and MEJA at the puzzle box task. Yet, we did find some evidence that MEJA are attending to social information. First, the hazard ratio comparing the rate of first interaction between species was greater than 2, indicating MEJA rate of first interaction was more than double that of CASJ. Secondly, MEJA in groups with trained demonstrators outperformed MEJA in control groups with no seeded demonstrators (Fig. 2). In general, control groups from both species showed lower motivation and were less willing to engage with the puzzle box, resulting in fewer trials and shorter trial times (Table 1). Lastly, we found that the rate of first interaction of a naïve MEJA at a given option was significantly affected by prior observations of group-mates either succeeding at that specific option (stimulus enhancement), or on that specific puzzle box (local enhancement), or both (imitation). However, this social information actually *negatively* affected the rate of first interaction of naïve jays at that option, indicating that they are avoiding options and locations where they observe group-mates succeed. Because we quantified the time to *first* interaction at a given puzzle box option, it is unlikely this result represents jays learning that a puzzle box compartment is depleted of peanuts after a group-mate succeeds. Moreover, naïve MEJA rate of first interaction at new options is hindered by their own previous success at other puzzle box options. Together, these results indicate that MEJA prioritize personal information and return to doors they have previously opened, but that social facilitation, where naïve jays are positively influenced by the motivation of knowledgeable group-mates (Zentall 2012), increases the likelihood of interacting with the novel foraging opportunity.

It is unexpected that MEJA would not use social learning to succeed at a complex foraging task because traits exhibited by this species, such as group-level coordination of behavior and prolonged development in a natal social group, are hypothesized to favor social learning (Coussi-Korbel and Fragaszy 1995). One potential explanation is that jays *did* learn, but did not have the opportunity to reproduce the behavior due to competition for access to that door. Future research with isolated jays is needed to discriminate between this possible explanation and the observed asocial enhancement leading to preference for doors previously opened by the focal jay. For example, after demonstrations in the social group, a locked puzzle box could be presented to solitary naïve jays. The researcher could then track the order in which jays open and empty the different puzzle box doors. Likely, jays will first go to the doors that they opened previously using trial-and-error individual learning (but that could be manipulated by filling the compartment with non-preferred food). However, if jays are subsequently able to replicate the behavior of the trained demonstrator quickly and efficiently to get food from the remaining locked door, then that could be evidence for social learning ability.

Another potential reason MEJA did not show social learning is that the foraging environment is not sufficiently complex or dynamic to merit the evolution of social learning of foraging behaviors (Kendal et al. 2005). MEJA and CASJ rely heavily on masting trees like oaks or pinyon pines for sustenance. While acorn shells and pinyon cones may be tough to breach, the motor patterns needed to succeed could be partially innately programmed (Curry et al. 2017), are not complex (Byrne and Russon 1998), and there is little risk of injury from trial-and-error learning (Hoppitt and Laland 2013). These two species are also opportunistic omnivores on arthropods and small vertebrates, especially during early spring through summer. However, it is possible that no animal prey are lethally toxic in a way that would favor social learning of which to eat over individual trial-and-error learning (Terhune 1977).

We found little evidence that social learning led to the ability to solve a novel foraging problem, indicating that complex social life is not sufficient for social learning cognitive mechanisms to evolve. This research provides no support for the SIH. However, we compared performance of only two species on a single socio-cognitive ability. Therefore to more explicitly understand the relationship between sociality and cognition, further experimental assessments of closely related species in the wild are needed. Future research could consider a non-foraging context for comparing social learning ability such as acquiring recognition of a novel predator as a threat after exposure to mobbing by a conspecific (Cornell et al. 2011). Alternatively, it may be more likely that the two species studied here would diverge in performance on a test of the socio-cognitive ability transitive inference (Bond et al. 2003), since access of an individual MEJA to resources is governed by within-group dominance hierarchies (Brown et al. 1997). To increase validity and repeatability of experiments testing the SIH, it is important for future research to reconsider the use of proxies for quantifying social behavior and cognition and to increase the ecological relevance of test assessments by extending experiments outside of the laboratory.

## Tables and figures

Table 1: Summary of experimental design. Numbers presented are treatment group means ± SE, and trial times are in minutes. California Scrub-Jay groups are labeled with the "SG" prefix and a group number, whereas Mexican Jay groups are labeled with a two-letter abbreviation.

| Treatment | Group names | Number of jays | Number of trials | Trial times |
|---|---|---|---|---|
| Control | CO, UC, SG4 | 8 ± 1.15 | 10.67 ± 1.2 | 902.3 ± 28 |
| Lock A | XM, SG1, SG2, SG5 | 9 ± 1.08 | 14 ± 0 | 1256.7 ± 98.8 |
| Lock B | KI, TA, SG3 | 9.33 ± 1.33 | 14.67 ± 0.67 | 1408.3 ± 55.7 |

Table 2: Ethogram used to code behaviors of jays at the puzzle boxes. Each behavior was then identified with a specific puzzle box and a specific door. Options described whether behavior was directed at the door or the lock and included A1, A2, B1, B2, C1, C2 and D

| Behavior | Definition | Option |
|---|---|---|
| Attempt_D | Attempt on the closed and locked door | 2 |
| Attempt_L | Attempt on the lock mechanism (stick or hook) | 1 |
| Success | Obtained a peanut after unlocking and opening the door | 1 |
| Scrounge | Obtained a peanut after another jay opened the door | 2 |
| Pry | Pried open the locked door just enough to get a peanut | 2 |
| Jiggle | Obtained a peanut by manipulating the door until the lock jiggled open or fell off | 2 |

Table 3: Cox proportional hazard model output for each species showing the influence of covariates on the rate of first interaction. Covariates include individual level variables, and social learning variables that describe the number of observed successes made by group-mates prior to first interaction of the focal jay. Significance values in bold face indicate those that fall below the alpha criterion of 0.05.

| CASJ | Hazard ratio | *z* | *p* |
|---|---|---|---|
| Rank | 0.85 | -1.96 | **0.05** |
| Previous Experience | 1.00 | 1.64 | 0.10 |
| Imitation | 0.98 | -0.04 | 0.97 |
| Stimulus enhancement | 0.69 | -1.12 | 0.26 |
| Local enhancement | 0.67 | -0.98 | 0.33 |
| Asocial enhancement | 0.99 | -1.75 | 0.08 |
| MEJA | Hazard ratio | *z* | *p* |
| Age (Juvenile) | 4.38 | 2.07 | **0.04** |
| Rank | 0.80 | -0.22 | 0.83 |
| Previous Experience | 1.01 | 3.43 | **<0.01** |
| Imitation | 0.83 | -2.68 | **0.01** |
| Stimulus enhancement | 0.88 | -2.09 | **0.04** |
| Local enhancement | 0.88 | -2.43 | **0.02** |
| Asocial enhancement | 0.97 | -3.55 | **<0.01** |

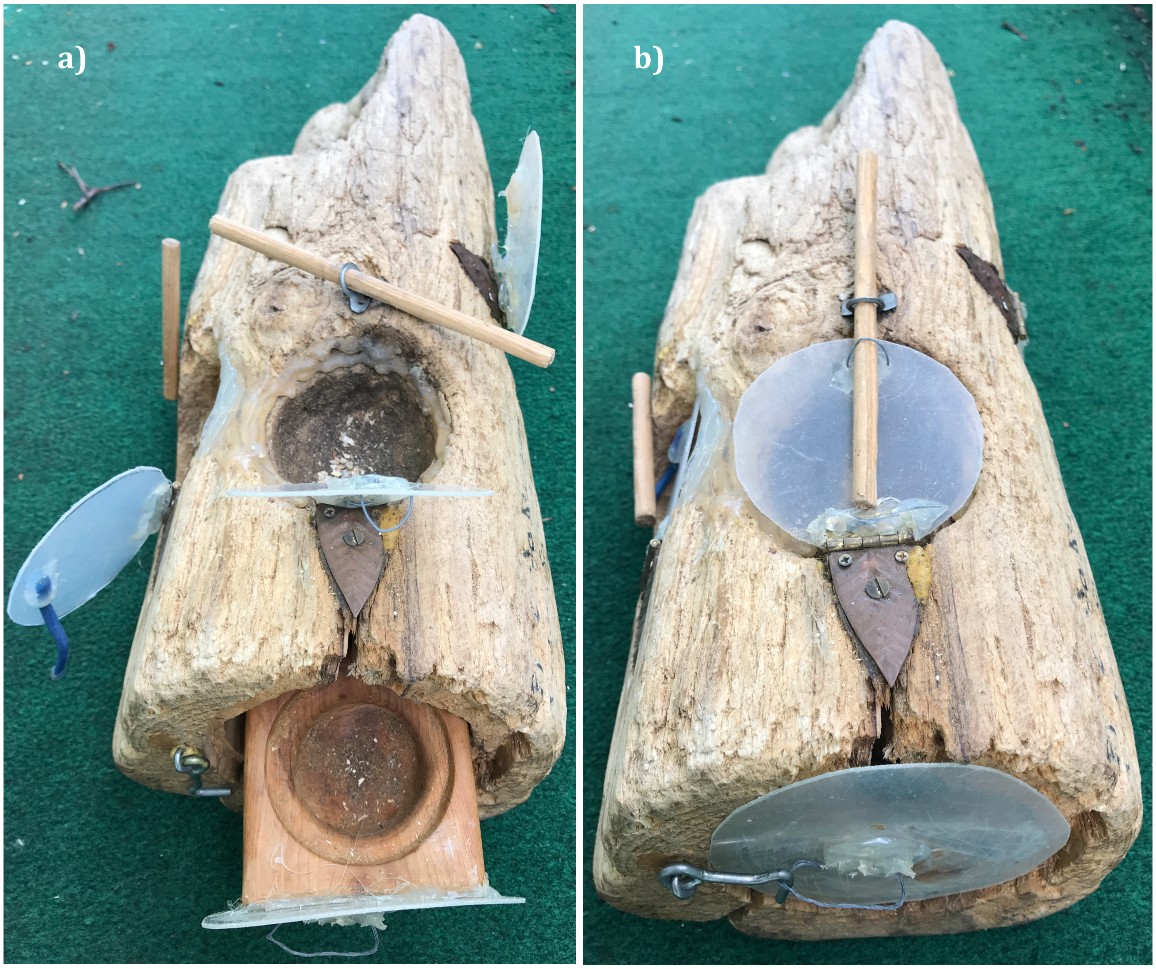

Figure 2: The puzzle box apparatus that we used to assess social learning of a novel foraging skill. A) shows the puzzle box with all doors open, while B) shows the closed and locked puzzle box. We trained demonstrator jays to open the top door (labeled A door), and the door on the left (B door).

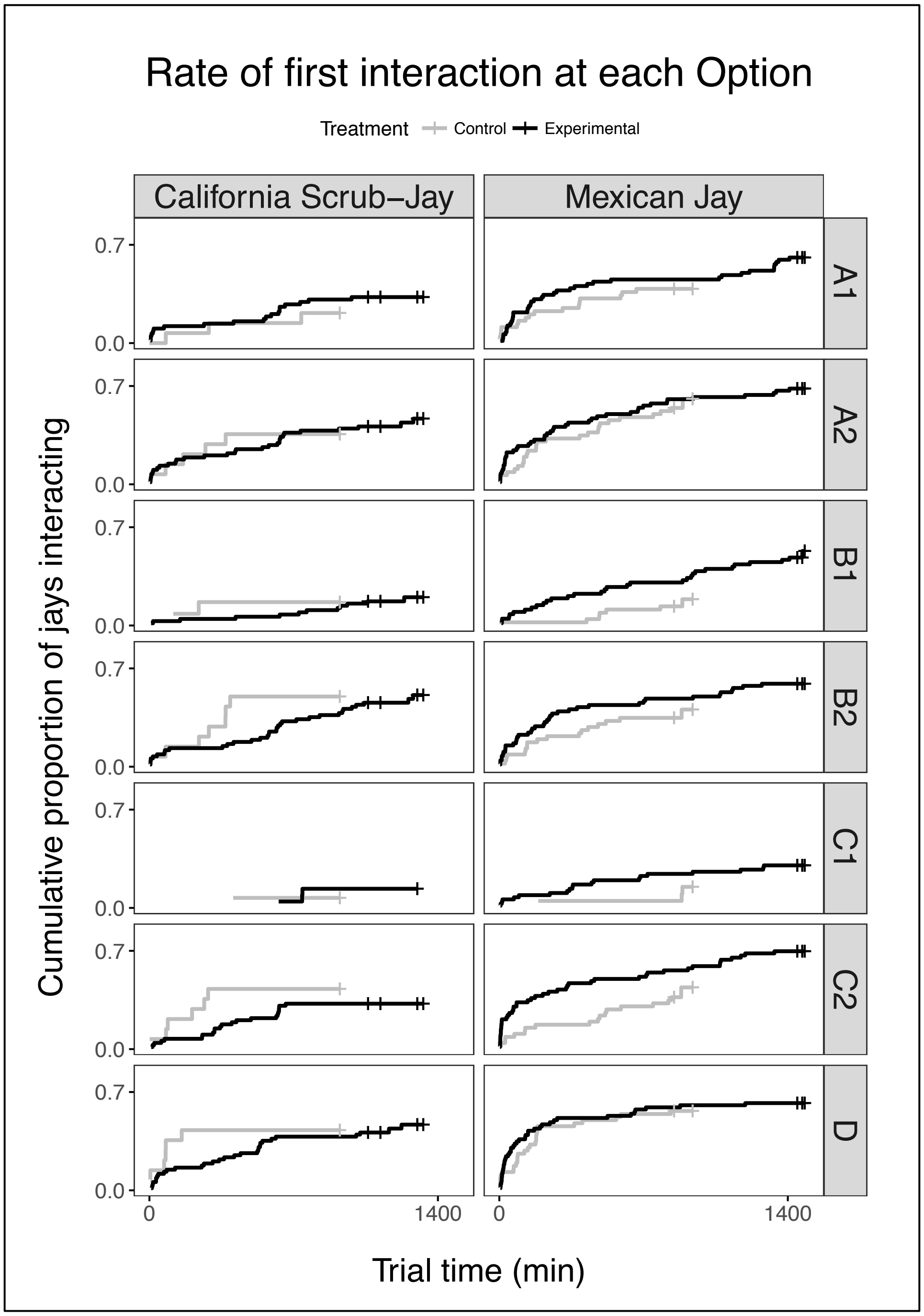

Figure 3: Survival curves illustrating the rate of first interaction at all puzzle box options for both species and treatments. Experimental groups had a seeded demonstrator trained on A1 and B1 options. If social learning occurs experimental groups should outperform control groups on this option. All jays had previous experience at option D so treatment types should perform equally.

## Literature cited

Barkan CPL, Craig JL, Strahl SD, Stewart AM, Brown JL. 1986. Social dominance in communal Mexican jays Aphelocoma ultramarina. Anim. Behav. 34:175–187; doi:10.1016/0003-3472(86)90021-7.

Benson-Amram S, Weldele ML, Holekamp KE. 2013. A comparison of innovative problem-solving abilities between wild and captive spotted hyenas, Crocuta crocuta. Anim. Behav. 85:349–356; doi:10.1016/j.anbehav.2012.11.003.

Bergman TJ, Beehner JC. 2015. Measuring social complexity. Anim. Behav. 103:203–209; doi:10.1016/j.anbehav.2015.02.018.

Bond AB, Kamil AC, Balda RP. 2003. Social complexity and transitive inference in corvids. Anim. Behav. 65:479–487; doi:10.1006/anbe.2003.2101.

Brown JL. 1994. Mexican Jay (Aphelocoma ultramarina). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Brown JL, Brown ER, Sedransk J, Ritter S. 1997. Dominance, age, and reproductive success in a complex society: A long-term study of the Mexican Jay. Auk 114: 279–286.

Byrne D. 1997. Machiavellian Intelligence. Evol. Anthropol. 172–180; doi:10.1007/s13398-014-0173-7.2.

Byrne RW, Bates L a. 2007. Sociality, evolution and cognition. Curr. Biol. 17:R714-23; doi:10.1016/j.cub.2007.05.069.

Byrne RW, Russon AE. 1998. Learning by imitation: A hierarchical approach. Behav. Brain Sci. 21:667–721; doi:10.1017/S0140525X98001745.

Cornell HN, Marzluff JM, Pecoraro S. 2011. Social learning spreads knowledge about dangerous humans among American crows. Proc. Biol. Sci. / R. Soc. 1–11; doi:10.1098/rspb.2011.0957.

Coussi-Korbel S, Fragaszy DM. 1995. On the relation between social dynamics and social learning. Anim. Behav. 50:1441–1453; doi:10.1016/0003-3472(95)80001-8.

Curry RL, Peterson AT, Langen TA, Pyle P, Patten MA. 2017. California Scrub-Jay (*Aphelocoma californica*), version 3.0. *In* The Birds of North American (PG Rodewald, Ed). Cornell Lab of Ornithology, Ithaca NY.

Dally JM, Emery NJ, Clayton NS. 2006. Food-caching Western scrub-jays keep track of who was watching when. Science 312: 1662–1665.

Davies NB, Welbergen J a. 2009. Social Transmission of a Host Defense Against Cuckoo Parasitism. Science (80-. ). 324:1318–1320; doi:10.1126/science.1172227.

Dunbar RIM. 1998. The social brain hypothesis. Evol. Anthropol. Issues, News, Rev. 6:178–190; doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

Dunbar RIM, Bever J. 1998. Neocortex size predicts group size in carnivores and some insectivores. Ethology 104:695–708; doi:10.1111/j.1439-0310.1998.tb00103.x.

Gittleman JL. 1986. Carnivore brain size, behavioral ecology, and phylogeny. J. Mammal. 67: 23–36.

Grambsch P, Therneau T. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika,* **81**, 515-26.

Griffin AS. 2008. Socially acquired predator avoidance: is it just classical conditioning? Brain Res. Bull. 76:264–71; doi:10.1016/j.brainresbull.2008.02.005.

Healy SD, Rowe C. 2007. A critique of comparative studies of brain size. Proc. R. Soc. B Biol. Sci. 274:453–464; doi:10.1098/rspb.2006.3748.

Hoppitt W, Laland KN. 2013. Social learning: An introduction to Mechanism, Methods, and Models. Princeton University Press, Princeton NJ.

Horner V, Whiten A. 2005. Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). Anim. Cogn. 8:164–81; doi:10.1007/s10071-004-0239-6.

Humphrey NK. 1976. The social function of intellect. In: *Growing Points in Ethology*. 303–317.

Jarvis ED, Yu J, Rivas M V., Horita H, Feenders G, Whitney O, et al. 2013. Global view of the functional molecular organization of the avian cerebrum: Mirror images and functional columns. J. Comp. Neurol. 521:3614–3665; doi:10.1002/cne.23404.

Jolly A. 1966. Lemur social behavior and primate intelligence. Science (80-. ). 153: 501–6.

Kendal RL, Coolen I, Bergen Y Van, Laland KN. 2005. Trade-offs in the aadaptive use of social and asocial learning. Adv. Study Behav. 35:333–379; doi:10.1016/S0065-3454(05)35008-X.

Kendal RL, Galef BG, van Schaik CP. 2010. Social learning research outside the laboratory: How and why? Learn. Behav. 38:187–94; doi:10.3758/LB.38.3.187.

Kenward B, Rutz C, Weir AAS, Kacelnik A. 2006. Development of tool use in New Caledonian crows: inherited action patterns and social influences. Anim. Behav. 72:1329–1343; doi:10.1016/j.anbehav.2006.04.007.

Kirsch JA, Güntürkün O, Rose J. 2008. Insight without cortex: Lessons from the avian brain. Conscious. Cogn. 17:475–483; doi:10.1016/j.concog.2008.03.018.

Laland KN. 2004. Social learning strategies. Anim. Learn. Behav. 32:4–14; doi:10.3758/BF03196002.

Lefebvre L, Reader SM, Sol D. 2004. Brains, innovations and evolution in birds and primates. Brain. Behav. Evol. 63:233–46; doi:10.1159/000076784.

Logan CJ, Breen AJ, Taylor AH, Gray RD, Hoppitt W. 2015. How New Caledonian crows solve novel foraging problems and what it means for cumulative culture. Learn. Behav. 1–15.

Mann J, Stanton M a, Patterson EM, Bienenstock EJ, Singh LO. 2012. Social networks reveal cultural behaviour in tool-using dolphins. Nat. Commun. 3:980; doi:10.1038/ncomms1983.

Marino L. 2002. Convergence of complex cognition in cetaceans and primates. Brain, Behav. Evol. 59:21–32; doi:10.1159/000063731.

Marino L. 1996. What can dolphins tell us about primate evolution? Evol. Anthropol. Issues, News, Rev. 5: 81–86.

Mathis A, Chivers DP, Smith RJF. 1996. Cultural transmission of predator recognition in fishes: Intraspecific and interspecific learning. Anim. Behav. 51:185–201;

doi:10.1006/anbe.1996.0016.

McCormack JE, Jablonski PG, Brown JL. 2007. Producer-Scrounger Roles and Joining Based on Dominance in a Free-Living Group of Mexican Jays ( Aphelocoma ultramarina ). Behaviour 144: 967–982.

Musgrave S, Morgan D, Lonsdorf E, Mundry R, Sanz C. 2016. Tool transfers are a form of teaching among chimpanzees. Sci. Rep. 6:34783; doi:10.1038/srep34783.

Olkowicz S, Kocourek M, Lučan RK, Porteš M, Fitch WT, Herculano-Houzel S, et al. 2016. Birds have primate-like numbers of neurons in the forebrain. Proc. Natl. Acad. Sci. 113:7255–7260; doi:10.1073/pnas.1517131113.

Pesendorfer MB, Koenig WD. 2017. Competing for seed dispersal: evidence for the role of avian seed hoarders in mediating apparent predation among oaks. Funct. Ecol. 31:622–631; doi:10.1111/1365-2435.12770.

Powell LE, Isler K, Barton RA. 2017. Re-evaluating the link between brain size and behavioural ecology in primates. Proceedings. Biol. Sci. 284:20171765; doi:10.1098/rspb.2017.1765.

Reader SM, Biro D. 2010. Experimental identification of social learning in wild animals. Learn. Behav. 38:265–83; doi:10.3758/LB.38.3.265.

Reader SM, Laland KN. 2002. Social intelligence, innovation, and enhanced brain size in primates. Proc. Natl. Acad. Sci. 99:4436–4441; doi:10.1073/pnas.062041299.

Reader SM, Sol D, Lefebvre L. 2005. Comparing cognition across species. Trends Cogn. Sci. 9:411; doi:10.1016/j.tics.2005.07.008.

Rice NH, Martínez-meyer E, Peterson AT. 2003. Ecological niche differentiation in the Aphelocoma jays: A phylogenetic perspective. Biol. J. Linn. Soc. 80: 369–383.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Seyfarth RM, Cheney DL. 2015. Social cognition. Anim. Behav. 103:191–202; doi:10.1016/j.anbehav.2015.01.030.

Shettleworth S. 2010. Cognition, evolution and behavior. Oxford University Press, Oxford UK.

Shultz S, Dunbar RIM. 2006. Both social and ecological factors predict ungulate brain size. Proc. Biol. Sci. / R. Soc. 273:207–215; doi:10.1098/rspb.2005.3283.

Slagsvold T, Wiebe KL. 2011. Social learning in birds and its role in shaping a foraging niche. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366:969–77; doi:10.1098/rstb.2010.0343.

Terhune EC. 1977. Components of a visual stimulus used by scrub-jays to discriminate a Batesian model. Am. Nat. 111: 435–451.

Terry M. Therneau (2018). coxme: Mixed Effects Cox Models. R package version 2.2-7. https://CRAN.R-project.org/package=coxme

Thornton A, Lukas D. 2012. Individual variation in cognitive performance: Developmental and evolutionary perspectives. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:2773–83; doi:10.1098/rstb.2012.0214.

Whitehead H. 2010. Conserving and managing animals that learn socially and share cultures. Learn. Behav. 38:329–336; doi:10.3758/LB.38.3.329.

Whiten A, Mesoudi A. 2008. Review. Establishing an experimental science of culture: animal social diffusion experiments. Philos. Trans. R. Soc. Lond. B. Biol. Sci.

363:3477–3488; doi:10.1098/rstb.2008.0134.

Zentall TR. 2012. Perspectives on observational learning in animals. J. Comp. Psychol. 126:114–28; doi:10.1037/a0025381.

# Evidence for personality conformity, not social niche specialization in social jays

Kelsey McCune[1*], Piotr Jablonski[2,3], Sang-im Lee[2,4], and Renee Ha[1]

1. Psychology Department, University of Washington, Seattle, WA, USA
2. Laboratory of Behavioral Ecology and Evolution, School of Biological Sciences, Seoul National University, Seoul, South Korea
3. Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
4. Daegu-Gyeongbuk Institute of Science and Technology School of Undergraduate Studies, Daegu, South Korea

*Corresponding author:
Kelsey McCune
Psychology Department, University of Washington
Box #351525
Seattle, WA 98105
kmccune@uw.edu

## Abstract

Animal personality traits are defined as consistent individual differences in behavior over time and across contexts. Occasionally this inflexibility results in maladaptive behavioral responses to external stimuli. However, in social groups inflexible behavioral phenotypes might be favored as this could lead to more predictable social interactions. Two hypotheses seek to describe the optimal distribution of personality types within groups. The social niche specialization hypothesis states that individuals within groups should partition social roles, like personality types, to avoid conflict. Whereas the conformity hypothesis states that individuals should assort with conspecifics of similar personality. However no research so far has compared these hypotheses using data from wild animal systems. We tested boldness in the wild on two species with vastly different social systems, the Mexican Jay and California Scrub-Jay. We found support for the

conformity hypothesis over the social niche specialization hypothesis because individuals within groups of the social species had more similar personalities, and consequently there was a statistically significant group effect. The most likely mechanism for this conformity is social learning of behaviors through development, but more explicit research on this is needed.

## Introduction

The evolution of animal personalities is not yet well understood. An animal personality trait is defined as a behavior that differs among individuals, but is consistent within an individual over time and across contexts (Gosling 2001). Much research has quantified personality traits in different populations and contexts, particularly along the bold/shy spectrum (Carter et al. 2013). At the population level, variation in a personality trait within a species can lead to population persistence under changing environmental conditions (e.g. Nicolaus et al. 2016), increased niche packing through microhabitat choice (e.g. Pruitt and Goodnight 2014), and alternative reproductive strategies (Smith and Blumstein 2008), or foraging decisions (Sih et al. 2004).

Individual personality results from genetic and experiential components (Brown et al. 2007; Edenbrow et al. 2017) leading to consistent, constrained behavioral responses, and this can cause non-adaptive individual behavior in certain situations. For example, high aggression could lead to successfully obtaining food or defending a territory, but could also result in failure to maintain a mate relationship (Johnson and Sih 2005). Additionally, individuals with a strong propensity to take risks (boldness) might be more likely to exploit a new foraging patch or resource, but might also be more likely to be in an exposed position and attacked by a predator (Bell 2005; Smith and Blumstein 2008). To explain how this occasionally maladaptive inflexibility evolved, researchers believe that constrained individual behaviors might be particularly beneficial to species that live in social groups (Bergmüller and Taborsky 2010; Wolf et al. 2011). If the behavior of a group mate is consistent over time and across different contexts then the result of future interactions can be predicted based on past experience with that individual (Dall et al. 2004).

The distribution of personality types in groups can affect fitness. For example, guppies put into mixed bold-shy groups in the lab experienced higher rates of feeding than fish in only bold, or only shy shoals (Dyer et al. 2009). Conflicting personality types within groups can have large negative effects on the fitness of the whole group in addition to the individuals involved in the conflict (Oliveira et al. 2001; Sapolsky and Share 2004; Wascher et al. 2008). Consequently, social feedback should exert strong pressure on the distribution of personality types within a group, causing individuals to assort non-randomly based on personality. There are two opposing hypotheses that describe the optimal assortment of personality types in group-living species. The *social niche specialization hypothesis* (hereafter SNS) states that systems in which individuals maintain relatively partitioned and stable social roles might show reduced conflict over resources such as food and reproductive opportunities (Bergmüller and Taborsky 2010; Montiglio et al. 2013). Alternatively, the *conformity hypothesis* (hereafter CON) states that groups will be composed of individuals with similar personality types (positive assortment). This could occur when individuals assort with conspecifics of similar personality type (Aplin et al. 2013; Both et al. 2005), or when individuals conform their behaviors to that of group mates (King et al. 2014).

A key prediction of both hypotheses is that the repeated interactions with group mates will lead to maintenance of assortment via selective immigration and emigration, or social learning of behaviors during development. As such, we would expect to find the best support for either of these hypotheses in animal groups that exhibit stable compositions, long-term interactions, and low levels of immigration or emigration. Furthermore, comparing two species that differ in the number of consistent social interactions experienced during development can elucidate the importance of this factor as a mechanism (Bergmüller and Taborsky 2010). Several studies have tested for SNS, but were conducted in lab environments with artificially manipulated, temporary social groups (Bierbach et al. 2017; Merten et al. 2017; Modlmeier et al. 2014). SNS was inferred from greater individual repeatability of behavior (larger between- than within-individual variance) in control groups of familiar individuals than in the manipulated unfamiliar groups. However, this experimental setup assumes short-term group dynamics approximate the natural development of groups assorted by personality. It's possible that unfamiliar groups of adults that cannot disperse would lead to individuals behaving unnaturally or erratically. In addition, individuals do not always react to a stimulus in the same way in social and asocial contexts (Koski and Burkart 2015). Therefore an important step towards a better understanding of the evolution of personality is to conduct research that quantifies individual personalities, and distribution of personalities in natural populations.

We tested for non-random assortment in wild animals by conducting two assays of boldness on Mexican Jays (*Aphelocoma wollweberi*; hereafter MEJA) and California Scrub-Jays (*Aphelocoma californica*; hereafter CASJ). These two species inhabit similar scrub-oak and pine habitat, but differ in social behavior. MEJA live in stable groups year-round on an all-purpose territory. Kin and non-kin help the breeding pairs feed nestlings and defend the territory against predators and territory intruders (Brown 1994). Low levels of immigration seem to be sufficient for inbreeding avoidance such that the average relatedness within a group is low (Brown and Brown 1981). There are between 5-25 jays in overlapping generations in each group, and the group territory is maintained indefinitely with a slow turnover in group members. This system could allow for increased stability and persistence of personality assortment because social interactions with group mates over several years of development may strongly shape personality traits of young jays. If so, there could be exaggeration in the degree of the trait within groups over time (Wolf et al. 2011). In contrast, CASJ are solitary monogamous breeders. Young CASJ disperse from the natal territory shortly after independence, so development of personality is likely to occur in the absence of consistent social interactions. Additionally, CASJ mate-pairs are unlikely to be genetically closely related, so any pattern in personality assortment more likely results from active mate choice (Gabriel and Black 2012). Once a territory and mate are acquired, pairs remain together for life (given successful nesting attempts), so it is also possible that limited plasticity of adult personality traits might result in weakly assortative or disassortative pairings by personality (Edenbrow et al. 2017; Westneat et al. 2015). As such, if the number and timing of consistent social interactions are an important influencing factor above and beyond genetic factors, then the two different social systems would lead to distinct effects on individual, and group or pair personality.

The SNS and CON hypothesis yield contrasting predictions about the direction of the effect of social interactions on personality development (Fig. 1). By definition, an individual personality trait results in decreased within-individual variance in behavior for increased response consistency (Dingemanse and Dochtermann 2013; Dochtermann and Dingemanse 2013). In addition, if the SNS hypothesis is supported there will be high between-individual variance, and no difference in group average personality score across groups. In contrast, the CON hypothesis would predict low between-individual variance in personality scores among group mates, and a significant difference in group average scores across groups. In our system, to find support for either hypothesis we specifically tested: 1) if there are significant differences in average boldness among groups (MEJA) or pairs (CASJ), 2) whether there is low between-individual variance in boldness scores within groups or pairs of jays and, 3) whether the number of early social interactions plays an important mechanistic role, as evidenced by MEJA showing smaller within-individual variance than CASJ, which have few consistent social interactions.

## Methods

*Subjects:* We conducted boldness assessments on 55 adult MEJA around the Southwestern Research Station in Portal, Arizona from May – November 2015. These individuals were naturally grouped into 7 distinct flocks ranging from 6–15 jays (mean 7.86). We also assessed boldness of 35 adult CASJ on 20 stable territories in Willamette Mission State Park in Keizer, Oregon from April 2014 – November 2016. All jay subjects of both species were trained to come to a whistle for peanuts, and received the same trapping and handling procedure to apply unique color band combinations. All methods were approved by the University of Washington Institutional Animal Care and Use Committee (protocol #4064-03), and the appropriate permitting agencies (Bird Banding Lab #22802, Oregon Department of Fish and Wildlife #015-16, Oregon State Parks #007-14, US Fish and Wildlife Service #MB51894B-0, Arizona Game and Fish Department #SP697293).

We experimentally measured jay boldness with novel object approach (NOA) and flight initiation distance (FID) assessments. We chose these two types of assessments because they are commonly used measures of boldness, but they measure boldness in two different contexts (e.g. Carter et al. 2010; Jolles et al., 2013). The NOA assessment measures boldness towards a potential threat in a foraging context, while the FID assessment measures boldness in a non-foraging threat context.

*Novel object approach:* NOA assessments for each species included both control and experimental trials that varied only in the novelty of the object. During experimental trials, a yellow plastic duck, approximately four inches in length and height, was placed on the ground within three concentric layers of peanuts on the territory of each subject (Fig. 2a). During control trials, the duck was replaced with a rock of similar size and shape, so that the only novel feature was the non-random arrangement of peanuts. The first layer of peanuts was placed touching the central object such that one peanut was touching each side (front, back, left, right). The next layer of peanuts was placed 8 inches (20cm) further away, but in line with the first layer. The third and last layer of peanuts was placed an additional 8 inches away from the middle layer, 16 inches (40cm) from the central object, and in line with the other two peanuts. We used peanuts that appeared to be the same size and shape, and placement of peanuts was random so there would be no correlation between nutritional value and distance from the central object.

To maintain the novelty of the assessment, we conducted only one trial a day per territory and no more than two total experimental and control trials on each territory. Trials lasted up to 15 minutes, or until all peanuts were taken, and we randomized the order of control and experimental trials on each territory. Before each trial we called jays in to the center of the territory and checked for motivation to engage in the task by tossing 2 peanuts on the ground. If jays quickly came down to get these peanuts then we set up the assessment and began the trial.

Jays were allowed to take multiple peanuts per trial as a group. Wild MEJA typically move around the territory as a unit and will predominantly encounter novelty as a group (Brown 1963; Koski and Burkart 2015), so our setup is evolutionarily and ecologically relevant. However, it is possible that MEJA approach to take peanuts not based on boldness, but rather dictated by a dominance hierarchy and resource monopolization behaviors (McCormack et al. 2007). To account for this possible confound, we quantified the order in which jays approached and the repeatability of first approach order. Additionally, for our measure of boldness we quantified the closest approach distance in centimeters to the central object, including jays that never took a peanut. Jays that never approached to 2 meters away were excluded from the analyses (n = 3) because at greater distances we could not exclude confounding possibilities for the decision not to approach, such as lack of food motivation or distraction. Similarly, CASJ pairs were allowed to simultaneously take peanuts during the NOA assessment. We repeated the NOA assessment on 18 CASJ, 9 months after the first trial. Due to a more time-limited field season we repeated NOA trials on 16 MEJA from 3 flocks 4 – 11 weeks after the first experimental trial.

*Flight initiation distance:* We also conducted up to 7 FID assessments on each banded jay throughout the year. When a solitary, color-banded jay was encountered sitting un-obscured at a low height (between 0 and 3 meters high), and not engaged with food, the experimenter directed his or her gaze towards the jay's face (Eason et al. 2006) and walked towards the jay at a normal pace. When the jay flushed, the experimenter stopped and measured the distance between his or her location and the jay's original location (Fig. 2b). Smaller FID values indicate a bolder jay that is comfortable letting a potential predator (human) approach closer. Only jays with more than one FID measure (n = 47 MEJA, n = 28 CASJ) were included in repeatability analyses.

*Statistical analyses*: FID values and residuals were normally distributed (Shapiro-Wilk $W = 0.96$, $p = 0.10$), so we used linear regression for all models with

FID as the dependent variable. Closest approach (in cm) to the novel object or rock is a count variable, so we used Poisson regression for all NOA analyses. All models included jay ID as a random effect, and all analyses were conducted in RStudio Version 1.0.136 (RStudio Team 2015). To ensure our NOA assessment was eliciting a neophobic response, we used a generalized linear mixed effect model in the *lme4* package (Bates et al. 2015) to measure the effect of treatment (duck versus rock) on closest approach.

To verify that our methods were capturing an aspect of personality, we quantified repeatability of individual responses in both assessments across time. We only used data from birds with more than one measure, and modeled repeatability using the rpt function in the package *rptR* (Nakagawa and Schielzeth 2010). This function uses mixed-effect modeling approaches to partition variance, and calculates repeatability as the ratio of variance among individuals divided by total variance (among individual plus residual variance). The rpt function also conducts parametric bootstrapping for confidence intervals, and likelihood ratio tests for p-values for each estimate. Repeatability analyses for each boldness method in each species resulted in 4 models (2 species and 2 boldness metrics) where the dependent variables were the repeated measures of individual boldness scores.

Within each trial on each MEJA flock or CASJ pair, the closest approach distance was significantly related to the order that peanuts were taken (MEJA: *ß* = -0.29, $z = -4.6$, $p < 0.01$; CASJ: *ß* = -1.94, $z = -11.17$, $p < 0.01$). Additionally, we found evidence that the dominance hierarchy affected the order that MEJA approached the

NOA assessment because order of first approach was significantly repeatable for individual MEJA across the two trials (n = 16, R = 0.52, $p$ = 0.02), but not for CASJ (n = 18, R = 0, $p$ = 1). It's possible that jays in larger groups are more bold in their responses (Stamps and Groothuis 2010). However, group size as a group-level covariate did not improve the fit of NOA or FID models already including flock ID (NOA: $X^2$ = 0.65, $p$ = 0.42; FID: $X^2$ = 1.97, $p$ = 0.16). Therefore we included a covariate of individual order to take a peanut in all NOA repeatability analyses.

We did not find a significant effect of sex on CASJ boldness (FID: $t$ = -1.24, $p$ = 0.23; NOA: $t$ = -1.70, $p$ = 0.10). Unfortunately, it was logistically impossible to determine sex in all MEJA individuals during the time of our field season because only a few jays per group build nests each breeding season, and there are no other behavioral signals of sex (Brown 1994). As such, we did not include fixed effects of sex in our models. Previous research has not shown a consistent effect of sex on animal personality (Koski and Burkart 2015; Titulaer et al. 2012; van Horik et al. 2017), and there is low sexual dimorphism in these jay species so it is unlikely we'd see consistent sex differences in boldness. Additionally, all data from hatch year birds were excluded from analyses because of the possibility that personality traits are relatively more plastic at this age (Stamps and Groothuis 2010).

To test whether our methods measured the same aspect of boldness across contexts, we compared the two types of boldness scores to each other using Pearson's correlation coefficient. In this analysis we treated each individual as a single observation, and calculated the average of the repeated boldness measures recorded on a single bird. We determined significance using the cor.test function in the *stats* base package.

To assess how boldness scores are distributed within groups (MEJA flocks and CASJ pairs) we again used the rpt function to build on our repeatability models by including group ID as an additional random effect. In this way we were able to partition variance in boldness into among group, among individuals within a group, and residual variance in order to test for support of the SNS or CON hypotheses, and to determine whether number of consistent social interactions relates to within-individual variance. Here we only used data from birds with more than one data point *and* we also only used CASJ pairs with data on both birds (n = 18 jays). We used the p-values and confidence intervals output by the rpt function to determine the repeatability of each random effect in these models. We also determined significance of the random group effect from Chi-squared values of likelihood ratio tests that compared the change in log-likelihood of a model with the random group effect to one without it.

## Results

*Novel object approach*: Our NOA assessment elicited a neophobic response. Individuals of both species had significantly smaller closest approach distances (MEJA $ß = 2.03$, $z = 33.6$, $p < 0.01$; CASJ $ß = 3.09$, $z = 20.4$, $p < 0.01$), indicating more bold responses, during the control trial when the rock was in the middle (MEJA mean ± se = 7.4 ± 1.82cm; CASJ = 1.73 ± 0.73cm) than in the experimental trial with the duck (MEJA n = 47, 56.34 ± 11.02cm; CASJ n = 35, 37.94 ± 11.1cm). This, combined with consistently jumpy and nervous behavior in the presence of the duck but not the rock (personal observation), confirms that the duck stimulus was aversive and jays classified it as a threat (Greggor et al. 2015). After including order of approach as a fixed effect to account for that potential confound, we found that NOA distance was repeatable for CASJ (n = 18, R = 0.59, $p = 0.01$) and for MEJA (n = 16, R = 0.49, $p = 0.02$).

*Flight-initiation-distance*: The number of FIDs collected per jay did not differ by species. We collected between 2 – 7 FIDs for MEJA (n = 45, mean (± se) = 3.39 ± 0.18) and 2 – 6 for CASJ (n = 28, mean = 3.11 ± 0.2; $z = 0.5$, $p = 0.62$). Surprisingly, when comparing average individual scores from the two boldness methods, our results suggest that the methods were not measuring the same aspect of personality. There was no significant correlation between FID and NOA for either species (MEJA n = 35, $r = -0.001$, $p = 1$; CASJ n = 28, $r = 0.24$, $p = 0.22$).

FID score did not change in relation to breeding season. We classified the breeding season as March – July, and non-breeding season as August – February. Looking only at jays with FIDs from both seasons, we found no difference in FID scores by breeding season (t-test: MEJA n = 18, $t = 1.23$, $p = 0.22$; CASJ n = 17, t = 0.77, $p = 0.44$), so FID data from both seasons were combined for each jay. Repeatability of CASJ FID was extremely low and close to zero (mean = 21.47 ± 0.74m; R = 0.01, $p = 0.13$; Fig. 3), indicating this may be an unsuitable paradigm for this particular species. In contrast, MEJA FID scores were highly repeatable (mean = 11.12 ± 0.63m; R = 0.47, $p < 0.001$).

*Boldness distribution:* We found support for the CON hypothesis over the SNS hypothesis. There was a significant group effect leading to group repeatability and clustering of scores within groups for both MEJA FID (R = 0.38; $\chi^2 = 16.5$, $p < 0.001$) and MEJA NOA (R = 0.57; $\chi^2 = 5.4$, $p = 0.02$) boldness scores. There was no significant group effect for either method in CASJ (FID: R = 0; $\chi^2 = 0$, $p = 0.5$; NOA: R = 0; $\chi^2 = 0$, $p = 0.5$). This indicates that MEJA overlap in individual boldness scores within a group, and groups differ in average boldness (Fig. 4). If frequent social interactions with consistent partners have an impact on personality development, then social species will have smaller within-individual variance components (more repeatable traits) than asocial species (Bergmüller and Taborsky 2010; Webster and Ward 2011). However we did not find a significant difference in species NOA within-individual variance because all confidence intervals overlap (Fig. 3), but species did significantly differ in the FID within-individual component. Due to the unrepeatability of CASJ FID resulting from the extremely large within-individual variance component, this result should be interpreted with caution.

## Discussion

In this experiment we are among the first to quantify the distribution of personality traits among and within groups of wild animals. These results indicate that the higher frequency of repeated social interactions in MEJA did not lead to partitioning and specialization of degree of boldness, but instead MEJA show evidence for conformity. In contrast, the distribution of CASJ mate-pair boldness scores shows no support for either hypothesis. Our conclusions are based on data from only two species, but this information adds to our understanding of how inflexible behavioral phenotypes could be favored through evolutionary time for group-living species. Lastly, our study extends previous research describing animal personality because our subjects were wild, free-flying individuals so it is likely the results from the natural behaviors we quantified are more ecologically relevant than lab-based studies of boldness (Webster and Ward 2011).

We found that both methods to measure boldness produced consistent, repeatable results in MEJA, but only the NOA assessment produced repeatable results in CASJ. FID was not repeatable for CASJ, potentially because the open and flat park environment introduced more variance in starting distance. Experimenters could sometimes see CASJ from further away than what was usual in MEJA, and were therefore walking towards jays for longer. One way to control for this in future research might be to quantify experimenter starting distance and number of steps taken before the animal moves, as well as the distance to the animal's original location after it moves. Additionally, species differences in FID repeatability might arise if individual behavioral consistency is more important in MEJA; because they evolved in larger groups with more coordinated behaviors, the need to predict behavioral responses of others may be higher (Wolf et al. 2011).

NOA and FID were not correlated, suggesting that these are not equivalent measures of boldness. NOA could potentially be a measure of exploration (Reale et al. 2007), or FID could be measuring activity level (Carter et al. 2012). These results indicate that future studies should take great care in choosing methods for assessing personality traits of interest, and emphasizes the need for researchers to use multiple different methods as well as ecological correlates of personality (territory size, daily activity patterns) to validate their measures (Carter et al. 2013).

There was a significant effect of MEJA group membership on boldness scores, and this indicates some level of conformity where group mates' boldness scores are similarly clustered around a specific flock mean (Fig. 4). In group living, cooperatively breeding species like MEJA, conformity and predictability of behavior could be adaptive for increasing and maintaining altruistic cooperation. The evolution of cooperation among unrelated individuals has been hypothesized to occur in species that exhibit long-term interactions, consistent individual variation in behavior, and the ability to detect and reject cheaters (McNamara et al. 2008). In species that exhibit conformity, there would be little need to develop costly mechanisms for detecting cheaters (Riolo et al. 2001). As such, personality conformity should facilitate cooperation in non-kin group living species, an effect we also see in cooperatively breeding common marmosets (Koski and Burkart 2015), as well as cooperative dyadic interactions of guppies (Croft et al. 2009), and chimpanzees (Massen and Koski 2014). Due to the evolutionary distance between fish, mammalian and avian taxa, this could indicate a general trend towards personality conformity in altruistic cooperative social groups or pairs.

We found no evidence that CASJ are specializing or conforming in their boldness with their mate (see Supplementary material). Due to the unrepeatability of CASJ FID we cannot assume we accurately measured personality with this metric, and so do not interpret these results with respect to our hypotheses. However, results from the CASJ NOA assessment can be interpreted, and the distribution of scores appears random; there is no significant group effect as expected under conformity, and the within-individual variance component is not significantly different than the between-individual variance, contrary to predictions under SNS. Unlike in other systems that exhibit personality conformity related to dyad-level cooperation (Croft et al. 2009; Massen and Koski 2014), the cooperation evident in CASJ is not altruistic. Behavioral coordination to raise young and defend the territory yields selfish, as well as cooperative benefits. Therefore, selection for consistent and similar boldness traits may be less strong (McNamara et al. 2008). Nevertheless, mate choice in other avian systems does show evidence for non-random positive assortment (Both et al. 2005; Gabriel and Black 2012) or negative assortment (Houtman and Falls 1994) by personality. Consequently, it is likely that choice of mate in our CASJ population occurs more opportunistically in response to a vacancy in a breeding position (Curry et al. 2002), or could be related to factors such as territory quality or availability, and physical appearance (Overeem et al. 2014).

Multiple mechanisms could explain the assortment of personality scores in MEJA. First, previous research has found that personality traits can be heritable (van Oers et al. 2005). Since young MEJA delay dispersal, it’s possible that groups are comprised of highly related individuals that share the genetic predisposition for degree of boldness. However, early research on this population found that relatedness within groups was low due to the presence of unrelated immigrant breeders, and ranged from 0.02 to 0.22 (Brown and Brown 1981). Similarly, the level of emigration and immigration seems to be too low to create a high level of conformity if individuals are directing dispersal towards groups with similar phenotypes. Variability in presence of certain ecological components across territories could also shape group-level personality (Bell 2005). At the time of the study the territories of all of the 7 flocks were contained within an approximately 3.5 km square area. As such, it is unlikely that the flocks experienced ecological differences between their territories large enough to shape distinct flock personalities. Therefore, the most parsimonious explanation is that the within-group conformity likely occurs through social learning and/or social facilitation of behavior (Edenbrow et al. 2017) during juvenile development. Boldness, or propensity to take risks, could be a trait that is particularly susceptible to a mechanism of learning so that individuals are optimally adapted to the particular risks in their local environment. For example, previous research across taxa has shown that observation of conspecific mobbing behavior leads to social learning of novel threats (Cook et al. 1985; Curio et al. 1978; Mathis et al. 1996). A coordinated mobbing response to a visible predator is important to successfully repel it, so it would be adaptive for young jays to learn similar behavioral responses towards novel and potentially threating objects from group members (Croft et al. 2009; Edenbrow et al. 2017).

Our study improves on previous research on the evolution of personality because we measured boldness in wild, free-flying animals, and compared results across closely related species with differing degrees of consistent social interactions. Future research should include additional assays to investigate the relationship between food-related and predator-related response to risk. Furthermore, experimental manipulation and longitudinal tracking of individuals would elucidate the relationship between development of individual and group-level variation in personality traits, as well as whether certain group average levels of boldness relate to higher fitness.

## Figures

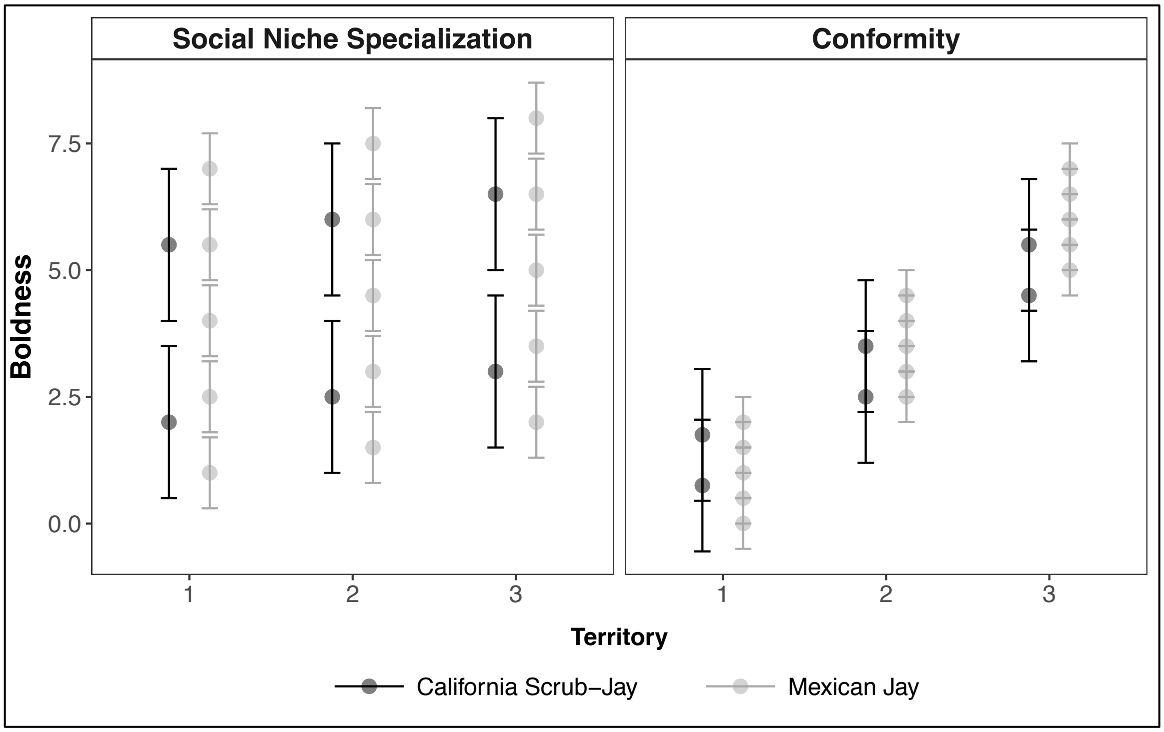

Figure 1: Predicted individual boldness estimates (points) and residual variance (whiskers) under each hypothesis.  Under the conformity hypothesis there would be a significant group effect and low between-individual variance.  Under social niche specialization there would be no between-group differences, and high between-individual variance.  For either hypothesis, species would differ in that California Scrub-Jays will have greater residual variance than Mexican Jays if social interactions are important for development of consistent personality traits.

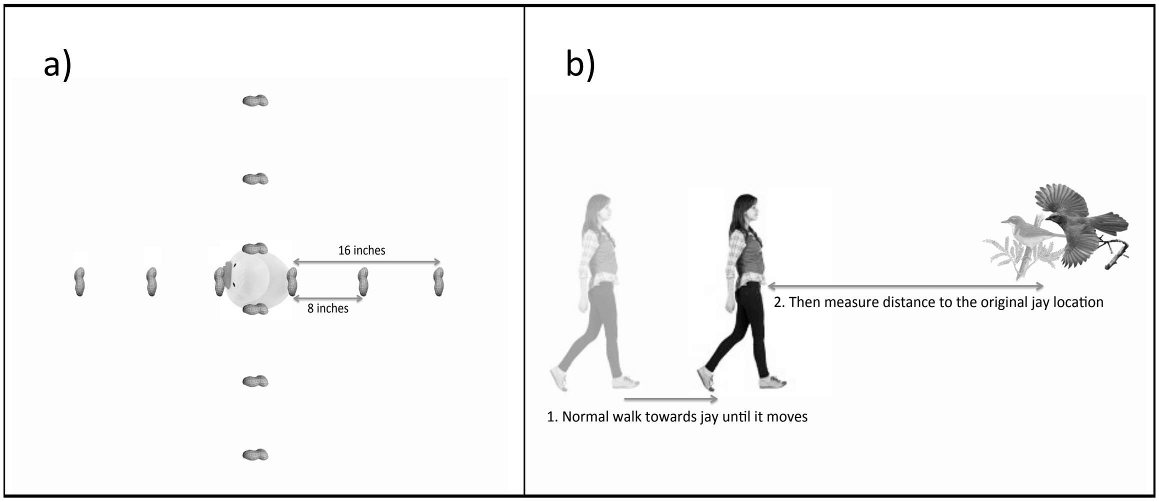

Figure 2: Visual depiction of the methods for measuring boldness: a) Novel object approach, bolder jays take peanuts from closer to the novel object (i.e. the rubber duck), and b) Flight initiation distance, jays that allow a human to walk closer before flushing are more bold.

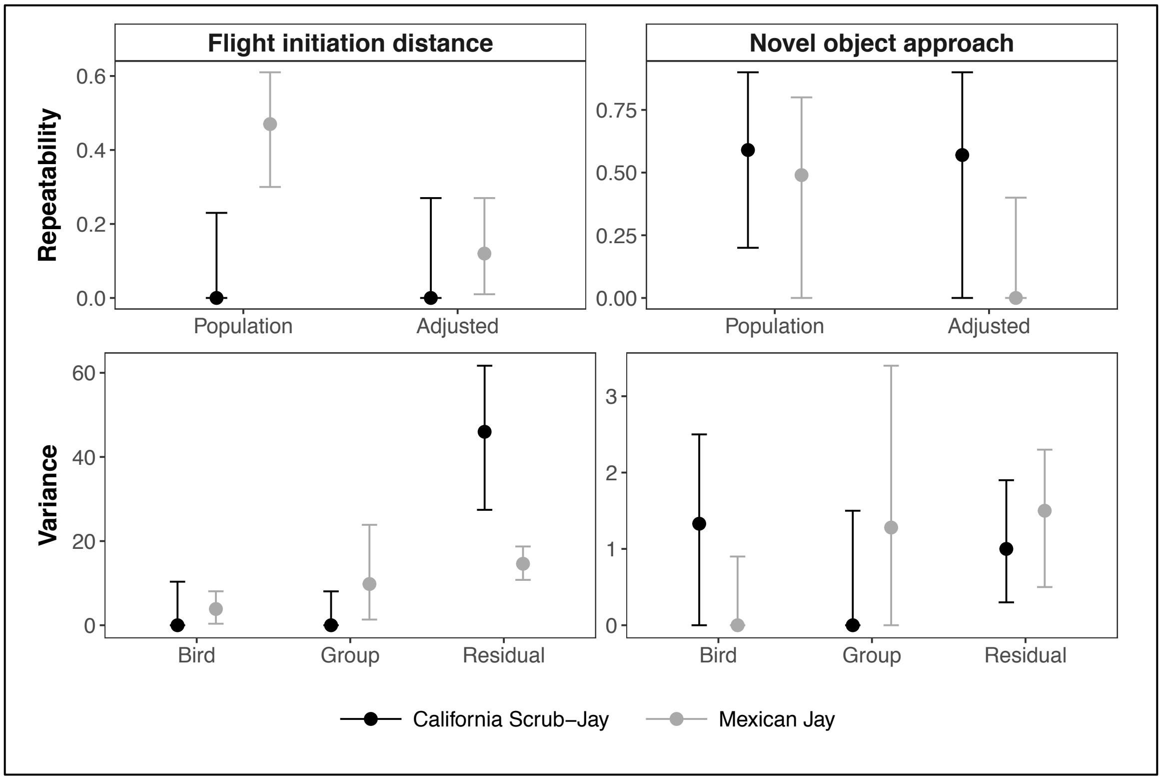

Figure 3: The top row of plots shows repeatability and 95% confidence intervals of each boldness method for each species (Population), and the change in repeatability after the group effect is included (Adjusted). The bottom row of plots shows the partitioning of variance components with 95% confidence intervals for the repeatability models that include the group effect. Mexican Jay repeatability drops after group is included because between-individual variance is now accounted for by group membership (support for Conformity hypothesis).

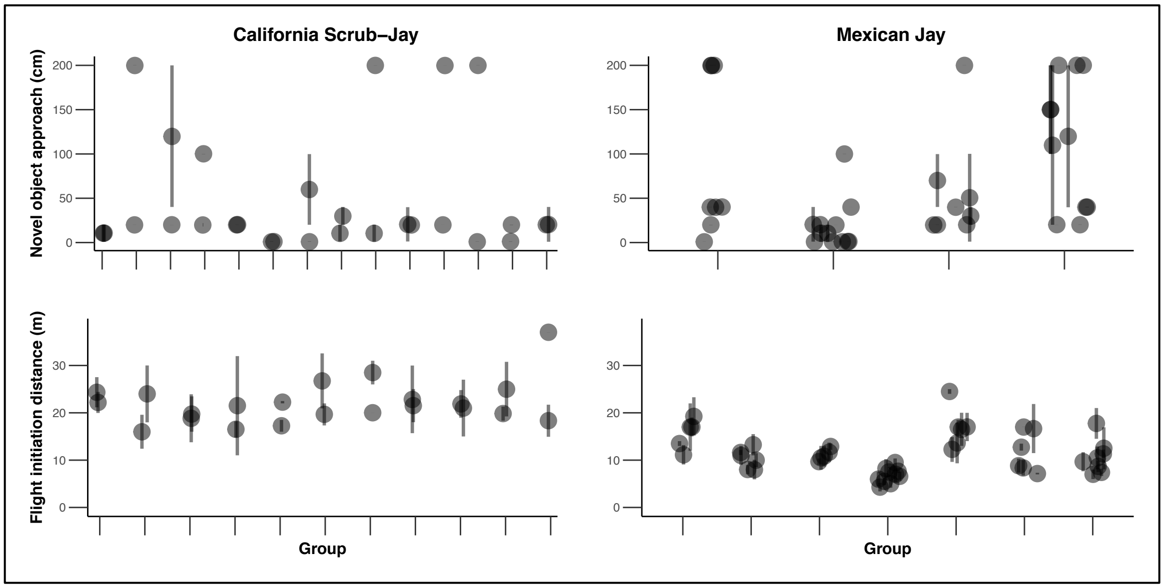

Figure 4: Scatterplot of jay boldness scores by group. This figure illustrates the clustering of boldness around a group-level mean in Mexican Jays, but not California Scrub-Jays. Dots represent the average score of an individual, and darker areas indicate multiple individuals overlapping in scores. Vertical lines through dots represent standard error of the mean for jays with multiple measures. Jays with only one measure are present in the novel object plots as circles with no vertical lines.

## Supplementary material

It's possible that we saw group-level conformity in boldness scores for MEJA but not CASJ because in the former species, cooperation occurs at the group-level. MEJA defend the territory, and communicate predator threats as a group (Brown 1994). Additionally, reproductive cooperation occurs at the group-level. During the breeding season, several jays in a flock will form relatively stable social mate-pairs, where they build a nest together and the male mate-guards the female. After the eggs hatch, all jays in the flock feed the nestlings and it becomes impossible to behaviorally identify which pair belongs to the nest (Brown 1972; Li and Brown 2002). MEJA are not genetically monogamous within a breeding season and often switch social mates each season, so while flock composition is stable across years mate-pairs are not. In contrast, long-term cooperation for reproduction, predator vigilance, and territory defense in CASJ all occurs at the level of the mate-pair. Other systems that exhibit dyad-level cooperation also show dyadic personality assortment (i.e. chimpanzees, Massen and Koski 2014; and guppies, Croft et al. 2009). Therefore, we took additional steps to further investigate for CASJ boldness assortment relative to the species differences in the level at which cooperation occurs. First, we compared the difference in CASJ mate-pair boldness scores to the difference in known MEJA social mate-pair boldness scores. Additionally we compared CASJ mate-pairs to MEJA boldness scores at the group level. If CASJ mate-pairs require the same level of cooperation as MEJA groups, than we would expect that the within-pair variability in boldness of CASJ would be the same as within-group variability in MEJA boldness, and greater than that observed in known MEJA social mate-pairs. Secondly, if positive or negative personality assortment is beneficial for CASJ cooperation, than the difference in observed pair boldness values should be significantly smaller or larger than that expected if jays were paired randomly with respect to boldness.

Finally, previous research in MEJA has found typical breeding pairs within groups to be comprised of at least one immigrant, such that relatedness in the flock is not different from the population average (Brown and Brown 1981). However, one alternative hypothesis to explain our findings of group-level conformity is that the genetically unrelated mate-pairs are the least similar in personality, and related group-mates inherited boldness scores that cluster around the mate-pair scores. To test for this, we also conducted randomization tests of MEJA pair boldness scores. If the difference in observed pair boldness scores is significantly higher than expected from random pairing within the flock, then the hypothesis is supported and our results are likely confounded by trait heritability.

**Methods:** To identify MEJA social mate-pairs, we noted nest building behaviors, copulation events, or mate-feeding behaviors while the female was incubating. In this way we identified 8 MEJA pairs from 4 flocks. We were unable to conduct NOA assessments in 2 of those flocks resulting in data from only 3 MEJA pairs, but we did have FID data from members of all 8 MEJA pairs. In CASJ we had NOA data for 13 pairs, and FID data for 11 pairs. For each pair, and for MEJA groups we calculated the observed coefficient of variation (hereafter CV), calculated as the standard deviation of boldness scores divided by the mean of boldness scores

(Whitehead 2008). We used Mann-Whitney tests to compare the CV between species because data were not normally distributed.

We also conducted randomization tests to determine assortment by boldness of mate-pairs within species. For CASJ, we compared the observed CV of pair difference scores (absolute value of female minus male boldness scores) to that resulting if males and females are randomized *among* territories. This randomization allowed us to test whether the observed mate-pair difference scores were smaller or larger than expected if jays were pairing randomly with respect to boldness. For MEJA, we compared observed CV of pair difference scores to that resulting if we randomized identity of mate-pairs *within* flocks. We determined significance as alpha less than or equal to 0.05 by quantifying the proportion of CV scores from randomized data that are more extreme than our observed value.

**Results:** We found no significant species differences (Fig. 5) when we compared mate-pair difference score CV for NOA ($W$ = 18.5, $p$ = 0.9, 95% confidence interval (CI) = -1.33, 1.16), or FID ($W$ = 40.5, $p$ = 0.80, CI = -0.18, 0.14). Additionally, the CV of CASJ mate-pair boldness was not different from the CV of boldness scores of MEJA groups (NOA: $W$ = 19, $p$ = 0.46, CI = -1.06, 0.32; FID: $W$ = 25, $p$ = 0.13, CI = -0.18, 0.04).

The difference in pair boldness scores in either species was not significantly different from scores of randomly paired conspecific jays. The CV for CASJ pair boldness did not differ from random (FID $p$ = 0.12, NOA $p$ = 0.18; Fig. 6). These results provide no support for either the SNS or CON hypotheses, but indicate that CASJ likely acquire mates opportunistically when territory vacancies occur.

Similarly, the CV for the difference in boldness of MEJA known mate-pairs did not differ from those resulting if jays are randomly paired within the flock (FID $p$ = 0.28, NOA $p$ = 0.29; Fig. 6). This demonstrates that the observed boldness conformity within MEJA groups is not due to kin-structured grouping of boldness scores around maximally different breeding pairs.

Due to the small NOA sample sizes (n = 3 MEJA pairs), the fact that CASJ FID is not a repeatable measure, and the ambiguous results of our species comparisons it is hard to reject the null hypothesis that CASJ pairs are forming randomly with respect to boldness. MEJA are likely not a system in which we should expect positive or negative assortment in mate-pair scores because jays are not genetically monogamous within a breeding season, or socially monogamous across breeding seasons. As such, it is not valid to conclude anything further from these data about species differences in mate-pair boldness scores. Further longitudinal research is necessary to better understand how MEJA social mate pairs form each season, and what traits, if any, CASJ attend to during mate choice.

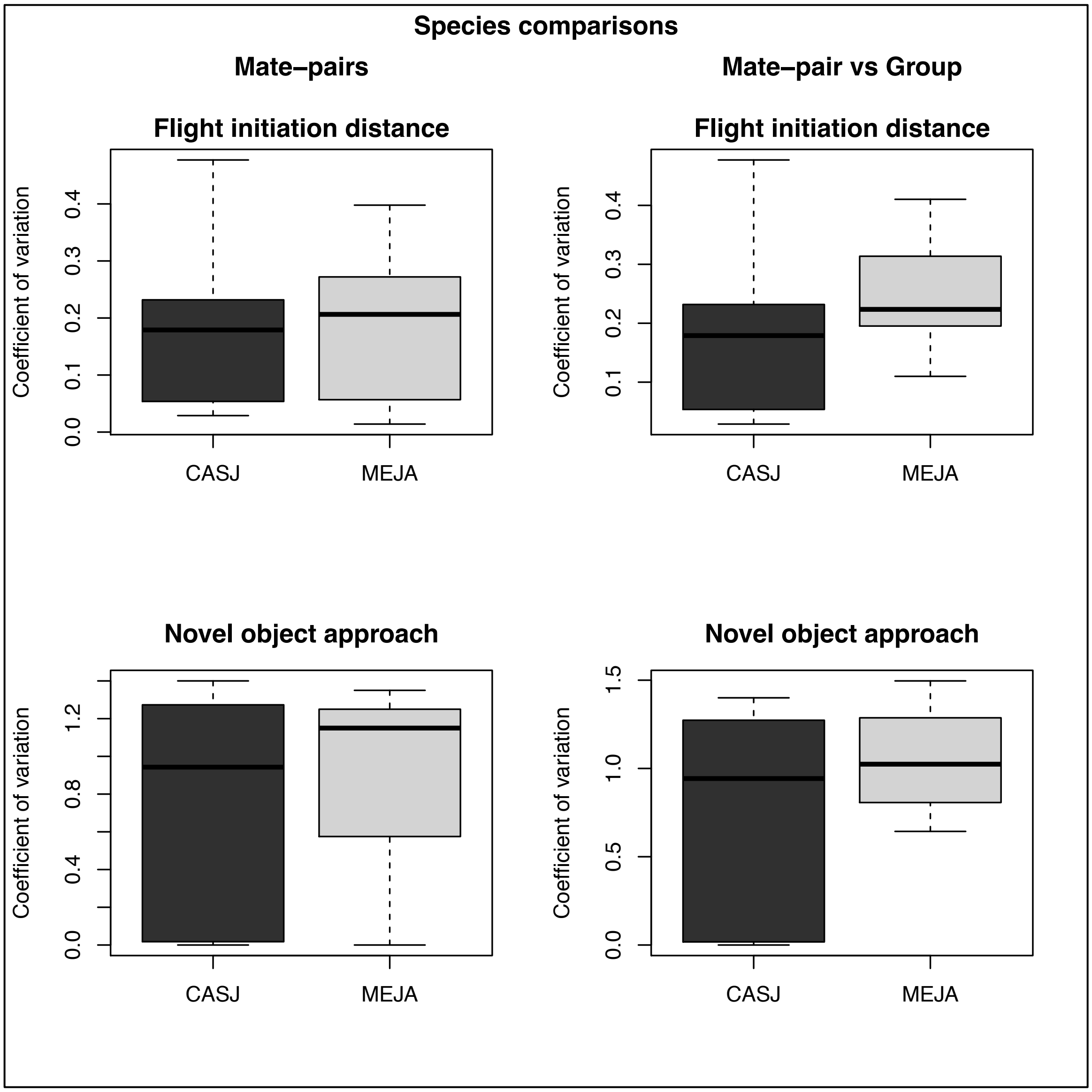

Figure 5: Comparisons of the coefficient of variation (CV) of boldness scores of California Scrub-Jay (CASJ; dark gray) mate-pairs to Mexican Jay (MEJA; light gray) mate-pairs on the left for both novel object approach and flight initiation distance. On the right, boxplots show the comparison of the CV of CASJ mate-pair boldness scores to the CV of boldness within MEJA groups for each measure.

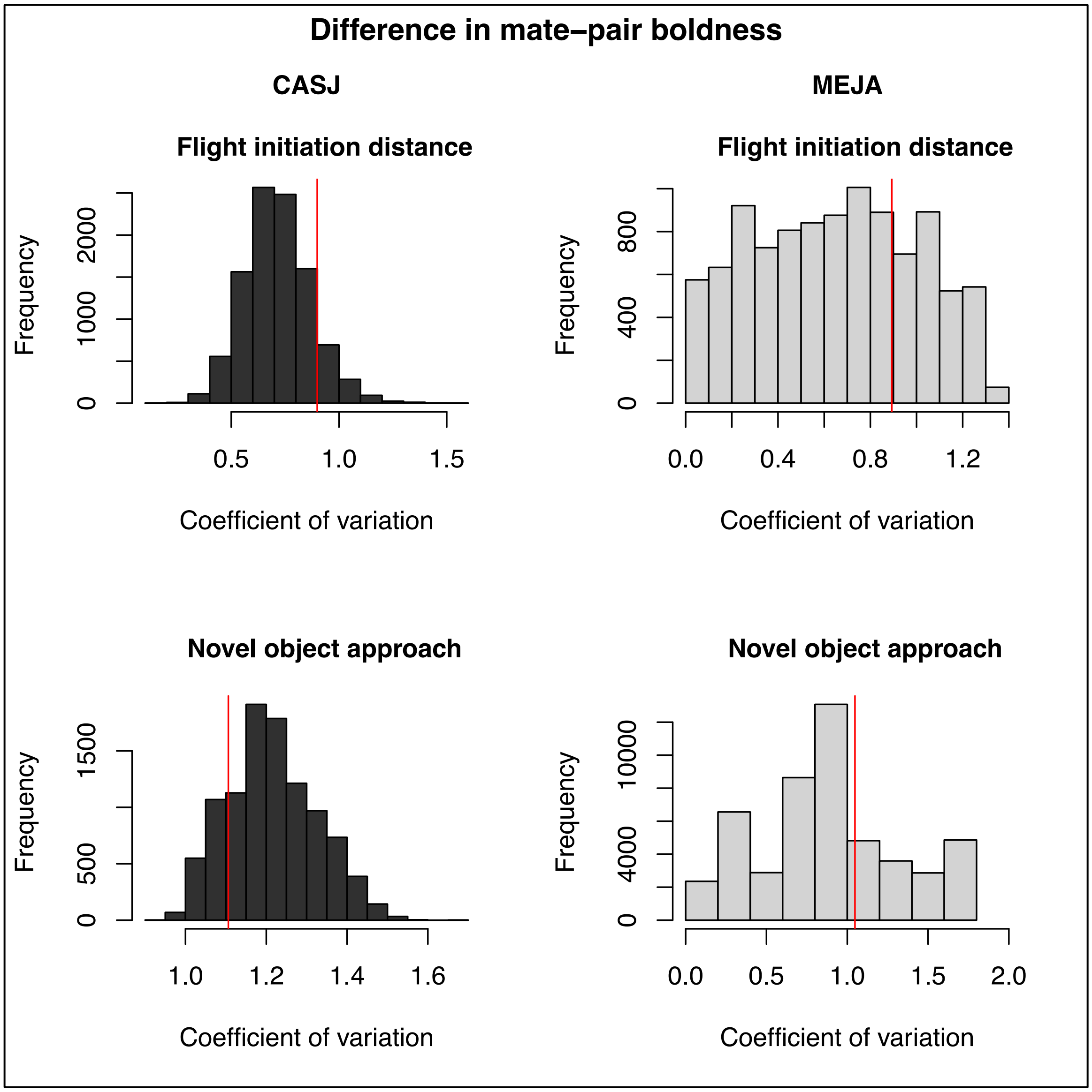

Figure 6: Results from the randomization tests. We calculated the coefficient of variation for observed differences (red vertical lines) in boldness scores of California Scrub-Jay (CASJ; dark grey) and Mexican Jay (MEJA; light grey) mate-pairs to those resulting from 10,000 permutations of our data. For both flight initiation distance and novel object approach, the difference between mate-pairs in boldness scores was neither larger, nor smaller than expected if jays paired randomly.

## Literature cited

Aplin LM, Farine DR, Morand-Ferron J, Cole EF, Cockburn A, Sheldon BC. 2013. Individual personalities predict social behaviour in wild networks of great tits (Parus major). Ecol. Lett. 16:1365–72; doi:10.1111/ele.12181.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J. of Stat. Soft. 67(1):1-48.

Bell AM. 2005. Behavioural differences between individuals and two populations of stickleback (Gasterosteus aculeatus). J. Evol. Biol. 18:464–473; doi:10.1111/j.1420-9101.2004.00817.x.

Bergmüller R, Taborsky M. 2010. Animal personality due to social niche specialisation. Trends Ecol. Evol. 25:504–511; doi:10.1016/j.tree.2010.06.012.

Bierbach D, Laskowski KL, Wolf M. 2017. Behavioural individuality in clonal fish arises despite near-identical rearing conditions. Nat. Commun. 8:1–7; doi:10.1038/ncomms15361.

Both C, Dingemanse NJ, Drent PJ, Tinbergen JM. 2005. Pairs of extreme avian personalities have highest reproductive success. J. Anim. Ecol. 74:667–674; doi:10.1111/j.1365-2656.2005.00962.x.

Brown C, Burgess F, Braithwaite VA. 2007. Heritable and experiential effects on boldness in a tropical poeciliid. Behav. Ecol. Sociobiol. 62:237–243; doi:10.1007/s00265-007-0458-3.

Brown JL. 1972. Communal feeding of nestlings in the Mexican jay (Aphelocoma ultramarina): Interflock comparisons. Anim. Behav. 20: 395–403.

Brown JL. 1994. Mexican Jay (Aphelocoma ultramarina). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Brown JL. 1963. Social organization and behavior of the Mexican Jay. Condor 65: 126–153.

Brown JL, Brown ER. 1981. Extended family system in a communal bird. Science (80-. ). 211: 959–960.

Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2013. Animal personality: What are behavioural ecologists measuring? Biol. Rev. 88:465–475; doi:10.1111/brv.12007.

Carter AJ, Goldizen AW, Tromp S a. 2010. Agamas exhibit behavioral syndromes: Bolder males bask and feed more but may suffer higher predation. Behav. Ecol. 21:655–661; doi:10.1093/beheco/arq036.

Carter AJ, Marshall HH, Heinsohn R, Cowlishaw G. 2012. How not to measure boldness: Novel object and antipredator responses are not the same in wild baboons. Anim. Behav. 84:603–609; doi:10.1016/j.anbehav.2012.06.015.

Cook M, Mineka S, Wolkenstein B, Laitsch K. 1985. Observational conditioning of snake fear in unrelated Rhesus monkeys. J. Abnorm. Psychol. 94: 591–610.

Croft DP, Krause J, Darden SK, Ramnarine IW, Faria JJ, James R. 2009. Behavioural trait assortment in a social network: Patterns and implications. Behav. Ecol. Sociobiol. 63:1495–1503; doi:10.1007/s00265-009-0802-x.

Curio E, Ernst U, Vieth W. 1978. Cultural transmission of enemy recognition: One function of mobbing. Science (80-. ). 202: 899–901.

Curry RL, Peterson AT, Langen TA. 2002. Western Scrub Jay (Aphelocoma californica). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Dall SRX, Houston AI, McNamara JM. 2004. The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. Ecol. Lett. 7:734–739; doi:10.1111/j.1461-0248.2004.00618.x.

Dingemanse NJ, Dochtermann NA. 2013. Quantifying individual variation in behaviour: Mixed-effect modelling approaches. J. Anim. Ecol. 82:39–54; doi:10.1111/1365-2656.12013.

Dochtermann NA, Dingemanse NJ. 2013. Behavioral syndromes as evolutionary constraints. Behav. Ecol. 24:806–811; doi:10.1093/beheco/art002.

Dyer JRG, Croft DP, Morrell LJ, Krause J. 2009. Shoal composition determines foraging success in the guppy. Behav. Ecol. 20:165–171; doi:10.1093/beheco/arn129.

Eason PK, Sherman PT, Rankin O, Coleman B. 2006. Factors affecting flight initiation distance in American robins. J. Wildl. Manage. 70:1796–1800; doi:10.2193/0022-541X(2006)70[1796:FAFIDI]2.0.CO;2.

Edenbrow M, Bleakley BH, Darden SK, Tyler CR, Ramnarine IW, Croft DP. 2017. The Evolution of Cooperation: Interacting Phenotypes among Social Partners. Am. Nat. 189:630–643; doi:10.1086/691386.

Gabriel PO, Black JM. 2012. Behavioural syndromes, partner compatibility and reproductive performance in Steller's Jays. Ethology 118:76–86; doi:10.1111/j.1439-0310.2011.01990.x.

Gosling SD. 2001. From mice to men: What can we learn about personality from animal research? Psychol. Bull. 127: 45–86.

Greggor AL, Thornton A, Clayton NS. 2015. Neophobia is not only avoidance: Improving neophobia tests by combining cognition and ecology. Curr. Opin. Behav. Sci. 6:82–89; doi:10.1016/j.cobeha.2015.10.007.

Houtman AM, Falls JB. 1994. Negative assortative mating in the white-throated sparrow, Zonotrichia albicollis: The role of mate choice and intra-sexual competition. Anim. Behav. 48: 377–383.

Johnson JC, Sih A. 2005. Precopulatory sexual cannibalism in fishing spiders (Dolomedes triton): A role for behavioral syndromes. Behav. Ecol. Sociobiol. 58:390–396; doi:10.1007/s00265-005-0943-5.

Jolles JW, Ostojić L, Clayton NS. 2013. Dominance, pair bonds and boldness determine social-foraging tactics in rooks, Corvus frugilegus. Anim. Behav. 85:1261–1269; doi:10.1016/j.anbehav.2013.03.013.

King AJ, Williams LJ, Mettke-hofmann C. 2014. The effects of social conformity on Gouldian finch personality. Anim. Behav. 99:25–31; doi:10.1016/j.anbehav.2014.10.016.

Koski SE, Burkart JM. 2015. Common marmosets show social plasticity and group-level similarity in personality. Nature 5:1–7; doi:10.1038/srep08878.

Li S-H, Brown JL. 2002. Reduction of maternal care: a new benefit of multiple mating? Behav. Ecol. 13: 87–93.

Massen JJM, Koski SE. 2014. Chimps of a feather sit together: Chimpanzee friendships are based on homophily in personality. Evol. Hum. Behav. 35:1–8;

doi:10.1016/j.evolhumbehav.2013.08.008.

Mathis A, Chivers DP, Smith RJF. 1996. Cultural transmission of predator recognition in fishes: Intraspecific and interspecific learning. Anim. Behav. 51:185–201; doi:10.1006/anbe.1996.0016.

McCormack JE, Jablonski PG, Brown JL. 2007. Producer-Scrounger Roles and Joining Based on Dominance in a Free-Living Group of Mexican Jays ( Aphelocoma ultramarina ). Behaviour 144: 967–982.

McNamara JM, Barta Z, Fromhage L, Houston AI. 2008. The coevolution of choosiness and cooperation. Nature 451:189–192; doi:10.1038/nature06455.

Merten S Von, Zwolak R, Rychlik L. 2017. Social personality: A more social shrew species exhibits stronger differences in personality types. Anim. Behav. 127:125–134; doi:10.1016/j.anbehav.2017.02.021.

Modlmeier AP, Laskowski KL, Demarco AE, Coleman A, Zhao K, Brittingham HA, et al. 2014. Persistent social interactions beget more pronounced personalities in a desert-dwelling social spider. Biol. Lett. 10:20140419; doi:10.1098/rsbl.2014.0419.

Montiglio P-O, Ferrari C, Reale D. 2013. Social niche specialization under constraints: Personality, social interactions and environmental heterogeneity. Philos. Trans. R. Soc. B Biol. Sci. 368:20120343–20120343; doi:10.1098/rstb.2012.0343.

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biol. Rev. 85:935–956; doi:10.1111/j.1469-185X.2010.00141.x.

Nicolaus M, Tinbergen JM, Ubels R, Both C, Dingemanse NJ. 2016. Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. Ecol. Lett. 19:478–486; doi:10.1111/ele.12584.

Oliveira RF, Lopes M, Carneiro LA, Canario AVM. 2001. Watching fights raises fish hormone levels. Nature 409: 2001.

Overeem KR, Gabriel PO, Zirpoli J a, Black JM. 2014. Steller sex: infidelity and sexual selection in a social Corvid (Cyanocitta stelleri). PLoS One 9:e105257; doi:10.1371/journal.pone.0105257.

Pruitt JN, Goodnight CJ. 2014. Site-specific group selection drives locally adapted group compositions. Nature 514:359–362; doi:10.1038/nature13811.

Reale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82:291–318; doi:10.1111/j.1469-185X.2007.00010.x.

Riolo RL, Cohen MD, Axelrod R. 2001. Evolution of cooperation without reciprocity. Nature 414: 441–443.

RStudio Team. 2015. RStudio: Integrated development for R. RStudio, Inc. Boston, MA.

Sapolsky RM, Share LJ. 2004. A pacific culture among wild baboons: Its emergence and transmission. PLoS Biol. 2:E106; doi:10.1371/journal.pbio.0020106.

Sih A, Bell A, Johnson JC. 2004. Behavioral syndromes: An ecological and evolutionary overview. Trends Ecol. Evol. 19:372–378; doi:10.1016/j.tree.2004.04.009.

Sloan Wilson D, Clark AB, Coleman K, Dearstyne T. 1994. Shyness and boldness in humans and other animals. Trends Ecol. Evol. 9:442–446; doi:10.1016/0169-5347(94)90134-1.

Smith BR, Blumstein DT. 2008. Fitness consequences of personality: A meta-analysis. Behav. Ecol. 19:448–455; doi:10.1093/beheco/arm144.

Stamps J, Groothuis TGG. 2010. The development of animal personality: Relevance, concepts and perspectives. Biol. Rev. 85:301–325; doi:10.1111/j.1469-185X.2009.00103.x.

Titulaer M, van Oers K, Naguib M. 2012. Personality affects learning performance in difficult tasks in a sex-dependent way. Anim. Behav. 83:723–730; doi:10.1016/j.anbehav.2011.12.020.

van Horik JO, Langley EJG, Whiteside MA, Madden JR. 2017. Differential participation in cognitive tests is driven by personality , sex , body condition and experience. Behav. Processes 134:22–30; doi:10.1016/j.beproc.2016.07.001.

van Oers K, Jong G De, Noordwijk AJ Van, Kempenaers B, Drent PJ. 2005. Contribution of genetics to the study of animal personalities: A review of case studies. Behaviour 142: 1185–1206.

Wascher CAF, Scheiber IBR, Kotrschal K. 2008. Heart rate modulation in bystanding geese watching social and non-social events. Proc. R. Soc. B Biol. Sci. 275:1653–1659; doi:10.1098/rspb.2008.0146.

Webster MM, Ward AJW. 2011. Personality and social context. Biol. Rev. 86:759–773; doi:10.1111/j.1469-185X.2010.00169.x.

Westneat DF, Wright J, Dingemanse NJ. 2015. The biology hidden inside residual within-individual phenotypic variation. Biol. Rev. 90:729–743; doi:10.1111/brv.12131.

Whitehead H. 2008. Analyzing animal societies: quantitative methods for vertebrate social analysis. Princeton University Press. Princeton, NJ.

Wolf M, Doorn GS Van, Weissing FJ. 2011. On the coevolution of social responsiveness and behavioural consistency. Proc. R. Soc. B Biol. Sci. 278:440–448; doi:10.1098/rspb.2010.1051.

## Conclusions and future directions

The research presented here has answered some important questions in the comparative cognition field. Previous studies seeking to explain the evolution of complex cognitive abilities have used broad proxy measures of sociality and cognition that are not appropriate for all taxa (Bergman and Beehner 2015). Those studies that do measure cognition from performance on experimental tasks are mainly conducted in lab environments with tightly controlled artificial conditions (Shettleworth 2010). Instead, we directly compared social complexity and cognitive abilities in two congeneric species in the wild. While there are many questions still to address, we believe this study is an important step towards critically assessing what we are measuring when designing experiments to test the SIH and related hypotheses.

First, we have shown that broad group-level definitions of complex sociality, such as group size or breeding system, do not accurately capture the quality of individual relationships. Selection for increased cognitive abilities likely occurs at the dyad level since the outcome of individual social interactions in complex groups can determine survival and fitness (Duckworth 2009; Taborsky and Oliveira 2012). Therefore in future research it will be important to account for social complexity at multiple scales (Whitehead 2008), and determine which cognitive mechanisms are likely important for navigating certain social interactions (Bergman and Beehner 2015).

Second, we found that performance on a cognitive assessment in captivity does not equate to performance of wild conspecifics. Previous research in diverse taxa have shown captive subjects outperform wild conspecifics on identical tasks (i.e. Benson-Amram et al. 2013; Bouchard et al. 2007; Gajdon et al. 2004; Webster and Lefebvre 2001). Yet here we have shown the opposite. Subjects temporarily held in captivity took significantly longer to solve a novel foraging task than wild conspecifics. Our results combined with those from previous research, indicate it is unlikely that studies of cognitive performance on subjects in the lab can be generalized to the species as a whole.

Young individuals of the social species (Mexican Jays, hereafter MEJA) experience prolonged development in their natal group (Brown 1994), a system in which we'd expect to see evidence for socio-cognitive mechanisms like social learning (Laland 2004). In contrast, even though asocial California Scrub-Jay (CASJ) adult mate-pairs have higher quality social interactions, young disperse after becoming independent (Curry et al. 2017) and so likely individually learn many behaviors. Therefore, if SIH is supported we expected MEJA to excel in social learning over CASJ. Our third experiment quantified social learning ability in these two congeneric species in the wild. We found some evidence for social facilitation on the motivation to engage with the task in MEJA. However, our results indicated that, on a complex foraging task, jays prioritize personal information over social learning from a knowledgeable conspecific. Therefore we found no support for the SIH because the species with a complex social system did not outperform the asocial species on our task.

It is possible that the foraging environment is not sufficiently complex or risky to merit the evolution of socio-cognitive mechanisms (Kendal et al. 2005).

Alternatively, it may be more likely that MEJA would socially learn appropriate behavioral responses towards threatening stimuli. Therefore we took one additional step to further examine the role of social facilitation on boldness behaviors. We found MEJA individuals within groups were similar in their responses towards two kinds of threats, and social group mean responses differed. In contrast CASJ showed no similarity in response with their mate. These results further indicate that social facilitation plays an important role in shaping behavior in this social species, and it is likely that conformity in behavior increases efficiency of cooperation (Edenbrow et al. 2017).

We did not find strong support for the SIH, but we did find evidence that MEJA attend to the behavior of group-mates. In social groups (like MEJA) where there is a dominance hierarchy, individuals would benefit from the ability to infer the rank of individuals by observing conspecific interactions (Bond et al. 2003). In fact, it's possible that the weaker quality, but greater quantity of social bonds in MEJA reflects generalized attention to all group-mates in order to attend to dominance interactions. Therefore, future research could compare MEJA and CASJ in the wild on a transitive inference task. If there is no difference in performance on this additional assessment, then we need to critically reconsider the factors important for cognitive evolution in birds.

## Literature cited

Benson-Amram S, Weldele ML, Holekamp KE. 2013. A comparison of innovative problem-solving abilities between wild and captive spotted hyenas, Crocuta crocuta. Anim. Behav. 85:349–356; doi:10.1016/j.anbehav.2012.11.003.

Bergman TJ, Beehner JC. 2015. Measuring social complexity. Anim. Behav. 103:203–209; doi:10.1016/j.anbehav.2015.02.018.

Bond AB, Kamil AC, Balda RP. 2003. Social complexity and transitive inference in corvids. Anim. Behav. 65:479–487; doi:10.1006/anbe.2003.2101.

Bouchard J, Goodyer W, Lefebvre L. 2007. Social learning and innovation are positively correlated in pigeons (Columba livia). Anim. Cogn. 10:259–66; doi:10.1007/s10071-006-0064-1.

Brown JL. 1994. Mexican Jay (Aphelocoma ultramarina). In: *The Birds of North America* (A. Poole and F. Gill, eds). The Birds of North America, Inc.:Philadelphia, PA.

Curry RL, Peterson AT, Langen TA, Pyle P, Patten MA. 2017. California Scrub-Jay (*Aphelocoma californica*), version 3.0. *In* The Birds of North American (PG Rodewald, Ed). Cornell Lab of Ornithology, Ithaca NY.

Duckworth RA. 2009. The role of behavior in evolution: A search for mechanism. Evol. Ecol. 23:513–531; doi:10.1007/s10682-008-9252-6.

Edenbrow M, Bleakley BH, Darden SK, Tyler CR, Ramnarine IW, Croft DP. 2017. The Evolution of Cooperation: Interacting Phenotypes among Social Partners. Am. Nat. 189:630–643; doi:10.1086/691386.

Gajdon GK, Fijn N, Huber L. 2004. Testing social learning in a wild mountain parrot, the kea (Nestor notabilis). Learn. Behav. 32: 62–71.

Kendal RL, Coolen I, Bergen Y Van, Laland KN. 2005. Trade-offs in the aadaptive use of social and asocial learning. Adv. Study Behav. 35:333–379; doi:10.1016/S0065-3454(05)35008-X.

Laland KN. 2004. Social learning strategies. Anim. Learn. Behav. 32:4–14; doi:10.3758/BF03196002.

Shettleworth S. 2010. Cognition, evolution and behavior. Oxford University Press, Oxford UK.

Taborsky B, Oliveira RF. 2012. Social competence: An evolutionary approach. Trends Ecol. Evol. 27:679–688; doi:10.1016/j.tree.2012.09.003.

Webster S, Lefebvre L. 2001. Problem solving and neophobia in a columbiform–passeriform assemblage in Barbados. Anim. Behav. 62:23–32; doi:10.1006/anbe.2000.1725.

Whitehead H. 2008. Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis. Chicago, IL: The University of Chicago Press.