

**Behaviour and Complex Cognition
in Kea (*Nestor notabilis*)**

Amalia Pinkusfeld Medeiros Bastos

*A thesis submitted in fulfilment of the requirements for the degree of Doctor of
Philosophy in Psychology, the University of Auckland, 2021.*

Abstract

Kea (*Nestor notabilis*) are a large-brained and highly social species of parrot adapted to living in the New Zealand alps. Both their ecology and evolutionary history suggest this species may have evolved flexible and complex cognition. Over the course of this thesis, I investigate several aspects of the physical, numerical, and social cognition of this species. First, I provide the first evidence of self-care tooling in a kea through repeated observations of a disabled individual. I then contextualise this behaviour phylogenetically across all parrot species, demonstrating how crowdsourcing can be used as a method to increase our detection of rare animal behaviours and trace their likely evolutionary history. Next, I find that, despite quickly learning to pull up baited vertical strings, kea have a poor understanding of connectivity in a horizontal loose-string task. The subsequent chapter demonstrates that kea simultaneously represent the identity and trajectory of two hidden objects and can predict the end points of incomplete object trajectories. I then build on these findings, and harness the signature-testing framework, to investigate kea's ability to judge probabilistic events. I find that kea make statistical inferences based on relative frequencies and integrate physical and social information into their predictions in a domain-general manner. My next chapter examines how naïve kea perceive virtual stimuli relative to the real world and finds that kea expect physical processes in the virtual and real worlds to be equivalent and continuous, helping validate the use of virtual stimuli to test animals' cognitive abilities. In my final two chapters, I overview how studies of avian cognition can contribute to important debates in comparative psychology and offer practical applications for how research on kea cognition can help the conservation of this endangered species.

Acknowledgements

This thesis would not have been possible without the support of the many people who helped, taught, and inspired me along the way. Firstly, I am deeply indebted to my supervisor, Dr. Alex Taylor, for seeing the potential in an enthusiastic but overly optimistic undergraduate student and putting in a phenomenal amount of time and effort to shape me into the scientist I am today. Thank you for providing me with every opportunity to learn the ropes of setting up and managing a research program, far beyond the requirements of a doctorate. Please know that all the meetings, calls, and manuscript comments over the past four years are greatly appreciated. I am also grateful to Professor Ximena Nelson, who warmly welcomed me into her lab during my time in Christchurch and always provided insightful feedback on my work. I am thankful to Professor Quentin Atkinson for his invaluable advice and collaboration on the third chapter of this thesis – the eventual manuscript will be infinitely better for it.

I am extremely grateful to all the colleagues, visiting students, and collaborators who have contributed to the work in this thesis and our many unrelated side projects. Amey, Rosalie, Billy, Jonathan, Kata, Patrick, Scott, Thanos, Martin, Lisa, Marieke, Conny, Chloe, Fred, Rebecca (and Regan), Guy, Kyle, Oliver, and Gabriela: thank you for being excellent co-workers, for the laughs, and for so often agreeing that one final board game was in order before calling it a night.

I very much appreciate the support of the amazing staff at Willowbank Wildlife Reserve. A particularly big thank you to Nick, Sam, and Kirsty for their help with, and enthusiasm for, the kea research. I am also eternally grateful to the extraordinary bird brains that so generously offered their time, patience, and half-eaten carrot pieces to make this research possible: Blofeld, Neo, Loki, Moriarty, Plankton, Cheeky, Taz, Bruce (Kati), Harley Quinn, Megatron, Joker, Mystique, Boh, Johnny, Spike, and Angelina. I am so lucky to have

so many feathered friends; I wish I could express my gratitude in a way they could understand. I also take this opportunity to thank the many other animals who have bettered my life and placed me on the path to studying animal cognition, particularly Avalon, Abott, Conrad, and Lani.

The last year or so of my studies were greatly enriched by the company and friendship of the brilliant Claudia Kaiser, who introduced me to the magical world of classical dressage. I always enjoyed her lessons and I like to think that, as a PhD herself, she also enjoyed hearing about my work. She left us much too soon, and the world is a poorer place for it. I am also incredibly grateful to Julia Lee for taking me on in Claudia's stead and helping to mend the wound with a good dose of laughter. Likewise, I thank my equine teachers, Belleza, Beata, Octavia, and Troy, for "keeping my ears forward".

Thank you to my wonderful husband, Patrick Wood, for his inordinate support throughout my studies. After so many discussions about experimental design and animal training over dinner, it was only a matter of time before you conducted some research yourself! I am incredibly fortunate that you not only joined our lab but also proved to be an exceptionally skilled technician and exemplary co-worker. I am proud but perhaps also slightly bitter that you are now undoubtedly the kea's favourite.

Finally, I am profoundly grateful to my parents, Claudia and Francisco, for their unwavering support and belief in me. I would not be where I am today without your love, encouragement, and sacrifices. I cannot thank you enough for being there for me, every step of the way. Whenever I faced an obstacle or needed advice, you listened carefully and guided me wisely. You are incredible parents.

Table of Contents

Chapter I: <i>General Introduction</i>	1
1.1. These Brains Were Made for Flying.....	1
1.2. What Makes Kea Special?	3
1.3. What Do We Know About Kea Cognition?	7
1.3.1. Physical Cognition.....	7
1.3.2. Social Cognition.....	11
1.3.3. Reasoning & Learning.....	15
1.3.4. Linking cognition and conservation.....	17
1.4. Structure of the Thesis	18
Chapter II: <i>Self-care tooling innovation in a disabled kea</i>	22
2.1. Abstract.....	22
2.2. Introduction.....	23
2.3. Materials and Methods.....	26
2.3.1. Subjects and Procedure	26
2.3.2. Video Coding & Analyses	30
2.4. Results.....	32
2.4.1. Do Bruce’s pebble manipulations co-occur with preening?	32
2.4.2. Are pebble tools valuable and effective to Bruce?	34
2.4.3. How does Bruce’s behaviour compare to that of other individuals?	36
2.5. Discussion.....	39
Chapter III: <i>Crowdsourcing and phylogenetic inference reveal the evolutionary history of a rare tool use behaviour</i>	42
3.1. Abstract.....	42
3.2. Introduction.....	43
3.3. Methods.....	46
3.3.1. Video Searches and Coding.....	46
3.3.2. Ethical Statement	47
3.3.3. Phylogenetic Data	47
3.3.4. Zero-inflated Phylogenetic Multilevel Model	48

3.3.5. Ancestral State Reconstruction	50
3.3.6. Encephalisation Data	50
3.4. Results.....	51
3.4.1. Description of Survey Findings	51
3.4.2. Phylogeny of Tool Use in Parrots.....	52
3.4.3. Identifying Additional Tool-Using Species.....	55
3.4.4. Ancestral State Reconstruction.....	57
3.4.5. Is Tool Use a Signal of Innovation?	59
3.5. Discussion.....	60
Chapter IV: <i>Kea fail a string-pulling connectivity task</i>	64
4.1. Abstract.....	64
4.2. Introduction.....	65
4.3. Methods.....	69
4.3.1. Subjects.....	69
4.3.2. Experiment 1	70
4.3.3. Experiment 2	71
4.3.4. Experiment 3.....	72
4.3.5. Video Coding and Analyses.....	72
4.4. Results.....	73
4.5. Discussion.....	78
Chapter V: <i>Kea represent object trajectory and identity</i>	82
5.1. Abstract.....	82
5.2. Introduction.....	83
5.3. Methods.....	84
5.3.1. Ethics Statement.....	84
5.3.2. Subjects.....	84
5.3.3. Materials	85
5.3.4. Training: Hand-tracking training protocol.....	86
5.3.5. Training: Food-search choice task for Experiment 3.....	87
5.3.6. Testing Procedure	90
5.3.7. Analyses.....	91
5.4. Results.....	91

5.4.1. Experiment 1	91
5.4.2. Experiment 2	95
5.4.3. Experiment 3	96
5.5. Discussion	100
Chapter VI: <i>Kea show three signatures of domain-general statistical inference</i>	103
6.1. Abstract	103
6.2. Introduction	104
6.3. Methods	107
6.3.1. Subjects and Apparatus	107
6.3.2. General Testing Procedures	108
6.3.3. Training for Experiment 1	109
6.3.4. Procedure for Experiment 1	111
6.3.5. Training for Experiment 2	113
6.3.6. Procedure for Experiment 2	114
6.3.7. Training for Experiment 3	115
6.3.8. Procedure for Experiment 3	117
6.3.9. Analyses	117
6.4. Results	118
6.4.1. Experiment 1	118
6.4.2. Experiment 2	122
6.4.3. Experiment 3	123
6.4.4. First Trial Performances	124
6.4.5. Learning Effect Analyses	124
6.4.6. Accidental Token Visibility Coding and Simulations	125
6.4.7. Clever Hans Control	128
6.5. Discussion	129
Chapter VII: <i>Are parrots naïve realists? Kea behave as if the real and virtual worlds are continuous</i>	133
7.1. Abstract	133
7.2. Introduction	134
7.3. Experiment 1	136
7.3.1. Subjects and Apparatus	136

7.3.2. Procedure	138
7.3.3. Coding and Analyses	139
7.3.4. Results.....	140
7.4. Experiment 2.....	141
7.4.1. Rationale	141
7.4.2. Subjects and Apparatus.....	141
7.4.3. Procedure	141
7.4.4. Coding and Analyses	142
7.4.5. Results.....	143
7.5. Discussion.....	144
Chapter VIII: Macphail’s null hypothesis of vertebrate intelligence: Insights from avian cognition	146
8.1. Abstract.....	146
8.2. Background.....	147
8.3. Differences in Intelligence Across Species.....	147
8.4. Intelligent Behaviour Beyond Pure Association.....	152
8.5. Echoes of Macphail’s Criticisms in the 21 st Century.....	157
8.6. Discussion.....	162
Chapter IX: Concluding Remarks	164
9.1. Summary of the Main Findings	164
9.2. Back to the Wild	170
9.3. Final Remarks	179
References.....	181

Chapter I

General Introduction

1.1. These Brains Were Made for Flying

The human brain is a remarkable machine: it contains about as many neurons as there are stars in the Milky Way, and about ten thousand times more connections (DeWeerdt, 2019). A single one of these machines is capable of reading and writing poetry, imagining the motion of invisible atoms, and acquiring a concept of death. For centuries, humans have selected different aspects of their extraordinary intellect to define their uniqueness relative to other species (Dawkins, 1993; Suddendorf, 2013), hand-waving the cognitive feats of other animals aside. When it came to our non-mammalian cousins, this seemed particularly warranted, given how much smaller and simpler their brains appeared.

The anatomy of avian brains was first described by Ludwig Edinger in the 19th Century (Edinger, 1899). Given the lack of a corrugated cortex similar to that found in human brains, the smooth structure of the avian brain was interpreted as a primitive organ, which controlled only instinctive behaviour and was incapable of thought and reasoning. Edinger suggested that the bird brain was analogous to an enlarged basal ganglion of the mammalian brain. It was thought that only a layered, corrugated, large, primate-like brain was capable of complex cognition.

Our understanding of brain anatomy has changed immensely since then. We now know that, although capable of impressive achievements, human forebrains are not anatomically extraordinary: they contain the number of neurons and synapses one would expect for a primate brain of its size (Herculano-Houzel, 2012; our cerebellums, on the other hand, might be extraordinarily large: see Barton & Venditti, 2014). Studies have also revealed that a considerable portion of the bird brain, known as the pallium, is functionally equivalent to the

human neocortex (Jarvis et al., 2005; Dugas-Ford, Rowell, & Ragsdale, 2012; Karten, 1997; Karten, 1969; Karten, 2013; Olkowicz et al., 2016; Reiner et al., 2004; Stacho et al., 2020), and can contain a comparable number of neurons to primate forebrains (Olkowicz et al., 2016). In fact, the pallial region of corvid and parrot brains can contain twice as many neurons as the equivalent area in monkey brains (Olkowicz et al., 2016). The crucial difference is that their neurons are smaller and more densely packed than in primates.

A likely reason for this difference is one of opposing selective pressures: while primates can carry around metabolically expensive large brains, birds require a more compact computing system that is suitable for flying (Chappell, 2017; Vincze, 2016; Winkler, Leisler, & Bernroider, 2004). This is only one of birds' many adaptations for flight, such as air sacs that hollow out and lighten their bones (Wedel, 2019), and efficient metabolic and digestive systems that rapidly dispose of waste (Lasiewski & Dawson, 1967; McWilliams & Karasov, 2014; Nagy, 1987; Price, Brun, Caviedes-Vidal, & Karasov, 2015).

Research on the cognitive abilities of birds has emerged hand in hand with the neuroscience of avian brains. An explosion in studies on corvid cognition, in particular, upended the meaning of the term 'bird brain' and led to their pronouncement as 'feathered apes' (Emery & Clayton, 2004; Emery, 2004; Emery & Clayton, 2005; Güntürkün & Bugnyar, 2016; Osvath, Kabadayi, & Jacobs, 2014; Taylor et al., 2007; Taylor, 2014; Taylor & Gray, 2014; Seed, Emery, & Clayton, 2009), given that they perform comparably to apes in a range of cognitive tests of mental time travel, tool use, and perspective-taking. These performances exemplify cases of complex cognition, that is, the ability to solve a problem through reasoning or thought, rather than simpler associative or instinctive mechanisms (Knauff & Wolf, 2010; Taylor, 2014). Parrots have also recently exhibited impressive performances in cognitive tasks, particularly in terms of their linguistic label comprehension (Pepperberg, 1999; Pepperberg, 2007), categorisation (Pepperberg, 1981), and physical cognition (Auersperg et al., 2014;

Auersperg, von Bayern, Gajdon, Huber, & Kacelnik, 2011; Auersperg, Szabo, von Bayern, & Kacelnik, 2012), and therefore gained increased attention in the field. Alongside corvids, parrots have now become focal species in the study of the convergent evolution of complex cognition between apes and birds (Auersperg & von Bayern, 2019; Lambert, Jacobs, Osvath, & von Bayern, 2019).

1.2. What Makes Kea Special?

The rasping contact call of the kea (*Nestor notabilis*) can be heard in the similarly harsh mountains of New Zealand's South Island. This endemic parrot inhabits rainforest and alpine environments ranging between 700 and 2,000 metres above sea level (Davis, 2008), and are generalist feeders (Greer, Gajdon, & Nelson, 2015), eating flowers, fruits, plant roots, invertebrates, and more (Diamond & Bond, 1999; Greer et al., 2015; Young, Kelly, & Nelson, 2012). Their ecology is unique among parrots, which largely inhabit tropical or subtropical forests (Olah et al., 2016; Vergara-Tabares, Cordier, Landi, Olah, & Nori, 2020; Zelenkov, 2016) rather than snowy alps (Diamond & Bond, 1999). This thesis focuses on the behaviour and cognition of this unique parrot.

Unlike many other parrot species, kea are sexually dimorphic, with males having longer upper mandibles and larger body size than females, they nest in burrows on the ground, and they have evolved largely in the absence of terrestrial predators (Diamond & Bond, 1999). Kea show a prolonged juvenile phase, reaching sexual maturity at around four years old (Higgins, 2002), are highly social, and exhibit playful behaviours throughout their whole lives (Diamond & Bond, 1999; Diamond & Bond, 2003; Diamond & Bond, 2004; Diamond, Eason, Reid, & Bond, 2006). Both object and social play behaviours are associated with a particular call – the warble (Schwing, Nelson, Wein, & Parsons, 2017). Just as laughter is contagious in humans

(Provine, 1992), kea that hear warble calls are more likely to produce the call and initiate play themselves (Schwing et al., 2017). Kea are one of just three bird species shown to produce a call associated with play behaviour (Diamond & Bond, 2003; Pellis, 1981; Schwing et al., 2017; Winkler & Bryant, 2021), and the only one for which this call has been demonstrated to be contagious. Their play often involves the exploration of unfamiliar objects, as kea are especially neophilic even among parrots (Diamond & Bond, 1999).

Several aspects of kea ecology – particularly their sociality, foraging ecology, and adaptability to changing environmental conditions – hint that this might be a particularly intelligent species. Within the scope of animal cognition and ethology, intelligence is commonly defined as an ability to problem-solve and innovate flexibly in response to changing environmental demands and conditions (e.g., Allen, 2017; Buckner, 2015). The theories that link kea's ecology and behaviour to intelligence are broadly divided into the social intelligence hypothesis and the ecological intelligence hypothesis.

The social intelligence hypothesis is exemplified by correlation observed across primate species: those that form larger group sizes also have larger relative neocortex sizes (Dunbar, 1998). This might indicate that sociality – measured as a function of group size – might be a driving force for the evolution of complex cognition. One reason for this might be that, as group size increases, individuals must recognise, monitor, and maintain an increasing number of social relationships between themselves and others, therefore requiring more sophisticated competitive or collaborative abilities (Byrne & Whiten, 1997; Dunbar, 1998; Humphrey, 1976). This social information about others can vary in its richness and complexity, involving knowledge about the individual's past interactions with each group member, their relative rank or dominance, and how group members relate or interact with each other. Large social groups might also display a high level of fission-fusion dynamics, where individuals that are part of a larger group will form smaller parties for foraging and travelling, often dictated

by the resources currently available (Aureli et al., 2008; Kummer, 1971). Living in such complex environments might lead to the evolution of sophisticated social cognition, for example, some primate species acquire social information about conspecifics indirectly through social eavesdropping (Cheney & Seyfarth, 2005). This third-party monitoring could also relate to non-social aspects of cognition: for example, an ability to make inferences about social rank after hearing the calls produced by out-of-sight conspecifics would require not only the ability to track multiple individuals and their relationships, but also to make transitive inferences about hierarchies more generally (Cheney & Seyfarth, 1999; Mennill, Ratcliffe, & Boag, 2002; Silk, 1999).

Additionally, some researchers have recently suggested that the social intelligence hypothesis could be expanded to include all conspecific interactions, including outgroup members (Ashton, Kennedy, & Radford, 2020). This extension could be particularly relevant when comparing the relative brain sizes and intelligence of species that engage in repeated between-group conflicts or form fluid social groups with continually shifting composition, comparable to those displaying fission-fusion dynamics.

Kea sociality fulfils many of the conditions proposed by the social intelligence hypothesis: they form large groups of up to thirty individuals (Diamond & Bond, 1999), display comparable fission-fusion dynamics to ape and monkey species (Schwing, Jocteur, Wein, Noë, & Massen, 2016), and regularly partake in social behaviours such as allopreening and social play (Diamond & Bond, 1999; Diamond & Bond, 2003; Diamond & Bond, 2004). Furthermore, kea are long-lived, reaching up to 50 years in captivity (Young, Hobson, Lackey, & Wright, 2012), and so wild groups are likely to span multiple generations, such that individuals must continually acquire novel social information about conspecifics throughout their entire lives. Therefore, in recent years they have become an important model species for

investigating social and cooperative cognition among parrots (Auersperg & von Bayern, 2019; Heaney et al., 2017a, b; Huber & Gajdon, 2006; O'Hara et al., 2012).

Beyond sociality, several aspects of kea's ecology are thought to promote flexible behaviour and complex cognition. The ecological intelligence hypothesis suggests that species may evolve greater intelligence in response to environmental challenges associated with the scarcity or difficulty of obtaining food in their natural environment (Duncan et al., 2005; Milton, 1981; Rosati, 2017; Sol, 2009). Three variations of this hypothesis are equally appropriate for kea ecology, namely: (i) the extractive foraging hypothesis, that links the difficulty of obtaining hard-to-access food resources such as underground roots or burrowing insects to the evolution of complex foraging behaviours and accompanying intelligence to support them; (ii) the spatial cognition hypothesis, that postulates greater relative brain size and cognitive ability for species that forage for scarce resources over large geographical ranges, therefore requiring the ability to remember the locations, travelling routes, and richness of different food patches; and (iii) the clever foraging hypothesis, which proposes that species living in more complex environments will evolve improved cognitive capacities (Bauchot, Bauchot, Platel, & Ridet, 1977; Park & Bell, 2010; Safi & Dechmann, 2005; Striedter, 2005).

Again, the kea and their ecology fulfil the requirements set out by these hypotheses. Firstly, kea exhibit extractive foraging, particularly digging, which may allow them to access underground roots (Clarke, 1970; Diamond & Bond, 1999; Greer et al., 2015). A recent study has also shown that a wild population of kea has developed stick tool use to interact with man-made predator traps, although it is unclear if this innovation serves an extractive foraging purpose, occurs as a form of object play, or comprises a combination of both (Goodman, Hayward, & Hunt, 2018). Second, their diet changes seasonally, for example, relying more heavily on fruits in the summer and autumn months (Greer et al., 2015). Kea's foraging behaviour also varies seasonally, with more movement between foraging patches in the winter

and less in the summer, likely as a direct consequence of changing resource availability (Greer et al., 2015). Individual kea have been found to navigate areas up to 4km² (Elliott & Kemp, 1999; Jackson, 1969), suggesting they cover large distances over the New Zealand alps, which requires appropriate spatial cognition. Finally, kea's rainforest and mountain habitats are complex in terms of both seasonality and the varying substrates, resources, and the species encountered at different altitudes.

The selective pressures imposed by their ecology, as well as their deeply social lives, therefore highlight kea as a species that is likely to exhibit complex cognition (Huber & Gajdon, 2006).

1.3. What Do We Know About Kea Cognition?

1.3.1. Physical Cognition

Physical cognition refers to animals' understanding of, and interactions with, the physical world (Shettleworth, 2010). Like human infants have innate expectations of how objects move and interact with their environment (e.g., Baillargeon, 1993; Hespos & Baillargeon, 2001; Hespos, Ferry, Anderson, Hollenbeck, & Rips, 2016; Piaget, 1954; Spelke, 1991), animals too may possess some level of naïve physics. This field of study was popularised by a series of experiments by Povinelli showing that chimpanzees have a poor grasp on concepts such as gravity, continuity, and connectivity (Povinelli, 2003). Since then, much work has attempted to uncover the extent of primates' physical cognitive abilities. Researchers have also started to explore the same questions in birds. This originally involved naturally tool-using species such as New Caledonian crows (*Corvus moneduloides*) and Galápagos woodpecker finches (*Cactospiza pallida*), under the assumption that tool users might have more sophisticated technical and physical knowledge (Byrne, 1997; Iwaniuk, Lefebvre, & Wylie, 2009; Kacelnik,

2009; Lefebvre, Nicolakakis, & Boire, 2002; Reader & Laland, 2002; Taylor et al. 2009), a hypothesis which has found little support due to studies demonstrating identical performances in physical cognition tasks by tool-users and non-tool-users (Bird & Emery, 2009; Miller et al., 2020; Teschke et al., 2013).

Among parrots, Goffin's cockatoos (*Cacatua goffiniana*) and kea are the most commonly studied species (Auersperg & von Bayern, 2019). Kea may be particularly good problem-solvers, as suggested by a string-pulling study where some subjects learned to retrieve a food reward hung on the end of a string in a matter of seconds (Werdenich & Huber, 2006). Following this performance, subjects also correctly identified and selected which of two strings contained the desired reward when they were slanted or crossed. These results suggest that kea can develop an understanding of physical connectivity. This is further supported by a study where kea had to select and pull on one of two wooden boards in a range of configurations, but in some cases only one of the two boards was connected to a food reward (Auersperg, Gajdon, & Huber, 2009). One of the six subjects in the study correctly selected a connected board over a board with a gap over the course of their first six trials. In a similar support task where kea had to distinguish between a board containing food and a board with food placed next to it, all subjects performed significantly above chance, outperforming infant chimpanzees (Auersperg et al., 2009; Spinozzi & Potí, 1993).

Despite not being habitual tool users in the wild, kea also readily learn to use tools in captive settings (Auersperg, Huber, & Gajdon, 2011a; Auersperg et al., 2011b; Auersperg, Gajdon, & Huber, 2010). For example, kea can deposit balls into tubes to obtain food trapped inside (Auersperg et al., 2010). Even if one side of the tube is closed shut, kea quickly learn to solve the problem by both using a ball to dislodge the food and then tipping the tube towards the opening, performing both behaviours sequentially and consistently (Auersperg et al., 2010). Despite having curved beaks that make manipulating sticks more difficult than it is for straight-

billed species such as New Caledonian crows, kea can also innovate stick tool use both in captivity (Auersperg et al., 2011a, b) and in the wild (Goodman et al., 2018). In a study that compared them directly against New Caledonian crows, kea were quicker to innovate ways to obtain a food reward from a multi-access puzzle box with four possible solutions (pulling a string attached to the food, inserting a ball to dislodge the food, using a stick to dislodge the food, and opening a window to access the food; Auersperg et al., 2011b). Kea innovated the window-opening and ball-inserting behaviours significantly faster than the crows, although they were slower to successfully use sticks to dislodge the reward. Unlike for the straight-billed crows, this often required kea to coordinate both their beak and feet to guide the tool accurately. An additional study on captive kea's stick tool use showed that they can learn to use stick tools following demonstrations and can direct them to make selections between two different options (Auersperg et al., 2011a). Wild kea have also recently innovated this behaviour, using branches and other objects to deactivate pest traps designed to catch invasive stoats and rats (Goodman et al., 2018).

Despite their ape-like performance in tests of technical, or physical, intelligence (Huber & Gajdon, 2006; O'Hara, Gajdon, & Huber, 2012), there is mixed evidence on how much kea understand of the physical processes encountered in these tasks. After learning to push a baited tube up a diagonal pole in order to remove it and access the food inside, kea were only able to use colour cues, but not functional information obtained during free object interactions, to distinguish between permanent and removable lids at the top of two poles (Gajdon, Ortner, Wolf, & Huber, 2013). Similarly, although kea can learn to drop stones into tubes filled with water in order to dislodge the water inside and bring a floating food reward into reach – a task known as the Aesop paradigm (Bird & Emery, 2009) –, they were unable to distinguish between a normal tube and a tampered tube that did not allow water level changes (Schwing, Weiss, Tichy, & Gajdon, 2019). This performance was similar to that of New Caledonian

crows, who perform well at a variety of Aesop fable tasks but that also do not attend to changes in water levels in the U-shaped tube task (Logan et al., 2014; Jelbert et al., 2014; Jelbert, Taylor, & Gray, 2015a; Taylor et al., 2011).

There is evidence that kea may be able to mentally simulate the steps required to solve complex problems. In one study where subjects had to open a challenging multi-step locked box to obtain the reward inside, kea made more corrections to their unproductive actions if they first had the opportunity to observe the problem through a transparent barrier, compared to trials where they observed an opaque barrier (Miyata, Gajdon, Huber, & Fujita, 2011). This suggests that, when faced with complex problems, kea might be able to plan several steps ahead, although the study also revealed that subjects often persisted in performing some unfruitful actions, which may have been a consequence of their naturally exploratory behaviour.

This exploratory drive is a central aspect of kea's interactions with objects and may emerge as a consequence of their innate playfulness and interest in novelty, or neophilia. Play is an interesting behaviour from both cognitive and evolutionary standpoints because it involves the repetition of apparently useless actions, providing no obvious advantage to the individual (Burghardt, 2006). Although play is relatively common in juveniles across many mammalian and bird species, its retention in adulthood is unusual (Burghardt, 2006). As kea retain play through all their lives, some work has investigated the possible mechanisms and function of play behaviour for this species. One study reported that kea show a natural disposition to inserting objects into empty tubes, and although this behaviour can be co-opted to extract food placed inside tubes (Auersperg et al., 2010), it emerges even in the absence of a reward, and individuals that have previously successfully inserted small objects into tubes will sometimes continue to attempt inserting too-large objects in an exploratory manner (Gajdon, Lichtnegger, & Huber, 2014). When provided with a range of novel objects, kea

produce a higher number of object-substrate combinations compared to five other parrot species, particularly insertions and recoveries of objects from holes, tubes, and poles (Auersperg et al., 2014). The frequency of their object-substrate combinations was rivalled only by Goffin's cockatoos (Auersperg et al., 2012). Finally, a study comparing kea with habitually tool-using New Caledonian crows found that both species' performance in two tool-selection tasks – dropping heavy blocks to collapse a baited platform and using rigid rope to extract food from a tube – improved following ten minutes of free exploration with the available object options (Lambert, et al., 2017). This suggests that object exploration is useful for acquiring knowledge about the functionality of novel objects, providing a potential explanation for kea's retention of play and neophilia into adulthood.

1.3.2. Social Cognition

As well as solitary object play, kea also partake in playful social interactions. These can include social object play, where multiple individuals interact with objects at the same time (Diamond & Bond, 1999), and three other categories of social play behaviours, namely play invitations, play chasing, and play fighting (Diamond & Bond, 2003). Compared to their closest relatives, kaka (*Nestor meridionalis*) and kakapo (*Strigops habroptilus*), kea exhibit more complex social play with greater behavioural variability and duration (Diamond & Bond, 1999; Diamond & Bond, 2003; Diamond & Bond, 2004; Diamond et al., 2006). Play fighting is vigorous and includes foot wrestling, pushing, jumping, wing flapping, rolling over, biting, and locking bills, with a single bout of play persisting for several minutes (Diamond & Bond, 1999; Diamond & Bond, 2004). In the wild, both social and object play are associated with the warble call, which may be emotionally contagious among individuals and triggers play behaviours in others (Schwing et al., 2017). Although very distinct in its components to play bouts, captive kea also appear to partake in ritualised aggressive displays that include cheek feather puffing,

synchronised walking, and object tossing, sometimes escalating to fights between two or more individuals (*pers. obs.*). These behaviours have thus far not been reported for wild kea.

The breadth and constancy of kea social behaviours are suggestive of a rich social life. This has sparked considerable interest in their social cognition, which originated with studies of social learning among wild birds. Huber and colleagues (2001) presented captive kea with an “artificial fruit” puzzle box similar to that presented to chimpanzees and capuchin monkeys in previous studies (Custance, Whiten, & Fredman, 1999; Whiten, Custance, Gomez, Teixidor, & Bard, 1996). Both males that were selected as demonstrators were successfully trained to open the apparatus within one minute. While observers were significantly better than birds in the non-observing control group at manipulating some of the locking mechanisms on the box, no kea learned to fully open the multi-step puzzle box from observation alone. The authors interpreted this as evidence of stimulus enhancement – increased attention to the apparatus or parts of the apparatus manipulated by the demonstrators – rather than the imitative learning observed in chimpanzees (Huber et al., 2001; Whiten et al., 1996).

Another social learning study with captive kea presented individuals with two collapsible platforms, only one of which was functional (Gajdon, Amann, & Huber, 2011). Demonstrators were trained to distinguish between the two platforms, and although adult observers did initially learn to discriminate between the two platforms, they eventually began to insert blocks indiscriminately into both platforms, suggesting that for kea, individual exploration and neophilia may override socially learnt information. This is supported by a looking-time study showing that, although adult kea’s looking times towards feeding conspecifics is comparable to those of three- to six-year-old children, they are less interested in others’ manipulation and searching behaviours, suggesting that they are less disposed to attend to and therefore imitate demonstrators’ foraging behaviours (Range, Horn, Bugnyar, Gajdon, & Huber, 2009).

Attempts at detecting social learning in wild kea have been less successful than captive studies. Although two wild individuals learned to correctly push a tube to the top of the pole to obtain a food reward, there was no evidence that their performance improved following observation of a trained demonstrator (Gajdon, Gyula, Fijn, & Huber, 2004), or that their strategy assimilated that of the demonstrator. It is unlikely that this result was a consequence of the task being too difficult, as there is also little evidence that wild kea copy others' naturally occurring foraging innovations such as bin opening (Gajdon, Fijn, & Huber, 2006). Put together, these studies suggest that kea might be capable of social learning, particularly in captive environments that are likely to offer longer uninterrupted observation time and less competition for apparatus access, but individual trial-and-error learning is likely to be their predominant innovative strategy in the wild (Gajdon et al., 2006).

Despite the lack of evidence for social learning in wild populations, kea may benefit from living in large social groups in other ways. One potential advantage of social living is an increase in opportunities for cooperative foraging, where individuals can work together to exploit food resources more effectively (Giraldeau & Caraco, 2000; Krause & Ruxton, 2008). Given their sociality and the seasonality and scarcity of food availability in their natural environments, kea offer a good candidate species for cooperation studies.

The first study of cooperation in kea presented captive dyads with a see-saw apparatus where one individual had to step on one end to lift a lid on the other end, therefore enabling the other subject access to the food reward (Tebbich, Taborsky, & Winkler, 1996). Although subjects did readily cooperate with each other, the more dominant individual was usually the only one to access the food rewards, suggesting that where kea dyads do cooperate, they do not share rewards equally. However, foraging in dyads would be unusual in wild kea, as they usually travel in large flocks. When given the same cooperative task in a larger group, all individuals in the study eventually obtained some of the food reward. This is likely a direct

consequence of kea's dominance structure, which is non-linear and based on individual relationships rather than overall rank (Tebbich et al., 1996), which led to all individuals eventually being displaced from the apparatus and therefore more even reward distribution across the group.

This work inspired a number of follow-up investigations of kea cooperation in captive settings, using variants of the loose-string cooperation task originally developed for and presented to chimpanzees (Hirata, 2003; Hirata, Satoshi & Fuwa, 2007; Melis, Hare, & Tomasello, 2006a; Melis, Hare, & Tomasello, 2006b). In this paradigm, two ends of a string are presented to the sides of a platform containing food rewards. If an individual approaches and pulls on one end of the string, they will draw the string closer, but will not move the platform. The platform can only be pulled closer if two individuals simultaneously pull on both ends of the string. This task therefore tests not only individuals' ability to cooperate, but also whether or not subjects understand that they require a partner to complete the task (Hirata, 2003).

In the loose-string task, kea cooperate preferentially with individuals that they spend the most time with (Schwing et al., 2016), can wait over a minute for a partner to arrive before touching the string, and can discriminate between situations where cooperation is needed and those where it is not (Heaney, Gray, & Taylor, 2017b). After receiving training to attend to the presence of partners, kea can also succeed at the loose-string task in cooperating triads (Schwing et al., 2020). Kea also cooperated in a modified version of this task where subjects had to simultaneously pull on between two and four strings to open a box containing food rewards, although this occurred only once the most dominant individuals in the group learnt that attempting to monopolise the apparatus could not lead to a successful outcome (Schwing, Meaux, Piseddu, Huber, & Noë, 2021). Despite ample evidence that kea can perform cooperative foraging in captive settings, there are only anecdotal reports of cooperative hunting

in wild kea (Heaney, Gray, & Taylor, 2017a). Interestingly, there is also no evidence at present that kea can detect unfairness in resource distribution (Heaney et al., 2017a), an ability which is thought to have evolved alongside cooperation as a mechanism to keep track of the relative contributions of unrelated individuals to joint tasks (Brosnan, 2011; Brosnan & de Waal, 2014).

Finally, a recent study investigating kea's prosociality and reciprocity found that although individuals often preferred to exchange a prosocial token that delivered food rewards to both themselves and a social partner over either a selfish or an unrewarding token, they continued to do so in control conditions where either the social partner was absent or could not make a choice (Heaney, Bastos, Gray, & Taylor, 2019). This suggests that kea may have selected the prosocial token because it was visibly associated with a larger reward, rather than because they understood the social nature of the task. Nevertheless, a single subject began to discriminate between test and control conditions in the second round of testing, suggesting that kea may be able to learn the contingencies of prosociality tasks with additional experience.

1.3.3. Reasoning & Learning

Kea's performance in tasks of social and physical cognition have also sparked interest in their more abstract cognitive abilities, such as reasoning and learning. Additionally, kea are part of the most ancestral clade of parrots (Suh et al., 2011; Wright et al., 2008), Strigopidae, which comprises only three extant species: kea, kaka (often subdivided into subspecies *N. m. meridionalis* and *N. m. septentrionalis*), and kakapo. This has led to claims that some of the cognitive abilities present in kea might be representative of other parrot lineages (Schwing, Weber, & Bugnyar, 2017; Wein, Gajdon, Schwing, & Wright, 2015).

Studies on inference by exclusion test subjects' ability to reason that if one of two mutually exclusive options is wrong, then the other must be correct (Call, 2006). This paradigm

has been studied twice in kea. The first attempt presented both kea and ravens (*Covus corax*) with pairs of baited and unbaited tubes and predicted that, if subjects used inference by exclusion, they should only need to investigate one of the two tubes to select the correct option (Schloegl et al., 2009). Ravens outperformed kea in this task, although kea's failure to inhibit their exploration of all available options was likely a reflection of their neophilic nature rather than an inability to perform inferential reasoning. This explanation is supported by the findings of a later study, where some kea successfully performed inference by exclusion on a touchscreen task that did not require such inhibition (O'Hara, Schwing, Federspiel, Gajdon, & Huber, 2016). This improved performance occurred despite evidence showing that, although kea readily transfer information learned about real objects to images of those objects in both pictures and touchscreens (Wein et al., 2015), they take significantly longer to learn tasks with virtual stimuli than with real-life stimuli (O'Hara, Huber, & Gajdon, 2015). Nevertheless, a recent study showed that kea perform well in a touchscreen reversal learning task where two stimuli alternate between being positively rewarding or neutral, demonstrating that they are capable of flexible and rapid application of a win-stay, loose-shift learning strategy on touchscreen tasks (Laschober, Mundry, Huber, & Schwing, 2021).

As well as reasoning by exclusion, kea also exhibit the ability to delay gratification (Schwing et al., 2017), that is, they can inhibit an impulse to obtain an immediately available reward in order to gain a better reward in the future. This is also known as the "marshmallow task" in humans, due to a seminal study that required children to refrain from eating a marshmallow for fifteen minutes in order to gain another one (Mischel, Ebbesen, & Raskoff Zeiss, 1972). Succeeding at this task may require both inhibition and future planning (Kacelnik, 2003), suggesting that despite their limited ability to inhibit exploratory actions with objects, they nevertheless exhibit impulse control in non-exploration-based tasks. In the kea study, subjects were able to wait several seconds to exchange a food reward they held in a container

for a higher quality reward, and waiting times increased as the difference in quality between the two rewards increased, demonstrating the capacity to make simple economic decisions about the future (Schwing et al., 2017).

1.3.4. Linking cognition and conservation

Work on kea intelligence has the potential to not only shed light on the evolution of avian intelligence, but also assist the conservation of this species, which is endangered in the wild (BirdLife International, 2017). The main threat to kea survival is predation by invasive species such as stoats, possums, and feral cats, that feed on the eggs and chicks in kea's burrows (Innes, Kelly, Overton, & Gillies, 2010). Efforts to control the populations of these invasive predators includes drops of aerial 1080 poison pellets (Elliott, Graeme & Kemp, 2016; Kemp, Mosen, Elliott, & Hunter, 2018), although a minority of kea may also ingest and succumb to the poison (Veltman & Westbrooke, 2011). Some current research efforts are therefore focused on using kea cognition to investigate methods to reduce non-target poisoning. One recent study showed that kea will continue to avoid bait to which they developed conditioned aversion for at least three weeks, suggesting that pellet drops following provision of irritant, but non-lethal baits could reduce wild kea's consumption of 1080 poison (Nichols, Bell, Mulgan, & Taylor, 2020). Additional studies are currently underway to establish the duration of aversion memory in kea. Alongside studies probing kea's perception of pellets and the effects of pellet characteristics on consumption (Brunton-Martin, Nichols, & Gaskett, 2021; Orr-Walker, Adams, Roberts, Kemp, & Spurr, 2012; Weser & Ross, 2013), this work could have real policy implications for conserving kea in the wild.

1.4. Structure of the Thesis

This thesis presents some of my work investigating kea's cognitive abilities, with the overall aim of exploring how this species can serve as a model for avian intelligence and the cognitive mechanisms it may have evolved convergently to the great apes. These studies encompass a broad range of abilities, including kea's technical intelligence in the context of tool use, their ability to mentally represent hidden objects and their trajectories, their numerical competencies in making statistical inferences, the physical cognition underlying their understanding of connectivity in a string-pulling task, and their ability to discriminate between physical processes in real and virtual environments. I explore these issues through a number of different approaches, including observational studies, binary success-testing, and a novel signature-testing framework, playing to the strengths of each these techniques to inform our knowledge of kea cognition. The following paragraphs give a brief overview of the relevance and contents of these chapters.

In **Chapter 2**, I provide evidence for an innovative self-care tool use behaviour by a kea. Despite not being a habitually tool-using species, kea can innovate the use of stick tools to forage in both wild and captive contexts (Auersperg et al., 2010; 2011a, b; Goodman et al., 2018). Tooling innovations in non-tool-using species has been taken as evidence of sophisticated problem-solving and flexible cognition (Emery & Clayton, 2009; Hunt, Gray, & Taylor, 2013; Lefebvre, Reader, & Sol, 2004), but researchers must be cautious not to overinterpret anecdotal object manipulations by animals (Auersperg, Schwing, Mioduszevska, O'Hara, & Huber, 2020; Dechaume-Moncharmont, 2020; Farrar, 2020; Sándor & Miklósi, 2020; von Bayern, Jacobs, & Osvath, 2020). Considering the recent criticisms of an anecdotal self-care tool use report in puffins (Fayet, Hansen, & Biro, 2020), this chapter uses repeated observations of both a focal individual that innovated pebble tool use for preening and non-tool-using conspecifics, to provide the first evidence of a self-care tooling innovation in a kea.

In **Chapter 3**, I expand on my discovery of innovative self-care tooling in a kea and use crowdsourced data of pet parrots to establish the prevalence of tool use behaviours across the parrot order. I demonstrate that self-care tool use is more widespread in the Psittaciformes than previously thought and use phylogenetic ancestral state reconstruction to demonstrate that the capacity for tool use is not only likely ancestral to parrots, but also lost in the recently radiating Psittaculidae clade. As well as serving as an example for how crowdsourced data can help researchers identify animal behaviours that are only rarely observed in the field, this work also outlines ways in which this data can help inform the evolutionary histories of these behaviours.

Chapter 4 investigates kea's understanding of connectivity in a horizontal string-pulling task where subjects must choose between pulling on a connected or a disconnected string to obtain a rewarding black token on the other end. The literature to date suggests that although some bird species can succeed at this horizontal connectivity task, this always follows evidence of vertical string-pulling and therefore experience of a perceptual-motor feedback loop, the mechanism thought to underlie string-pulling behaviour (Taylor et al., 2010; Taylor, Knaebe, & Gray, 2012). In this chapter, I present kea with this task both before and after gaining experience of vertical string-pulling, to assess whether experience of the perceptual-motor feedback loop modulates attention to the physical properties of strings, therefore enabling success at the horizontal connectivity task. Further, the three experiments presented also test for some of the other possible mechanistic explanations for string-pulling behaviour, namely insight, planning, and means-end understanding.

In **Chapter 5**, I develop a novel hand-tracking training technique and a series of three experiments to investigate kea's ability to mentally represent the locations and trajectories of two hidden objects. In humans, this cognitive ability is required in more complex mental simulations, for example reasoning about hidden causal mechanisms (Kushnir, Tamar, Gopnik,

Lucas, & Schulz, 2010; Wilcox & Baillargeon, 1998). I also examine whether kea can successfully predict an object's incomplete trajectory and therefore make predictive choices about likely future events.

Chapter 6 builds on the findings of the previous chapter and uses kea's abilities to make predictive choices based on mental simulations to investigate probabilistic reasoning in this species. Over the course of three experiments, I identify three signatures of domain-general human statistical inference – that is, an ability to judge the relative likelihoods of two events by using frequency information – in kea, namely: (i) using relative frequencies to make probabilistic judgements (Denison & Xu, 2014), (ii) integrating a physical constraint into their probabilistic predictions (Denison & Xu, 2010; Téglás, Girotto, Gonzalez, & Bonatti, 2007; Téglás et al., 2011), and (iii) integrating a social agent's bias for particular objects into their predictions (Gweon, Tenenbaum, Schulz, & Carey, 2010; Kushnir, Xu, & Wellman, 2010; Ma & Xu, 2011; Wellman, Kushnir, Xu, & Brink, 2016; Xu & Denison, 2009). These three signatures suggest that kea are capable of cross-domain integration, providing the first empirical evidence for domain-general intelligence in a bird.

Given kea's sophisticated physical cognition, **Chapter 7** investigates this species' innate perception and understanding of physical processes in the virtual and real worlds. In this chapter, I test whether kea are “naïve realists”, that is, whether they expect events in virtual environments to be equivalent and continuous with real-life. This work is based upon a human study showing that 19-month-old infants expect physical events that take place within a screen to be discontinuous from the real objects surrounding the screen (Revcu & Csibra, 2020), and therefore are not naïve realists. I adapt this study for within-individual animal testing, highlighting how testing animals' expectations of continuity between virtual and real stimuli is essential to validating the growing use of touchscreen and virtual reality (VR) technologies in animal testing.

Chapter 8 is a review addressing how studies from avian cognition such as the ones presented in previous chapters can inform central debates in comparative psychology. In this chapter, I present evidence from the avian cognition literature that push against Euan Macphail's claim in the 1980's that there are no quantitative or qualitative differences in intelligence between species. Although this claim was made four decades ago, scholars still disagree on the extent to which ecology shapes the cognitive processes of different species (for divergent views from my own, see: Colombo & Scarf, 2020; Zentall, 2021). This chapter also highlights how moving from binary success-testing tasks that establish if animals succeed or fail to solve particular problems, to a signature-testing framework that instead attempts to pinpoint the cognitive processes being used by different species confronted with ecologically relevant tasks, as well as setting comprehension baselines as part of pre-test training protocols, and interpreting animal performances within a Bayesian framework, can all contribute to this debate.

Finally, **Chapter 9** provides an overall summary of the studies presented in this thesis and discusses how studies of kea behaviour and cognition might be used to the benefit of the species. In particular, I focus on the practical applications of this work for the conservation of kea in the wild and propose a series of potential real-world interventions that could help conserve this charismatic species.

Chapter II

*Self-care tooling innovation in a disabled kea*¹

2.1. Abstract

Tooling is associated with complex cognitive abilities, occurring most regularly in large-brained mammals and birds. Among birds, self-care tooling is seemingly rare in the wild, despite several anecdotal reports of this behaviour in captive parrots. Here, we show that Bruce, a disabled parrot lacking his top mandible, deliberately uses pebbles to preen himself. Support for this claim comes from five lines of evidence: (i) in over 90% of instances where Bruce picked up a pebble, he then used it to preen; (ii) in 95% of instances where Bruce dropped a pebble, he retrieved this pebble, or replaced it, in order to resume preening; (iii) Bruce selected pebbles of a specific size for preening rather than randomly sampling available pebbles in his environment; (iv) no other kea in his environment used pebbles for preening; and (v) when other individuals did interact with stones, they used stones of different sizes to those Bruce preened with. Our study provides novel and empirical evidence for deliberate self-care tooling in a bird species where tooling is not a species-specific behaviour. It also supports claims that tooling can be innovated based on ecological necessity by species with sufficiently domain-general cognition.

¹ Based upon: Bastos, A. P. M., Horváth, K., Webb, J. L., Wood, P. M., Taylor, A. H. (2021). Self-care tooling innovation in a disabled kea (*Nestor notabilis*). *Scientific Reports*, 11, 18035. doi: 10.1038/s41598-021-97086-w

2.2. Introduction

Tooling – deliberately generating a mechanical interface by using an object to manipulate another target or surface (Fragaszy & Mangalam, 2018, 2020) – is a complex embodied form of tool use which has been documented in only a few species (Fragaszy & Mangalam, 2018; Hunt, Gray, & Taylor, 2013; Shumaker, Walkup, Beck, & Burghardt, 2011). When tooling, an animal and the tooling object are transformed into a single body-plus-object system (Fragaszy & Mangalam, 2018, 2020), performing contextually appropriate, embodied problem-solving (Cabrera-Alvarez & Clayton, 2020; Fragaszy & Mangalam, 2018, 2020). Flexible tool use, including tooling, often emerges in the form of behavioural innovations, where animals develop a novel behaviour in response to situational changes (Ramsey, Bastian, & van Schaik, 2007). Tool use behaviour – including but not limited to tooling – has been regarded as a marker of complex cognition across birds and mammals (Emery & Clayton, 2009; Hunt et al., 2013; Lefebvre, Reader, & Sol, 2004), a link which has led to considerable interest in the field (Bentley-Condit & Smith, 2010; Hunt, 1996; Lefebvre, Whittle, Lascaris, & Finkelstein, 1997; Shumaker et al., 2011). Within birds, flexible tooling has been found most commonly in clades with large relative brain sizes, such as corvids (Bentley-Condit & Smith, 2010; Gruber et al., 2019; Hunt & Gray, 2002; Lefebvre, Nicolakakis, & Boire, 2002; Rutz et al., 2016; Shumaker et al., 2011; Taylor, Hunt, & Gray, 2012; Weir, Chappell, & Kacelnik, 2002) and parrots (Auersperg, Szabo, von Bayern, & Kacelnik, 2012; Borsari & Ottoni, 2005; Heinsohn, Zdenek, Cunningham, Endler, & Langmore, 2017; Lambert, Seed, & Slocombe, 2015; Lefebvre et al., 2002; Shumaker et al., 2011).

Most reports of tooling in birds revolve around foraging (Bentley-Condit & Smith, 2010; Borsari & Ottoni, 2005; Hunt, 1996; Hunt & Gray, 2002; Kacelnik, Chappell, Weir, & Kenward, 2006; Lambert, Megan L. et al., 2015; Lefebvre et al., 1997; Lefebvre et al., 2002; Rutz et al., 2016; Weir, Chappell & Kacelnik, 2006). Among parrots this is more common in

captive settings, for example, greater vasa parrots use small stones to scrape or break up shells, which they then ingest (Lambert et al., 2015), hyacinth macaws use wedges to manipulate nuts (Borsari & Ottoni, 2005), and Goffin's cockatoos innovate and manufacture stick tools to retrieve out-of-reach food (Auersperg, et al., 2012). Despite not habitually using tools in the wild, kea parrots (*Nestor notabilis*) learn to insert sticks and other objects into traps designed for pest species such as stoats (Goodman et al., 2018), which allows them to safely access egg bait placed inside. Kea also probe at and set off unbaited traps, an apparently non-functional behaviour which may be playful or exploratory in nature (Goodman et al., 2018). In captivity, kea also readily learn to use stick tools for extractive foraging in experimental settings (Auersperg et al., 2010, 2011).

In the wild, tooling in a self-care context appears to be rarer than in foraging contexts (Fayet et al., 2020; McGrew, 2013). However, there are anecdotal and video reports of several parrot species innovating self-care tooling in captivity, primarily by holding sticks or other objects with their feet to scratch themselves (Bentley-Condit & Smith, 2010; Janzen, Janzen, & Pond, 1976; Lefebvre et al., 2002). These behaviours have not been rigorously tested and so it is unclear how they were innovated, how frequently they occur, or if reduced interactions with conspecifics in captivity reduces allopreening and so drives the individual innovation of self-care tooling.

Recently, a study reported self-care tool use in Atlantic puffins (*Fratercula arctica*), which were observed holding sticks to their bodies, possibly in order to scratch themselves (Fayet et al., 2020). This claim rests on two observations across four years of two puffins living in colonies over 7,000km apart (in Wales and Iceland), for which only one tool use instance was recorded on video. This claim has garnered significant attention from the scientific community, which has been sceptical that the study provides sufficient evidence of tool use (Auersperg et al., 2020; Dechaume-Moncharmont, 2020; Farrar, 2020; Sándor & Miklósi,

2020; von Bayern, Jacobs, & Osvath, 2020). Not only is the single recorded incident short in duration, lasting approximately one second, but the touching of the stick to the puffin's chest may have been an accidental combination of two other behaviours, namely holding a stick and attempting to scratch itself with its beak (Auersperg, et al., 2020; Farrar, 2020; Sándor & Miklósi, 2020). Critics of this study (Auersperg, et al., 2020; Dechaume-Moncharmont, 2020; Farrar, 2020; Sándor & Miklósi, 2020; von Bayern et al., 2020) have argued for a hypothesis-testing approach to anecdotal reports such as these, suggesting that convincing evidence for tool use in puffins should include: (i) comparisons between the number of instances where sticks are picked up and used for scratching, and instances where they are picked up but not used for scratching; (ii) evidence of subjects' intention to scratch with a stick, for example by showing that they exhibit preferences for sticks with favourable characteristics, value previously useful tools, or perform the same behaviour repeatedly over multiple days; and (iii) comparisons between tool-using and non-tool-using individuals, where the latter should be more likely to pick up sticks without performing scratching actions than the former.

These criticisms are pertinent to our observations of apparent innovation of self-care tooling by a disabled kea parrot, a species where tooling is not a species-specific behaviour. Bruce, a wild-born male with an estimated age of 8 years, is disabled due to him missing the upper part of his bill. This means that he struggles to perform basic functions kea use their powerful beaks for, such as eating and preening. Bruce appears to use small pebbles (which he typically takes from a gravel path in the aviary, using his lower mandible to scoop them up) to preen himself. Pebbles are wedged between his lower mandible and tongue and moved along his feathers. This appears analogous to other subjects' clasp and grinding of feathers between their upper and lower mandibles.

Here, we aim to provide evidence for Bruce's deliberate self-care tooling behaviour from repeated observations, by creating a scientifically rigorous way to report self-care tooling.

In line with critiques of the puffin report, we hypothesise that this behaviour could be considered deliberate rather than incidental if our data indicate that: (i) Bruce's instances of pebble manipulations more often than not occur simultaneously with preening; (ii) he performs this preening tooling behaviour repeatedly and specifically with pebbles, retrieving or replacing lost preening pebbles; (iii) he uses pebbles with particular characteristics which can afford this preening function; (iv) he preens with a pebble more often than other individuals with typical beak morphology housed in the same aviary, who should be less likely than him to combine object manipulation and preening behaviours into a single action; and (v) the types of objects he interacts with are different to those selected by other individuals, who do not use pebbles as preening tools.

2.3. Materials and Methods

2.3.1. Subjects and Procedure

All observations were of a captive population of kea (*Nestor notabilis*) housed in a large, naturalistic outdoor aviary at Willowbank Wildlife Reserve in New Zealand. This population comprised thirteen individuals (10 males, 3 females) with ages ranging from 8 months to 25 years (Table 1). Food and water were available *ad libitum*, and subjects were free to behave and interact normally in their environment throughout observation sessions. Observation sessions were at least one hour long and took place in the mornings, when birds were most active. This research was conducted under ethics approval from The University of Auckland Ethics Committee (reference number 001816).

Subject	Hatch Date (Known or estimated)	Sex
Angelina	October 2006	F
Blofeld	August 2013	M
Boh	September 2012	F
Harley Quinn	August 2014	F
Johnny	Unknown	M
Bruce	September 2012	M
Loki	August 2014	M
Megatron	October 2019	M
Moriarty	August 2014	M
Neo	September 2012	M
Plankton	August 2014	M
Spike	October 1995	M
Taz	September 2012	M

Table 1. Hatch dates and sex of individuals observed as a part of this study. All subjects were parent reared.

Observations took place in two stages. The first stage involved 20 hours of focal observations of Bruce over nine days, where any interactions with objects and any preening episodes were recorded on video, until 30 seconds after the end of the preening or object manipulation episode. In every preening episode, experimenters either approached Bruce or zoomed in with their video camcorders to establish if an object was being held in his beak during the preening episode. This was unlikely to affect his behaviour, given that Bruce was familiar with all four experimenters and has an extensive history of participating in cognitive studies where he stands in close proximity to humans (Bastos & Taylor, 2019, 2020; Heaney, Bastos, Gray, & Taylor, 2019). Bruce’s preening episodes were never interrupted, so that we

could compare the total length of preening episodes with and without pebbles to assess whether pebble use affected preening efficiency.

The second stage involved 20 hours of observations of the remaining 12 subjects in the aviary, over thirteen days. The experimenters watched the group and, whenever an individual began to preen or interact with objects, these episodes were recorded on video. At the end of each preening episode, or after two minutes of preening (whichever occurred first), the subject was interrupted and offered a black token, which they could exchange for a piece of food. This was done to inspect the subject's mouth for any objects. We did not record entire preening episodes for other subjects, as we did not plan to compare the length of their preening episodes to Bruce's. Most importantly, the observation of the other subjects allowed the experimenter to verify whether the subject had been holding any objects in their beak as they preened. Where the subject was perched out of reach, the video was zoomed into its beak until the experimenter could determine whether the subject was holding an object in its beak.

Wherever possible, across both observation stages, experimenters attempted to retrieve any objects subjects interacted with, either for preening or during other object manipulation behaviours. Experimenters also collected thirteen randomly selected pebbles and stones from across the aviary which were smaller than or equal to the largest stone manipulated by kea during our observation sessions. To ensure that pebbles were selected randomly, we drew a map of the aviary divided into 1m^2 squares with a grid. Thirteen coordinate locations were randomly generated, and the experimenter visited each of those in turn. Upon arrival in the middle of the pre-determined square, they closed their eyes and spun around. After opening their eyes, they took the pebble or stone closest to their right foot, provided it was smaller than or equal to the largest stone other subjects had interacted with in the past (Figure 1). If it was bigger than the largest stone any of the kea had interacted with during observation sessions, the procedure was repeated until an appropriately sized stone was sampled. This ensured that we

did not select any disproportionately large stones or rocks, which not even subjects with complete beaks would have been able to hold. Their surface areas were then measured digitally from photographs to compare Bruce's pebble selections to a random subset of those available in the aviary.



Figure 1. Photographs of objects that subjects other than Bruce interacted with and subsequently dropped, which experimenters were able to retrieve. There were no instances of these subjects preening whilst manipulating any objects.

2.3.2. Video Coding & Analyses

Video data was coded using BORIS v.7.9.15 (Behavioral Observation Research Interactive Software; Friard & Gamba, 2016). Behaviours were classed using a purpose-designed ethogram (Table 2).

Behaviour	Description
<i>Pebble Manipulation</i>	Start and end times for all interactions with pebbles (defined as stones < 1cm diameter).
<i>Non-pebble Object Manipulation</i>	Start and end times for all interactions with objects other than pebbles.
<i>Preening</i>	Start and end times for instances of preening, that is, instances where a subject's beak was in contact with any part of its body. Further labelled to distinguish preening in one of five body parts: wing, back, neck, chest, or tail.

Table 2. Videos were coded with three different state events, for which coders identified start and end times. The raw time stamp data was then used to calculate the duration of preening episodes with and without pebbles, as well as counts for instances of non-preening pebble manipulations, non-pebble object manipulations, and Bruce's retrieval or replacement of preening pebbles within preening episodes.

Preening episodes were defined as lengths of time where the subject repeatedly touched their beak to any part of their body and ended when there was no beak contact for >30 seconds. All observations were coded by one experimenter, while a second experimenter independently coded a random 10% sample of all recorded observations. Inter-coder agreement between experimenters was high for both stages, both in terms of preening behaviours and object manipulations (Table 3).

Observation Stage	Behaviour category	ICC	95% Confidence Intervals
1 and 2	Preening	0.894	0.767 - 0.952
1*	Pebble and Non-pebble Object Manipulations	0.925	0.790 - 0.974

Table 3. Intraclass correlation coefficients for coder’s identification of preening behaviours (all body parts, with and without pebbles) and object manipulation behaviours (including pebbles and non-pebble objects). *ICCs could not be generated for object manipulation in stage 2 due to limited variance.

Inter-coder reliability was analysed using the irr package (Gamer, Lemon, & Singh, 2012) within R version 3.6.0 (R Core Team, 2019). To assess the reliability of observations made by two experimenters on the random 10% subset of shared videos, we calculated a two-way random, absolute, average-measures intraclass correlation coefficient (*ICC*; Hallgren, 2012). Under this analysis, the degree of agreement between experimenters is evaluated using strict absolute differences and results can be generalised to any other potential observers possessing similar characteristics (Koo & Li, 2016).

ICCs were generated for two pooled behavioural categories to maximise the sample size available: preening behaviour (comprising preening any of five body parts: wing, back, neck, chest, or tail) or object manipulation (comprising pebbles and non-pebble objects). Outputs were interpreted using previously cited thresholds (Cicchetti, 1994), where $ICC \geq 0.75$ suggests excellent inter-coder agreement, $0.6 \leq ICC < 0.75$ good agreement, $0.4 \leq ICC < 0.6$ moderate agreement, and $ICC < 0.4$ poor agreement.

To assess whether Bruce’s pebble manipulations co-occurred with preening, we used a Bayesian contingency table test with a Poisson sample distribution to compare his object manipulation frequency when object manipulation occurred alone or concurrently with preening. Second, we investigated whether the thirteen pebble tools selected by Bruce and

recovered by the experimenters were deliberately selected for their properties by comparing their sizes to those of thirteen randomly selected pebbles from across the aviary, with a Bayesian independent samples t-test. We also assessed the frequency and duration of Bruce's preening of different body parts with and without a pebble across a subset of observations where we could identify which body part was being preened. The frequency data were analysed by a Bayesian contingency table test with a Poisson sample distribution. The duration was analysed by a Bayesian paired samples t-test. Finally, we directly compared the objects manipulated by Bruce to those manipulated by other individuals using a Bayesian independent samples t-test. All main analyses were carried out in JASP v.0.14.1.0 (Jasp Team, 2019) using default priors and all t-tests were two-sided.

2.4. Results

2.4.1. Do Bruce's pebble manipulations co-occur with preening?

Bruce's pebble manipulations typically began with rolling a small pebble (Figure 2) around with his tongue. He then wedged the pebble in his lower mandible, and either rolled it or ran it along his feathers, which were held between the pebble and his tongue. This latter motion appears analogous to other subjects' clasping and grinding of feathers between their upper and lower mandibles, which helps remove ectoparasites lodged between their barbs (Clayton et al., 2005).



Figure 2. Photographs of pebbles Bruce manipulated and preened with which could be retrieved by the experimenters. Tools were retrieved by the experimenters only after Bruce dropped them.

We first examined whether or not Bruce's pebble manipulations were followed by preening. Across 20 hours of video observations recorded over 9 days, we recorded 30 videos where Bruce picked up a pebble and subsequently preened with it at least once (93.75% of cases), while there were only two instances where he picked up a pebble and did not preen with it for the duration of that observation. In one of these two instances, Bruce picked up a pebble and then took part in an aggressive display against another male, subsequently dropping it. In the other, he attempted to preen with the pebble but dropped it, then began interacting with another male. The frequency of his pebble interactions and preening behaviours provides strong evidence that Bruce's preening behaviour was associated with his pebble manipulations; namely, when he interacted with a pebble, he was 1579.72 times more likely to preen with a pebble than not (Bayesian contingency table, $n = 105$, $BF_{10} > 100$, \log_e odds ratio = 7.365).

2.4.2. Are pebble tools valuable and effective to Bruce?

To determine whether Bruce's pebble tooling was deliberate, we examined whether he attempted to retrieve or replace preening pebbles that he dropped during preening. We recorded 250 events where he dropped his preening pebble and subsequently retrieved or replaced it before he resumed preening, out of a total 262 retrieval and replacement events (95.42%). It is also notable that all pebbles used were similar in size ($19.70 \pm 9.14 \text{ mm}^2$; Figure 2), suggesting that his choice of pebble tools was deliberate and functional. To test this, we compared the pebble sizes Bruce used to a random sample of pebbles and stones in his environment ($98.39 \pm 75.62 \text{ mm}^2$; Figure 3) and found that strong evidence to suggest that Bruce selected smaller pebbles than those randomly selected from the aviary (Bayesian independent samples t-test, $BF_{10} = 29.41$). This suggests that Bruce deliberately selected very small pebbles relative to the population of pebbles and stones present in the aviary.



Figure 3. Photographs of pebbles and stones randomly sampled from the aviary by the experimenters.

In total, we observed 103 preening episodes by Bruce, 30 of which involved a pebble. This raises the question of whether Bruce preferred to preen with a pebble only in certain contexts. To this end, we compared the percentage of the time he spent preening different body parts (divided into wings, back, neck, chest, legs, and tail) with and without a pebble. Bruce employed pebbles for preening differentially throughout his body (Bayesian contingency table test, $BF_{10} > 100$; Table 4). He preened most of his body parts without a pebble more frequently than with a pebble (wing: 29.72% vs. 9.94%; back: 8.39% vs. 5.00%; neck: 6.33% vs. 1.44%; chest: 11.22% vs. 3.33%; tail: 2.56% vs. 1.11%). However, he preened his legs without a pebble only 5.78% of the time, but with the pebble this constituted 15.17% of all observed preening behaviours suggesting that this body part was a specific focus of his pebble preening.

Modifiers		Pebble Use		Total
		No	Yes	
Back	Count	151	90	241
	% of total	8.39 %	5.00 %	13.39 %
Chest	Count	202	60	262
	% of total	11.22 %	3.33 %	14.56 %
Leg	Count	104	273	377
	% of total	5.78 %	15.17 %	20.94 %
Neck	Count	114	26	140
	% of total	6.33 %	1.44 %	7.78 %
Tail	Count	46	20	66
	% of total	2.56 %	1.11 %	3.67 %
Wing	Count	535	179	714
	% of total	29.72 %	9.94 %	39.67 %
Total	Count	1152	648	1800
	% of total	64.00 %	36.00 %	100.00 %
BF ₁₀ Poisson		1.321e+58		
N		1800		

Table 4. Bayesian contingency table including counts (total n = 1800) of all of Bruce’s preening bouts (individual instances of contact between his beak and body) across 103 preening episodes, as categorised by body parts (back, chest, leg, neck, wing) and use of a pebble tool (yes or no). We found strong evidence ($BF_{10} > 100$) for differences in the distribution of preening behaviour across the five categories.

Finally, we compared the time Bruce spent preening with and without a pebble, to establish whether pebble use affected his preening time. Bruce preened himself for slightly longer with a pebble than without one (Bayesian independent samples t-test, $BF_{10} = 1.26$; with a pebble: 176.15 ± 208.51 s, without a pebble: 114.28 ± 133.89 s).

2.4.3. How does Bruce's behaviour compare to that of other individuals?

We compared Bruce's manipulations of pebbles and non-pebble objects with those of other individuals. We observed no instances of any other subjects preening with pebbles or other objects, despite them regularly manipulating objects for purposes other than preening. Non-preening object manipulations occurred 202 times across eight other subjects (averaging 25.25 ± 37.53 instances each), compared to 18 times for Bruce. Therefore, Bruce appeared to manipulate non-pebble objects at a comparable or lower rate to other subjects (Table 5) but was the only individual that used pebbles as preening tools. Interestingly, the size and types of objects that other subjects interacted with were also different, appearing to favor larger stones and non-pebble objects which were never used for preening (Bruce: 19.70 ± 9.14 mm²; other subjects: 184.92 ± 172.15 mm²; Bayesian independent samples t-test $BF_{10} = 18.11$). This was despite the fact that Bruce can competently hold larger objects (Figure 4).

Individual	Number of Object Manipulations Observed	BF ₁₀
Boh	1	0.667
Harley	1	0.667
Loki	125	1.661e +32
Megatron	24	7647.734
Moriarty	8	2.069
Neo	39	6.385e +7
Plankton	3	0.533
Taz	1	0.667
Angelina	0	-
Blofeld	0	-
Johnny	0	-
Spike	0	-
Bruce	2*	

Table 5. To examine whether the frequency of Bruce’s non-pebble object manipulations was comparable to those of other birds living in the aviary, we conducted separate Bayesian binomial tests between Bruce and each individual. Out of the 12 other birds, four were not seen to manipulate objects, four manipulated objects at a similar frequency to Bruce ($BF_{10} < 1$), one showed a slightly higher number of object manipulations ($1 < BF_{10} < 3$), and three manipulated objects much more frequently than Bruce ($BF_{10} > 100$; shaded in green). *To standardise the observation duration for all individuals across Stages 1 and 2, as only Bruce was observed for Stage 1 whereas 12 birds were observed for Stage 2, we divided the total number of Bruce’s object manipulations (18) by 12 (rounded to a whole number). This was performed under the assumption that object manipulations were equally likely to occur across Stages 1 and 2, and that each individual in Stage 2 was observed for an equal amount of time.



Figure 4. Photographs of Bruce handling objects larger than his preening pebbles, namely: (a) a slice of carrot, (b) a stone, (c) a piece of bark, (d) a black token used in previous cognitive experiments he was a part of. We also provide a close-up image in (e) demonstrating how he uses his tongue and lower mandible to hold these objects.

2.5. Discussion

We show that kea are capable of innovating self-care tooling, which is reportedly rare in birds (Fayet et al., 2020; McGrew, 2013). Our observations suggest that the preening pebble may be used as a flat surface against which Bruce can grind his feathers, such that the lower mandible and pebble act analogously to the upper and lower mandibles of other individuals. Bruce's pebble preening behaviour strongly suggests deliberate tooling (in line with the criticisms by Auersperg et al., 2020; Dechaume-Moncharmont, 2020; Farrar, 2020; Sándor & Miklósi, 2020; von Bayern et al., 2020). First, Bruce's manipulations of pebbles were almost always followed by preening, suggesting that he picked up the pebble with the intent of using it as a preening tool. Second, Bruce often retrieved or replaced pebbles he lost or dropped, suggesting that he valued the pebble tool during preening. Third, Bruce selected pebbles of a particular size for use as tools, given that his pebble tools differed in size significantly from a randomly drawn sample of stones available in his environment. Fourth, while Bruce manipulated non-pebble objects at a similar rate to other individuals, he was the only subject ever observed preening with a pebble. Finally, when other kea did interact with stones and pebbles, they were significantly greater in size than those selected by Bruce. We therefore show that Bruce innovated pebble preening as a self-care tooling behaviour, likely as a direct consequence of his disability, by systematically testing the predictions for deliberate tooling in a bird species where tooling is not a species-specific behaviour.

Within the framework of tooling (Fragaszy & Mangalam, 2018, 2020), Bruce's pebble preening can be described as an egocentric (self-directed) behaviour, with a single dynamic relation between the tool and his feathers or skin. The placing of the pebble is probably crucial to its functional use. Given that Bruce often repeatedly moves the pebble until he rests it on his lower mandible before he begins to preen with it, it is possible that its orientation is also important. However, this could also serve the purpose of lodging it comfortably within the

mandible, unrelatedly to the function of the pebble. It is unclear if the alignment of the pebble is important. The properties of Bruce's pebble tooling may therefore be different to those observed in foraging stick tooling by both kea and other parrot species (Auersperg et al., 2010, 2011, 2012; Goodman et al., 2018), which is allocentric and alignment-crucial, and the egocentric self-care tooling observed in a range of captive parrot species (Bentley-Condit & Smith, 2010; Janzen et al., 1976; Lefebvre et al., 2002).

Our results suggest that Bruce's tooling is flexible and deliberate. Evolved stereotyped tooling is innate, involving fixed action patterns that cannot be adapted to novel situations (Hunt et al., 2013). It is highly unlikely that Bruce's pebble use is an evolved stereotyped action, given that, to the best of our knowledge, it is unique to him as an individual, and therefore likely a flexible, context-dependent innovation. Furthermore, Bruce does not attempt to preen with objects of a similar size in his environment to the pebble tools or with larger stones, as might occur with a fixed action pattern, despite being able to competently hold larger objects either between his tongue and lower mandible, or between the remnants of his upper mandible and lower mandible.

It is unclear if the pebble tool employed by Bruce improves his preening efficiency, or whether it is adopted for some other function, given that preening episodes with the pebble were longer than those without. It might be that the pebble tool improves preening efficacy but not efficiency, for example by providing increased success in dislodging parasites at the cost of greater time expenditure. Bruce appeared to be selective in terms of the properties of the pebbles he chose, given both that he discarded some pebbles before attempting preening with them, and our data showing that he did not randomly sample from the environment.

Together with kea's other tooling innovations reported in the wild (Goodman et al., 2018) and in captivity (Auersperg et al., 2010, 2011), Bruce's deliberate self-care tooling suggests that kea may excel at innovating context-appropriate tools. This provides additional

evidence for kea's highly flexible problem-solving abilities, as evidenced in previous experiments including both captive and wild populations (Auersperg, Gajdon, & Huber, 2009; Auersperg et al., 2011a, 2011b; Bastos & Taylor, 2019; Goodman et al., 2018; Heaney, Gray, & Taylor, 2017; Huber & Gajdon, 2006; Werdenich & Huber, 2006). The ability to flexibly combine information in a domain-general manner (Bastos & Taylor, 2020), when combined with playfulness and neophilic exploration of the environment, may be an important driver of technical innovations in kea and other species (Gajdon, Lichtnegger, & Huber, 2014; Lambert et al., 2017). These results therefore support recent claims that tool use, including tooling, arises not only from the evolution of specialized physical cognition (Kacelnik et al., 2006; Rutz et al., 2016; Teschke, Cartmill, Stankewitz, & Tebbich, 2011; Weir, Chappell & Kacelnik, 2006), but can be innovated, when ecologically necessary, by species with sufficiently domain-general cognition (Bird & Emery, 2009; Miller et al., 2020; Taylor & Gray, 2014; Teschke et al., 2011).

Chapter III

*Crowdsourcing and phylogenetic inference reveal the evolutionary history of rare tool use behaviour*²

3.1. Abstract

Some animal behaviours are only rarely observed and therefore challenging to study. Tool use, for example, often occurs at a low frequency both within individuals' behavioural repertoires and across individuals of a species. Using tool use in parrots as a case study, we demonstrate two approaches for investigating the phylogenetic prevalence and evolutionary history of a rare behaviour. First, we crowdsource video data to identify instances of parrot tool use not reported in the existing scientific literature, increasing the number of parrot species known to use tools from 11 to 28. We then map the phylogenetic spread of tool use in parrots and reconstruct the ancestral state of this group, both using our data and a zero-inflated phylogenetic multilevel model, to infer which additional species are most likely to display tool use despite a lack of observations of this behaviour. Finally, we test whether encephalisation predicts tool use across parrots, to determine whether tool use is a suitable signal for behavioural innovation and flexibility in this clade. Our results show that tool use distribution is explained by phylogenetic relatedness in parrots and is an ancestral trait, with a later loss occurring in the Psittaculidae family. We also show that tool use is not predicted by encephalisation within parrots. These findings not only change our understanding of the evolution of tool use in birds, but also demonstrate how data crowdsourcing and phylogenetic analyses can be combined to deepen our understanding of the evolution of rare behaviours.

² Based upon: Bastos, A. P. M., Claessens, S., Nelson, X. J., Atkinson, Q. D., Taylor, A. H. (*in prep*) Crowdsourcing and phylogenetic inference reveal the evolutionary history of rare tool use behaviour.

3.2. Introduction

Tool use is present in only a few species spread across distantly related taxa, from octopodes (Finn, Tregenza, & Norman, 2009) to the great apes (Whiten et al., 1999). Widely regarded as a marker of intelligence (Byrne, 1997; Hunt, Gray, & Taylor, 2013; Kacelnik, 2009; Seed & Byrne, 2010; Taylor & Gray, 2014), some researchers suggest that using tools may require complex cognitive skills, including planning, insight, and an understanding of causality, rigidity, and connectedness (Seed & Byrne, 2010). Research with apes and corvids suggests that a sophisticated understanding of physical cognition may underpin flexible and context-appropriate tool use (Boeckle et al., 2020; Gruber et al., 2019; Mulcahy & Call, 2006; Taylor et al., 2009, 2012; von Bayern et al., 2009, 2018). Although parrots perform comparably to the great apes in a range of physical cognition tasks (Auersperg et al., 2009; Bastos & Taylor, 2020; Krasheninnikova, Berardi, Lind, O'Neill, & von Bayern, 2019), their tool use is relatively understudied.

To date, tool use has only been documented in 11 species of parrots. In the wild, palm cockatoos (*Probosciger aterrimus*) use sticks to drum on hollow trees in their courtship displays (Heinsohn et al., 2017), hyacinth macaws (*Anodorhynchus hyacinthinus*) use wedges to open nuts (Borsari & Ottoni, 2005), and kea (*Nestor notabilis*) prod pest control traps with sticks to safely obtain the bait inside, though they do not display tool use in their natural foraging behaviours (Goodman et al., 2018). In captivity, greater vasa parrots (*Coracopsis vasa*) use date pits and pebbles to break up and scrape calcium powder from seashells (Lambert et al., 2015), kea use sticks in cognitive tasks (Auersperg et al., 2009, 2010, 2011a, 2011b) and can innovate the use of pebbles for preening (Bastos, Horváth, Webb, Wood, & Taylor, 2021a) and Goffin's cockatoos (*Cacatua goffiniana*) can innovate the use of a stick tool to bring food within reach (Auerperg et al., 2012). Goffin's cockatoos have also since been shown to use tool sets to forage in the wild (O'Hara et al., 2021). Anecdotal accounts also report tool use in

African grey parrots (*Psittacus Erithacus*; Janzen et al., 1976), yellow-crowned amazon parrots (*Amazona ochrocephala*; Lefebvre et al., 2002), and four additional cockatoo species (*Cacatua galerita*, *C. moluccensis*, *C. sanguinea* and *C. sulphurea*; Bentley-Condit & Smith, 2010; Lefebvre et al., 2002). Among these anecdotal accounts, four include self-care tool-use (in these cases, the use of tools for scratching) in African grey parrots and three cockatoo species (*Cacatua moluccensis*, *C. sanguinea*, and *C. sulphurea*; Bentley-Condit & Smith, 2010; Lefebvre et al., 2002). Together with a recent study in kea (Bastos et al., 2021a), these comprise the only existing reports of self-care tool use in parrots. Some authors have claimed this form of tool use to be rare in birds (Fayet et al., 2020; McGrew, 2013), with the notable exception of ‘anting’, a highly stereotyped behaviour where birds wipe insects onto themselves (Morozov, 2015), that occurs only in passerines (Bentley-Condit & Smith, 2010). Although the literature reports tool use behaviours in only a few parrot species, whether this is a true reflection of its prevalence, or whether it is only rarely observed and therefore underreported, is unclear.

Crowdsourcing offers one avenue through which researchers can investigate rare or underreported behaviours such as parrot tool use. Over 70% of all extant parrot species are bred in the aviculture industry and are kept as pets across the world (Anderson, 2003; Carrete & Tella, 2008; Drews, 2001; Kelly, McCarthy, Menzel, & Engebretson, 2014; Li & Jiang, 2014; Su, Cassey, Vall-Ilosera, & Blackburn, 2015; Young et al., 2012). Given their popularity as household pets and the limited scientific literature on the behaviour and cognition of most parrot species, here we use crowdsourcing of videos shared on a visual media platform to search for additional evidence of tool use in parrots. Crowdsourced data can provide insights into several aspects of animal behaviour and cognition, especially when instances of behaviour are rare (Klump et al., 2021; Krueger, Esch, & Byrne, 2019; Nelson & Fijn, 2013). Detecting additional observations of rare tool use behaviours in parrots would provide a first

opportunity to phylogenetically map how widespread these behaviours are. Additionally, determining the prevalence of rare behaviour across a wider range of species can also inform our understanding of that behaviour's evolutionary history and increase the certainty with which we make inferences about the ancestral state of these behaviours at different taxonomic levels. To this end, we use Bayesian comparative phylogenetic methods to model the prevalence and evolution of the capacity for tool use in Psittaciformes.

Finally, given our increased dataset on parrot tool use, we test whether encephalisation predicts tool use in parrots, to determine whether this behaviour is predictive of behavioural innovation and flexibility within the Psittaciformes order, rather than a phylogenetically determined behaviour (Auersperg et al., 2012; Bird & Emery, 2009; Emery, 2006; Kacelnik, 2009).

In sum, this study aims to compile crowdsourced observations of tool use behaviours in parrots in order to establish additional parrot species not previously identified as tool users, and model the evolution of tool use across the parrot order and use this information to: (i) infer the evolutionary trajectory of tool use in parrots, (ii) test the hypothesised role of encephalisation in the evolution of tool use, and (iii) identify the probability that there are additional unidentified tool using species which should be the effort of future research.

3.3. Methods

3.3.1. Video Searches and Coding

Our video search was conducted on YouTube. Search terms included “parrot using tool” and variants (e.g., “macaw using tool”, “lorikeet using tool”, “parakeet using tool”), “tool use in parrot”, “parrot tool use”, “parrot scratching itself” (included after we found several videos demonstrating self-care tool use in previous searches) and equivalent terms (e.g. “parrot preening itself”, “parrot grooming itself”, “parrot scratching”). For all species that did not display results including object manipulation or scratching behaviours, we also searched the species’ common name(s) + “tool use”, as well as the species’ common name(s) + “scratching”. We also searched for translations of the terms “parrot tool use” and “parrot scratching” in languages for all countries where bird ownership was reported >5% (*Global GfK survey: Pet ownership, 2016*), namely, Turkish, Czech, Polish, French, Italian, Dutch, German, Russian, Spanish, Portuguese, and Mandarin.

When we found a relevant video, we also searched for similar content uploaded by the same person/channel. For each YouTube search conducted, we watched all relevant videos until we reached five consecutive videos that did not feature any parrots. At this point, we ended that search and initiated the next. In line with the recommendations made by Nelson and Fijn (2013), we planned to exclude any videos that consisted of four or more shots edited together so as to ensure the behaviours being observed were not edited or manipulated, but none of the videos obtained qualified for exclusion.

All videos featuring parrots manipulating objects were investigated for potential tool use or borderline tool use. We defined tool use as the manipulation of an unattached object as an extension of the beak or foot to achieve a goal towards another object, individual, or oneself (Shumaker et al., 2011). Borderline tool use was similarly defined, except that it involved the use of an object that was still attached to a substrate (Seed & Byrne, 2010). For example, if

individuals used a fallen feather or stick for self-scratching this was defined as tool use, but using one's currently attached tail feathers or cage furnishings for the same purpose was defined as borderline tool use. Self-scratching had to involve slow and repeated movements of touching an object to one's body (or rubbing repetitively against an attached object, in the case of borderline tool use; Sándor & Miklósi, 2020).

All relevant videos were coded for video length, species, tool use presence (yes/borderline), tool use type (e.g., scratching, feeding), the object being used (e.g., feather, stick), tool use target, human interference with the action (talking or handing object to parrot, holding parrot), and the number of shots within each video. Our complete dataset also includes the name for each video, link, subject name, sex (as declared by owner, as most parrot species are not sexually dimorphic), publishing date, and date found (and coded).

3.3.2. Ethical Statement

This study did not require ethics approval; videos are publicly available under either a CC BY licence or a standard YouTube license.

3.3.3. Phylogenetic Data

We used the phylogenetic tool at www.birdtree.org (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012) to compile 1000 posterior draws of phylogenetic trees for 174 parrot species. These included 25 of the tool-using species identified in our study (exceptions were *Aratinga nenday*, *Psittacara erythrogenys*, and *Psittacus timneh*, whose genomes were not available in the database). A single maximum clade credibility tree was generated from these posterior draws for visualisation purposes. In our main phylogenetic analyses, we iterated over posterior draws of the phylogeny to account for phylogenetic uncertainty. We calculated Pagel's λ for tool use

as a binary trait (present vs. absent) by using the *fitDiscrete* function from the *geiger* R package (Pennell et al., 2014).

3.3.4. Zero-inflated Phylogenetic Multilevel Model

In our data, there are many parrot species that have no record of tool use in the scientific literature and/or have zero videos of tool use on YouTube. There are two ways in which such zeroes could arise in our data. First, the parrot species might not use tools. Second, the parrot species might use tools, but this tool use has not yet been recorded in the scientific literature or on YouTube. These two ways of generating zeroes in our data imply that, even with our extensive literature review and video survey, several tool-using parrot species may have yet gone undetected. To identify species that are likely undetected tool users, we fitted two Bayesian zero-inflated phylogenetic multilevel models simultaneously to the data, one modelling the presence/absence of tool use in the literature for each species, and the other modelling the number of tool use videos identified on YouTube for each species.

For the presence/absence of tool use in the literature, we fitted a zero-inflated Bernoulli model to the presence/absence binary outcome variable. This model assumes that each parrot species has its own latent probability of tool use. Observed zeroes (absences in the literature) can arise in one of two ways: either tool use is absent, or tool use is present and has not yet been identified. Thus, we also estimate the probability with which tool use is correctly identified for tool-using species in the literature, and allow “research effort” (proxied as log number of search hits on YouTube) to predict this probability.

For the number of tool use videos identified on YouTube, we fitted a zero-inflated Poisson model to the “number of videos” count variable. This model assumes that each parrot species’ latent probability of tool use is identical to the zero-inflated Bernoulli model. Observed

zeroes (no videos recorded on YouTube) suggest either tool use is absent, or tool use is present but no videos have yet been recorded. Thus, we also estimate the rate at which videos of tool using parrot species are correctly produced for tool-using species, and allow “research effort” (log number of search hits on YouTube) to predict this rate.

In both zero-inflated models, fitted simultaneously, latent probabilities of tool use are allowed to covary between species according to phylogenetic relatedness. We modelled this using an Ornstein-Uhlenbeck Gaussian process (McElreath, 2020). To incorporate phylogenetic uncertainty, we iterated over 100 posterior draws from the phylogeny. We used weakly informative priors for all parameters. The full model is as follows:

$$\begin{aligned}
\text{LitExists}_i &\sim \text{ZIBernoulli}(p_i, \gamma_i) \\
\text{VideosFound}_i &\sim \text{ZIPoisson}(p_i, \lambda_i) \\
\text{logit}(\gamma_i) &= \alpha_\gamma + \beta_\gamma \log(\text{SearchHits}_i) \\
\log(\lambda_i) &= \alpha_\lambda + \beta_\lambda \log(\text{SearchHits}_i) \\
\text{logit}(p_i) &= \alpha_p + \kappa_{\text{SPECIES}[i]} \\
\kappa &\sim \text{MVNormal}((0, \dots, 0), \mathbf{K}) \\
\mathbf{K}_{ij} &= \eta^2 \exp(-\rho^2 D_{ij}) \\
\alpha_\gamma, \alpha_\lambda, \alpha_p, \beta_\gamma, \beta_\lambda &\sim \text{Normal}(0, 1) \\
\eta^2 &\sim \text{Exponential}(0.5) \\
\rho^2 &\sim \text{HalfNormal}(3, 1)
\end{aligned}$$

where LitExists_i is the binary presence/absence of tool use in the literature for each species, VideosFound_i is the number of tool use videos recorded for each species, SearchHits_i is the number of search hits on YouTube, and D_{ij} is the phylogenetic patristic distance between species i and j . Our main quantity of interest is $1 - p_i$, the latent probability of tool use for each species.

3.3.5. Ancestral State Reconstruction

To infer the presence or absence of tool use for different nodes in the phylogenetic tree, we used ancestral state reconstruction methods as implemented in the software *BayesTraits* v3.0 (Pagel et al., 2004). We fitted three Bayesian Multistate models with different binary dependent variables: (i) presence vs. absence of tool use in literature only, (ii) presence vs. absence of tool use in literature and/or videos, and (iii) presence vs. absence of tool use in literature, videos, and/or additional species identified from our zero-inflated phylogenetic multilevel modelling. For each model, we estimated the posterior probability of tool use presence for every ancestral node in the tree. These models were iterated over all 1000 posterior draws from the phylogeny.

3.3.6. Encephalisation Data

To examine the relationship between encephalisation and the presence of tool use across parrot species, we collated data from the literature (Flammer, Whitt-Smith, & Papich, 2001; Iwaniuk, Dean, & Nelson, 2004; Mazenganya, Bhagwandin, Manger, & Ihunwo, 2018; Olkowicz et al., 2016; Schuck-Paim, Alonso, & Ottoni, 2008; Silva, Guzmán, Urantówka, & Mackiewicz, 2017) for all known body mass (g) and brain mass (g) values across parrots. We calculated the encephalisation quotient (EQ), using the Jerison's (1973) formula: $EQ = \text{brain weight} / (0.12 \times \text{body weight}^{(2/3)})$, for all species for which we obtained body and brain mass data. These data were compiled into a list of all 381 extant species (Gill, Donsker, & Rasmussen, 2020). We then included tool use presence/absence data, including all cases of true tool use behaviour, either reported in the literature (Bentley-Condit & Smith, 2010; Borsari & Ottoni, 2005; Goodman et al., 2018; Heinsohn et al., 2017; Janzen et al., 1976; Lambert et al., 2015; Lefebvre et al., 2002) or observed in our video survey. We found body mass and brain mass data for all tool-using species in our dataset, with the exception of three species: *Diopsittaca nobilis*,

Psittacara erythrogenys and *Coracopsis vasa*. For the latter, we used values for the closely related *Coracopsis nigra*. The other two species were excluded from the final dataset, which comprised 168 species for which we found no evidence of tool use, and 26 species for which we did find evidence of tool use (either from the literature or from our video search), totalling 194 species.

To determine whether encephalisation predicted the probability of tool use after controlling for shared phylogenetic history and research effort, we fitted a Bayesian phylogenetic logistic regression with encephalisation quotient and log number of YouTube search hits as predictors. To account for phylogeny, we allowed species varying intercepts to covary according to an Ornstein-Uhlenbeck Gaussian process (McElreath, 2020). To account for phylogenetic uncertainty, we iterated over 100 posterior draws from the phylogeny.

3.4. Results

3.4.1. Description of Survey Findings

We found 116 videos of 104 individuals from 25 species performing either true tool use ($n = 100$ videos; 22 species) or borderline tool use ($n = 16$ videos; across 7 species). All borderline tool use cases were excluded from further analyses. For species showing true tool use, 15 (60%) were represented by two or more individuals over multiple independent observations. All videos featured pet parrots in captive settings. True tool use always involved the subject using an object for self-scratching (95.7% involved scratching the head and/or neck). The most common tool ($n = 53$, 45.7%) was a moulted feather. Human-made objects (e.g. pens, spoons, pieces of wood, cardboard) were also common.

According to YouTube video descriptions and owner comments, 53 of the individuals were males and 23 were females. No sex information was provided for the remaining 28

subjects. As owners provided no information on whether sex had been established through genetic testing, and sexual dimorphism in parrots is rare (Bercovitz, 1987; Hoyo & Bierregaard, 2011), we could not typically ascertain if descriptions were accurate. It is unclear if the disproportionately large number of males in the sample is a consequence of (i) owners more readily assuming their parrots are male when they have not been genetically tested, (ii) owners being more likely to own or film male parrots, or (iii) male parrots exhibiting more tool use behaviours than female parrots.

Owners did not appear to interfere with the subjects' actions in 68/116 observations (58.6%). In 43 cases (37.1%), there was potential interference, either through being in close physical contact with the bird (e.g., bird perching on hand), talking to the bird, or handing it the tool. We could not establish the degree of interference in the remaining 5 videos (4.3%), as sound had been removed or was substituted by music.

3.4.2. Phylogeny of Tool Use in Parrots

Our video survey resulted in true tool use observed in 22 species, all of which were cockatoos (Cacatuidae), Old World parrots (Psittacinae), or neotropical parrots (Arinae). The most common species in our dataset, accounting for 41 of our observations ($n = 37$ unique individuals), was the green-cheeked conure (*Pyrrhura molinae*). We found no evidence of tool use in any species of Psittaculidae, despite this family containing some of the most commonly kept pet species (including lovebirds, lorikeets, and Asian parakeet species).

Among species without video evidence for tool use, the scientific literature identifies another six species for which there is evidence of true tool use in the literature; these were hyacinth macaws (*Anodorhynchus hyacinthinus*), palm cockatoos (*Probosciger aterrimus*), greater vasa parrots (*Caracopsis vasa*), kea (*Nestor notabilis*), yellow-crowned amazons (*Amazona ochrocephala*), and little corellas (*Cacatua sanguinea*). Again, the literature

contains no reports of tool use among the Psittaculidae. Overall, we identified 28 tool-using species, compared to the 11 previously reported in the scientific literature (Figure 5). There was evidence for tool use from both the literature and crowdsourced data for only five species.

The identification of new tool-using species in our video survey changes our inferences about the evolution of tool use in parrots. We estimated phylogenetic signal (Pagel's λ) of tool use using both the pre- and post-survey data. Pagel's λ varies between 0 and 1, where 0 implies that the distribution of a trait across species is unexplained by phylogenetic relatedness and 1 implies that the distribution of a trait across species is fully explained by phylogeny. Using the evidence of tool use from the literature alone (pre-survey data), we estimated $\lambda = 0.25$, a value which likelihood-ratio testing revealed was no different from zero ($p = 0.85$). In comparison, combining the evidence from both the literature and the video survey (post-survey data) resulted in higher phylogenetic signal. With these data, we estimated $\lambda = 0.58$. A likelihood ratio test revealed that this value was significantly greater than zero ($p < 0.01$). Thus, the results of our video survey increase the extent to which the distribution of tool use across parrot species can be explained by phylogenetic relatedness.

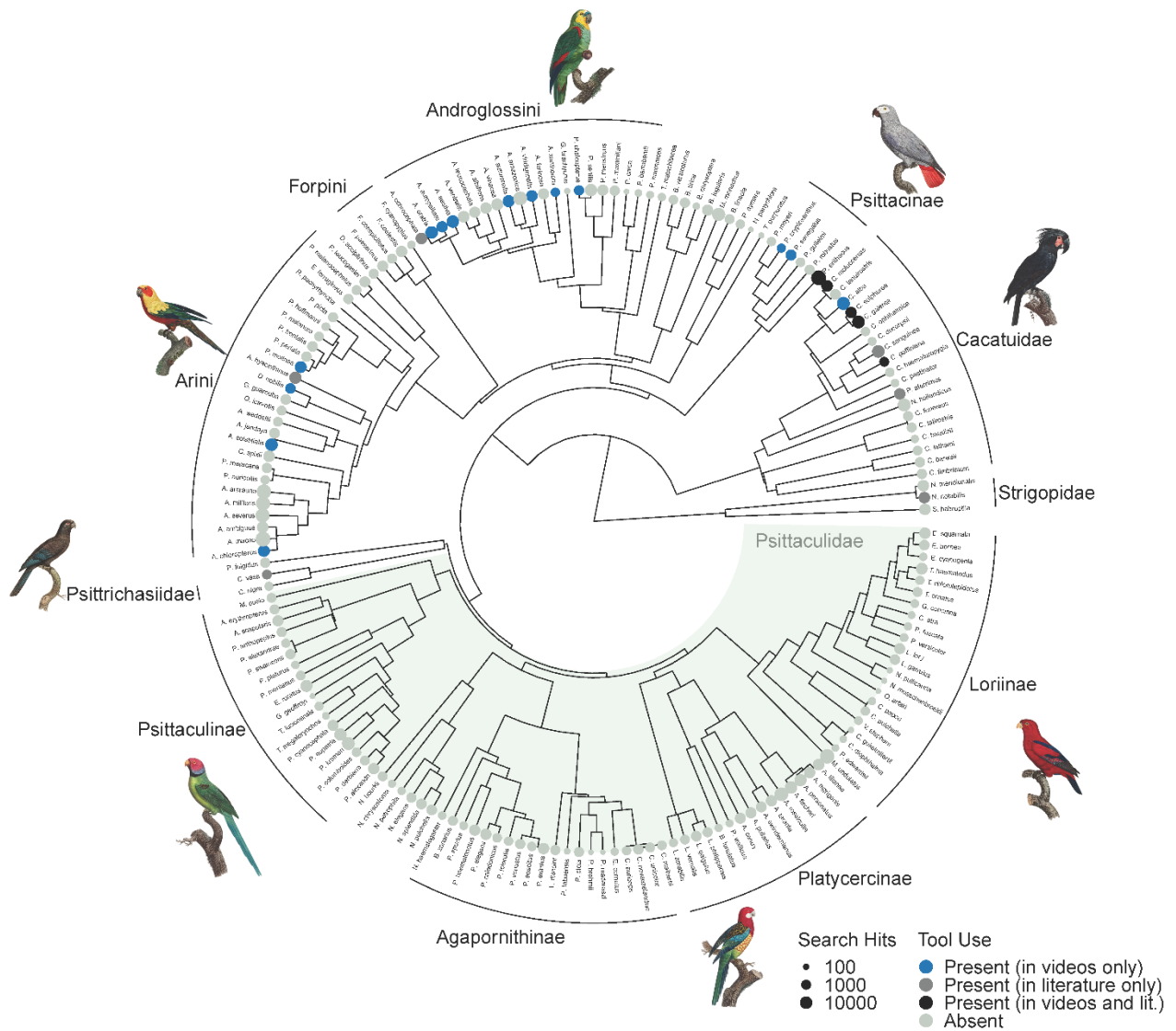


Figure 5. Phylogenetic tree of 174 parrot species representing all parrot subfamilies. 25 of the 28 tool-using species identified here are marked by either grey, blue, or black points at the tips of the tree. Grey points represent species in which tool use was identified in the literature, but not the video survey. Blue points represent species in which tool use was identified in the video survey, but not the literature. Black points represent species in which tool use was identified in both the literature and the video survey. Green points represent absences of tool use. Points are scaled according to the number of YouTube search hits for that species. Three species are not shown because their genomic data was unavailable, namely: *Aratinga nenday* (Arini), *Psittacara erythrogenys* (Arini), *Psittacus timneh* (Psittacinae). We found no evidence for tool use in any species from the family Psittaculidae (shaded in green).

3.4.3. Identifying Additional Tool-Using Species

In addition to the 28 tool-using species identified in the literature and our video survey, we also fitted a Bayesian zero-inflated phylogenetic multilevel model to identify further species that are likely to be undetected tool-users (i.e., tool-using species with no tool use reported in the literature and no videos recorded). Figure 6 visualises the posterior log odds of tool use (fitted values) for the top 50 species, ranked by the likelihood of tool use. As expected, the model predicts that species with evidence of tool use in both the literature and the video survey are the most likely to be tool users, and species with evidence from only one source are relatively less likely. In addition, 17 species of parrot with an absence of evidence are ranked among the species likely to display tool use, largely due to phylogenetic inertia from closely related tool-using species. For example, the five members of the *Cacatua* genus without evidence for tool use are ranked the highest, since the other six members of the genus have all been recorded as tool-users. Similar phylogenetic effects can explain the high rankings of members of *Amazona* and *Poicephalus* genera.

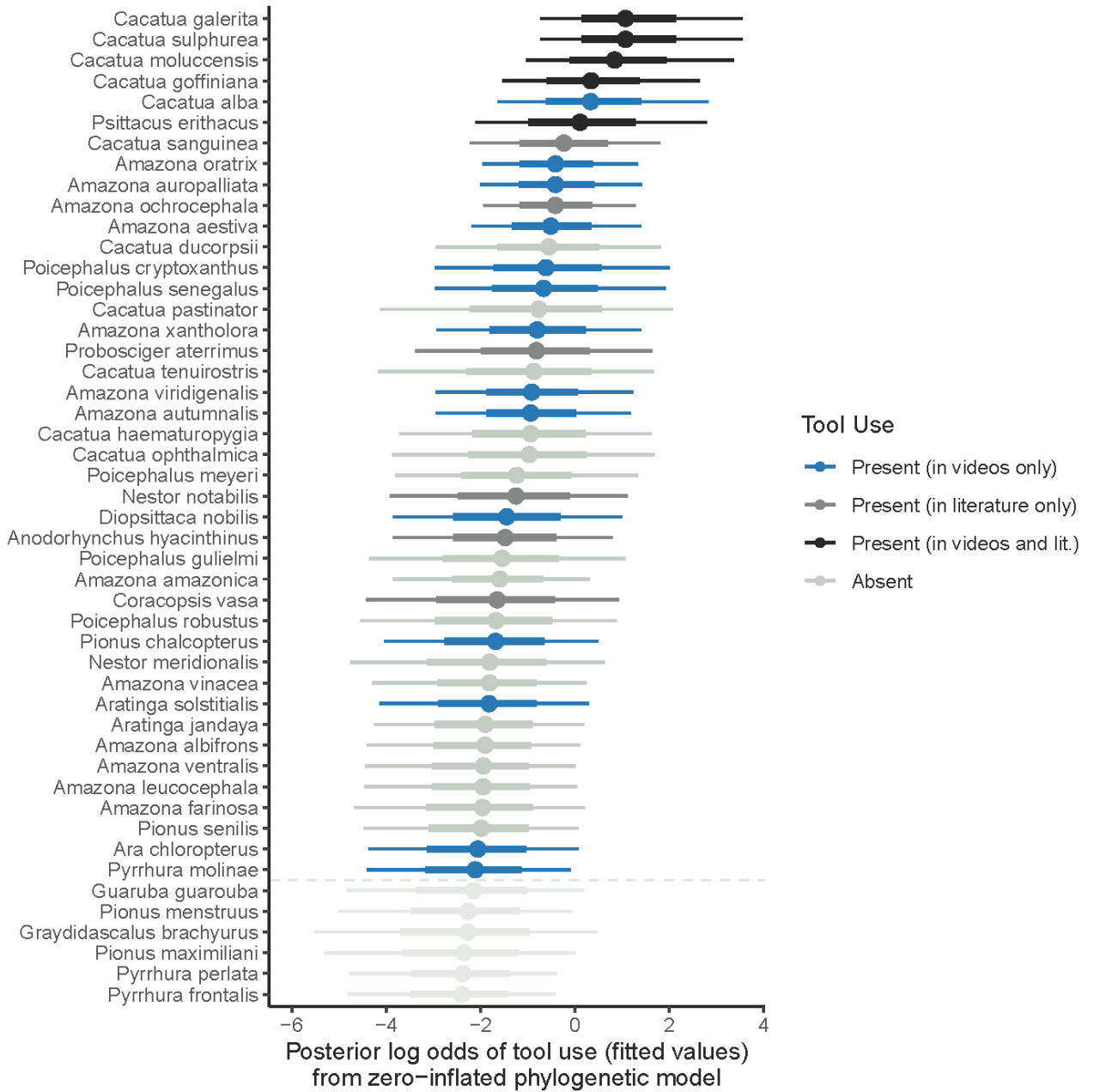


Figure 6. Results of zero-inflated phylogenetic multilevel model. The 50 most likely tool-using species are ranked according to their estimated probability of tool use (fitted values). For the remaining 124 species, tool use is absent in both the literature and the video survey, and they lie below the dashed green line. Points represent median posterior log odds of tool use for each species. Lines represent 50% and 95% credible intervals.

3.4.4. Ancestral State Reconstruction

Given the distribution of tool use on the tips of the phylogeny, we can also make inferences about the evolutionary sequence of tool use in parrots. For example, the lack of tool use across the entire Psittaculidae family raises the possibility that tool use was present in the ancestor of all parrots but was lost in the common ancestor of the Psittaculidae family. To explore the evolutionary history of tool use in parrots, we conducted ancestral state reconstructions for two nodes in the phylogenetic tree – the ancestor of all parrots in the tree, and the ancestor of Psittaculidae – estimating for each node the probability that tool use was present or absent (Figure 7). When using evidence of tool use from the literature alone (pre-survey data), the model is equivocal regarding the presence of tool use at these two nodes. For the pre-survey data, the median posterior probability of tool use at the root of the tree is 0.50 (95% CI [0.50 0.51]), and the median posterior probability of tool use in the ancestor of Psittaculidae is also 0.50 (95% CI [0.14 0.50]). However, when introducing the species identified in the video survey (post-survey data), the model begins to favour the loss of tool use in the Psittaculidae. For the post-survey data, the median posterior probability of tool use at the root of the tree increases to 0.52 (95% CI [0.45 0.95]), and the median posterior probability of tool use in the ancestor of Psittaculidae decreases to 0.11 (95% CI [0.00 0.50]). Further supporting this pattern, when additionally including the 17 species predicted by the zero-inflated phylogenetic multilevel model, the median posterior probability of tool use at the root of the tree increases to 0.93 (95% CI [0.13 0.99]) and the median posterior probability of tool use in the ancestor of Psittaculidae decreases to 0.001 (95% CI [0.000 0.007]). Although these latter results should be interpreted with caution, since this model does not incorporate the uncertainty surrounding the 17 additionally predicted species, these results generally support a sequence of evolution in which the ancestor of all parrots used tools, followed by the subsequent loss of this trait in the Psittaculidae family.

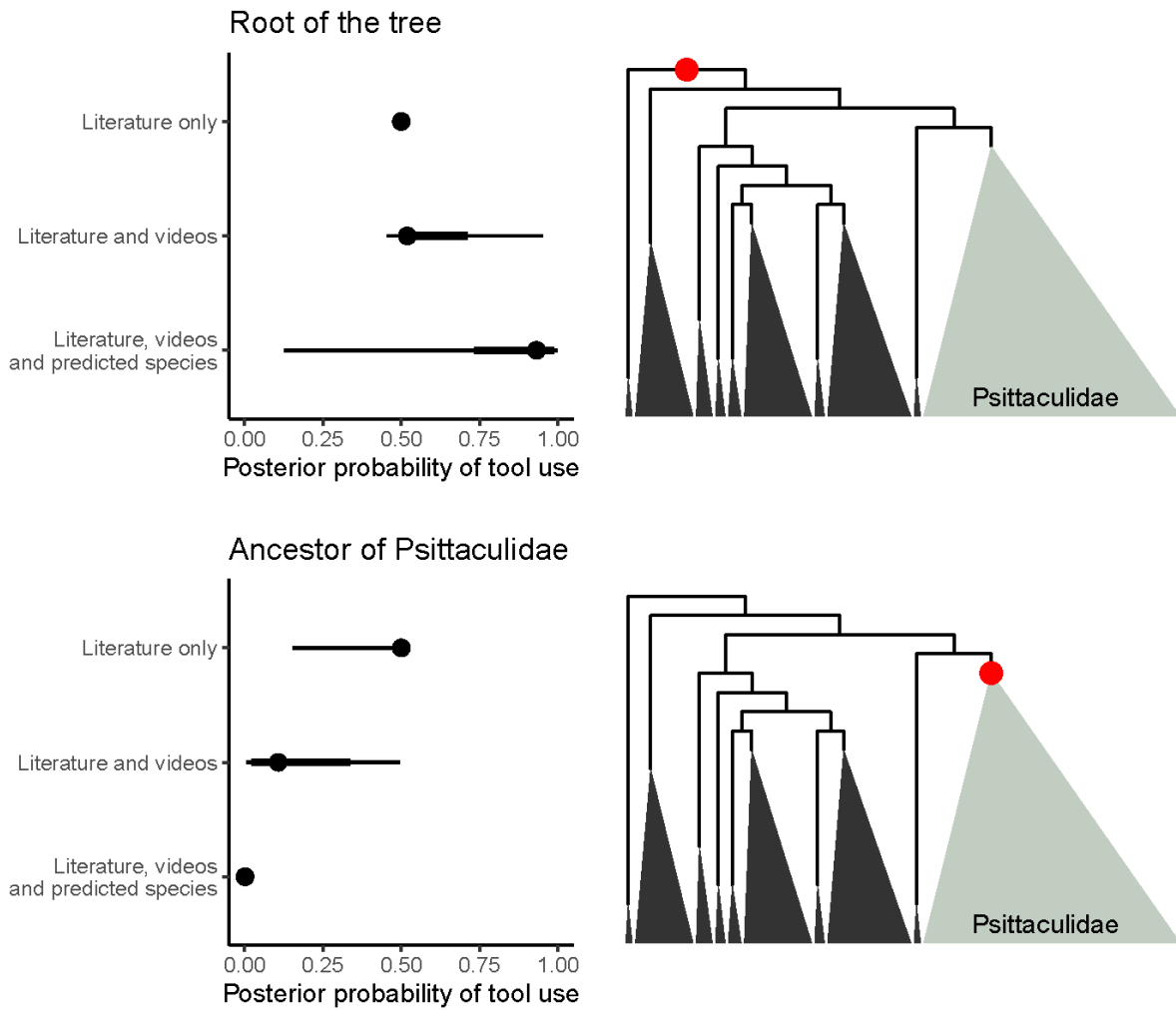


Figure 7. Ancestral state reconstructions for tool use. Posterior probabilities of tool use at two nodes on the tree: (i) the ancestor of all parrots, and (ii) the ancestor of Psittaculidae. Points reflect median posterior probabilities, thick lines represent 50% credible intervals, and thin lines represent 95% credible intervals.

3.4.5. Is Tool Use a Signal of Innovation?

Results of our phylogenetic modelling suggests that tool use in parrots is predicted by phylogeny. However, previous literature has considered parrot tool use as a signal of behavioural innovation and flexibility, rather than an evolutionarily ancient, phylogenetically determined behaviour (Auersperg et al., 2012; Bird & Emery, 2009; Emery, 2006; Kacelnik, 2009). This innovation hypothesis suggests that, given the established correlations between encephalisation behavioural innovation, encephalisation should be a better predictor for tool use than phylogeny. We therefore examined if encephalisation predicted tool use in the parrot order, controlling for phylogenetic relatedness and research effort. In a phylogenetic logistic regression, we found that encephalisation quotient did not predict the probability that a species had evidence of tool use in either the literature or the video survey (median posterior log odds slope = 0.17, 95% credible interval [-1.23 1.67]). Thus, we find no support for the hypothesis that tool use occurs primarily in parrot species with greater encephalisation.

3.5. Discussion

Our video survey increases the number of known tool-using parrot species by 17 species, from 11 to 28. Phylogenetic mapping shows that tool use is widespread in the parrot order and its distribution is predicted by phylogeny, but not by encephalisation. Ancestral state reconstruction modelling suggests that a capacity for tool use is an ancestral trait in the parrot order, while the results of our zero-inflated modelling suggest, based on phylogenetic relatedness, that there are still additional parrot species that are likely capable of tool use but for whom this rare behaviour has thus far gone undetected, highlighting potential species of interest for future studies on parrot tool use.

Our observations usually consisted of individual data points, or anecdotes, which are sometimes unreliable sources from which to make claims about animal cognition (Auersperg et al., 2020; Bastos et al., 2021a; Dechaume-Moncharmont, 2020; Farrar, 2020; Sándor & Miklósi, 2020; von Bayern et al., 2020). Nevertheless, anecdotes and case studies of rare behaviours can provide valuable insights into animal cognition and behaviour, provided they are substantial and cautiously interpreted (Bastos et al., 2021a; Sándor & Miklósi, 2020; von Bayern et al., 2020). Even though the ideal method for determining each individual's tool use is intentional is through systematic hypothesis testing (as in Bastos et al., 2021a), this approach is not possible with crowdsourced data. Nevertheless, there are four reasons to suggest that these observations were mostly, if not all, examples of intentional tool use. First, in our data, tools were used slowly and repetitively over long periods, suggesting that parrots were making intentional use of their scratching tools (see: Sándor & Miklósi, 2020). Second, we observed multiple videos of the same or equivalent behaviours being performed by the same individual, which suggests that at least in some pet parrots this behaviour is repeated, suggesting intentionality (Bastos et al., 2021a). Third, for 60% of the species we report on, we also found two or more videos of repetitive and sustained scratching by different individuals in separate

households, suggesting that the likelihood of such behaviours being accidental combinations of its component parts (holding an object and attempting to scratch oneself) is unlikely. Finally, and most importantly, if all our observations of self-care tool use in parrots were the consequence of accidental combinations of two unrelated actions, then we should not expect to see the phylogenetic signal observed in our data.

Another potential concern with crowdsourced data is its reliability (Auersperg et al., 2020; Dechaume-Moncharmont, 2020; Farrar, 2020; Nelson & Fijn, 2013; Sándor & Miklósi, 2020; von Bayern et al., 2020). In particular, the self-care tool use instances detected in our study might be attributable to training or unintentional cueing by the birds' owners. However, we found little evidence to suggest this. Over half of the videos coded contained no evidence of human interference with the parrots' tool use outside of filming the behaviour. Humans only handed parrots their tools in two of the videos (1.7%), and none of the videos featured owners directly rewarding tool use behaviours with food. These observations are supported by two patterns in our data. First, self-care tool use was not seen in Psittaculidae species, which would be expected if this behaviour was being trained or cued by humans, given that this family comprises some of the most popular pet species (such as lovebirds, lorikeets, and ringneck parakeets). Second, if self-care tool use behaviours were a consequence of training, we should not detect a phylogenetic signal for this behaviour.

Our crowdsourcing revealed that a wide range of parrot species exhibit self-scratching behaviours in captive environments. It is possible that these self-care behaviours may emerge more often in captivity because pet parrots are kept with few or no conspecifics, and therefore must innovate tool use to preen inaccessible body areas, such as their heads and necks, which would have been otherwise preened by flock members (Bastos et al., 2021a). This may indicate that human interaction and allopreening in the absence of a conspecific is insufficient for pet parrots, as suggested by studies on the welfare and chronic stress in lone captive parrots

(Aydinonat et al., 2014; Engebretson, 2006; Mason, 2010; Meehan, Garner, & Mench, 2003). These behaviours might therefore emerge in individuals when ecologically necessary.

Our phylogenetic analyses combined both findings of tool use from the literature and from crowdsourced data. These two sources of data are distinct in the types of tool use they report. In the scientific literature, most examples involve foraging tool use (e.g., Auersperg et al., 2012; Borsari & Ottoni, 2005; Goodman et al., 2018; O'Hara et al., 2021), where individuals use objects to extract food that is otherwise not readily accessible. Our crowdsourcing data, on the other hand, revealed only examples of self-care tool use, where birds use objects to scratch themselves. This distinction might suggest that animal cognition researchers are more interested in or more attentive to foraging tool use than self-care tool use, but this is unlikely given suggestions that self-care tool use in birds is rare and therefore worthy of investigation (Fayet et al., 2020; McGrew, 2013). Instead, we suggest that ecological necessity explains this pattern: in the wild or in laboratory settings, subjects will more often be faced with situations where food is difficult to access, and therefore innovate an appropriate tool for foraging; whereas in human households, pet parrots are less likely to encounter situations where food is not readily available or accessible but are more likely to lack conspecifics or sufficient allopreening from their human caretakers. As such, the results of our phylogenetic analysis suggest that the capacity for tool use is ancestral to all parrots, but the forms of tool use observed in the scientific literature compared to our crowdsourced data suggests that tool use expression varies with ecological necessity.

This finding challenges a key assumption made in the literature to date, that only a minority of especially large-brained parrots innovate tool use (e.g., Auersperg et al., 2012; Auersperg, Borasinski, Laumer, & Kacelnik, 2016; Bastos et al., 2021a; Lambert et al., 2013, Goodman et al., 2018; O'Hara et al., 2021). Past work has suggested that, since tool behaviours do not play a major role in the observed ecology of parrot species, and parrots appear to lack

morphological specialisations for tool use, evidence for innovative tool behaviours in parrots signals the presence of sophisticated physical cognition or domain-general intelligence. Our work casts doubt on these assumptions. In the Psittaciformes, tool use is predicted by phylogeny, but not by encephalisation, and ancestral trait reconstruction suggests that tool use may be ancestral to all parrots. These findings therefore suggest that tool use is not a signal of innovation, or cognitive ability, in the parrot order. Rather, tool use is widespread among the parrot order, with 28 species now known to use tools and the ancestor of the group also likely capable of such behaviour. We therefore provide a clear demonstration of how crowdsourcing can be combined with phylogenetic analyses to identify and trace the evolutionary history of rarely observed animal behaviours. Finally, our results allow us to conclude that, when it comes to parrot tool use, necessity is not the mother of invention. Rather, necessity leads to the expression of a phylogenetically ancient capacity for tool use.

Chapter IV

*Kea fail a string-pulling connectivity task*³

4.1. Abstract

Naïve individuals of some bird species can rapidly solve vertical string-pulling tasks with virtually no errors. This has led to various hypotheses being proposed which suggest that birds mentally simulate the effects of their actions on strings. A competing embodied cognition hypothesis proposes that this behaviour is instead modulated by perceptual-motor feedback loops, where feedback of the reward moving closer acts as an internal motivator for functional behaviours, such as pull-stepping. To date, kea have produced some of the best performances of any bird species at string-pulling tasks. Here, we tested the predictions of the four leading hypotheses for the cognition underpinning bird string-pulling by presenting kea with a horizontal connectivity task where only one of two loose strings was connected to the reward, both before and after receiving perceptual-motor feedback experience. We find that kea fail the connectivity task both before and after perceptual-motor feedback experience, suggesting not only that kea do not mentally simulate their string-pulling actions, but also that perceptual-motor feedback alone is insufficient in eliciting successful performance in the horizontal connectivity task. This suggests a more complex interplay of cognitive factors underlies this iconic example of animal problem-solving.

³ Based upon: Bastos, A. P. M., Wood, P. M., Taylor, A. H. (2021) Kea (*Nestor notabilis*) fail a string-pulling connectivity task. *Scientific Reports*, 11, 15492. doi:10.1038/s41598-021-94879-x

4.2. Introduction

String-pulling behaviour occurs naturally in birds and has been reported in a range of species, being especially prevalent across passerines – particularly corvids –, and parrots (Jacobs & Osvath, 2015). This behaviour describes subjects pulling on a baited string to bring a reward closer, and holding it down with their foot, preventing the reward from falling back down. The swiftness and efficiency of this behaviour in naïve individuals has been attributed by different researchers as evidence for various cognitive mechanisms including planning, means-end understanding, and insight (an immediate understanding of the problem without prior experience; Heinrich, 1995; Heinrich, Bernd & Bugnyar, 2005; Jacobs & Osvath, 2015). All three of these hypotheses would predict that, upon seeing a string with a reward at one end, subjects can mentally simulate the effects of their pulling actions on the string, imagining how they might move the reward closer to themselves. Therefore, upon seeing two strings, only one of which is baited, the same hypotheses predict that animals can mentally simulate their pulling actions on both options, and therefore select the correct string above chance, without first requiring trial-and-error learning.

Contrary to these hypotheses, the perceptual-motor feedback loop hypothesis suggests that, upon encountering a baited string for the first time, subjects will perform exploratory behaviours, and when a set of behaviours – such as pull-stepping – bring the reward on the end of the string closer, or keep it in place, they continue to perform these behaviours, such that they become self-reinforcing and increase in both frequency and efficiency over time (Cheke, Bird, & Clayton, 2011; Taylor et al., 2010; Taylor, Knaebe, & Gray, 2012). Therefore, unlike several other cognitive mechanisms suggested for string-pulling, this embodied cognition hypothesis does not require mental modelling or simulating the effects of subjects' actions on a string (Taylor et al., 2010). Rather, it relies upon subjects having sufficiently large associative

brain areas or more connected perceptual and motor pathways to detect and link the effects of their actions while coordinating pull-step actions, or both (Taylor et al., 2010).

The perceptual-motor feedback loop account therefore predicts that subjects without string-pulling experience should fail at tasks where pulling on the string provides no feedback, that is, does not cause the food reward to move closer to the subject (Taylor et al., 2010, 2012). To date, only naïve New Caledonian crows (*Corvus moneduloides*) have been presented with tasks that directly test this claim, as studies with other species have given subjects experience of this feedback loop before test (Danel, von Bayern, & Osiurak, 2019; Krasheninnikova, Bräger, & Wanker, 2013; Krasheninnikova, 2013; Logan, 2016; Obozova, Bagotskaya, Smirnova, & Zorina, 2014; Obozova & Zorina, 2013; Schmidt & Cook, 2006; Schuck-Paim, Borsari, & Ottoni, 2009; Torres Ortiz, Maxwell, Krasheninnikova, Wahlberg, & Larsen, 2019; Wakonig, Auersperg, & O'Hara, 2021; Wang et al., 2019; Werdenich & Huber, 2006). In one study, naïve New Caledonian crows struggled to pull up a string when the end of the string was visually restricted, which prevented them from perceiving the distance between themselves and the reward (Taylor et al., 2010). In contrast, crows that had previously experienced the feedback loop in a typical string-pulling task learned to solve this problem.

Two studies that have challenged the perceptual-motor feedback hypothesis used the pull-down test with ravens (Heinrich, Bernd & Bugnyar, 2005) and green-winged macaws (Gaycken, Picken, Pike, Burman, & Wilkinson, 2019). In both cases, groups of individuals that were either naïve to or experienced with vertical string-pulling were presented with a string-pulling task that required a counter-intuitive motor action: subjects had to pull down on a string looping over a higher perch before coming back down. Inexperienced ravens and both groups of macaws failed at this task, despite the food at the end of the string moving up when they performed a downward pull. Although both studies interpret this as evidence that a perceptual-motor feedback loop is not sufficient for acquiring string-pulling behaviour (Gaycken et al.,

2019; Heinrich, Bernd & Bugnyar, 2005), it is unclear if subjects experienced it as such: when reaching up with their bills to pull down on the string, their eyes would have been directed upwards, and so it is possible that subjects never observed the reward moving closer. Therefore, these studies could also be interpreted as providing evidence for the perceptual-motor feedback hypothesis.

Finally, support for the feedback loop hypothesis has also emerged from a connectivity task with naïve New Caledonian crows. When crows with no experience of vertical string-pulling chose between a meandering horizontal continuous string and a similarly positioned broken string, both of which contained food at their ends, they showed no preference for either string (Taylor et al., 2012). This horizontal loose-string connectivity task, or broken-string task, is particularly interesting because in all cases where avian subjects have succeeded at this task, they previously experienced a perceptual-motor feedback loop with vertical strings (Krasheninnikova et al., 2013; Obozova et al., 2014; Obozova & Zorina, 2013; Schmidt & Cook, 2006; Schuck-Paim et al., 2009; Torres Ortiz et al., 2019; Wang et al., 2019). This pattern also appears to extend beyond birds to other string-pulling tasks devoid of feedback: for example, bumblebees with feedback-loop experience successfully pull on coiled strings, but naïve bees and bees that have only observed string-pulling by others (and therefore did not experience feedback themselves), do not (Alem et al., 2016).

A key question that remains unanswered is whether naïve New Caledonian crows' failure at the loose-string connectivity task is directly attributable to their lack of perceptual-motor feedback loop experience. Given the pattern of results present in the literature, we raise a feedback experience hypothesis: that birds must first experience perceptual-motor feedback with strings to learn to attend to the ends of the strings, which in turn could allow them to succeed at two-string discrimination tasks such as those testing connectivity, contact, and continuity.

Kea (*Nestor notabilis*) are an ideal model species for testing this feedback experience hypothesis. Kea spontaneously solve the single vertical string task on their first trial and succeed in the crossed-strings task, where the end of the baited string is located under the starting point of the unbaited one (Huber & Gajdon, 2006; Werdenich & Huber, 2006). In contrast, experienced New Caledonian crows perform at chance at this test of causal understanding (Taylor et al., 2010), suggesting kea might outperform them at string-pulling tasks generally. Furthermore, a previous study showed that two of five kea selected a connected over an unconnected board within their first trial, and one of them selected the correct option in all of their first 10 trials (Auersperg, Gajdon, & Huber, 2009). Although the other subjects took longer to understand the nature of this task, this result suggests that kea are capable of understanding connectivity without extensive training, at least in contexts unrelated to string-pulling.

We therefore tested naïve kea (*Nestor notabilis*) on three experiments to test both the predictions of the three mental simulation hypotheses, and the feedback experience hypothesis we propose. In Experiment 1, kea were presented with the horizontal connectivity discrimination task previously presented to New Caledonian crows, to establish whether kea without any perceptual-motor feedback loop experience in string-pulling contexts were capable of understanding the nature of this task. In Experiment 2, kea experienced 10 vertical string-pulling trials, replicating a previous study with another captive kea population. Then, following an additional 10 trials of vertical string-pulling experience, in Experiment 3 kea were presented with a repeat of the first experiment. If experience of perceptual-motor feedback explains the pattern of results in the literature up to this point, enabling birds to attend to the ends of strings, then kea should fail at Experiment 1 but succeed at Experiment 3, following experience of vertical string-pulling.

4.3. Methods

4.3.1. Subjects

Our subjects were eight captive kea housed at Willowbank Wildlife Reserve (Table 6). In both experiments, strings were attached to black tokens, which kea had been previously trained to exchange for a food reward (Bastos & Taylor, 2019, 2020). None of the subjects had any prior experience with strings. Research was carried out with approval from the University of Auckland ethics committee (reference number 001816) and all methods were carried out in accordance with the relevant guidelines and regulations.

Subject	Hatch Date (Known or estimated)	Sex	Participation
Blofeld	August 2013	M	Experiments 1 – 3
Bruce	September 2012	M	Experiments 1 & 2
Loki	August 2014	M	Experiments 1 – 3
Megatron	October 2019	M	Experiment 2
Moriarty	August 2014	M	Experiments 1 – 3
Neo	September 2012	M	Experiments 1 – 3
Plankton	August 2014	M	Experiments 1 – 3
Taz	September 2012	M	Experiments 1 – 3

Table 6. Individuals' hatch dates, sex, and participation in the two experiments. All subjects were parent reared.

4.3.2. Experiment 1

Trials were conducted on a rectangular platform measuring 65cm x 130cm, with a centrally protruding 30cm x 30cm shelf where subjects could stand behind an acrylic sheet between trials, serving as subjects' starting position in all trials (Figure 8). Two 20cm x 30cm curved acrylic shields were placed diagonally on the platform, equidistant to the starting position. Kea were first habituated to the apparatus and trained to make choices between options presented under the two shields, neither of which contained strings or required pulling. At test, two horizontal string-pulling options were presented under the two shields, with only their tips accessible at the front. As in the study with New Caledonian crows (Taylor et al., 2012), kea had to choose between a continuous piece of string attached to a rewarding black token at the far end, or a broken string missing a 10cm section which was otherwise identically arranged. The side on which the continuous (correct) option was presented was pseudorandomised and counterbalanced, with no more than two trials in a row where the continuous string was placed on the same side, for a total of 20 trials, presented in blocks of 10 trials. The experimenter wore mirrored sunglasses and was kept blind to experimental hypotheses across all trials, so as to not unintentionally cue subjects at test.

Each trial began with the subject standing at the starting position, behind the closed plexiglass barrier, for an observation period of 20 seconds, during which subjects could see but not approach the two options. After 20 seconds, the experimenter opened the plexiglass barrier and stepped back, allowing subjects to step towards either plexiglass shield. Subjects were allowed to make only one choice, and the trial ended when the kea either: (a) obtained the black token from the continuous string, which they were then allowed to exchange for a food reward, (b) interacted with one string and then attempted to interact with another without obtaining either token, (c) interacted with either string without obtaining the continuous string's token for 1 minute, or (d) refused to touch or interact with either choice for 3 minutes.

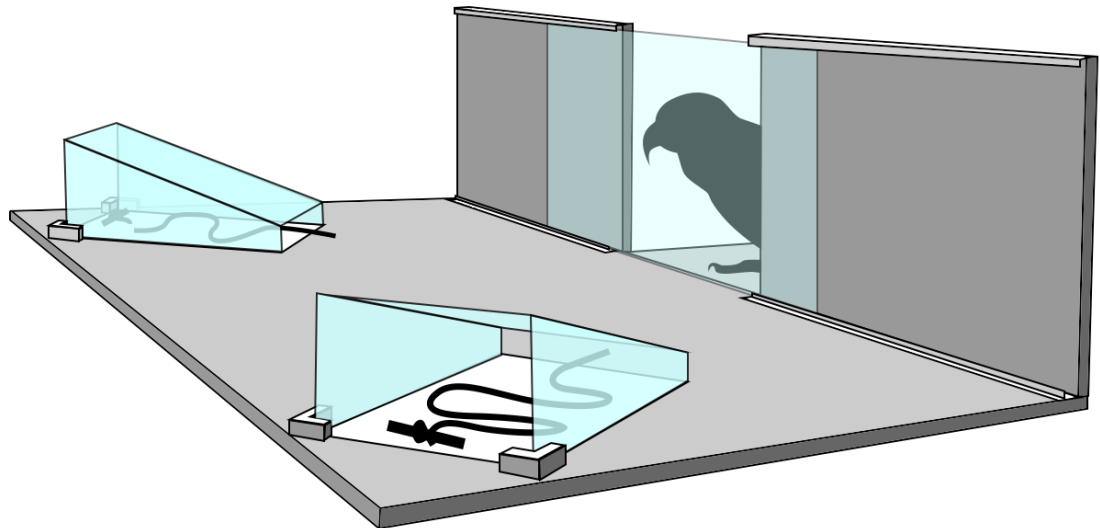


Figure 8. Diagram of setup used in Experiment 1, where kea had to choose between a continuous and a broken string, both of which were placed under sloping acrylic shields and were attached to rewarding black tokens.

4.3.3. Experiment 2

To ensure that our population of kea performed typically in a vertical string-pulling task, subjects experienced 10 trials of a single continuous vertical cotton string. The platform for this experiment consisted of a 30cm x 30cm base with a \varnothing 40mm perch which overhung from one side by 41cm. A 70cm long piece of string (the same length as used in a previous study; Werdenich & Huber, 2006) with a black token attached at one end hung 20cm from the edge. Each trial was 3 minutes long, during which time the subject was allowed to interact with the perch and the string. Failure to pull the token up within the duration of a trial was counted as a failure and leaving the platform mid-trial was counted as a refusal. If subjects failed or refused a trial, that block was interrupted and resumed at a later time. Where subjects succeeded at a trial, they were given the next trial immediately, for a total of up to 10 trials.

4.3.4. Experiment 3

Experiment 3 was an exact replication of Experiment 1, excluding the only individual (Bruce) that did not experience the perceptual-motor feedback loop in Experiment 2 due to being physically unable to pull up the string.

4.3.5. Video Coding and Analyses

Trials were filmed and coded in terms of subjects' binary choices (Experiments 1 and 3) or the duration of successful trials (Experiment 2). Successful trials in Experiment 2 (where the subject retrieved the black token from the vertical string) were also coded for number of pulls, steps, and errors. Errors occurred when subjects: (a) failed to step on the string following a pull, (b) failed to otherwise secure the string following a pull, (c) mis-coordinated a step after a pull, which failed to secure the string, and (d) stopped pull-step actions and released the string before the token was successfully retrieved. Their behaviours were used to calculate a pull-step ratio (Taylor et al., 2010), which consisted of the number of correct pull-steps over the total number of pulls attempted by the subject. Any manipulation of the string prior to the first pull were ignored (Taylor et al., 2010), as these usually consisted of exploratory behaviours such as touching or biting the string where it was attached to the perch. Unlike New Caledonian crows, kea used other forms of string attachment as an alternative to stepping (such as swinging the string over the side of the perch, which also held it in place). These were counted as steps for the purposes of the pull-step ratio.

Success in Experiment 1 was analysed at the individual level, with Bayesian two-tailed binomial tests with default beta priors set at 0.5. We also correlated successful trial number (counting from the first successful trial onwards) and trial duration for Experiment 2 at the group level, using a Bayesian correlation with default stretched beta prior width of 1. We used

two-tailed Bayesian t-tests to compare performances both within Experiment 2 and between our results and those of a previous study (Werdenich & Huber, 2006). Bayes Factors below 0.33 and above 3 were taken as substantial evidence for the null hypothesis or alternative hypothesis, respectively (Wagenmakers et al., 2018). All analyses were carried out in JASP v.0.13.1.0 (Jasp Team, 2019).

4.4. Results

None of our subjects performed above chance across the 20 trials of Experiment 1 (Table 7). Overall, kea attempted to change their original choice in 41 of 140 trials (29.3% of the time), and of these switches, only 9 were attempted in the wrong direction (attempting to switch to the incorrect choice after selecting the correct choice; 22.0%). Unlike New Caledonian crows (Taylor et al., 2012), kea persisted in their interactions with the apparatus and strings, even when they did not completely retrieve them from the apparatus. Strings were fully retrieved by subjects in 112 of 140 trials (80.0%) and across all trials subjects spent an average of 18.54 ± 19.70 seconds interacting with the strings. When strings were not fully retrieved, there was no consistent pattern across individuals as to whether this was the correct or incorrect string. Unlike New Caledonian crows, all subjects continued to interact with strings even in failed trials (averaging 22.36 ± 21.35 seconds in incorrect trials) and completed the experiment with no refusals.

In Experiment 2, our subjects performed similarly to the kea population used in a previous study (Werdenich & Huber, 2006). All subjects succeeded in retrieving the rewarding token hanging on the end of the vertical string, with the exception of Bruce, whose missing upper mandible made manipulating the vertical string too challenging (Table 8). As in previous research (Werdenich & Huber, 2006), most kea rapidly solved the task from their first or second

trial, with the average (mean) duration of individuals' first successful trial being higher than that for subsequent successful trials (Table 9). Solution times were similar to those in the previous study (83.10 ± 128.39 seconds in previous study; 71.36 ± 42.88 seconds in present study; Bayesian two-tailed independent samples t-test, $BF_{10} = 0.705$).

Subject	Correct Choices	Correct Switches	Incorrect Switches	Unretrieved Strings	Interaction Time
Blofeld	10/20 (BF = 0.270)	1	4	10 (6 correct)	32.68 ± 23.93 seconds
Bruce	11/20 (BF = 0.297)	4	6	4 (4 correct)	11.74 ± 12.76 seconds
Loki	12/20 (BF = 0.396)	0	4	0	9.01 ± 16.77 seconds
Moriarty	9/20 (BF = 0.297)	0	3	2 (0 correct)	24.81 ± 22.97 seconds
Neo	7/20 (BF = 0.644)	2	7	4 (2 correct)	18.68 ± 18.02 seconds
Plankton	7/20 (BF = 0.644)	1	5	7 (2 correct)	18.66 ± 19.33 seconds
Taz	11/20 (BF = 0.297)	1	3	1 (1 correct)	14.19 ± 13.49 seconds

Table 7. Subjects' performance in Experiment 1, with columns showing, in order: number of correct choices (measured as first touch) for the continuous string (Bayesian binomial test values in parentheses), number of times subjects first made a correct choice and then switched to the incorrect choice (counted as correct and ended the trial), number of times subjects made the incorrect choice and then switched to the correct choice (counted as incorrect and ended the trial), number of times subjects did not fully retrieve the chosen string from the apparatus and how many of these unretrieved strings were the correct choice, and average time spent interacting with strings.

Subject	Successful Retrievals	Mean Trial Duration	Pull-Step Ratio
Blofeld	6	25.14 ± 31.12 seconds	89.17 ± 12.01%
Bruce	0	n/a	n/a
Loki	9	9.31 ± 2.78 seconds	91.76 ± 10.29%
Megatron	10	20.76 ± 29.09 seconds	82.00 ± 14.14%
Moriarty	10	12.25 ± 13.57 seconds	86.67 ± 14.27%
Neo	10	19.56 ± 41.98 seconds	95.50 ± 9.56%
Plankton	10	16.49 ± 14.12 seconds	85.81 ± 12.97%
Taz	10	15.38 ± 26.16 seconds	95.50 ± 9.56%

Table 8. Individual performances in Experiment 2, namely: the number of successful retrievals performed, the average time taken to retrieve the token across all of their successful trials, and each individual's pull-step ratio across all their successful trials.

Trial Number	Mean Duration of Successful Trials	Pull-Step Ratio of Successful Trials
1	71.36 ± 42.88 seconds	83.10 ± 12.49%
2	10.66 ± 3.82 seconds	88.33 ± 11.18%
3	8.37 ± 2.67 seconds	93.57 ± 11.07%
4	9.99 ± 4.57 seconds	83.10 ± 17.01%
5	10.45 ± 3.70 seconds	84.76 ± 10.82%
6	10.30 ± 4.75 seconds	94.13 ± 11.04%
7	6.95 ± 3.11 seconds	95.83 ± 10.21%
8	7.08 ± 1.22 seconds	100.00 ± 0.00%
9	10.13 ± 4.78 seconds	83.61 ± 13.60%
10	7.35 ± 1.16 seconds	90.00 ± 13.69%

Table 9. The first column shows the average duration across all individuals' successful trials in Experiment 2, measured as the time taken from touching the string for the first time to holding the black token. Unsuccessful trials where individuals failed to retrieve the rewarding token are not included. The second column contains the pull-step ratios across individuals their successful trials, calculated as the percentage of correct pulls followed by steps (or other string attachments) over all attempts to pull the vertical string.

On average across all trials, subjects performed 3.35 ± 0.93 correct pull-steps, and only 0.49 ± 0.59 errors per trial. Kea therefore exhibited high pull-step ratios in their successful trials (mean across all subjects in all successful trials: $89.47 \pm 12.44\%$), which was similar to the equivalent measure in New Caledonian crows ($90.2 \pm 2.42\%$ in Taylor et al., 2010). Pull-step ratios were high from the first successful trial (mean across all subjects' first trials: $83.10 \pm 12.49\%$). The number of pull-steps performed in their first successful trial did not differ significantly from the number of pull-steps in their last successful trial (mean for first trial across all subjects: 3.57 ± 0.98 pull-steps; mean for final trial: 2.80 ± 0.45 pull-steps; Bayesian two-tailed paired samples t-test, $BF_{10} = 0.522$). We found no evidence to suggest that the group's performance improved with experience across all test trials (Bayesian correlation, $n = 7$, Pearson's $r = -0.553$, $BF_{10} = 1.308$), despite the reduced solution time between the first two successful trials, mirroring the results from previous work on another kea population (Werdenich & Huber, 2006). Therefore, trial duration averages indicate that kea rapidly achieved ceiling performance from their second vertical string-pulling trial onwards, rather than making gradual improvements to their performance over several trials.

Experiment 3 was a direct replication of Experiment 1 with the five individuals that experienced both the loose-string connectivity task of Experiment 1 and successfully retrieved the vertical string in at least one trial of Experiment 2. This did not include one subject, Bruce, that was unable to pull up the string due to his missing upper mandible. None of the five subjects in this study performed above chance following experience of the perceptual-motor feedback loop (Table 10). Subjects attempted to switch their choices in 28 out of 120 trials (23.3%), with only one of these switches being from a correct to an incorrect choice (3.57% of switches, compared to 21.95% in Experiment 1). As in Experiment 1, subjects continued to interact with the strings even when they made incorrect choices, averaging 20.20 ± 12.91 seconds across all trials.

Subject	Correct Choices	Correct Switches	Incorrect Switches	Unretrieved Strings	Interaction Time
Blofeld	10/20 (BF = 0.270)	1	1	0	33.89 ± 22.61 seconds
Loki	11/20 (BF = 0.297)	0	8	0	5.26 ± 12.91 seconds
Moriarty	9/20 (BF = 0.297)	0	2	0	26.21 ± 23.22 seconds
Neo	10/20 (BF = 0.270)	0	7	0	20.09 ± 19.90 seconds
Plankton	13/20 (BF = 0.644)	0	3	1 (1 correct)	23.47 ± 22.98 seconds
Taz	11/20 (BF = 0.297)	0	6	0	12.29 ± 16.64 seconds

Table 10. Subjects' performance in Experiment 3, with columns showing, in order: number of correct choices for the continuous string (Bayesian binomial test values in parentheses), number of times subjects first made a correct choice and then switched to the incorrect choice (counted as correct and ended the trial), number of times subjects made the incorrect choice and then switched to the correct choice (counted as incorrect and ended the trial), number of times subjects did not fully retrieve the chosen string from the apparatus, and average time spent interacting with strings.

4.5. Discussion

Our study tested whether experience of a perceptual-motor feedback loop during string-pulling might explain the pattern of results observable in the bird literature: naïve subjects fail at loose-string connectivity tasks (Taylor et al., 2012), and experienced birds sometimes succeed at it (Krasheninnikova et al., 2013; Obozova et al., 2014; Obozova & Zorina, 2013; Schmidt & Cook, 2006; Schuck-Paim et al., 2009; Torres Ortiz et al., 2019; Wang et al., 2019). Kea failed this task both before and after receiving experience of vertical string-pulling, that is, regardless of their experience with perceptual-motor feedback loops. This result both opposes the

predictions of the insight, planning, and means-end understanding hypotheses which predict an immediate understanding of string-pulling problems without perceptual-motor feedback experience, but also shows that feedback experience alone cannot elicit success at the horizontal loose-string connectivity task.

It is unlikely that the failure of kea at both connectivity experiments reflects a general discrepancy in cognitive or motor abilities in our population of kea. In fact, the results of Experiment 2 reveal that our population of kea behaved very similarly to conspecifics in another captive population (Huber & Gajdon, 2006; Werdenich & Huber, 2006) and wild New Caledonian crows (Taylor et al., 2010, 2012). As in the original kea study (Werdenich & Huber, 2006), our subjects quickly learned to pull up vertical strings, showing few errors and ceiling performances after their first successful trial.

Our results also demonstrate that kea's failure was not a result of hesitancy to interact with the strings. Unlike crows (Taylor et al., 2012), kea persisted in their interactions with the strings throughout Experiments 1 and 3, and we never observed choice refusals in any trials. As a highly neophilic species (Diamond & Bond, 1999), it is possible that, even after having failed to obtain the reward at the end of the string, kea were still interested in the properties of this novel material and therefore continued to interact with it. This is consistent with our observation that subjects often chewed on the ends of the strings even following incorrect choices, suggesting that not all of their interactions with the strings were made with the intent to retrieve the out-of-reach black token. This issue was not observed in Experiment 2 presumably because either goal (obtaining the black token or playing with the string) could have equally resulted in accurate and efficient step-pulling behaviour by the subjects. Kea became more persistent in Experiment 3 than they were in Experiment 1, leaving fewer strings unretrieved in the final experiment. This may have been a consequence of their increased experience performing multiple pulling actions on the vertical string presented in Experiment

2. However, their vertical string experience did not improve their ability to discriminate between the connected and unconnected strings in their second attempt at the horizontal connectivity task. Their inability to make this discrimination is unlikely to be a consequence of kea finding it difficult to make choices between the two options presented. This same population of kea have previously demonstrated an ability to make binary choices between highly similar stimuli (Bastos & Taylor, 2019, 2020; Heaney, Bastos, Gray, & Taylor, 2019), and subjects were specifically trained to make choices using the apparatus of these experiments prior to test.

Our results demonstrate that although kea made fewer switches from correct to incorrect choices in their second attempt at the loose-string connectivity task, they still initially selected the correct string at chance. This would suggest that although kea realised that they made an incorrect choice after attempting to pull on the unconnected string, they did not simulate their actions on horizontal strings prior to their first touch, even after ample experience with the setup. Therefore, kea's persistent failure at the connectivity task provides evidence against the insight, planning, and means-understanding hypotheses for string-pulling behaviour, all of which would predict an ability to mentally simulate the consequences of their actions on strings both with and without experience of perceptual-motor feedback loops. This finding is notable in light of research suggesting that kea are capable of mentally representing objects in other contexts (Bastos & Taylor, 2019; 2020).

Despite the pattern of results in the literature showing that individuals of several bird species succeed at string connectivity tasks following perceptual-motor feedback with vertical strings (Krasheninnikova et al., 2013; Obozova et al., 2014; Obozova & Zorina, 2013; Schmidt & Cook, 2006; Schuck-Paim et al., 2009; Torres Ortiz et al., 2019; Wang et al., 2019), while naïve individuals fail (Taylor et al., 2012), we did not find evidence that experience of feedback improves performance on a string connectivity task in kea. This is particularly puzzling given

that previous studies have shown that kea can successfully distinguish connected from unconnected wooden boards (Auersperg et al., 2009). As such, further work is required to establish whether birds' understanding of connectivity is context-dependent, by comparing performance before and after feedback experience both during string-pulling and in tasks using materials other than string. This could shed light on the interplay of contextual variables and cognitive mechanisms that underpin performance on this iconic problem-solving task.

Chapter V

*Kea represent object trajectory and identity*⁴

5.1. Abstract

The ability to represent both the identity and trajectory of hidden objects underlies our capacity to reason about causal mechanisms. However, to date no studies have shown that non-human animals are capable of representing these two factors simultaneously. Here, we tested whether kea can represent out-of-sight object trajectories and identities by presenting subjects with three tasks, each of which involved tracking or predicting hand trajectories as they moved behind a screen. Taken together, our results suggest that kea have the capacity for mental simulation in complex tasks involving moving hidden objects.

⁴ Based upon: Bastos, A. P. M. & Taylor, A. H. (2019). Kea (*Nestor notabilis*) represent object trajectory and identity. *Scientific Reports*, 9, 19759. doi:10.1038/s41598-019-56380-4

5.2. Introduction

Humans are capable of representing both the trajectory and identity of moving out-of-sight objects (Moore, Borton, & Darby, 1978; Kellman & Spelke, 1983; Baillargeon, 1986; Baillargeon, 2004). For example, if we see a bird fly behind a tree, we expect the same bird to emerge from behind the tree on the same trajectory. This ability is a key foundation underlying our ability to simulate complex out-of-sight interactions, including hidden causal mechanisms (Wilcox & Baillargeon, 1998; Kushnir et al., 2010).

Research on non-human animals has largely focused on object trajectory, using invisible displacement and spatial transposition tasks, rather than the combination of object trajectory and identity. In invisible displacement tasks, an object is placed in a displacement device, and this device is then moved behind a screen. The device is then shown to be empty. This requires subjects to infer, based on the observed trajectory of the device, that the object must have been left behind the screen. In spatial transposition tasks, an object is placed under one of several displacement devices, which are moved to different positions. Subjects have to track the visible trajectory of the device containing the object. These tasks have been solved by children (Sophian & Sage, 1983; Sophian, 1984; Barth & Call, 2006), corvids (Pollok et al., 2000; Zucca et al., 2007; Hoffmann et al., 2011), psittacines (Pepperberg & Funk, 1990; Pepperberg & Kozak, 1986; Funk, 1996; Pepperberg et al., 1997; Auersperg et al., 2014), and great apes (Natale et al., 1986; Call, 2001; Collier-Baker & Suddendorf, 2006; Barth & Call, 2006), although it remains unclear whether any species other than humans and the great apes have succeeded in these tasks using mental representation, rather than simpler associative learning strategies (Natale et al., 1986; Jaakkola, 2014; but see Pepperberg, 2015). Nevertheless, there is convincing evidence that several species, including dogs, monkeys, parrots, corvids, and chickens can represent object identity in other tasks (Regolin & Vallortigara, 1995; Regolin et al., 1995, 2004; Pepperberg et al., 1997; Clayton & Dickinson,

1998; Bräuer & Call, 2011; Phillips & Santos, 2007; Chiandetti & Vallortigara, 2011; Fontanari et al., 2011), and chickens have also been shown to represent an object's trajectory, correctly predicting the end-destination of an object behind one of two screens (Vallortigara et al., 1998; Freire & Nicol, 1999; Regolin et al., 2005).

So far, none of these tasks have required an animal to simultaneously represent both the hidden trajectories of multiple objects, and the identity of these objects, as humans can (Moore et al., 1978). Kea (*Nestor notabilis*) are an ideal candidate species for examining this possibility, due to their sophisticated sensorimotor intelligence (Auersperg et al., 2009; Huber & Gajdon, 2006). Here, we tested whether kea could simultaneously represent the trajectory and identity of objects, using tasks where subjects had to represent the identity of hidden tokens while making predictions about their trajectories.

5.3. Methods

5.3.1. Ethics Statement

This research was conducted under ethics approval from The University of Auckland Ethics Committee (reference number 001816). Our research was conducted in accordance with the New Zealand National Animals Ethics Advisory Committee guidelines.

5.3.2. Subjects

We tested ten kea, housed in a large outdoor aviary in Willowbank Wildlife Reserve (Table 11). Food and water were available *ad libitum* within the aviary. All participation was voluntary, and subjects were free to leave the testing platform at any time.

Subject	Hatch Date (Known or estimated)	Sex	Origin	Participation (Experiments)
Blofeld	August 2013	M	Captive bred	All experiments
Bruce	September 2012	M	Wild	All experiments
Cheeky	October 2006	M	Captive bred	Experiments 1 and 2
Harley Quinn	August 2014	F	Captive bred	Experiment 1
Loki	August 2014	M	Captive bred	All experiments
Moriarty	August 2014	M	Captive bred	Experiments 1 and 3
Neo	September 2012	M	Captive bred	All experiments
Plankton	August 2014	M	Captive bred	Experiments 1 and 3
Taz	September 2012	M	Captive bred	Experiments 1 and 3
Spike	October 1995	M	Wild	Experiment 1

Table 11. All subjects were parent reared. Hatch dates were either known (for captive bred subjects) or estimated from veterinarian evaluations (wild subjects). One further subject, Johnny, was tested but refused to complete all required trials for Experiment 1, and so was excluded from the dataset.

5.3.3. Materials

Subjects were tested on individual training platforms (42 cm × 42 cm) within the aviary and rewarded with Hill’s Science Diet pellets during trials. For training and testing, a small wooden shelf (60 cm × 20 cm) was used with a plexiglass screen separating the kea from the experimenter.

5.3.4. Training: Hand-tracking training protocol

Subjects were trained to exchange black tokens for a reward, whereas orange tokens were unrewarding. Kea were then trained to track human hands which moved behind a fully transparent plexiglass screen over the course of three training stages.

All training stages were provided in blocks of 10 or 20 trials, using the testing apparatus. All trials began with subjects being shown two empty hands facing them, which then performed different actions in turn. Once the hands had reached their final positions at the top of the screen as closed fists, subjects made a choice by touching their beak or cere to one of the experimenter's fists. Only one choice was allowed per trial. In trials where a black token was chosen, the token was handed to the subject, which then had to hand it back to the experimenter in exchange for a pellet. Where the subject made the incorrect choice of picking an empty hand or a hand containing an orange token, the contents of the hand were shown and then the hand was retracted, and the next trial was started immediately. Criterion is specified for each training stage in parentheses in the descriptions below.

Stage 1. Kea must choose the correct hand when one hand holds a pellet, and one hand is empty:

- a. When both hands move up simultaneously and in parallel. (17/20 trials)
- b. When two trial types are interspersed: either both hands move up simultaneously and in parallel, or when both hands are moved simultaneously and in parallel, then crossed over above the plexiglass such that they end up on the opposite sides. During this stage, subjects could touch one or both hands at multiple times as they were crossed over. (17/20 trials)
- c. When two trial types are interspersed: either both hands move up simultaneously and in parallel or when both hands are moved simultaneously and crossed over behind the plexiglass such that they end up on the opposite sides. At this stage,

subjects were required attend to the motion of both hands taking place behind the screen, without touching either hand. (17/20 trials)

Stage 2. Kea must choose the correct hand when one hand holds the black token, and one hand is empty, in interspersing parallel and crossed trial types as in Stage 1c. This ensured that subjects remembered the black token was rewarding and still made the correct choice when the pellet was substituted for a rewarding token. (10/10 trials)

Stage 3. Kea must choose the correct hand when hands pick up, show, and then enclose either the rewarding black token, or the unrewarding orange token, in interspersed trials with parallel and crossed trajectories, as in Stages 1c and 2. In this case, subjects had to keep track of two hands holding two different objects, only one of which was rewarding. (17/20 trials)

5.3.5. Training: Food-search choice task for Experiment 3

Before taking part in Experiment 3, subjects were habituated to putting their beaks through a small window ($\emptyset 6$ cm) and trained to make functional choices between two options in a food-searching task. These new behaviours were easy to interpret (i.e., if a subject placed their beak through a window, this was considered their final choice) and encouraged subjects to make a clear choice at test. All training steps (and subsequent testing for all subjects) were conducted by experimenters blind to hypotheses, wearing mirrored sunglasses. This was done as a precaution against Clever Hans cueing, i.e., subjects using experimenter's unintentional cues to guide their choices in a task. All training steps required a 17/20 criterion to be reached before proceeding to the next stage of training. They are detailed below:

1. Touch visible fist with beak: Subjects were habituated to standing on a wooden step and placing their beak through a small round window ($\emptyset 6$ cm) in the middle of a piece of plexiglass ($45 \text{ cm} \times 25 \text{ cm}$) to touch the experimenter's closed fist. The

experimenter's fist contained a Science Diet pellet which was given to the subject after they touched the fist.

2. Search for experimenter's hand: In this step, the same piece of plexiglass was used, with a cardboard square (25 cm × 25 cm) on the back, surrounding the central window. Now, the experimenter placed their closed fist behind the cardboard in any position. Subjects were required to place their beak into the window and "search" for the experimenter's hand. Once their beak touched the experimenter's fist, they were rewarded with a Science Diet pellet. This ensured that subjects were still motivated to search for the experimenter's fist and put their beak through the window even when the hand was not visible.
3. Search for a piece of food: This was identical to the previous step, but the Science Diet pellet was directly attached to the back of the cardboard, so the subjects would search for it directly with their beak. This required a much longer search time, ensuring that subjects were fully comfortable placing their beaks through the window in the plexiglass screen.

The plexiglass screen was then substituted for a larger version (45 cm × 45 cm), with two identical windows (Ø6 cm) drilled into the left side. A large cardboard occluder (45 cm × 20 cm) could slide over the left side to hide the two windows. The windows were equidistant from each other (21 cm) and from the edges of the plexiglass screen (6 cm). Two identical semi-circular transparent plastic baskets (approximately 6.5 cm wide × 2 cm tall, 3 cm radius) were positioned under each of the two windows on the experimenter's side of the plexiglass. The final two training steps were done using this new apparatus. Again, these training steps required 17/20 criterion:

4. Making a functional choice between two options: At the start of each trial, the experimenter slid the cardboard occluder in front of the two windows in the plexiglass.

Then, they shuffled a Science Diet pellet between their hands behind their backs, enclosing it into one fist and leaving another closed fist empty. Both hands were simultaneously moved towards the two windows at the same time. One of the baskets below a window was baited, and the other was sham-baited. Both these baiting episodes occurred simultaneously. Following this, both hands slid the cardboard occluder, so the kea could see which basket contained the food pellet. Baskets and hands baited were pseudorandomised and counterbalanced within blocks of 10 trials. During this step, hand motions for both hands were identical so as to avoid teaching the subjects any additional information on hand-trajectories. Subjects simply had to look into both baskets and place their beak through the correct window to obtain the food reward. Only one choice was allowed. If subjects placed their beak through the wrong window, the cardboard occluder was immediately slid back to the left and subjects were not allowed to change their choice.

5. Remembering the location of hidden food: This step was identical to the last, but an additional piece of plexiglass (25 cm × 20 cm) was placed behind the cardboard. Instead of allowing the subjects to take the piece of food from the basket immediately once the cardboard occluder was slid to the right, this plexiglass was left in place. Subjects could therefore see, but not retrieve, the Science Diet pellet in one of the two baskets. After subjects had seen the Science Diet pellet, the windows were occluded by the cardboard again. The experimenter then placed two small cardboard rectangles (6.5 cm × 2 cm) into each basket simultaneously, so the pellet would no longer be visible to the subject. Both the plexiglass and cardboard occluder were now slid back to the right, and the subject was allowed to place their beak through their chosen window and remove the cardboard rectangle. If they made the correct choice, this would reveal the hidden food pellet, which they were allowed to take. If they made the wrong choice, this would

reveal an empty basket, which would be immediately covered up again, and the subject would not receive a reward.

During the fourth training step, two subjects (Harley and Spike) would leave the working platform each time the large plexiglass screen was introduced. We interpreted that as a sign of fear and excluded these subjects from any further testing with this apparatus, in compliance with our lab's code of conduct and ethics protocols. Participation in all experiments is voluntary and subjects are free to leave the platform if they do not wish to work at any given time or with a given apparatus.

5.3.6. Testing Procedure

All trials began with the experimenter showing their empty hand(s) to the subject, then picking up a token and enclosing it in their fist. This was held up momentarily before the start of each trajectory. Tokens were picked up one at a time. After the trajectory was performed, the experimenter either held out both closed fists simultaneously (Experiments 1 and 2), waiting for the subject to make their choice, or hid their fist behind the occluder until the subject made a prediction about its likely trajectory (Experiment 3). When subjects picked the incorrect hand, they were shown the unrewarding (orange) token, which was then taken away, and the next trial was started immediately. When they made the correct choice, they received a rewarding (black) token, which they subsequently exchanged for a food reward.

We took several measures to ensure kea could not use body movements as cues to identify hidden trajectories. In Experiments 1 and 2, experimenters' shoulders were kept static, with elbows held together in the centre (so as to be fully hidden behind the rectangular occluder), and all movement was performed by the forearms. In Experiment 3, the hand appeared to follow the initial trajectory when fully visible, but always stopped half-way

through, behind the occluder, rather than reaching its end destination. It only continued onto the end of the trajectory after the subject had placed their beak through the window if they made the correct choice. Furthermore, the experimenter stood behind the large occluder (45 cm × 45 cm) in all trials, which hid their right shoulder's movements from the subject. Experimenters also wore mirrored sunglasses so as to not provide any gaze cues to subjects.

5.3.7. Analyses

All trials were coded *in situ* and filmed. Performance for every set of 20 trials of each condition were analysed at the individual level, using two-tailed Bayesian binomial tests. The test value was set at 0.5 using default beta priors. All analyses were carried out using JASP (Jasp Team, 2020). We followed the convention that a Bayes factor (BF) < 0.33 shows substantial support for the null hypothesis, whilst a BF > 3 shows substantial support for the competing hypothesis (Wagenmakers et al., 2018).

5.4. Results

5.4.1. Experiment 1

Experiment 1 used three conditions to test if kea could predict hand trajectories after they moved behind a screen, based only on their initial movement. In all conditions, trials began when the experimenter demonstrated that one of their hands held a rewarding black token and the other an unrewarding orange token. Kea then observed the hands either moving parallel to a cardboard occluder (approximately 20 cm × 20 cm) attached to a plexiglass screen (43 cm × 29 cm) in the parallel trials, or behind the occluder, in crossed and split trials (Figure 9). In crossed trials, the experimenter's hand moved towards the occluder from opposite sides, crossed behind the occluder, and then emerged on the opposite sides. In split trials, the hands

moved identically at the start of the trial, but after moving behind the occluder, instead of crossing over, they were returned to their initial starting positions. Parallel trials served as a control condition, as throughout these trials both hands were visible at all times. They also ensured subjects were motivated to carry out testing, increasing the number of solvable trials in each block. At the end of each trial, kea were allowed to choose one of the hands to indicate where they believed the rewarding token must be. Ten kea experienced up to two sets of sixty trials, presented over the course of multiple testing sessions. Only fifteen trials were presented in each testing session, consisting of five trials for each of the three trajectory types (crossed, parallel, and split trajectories).

If subjects were predicting hand trajectories behind the screen from their initial movement, we expected them to perform above chance at the crossed trials, and at or below chance for split trials. This pattern could not be explained by subjects using hand identity or simple associative rules, such as searching the location on the same side as where the token was first picked up. It also ruled out the use of unintentional cues provided by the experimenter on the location of the token, given that in all trial types, the experimenter knew in which hand the rewarding token was located. Subjects that succeeded in parallel and crossed trial types, but not split trial types, within either their first or second set of sixty trials (20 trials of each trajectory type), proceeded to Experiment 2. Subjects that succeeded in parallel and split trials, but not crossed trials, did not participate in Experiment 2, and instead proceeded to Experiment 3. All subjects were given up to two sets of sixty trials to learn the contingencies of the tasks.

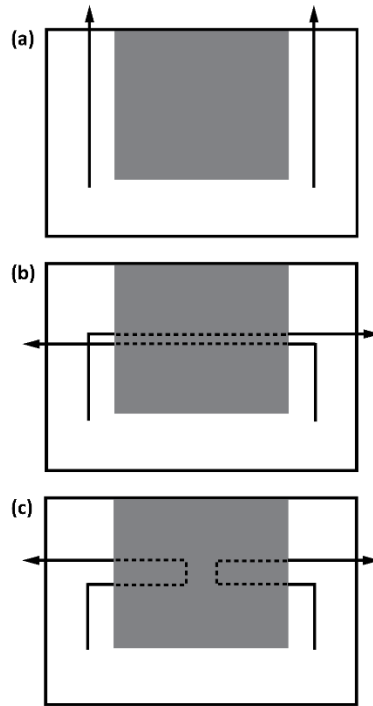


Figure 9. Experiment 1 (a) parallel, (b) crossed, and (c) split trajectories of Experiment 1. The area covered by the occluding screen is shown in grey. Solid lines indicate the visible parts of the trajectory, and dotted lines indicate the paths followed by the hands behind the occluder.

Four of the ten kea tested showed the pattern of results predicted if kea were able to represent object trajectory and identity, performing better than chance in parallel trials and crossed trials and at chance in the split trials (two-tailed Bayesian binomial tests, relative preference 0.5, $BF > 3$, Table 12). Whilst two subjects (Bruce and Cheeky) achieved this within their first set of trials, two other subjects (Loki and Neo) succeeded in their second set of trials, suggesting that they took some time to learn the affordances of the task. In contrast, three subjects (Harley Quinn, Moriarty, and Spike) appeared to adopt a strategy relying on the proximity of the two hands to the trajectory starting points to make their choices, as they performed better than chance in split but not crossed trajectory trials. No subjects performed above chance in both split and crossed conditions, suggesting that they did not use hand identity or unintentional cues by the experimenter to guide their choices.

	Experiment 1		Experiment 2
Subject	Performance in First 60 Trials (Number of times black token was chosen)	Performance in Following 60 Trials (Number of times black token was chosen)	Performance over 40 Trials (Number of correct choices)
Blofeld	Parallel: 20/20 Split: 10/20 Crossed: 10/20	Parallel: 20/20 Split: 12/20 Crossed: 10/20	NA
Bruce	Parallel: 20/20 Split: 5/20 Crossed: 19/20	NA	Hidden: 9/20 Visible: 20/20
Harley Quinn	Parallel: 19/20 Split: 20/20 Crossed: 6/20	NA	NA
Cheeky	Parallel: 19/20 Split: 3/20 Crossed: 16/20	NA	Hidden: 17/20 Visible: 18/20
Loki	Parallel: 20/20 Split: 9/20 Crossed: 12/20	Parallel: 20/20 Split: 8/20 Crossed: 19/20	Hidden: 19/20 Visible: 20/20
Moriarty	Parallel: 17/20 Split: 14/20 Crossed: 9/20	Parallel: 15/20 Split: 19/20 Crossed: 9/20	NA
Neo	Parallel: 19/20 Split: 6/20 Crossed: 14/20	Parallel: 20/20 Split: 5/20 Crossed: 19/20	Hidden: 20/20 Visible: 20/20
Plankton	Parallel: 18/20 Split: 13/20 Crossed: 8/20	Parallel: 20/20 Split: 14/20 Crossed: 7/20	NA
Taz	Parallel: 20/20 Split: 11/20 Crossed: 11/20	Parallel: 20/20 Split: 10/20 Crossed: 11/20	NA
Spike	Parallel: 19/20 Split: 18/20 Crossed: 3/20	NA	NA

Table 12. Performance of all ten kea in Experiment 1, followed by performance of the four subjects in Experiment 2. Subjects highlighted in green showed performances consistent with using a mental representation strategy to solve the tasks within each of the two experiments. Bold performances denote Bayes Factor > 3.

5.4.2. Experiment 2

Experiment 2 tested whether the four successful kea from Experiment 1 had succeeded by using an associative strategy: when hands move towards the middle of the screen, switch sides. If kea followed this rule, they should make a specific error: after seeing a hand move towards the middle of the screen they should then choose the hand on the opposite side of the screen, even if they had previously seen that this hand was holding the non-rewarding token. In contrast, if kea represented the hand containing the rewarding token as it moved behind the screen, then we expected them to be able to ignore the hand containing the non-rewarding token.

Subjects were given two blocks of 20 trials where one hand moved behind, and followed the trajectory of, an opaque U-shaped piece of cardboard behind the plexiglass screen, whilst the other hand visibly moved towards the opposite end of the U-shape, and then away to the other side of the plexiglass, as if continuing the motion of the hidden hand (Figure 10). The side tokens were placed on, the side the U-shape was positioned on, and whether trajectories for the rewarding black token were visible or hidden (behind the U-shape) were counterbalanced within blocks. Subjects experienced 20 visible trajectory trials, and 20 hidden trajectory trials.

Of the four kea tested, three performed above chance within their first 20 trials of both visible and hidden trajectory conditions (two-tailed Bayesian binomial tests, relative preference 0.5, $BF > 3$; Table 12), indicating these kea were not using the associative rule of searching in the location opposite to where the token was first picked up.

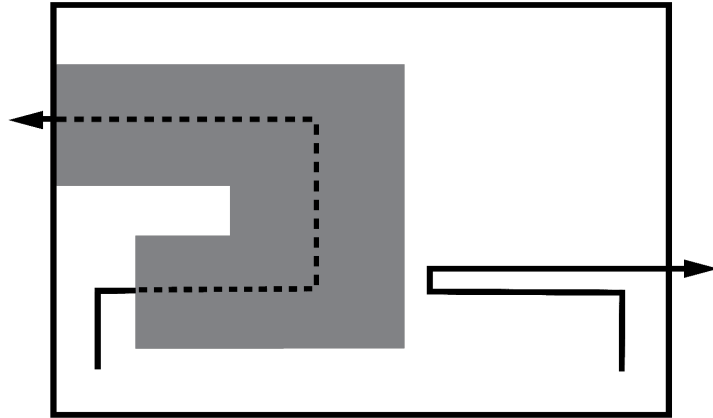


Figure 10. The two trajectories for Experiment 2, with one hand moved behind the U-shaped screen (hidden trajectory), and the other moved simultaneously on the opposite side of the screen (visible trajectory). Solid lines indicate the visible parts of the trajectory, and dotted lines indicate the paths followed by the hands behind the occluder, represented by a grey rectangle.

5.4.3. Experiment 3

Experiment 3 expanded on the results from Experiments 1 and 2 and was designed to test whether kea are capable of predicting the trajectory of a single token, rather than guessing their likely trajectory after both hands were presented. Unlike the previous experiments, this involved only a single trajectory carried out by one hand, and the kea never observed its end-destination until they had made a choice. Besides testing for kea's ability to predict trajectories given no information of their potential endpoints, this also removed any hand identity or timing-associated cues that might have been present from tasks using two simultaneously moving hands. Therefore, this experiment was designed to confirm that kea are capable of representing novel trajectories and had not relied on these associative rules to succeed in the two previous experiments.

Kea were first trained to search behind two windows ($\emptyset 6$ cm, 21 cm apart) on a large plexiglass screen (45 cm \times 45 cm) in order to find food but were not given any information on

potential hidden trajectories that the hands might take. At test, kea had to predict which of two windows in a screen they should look behind to find a hand containing a token, after observing the hand moving towards the screen along one of four different paths, and then following an occluded trajectory (Figure 11). Kea were presented with 20 trials of each trajectory type (top horizontal, bottom horizontal, top-bottom diagonal, bottom-top diagonal), which were pseudorandomised and counterbalanced. Testing was carried out by an experimenter blind to hypotheses, wearing mirrored sunglasses. Seven kea were presented with Experiment 3 (one kea had been transferred to another facility and two subjects did not habituate to the apparatus).

A control condition was then performed where we tested whether subjects would flexibly change their prediction of an object's trajectory when given contradictory information. Subjects experienced trials where the experimenter's hand appeared to start on one of the four previously seen trajectories before becoming occluded, but this time part of the occluding barrier was removed, suggesting that the hand could not have reached its expected destination. Subjects had to realise that the previously rewarded trajectory was no longer possible, and so the experimenter's hand must have followed an alternate path and ended at the opposite window, flexibly adjusting their predictions of these trajectories given this novel information.

Four of the seven subjects performed better than chance within their first 20 trajectory prediction trials (two-tailed Bayesian binomial tests, relative preference 0.5, $BF > 3$, Table 13). Their performance shows that subjects readily represented and predicted the end-destination of four novel object trajectories. Interestingly, a subject that had previously failed to pass Experiment 1 (Moriarty), was successful in Experiment 3, suggesting that predicting a single trajectory might be less cognitively taxing than simultaneously representing two object identities and their trajectories. All seven subjects succeeded within their first 20 control trials, showing that they flexibly updated their predictions when they were given information to suggest that the previously rewarding trajectory had become impossible.

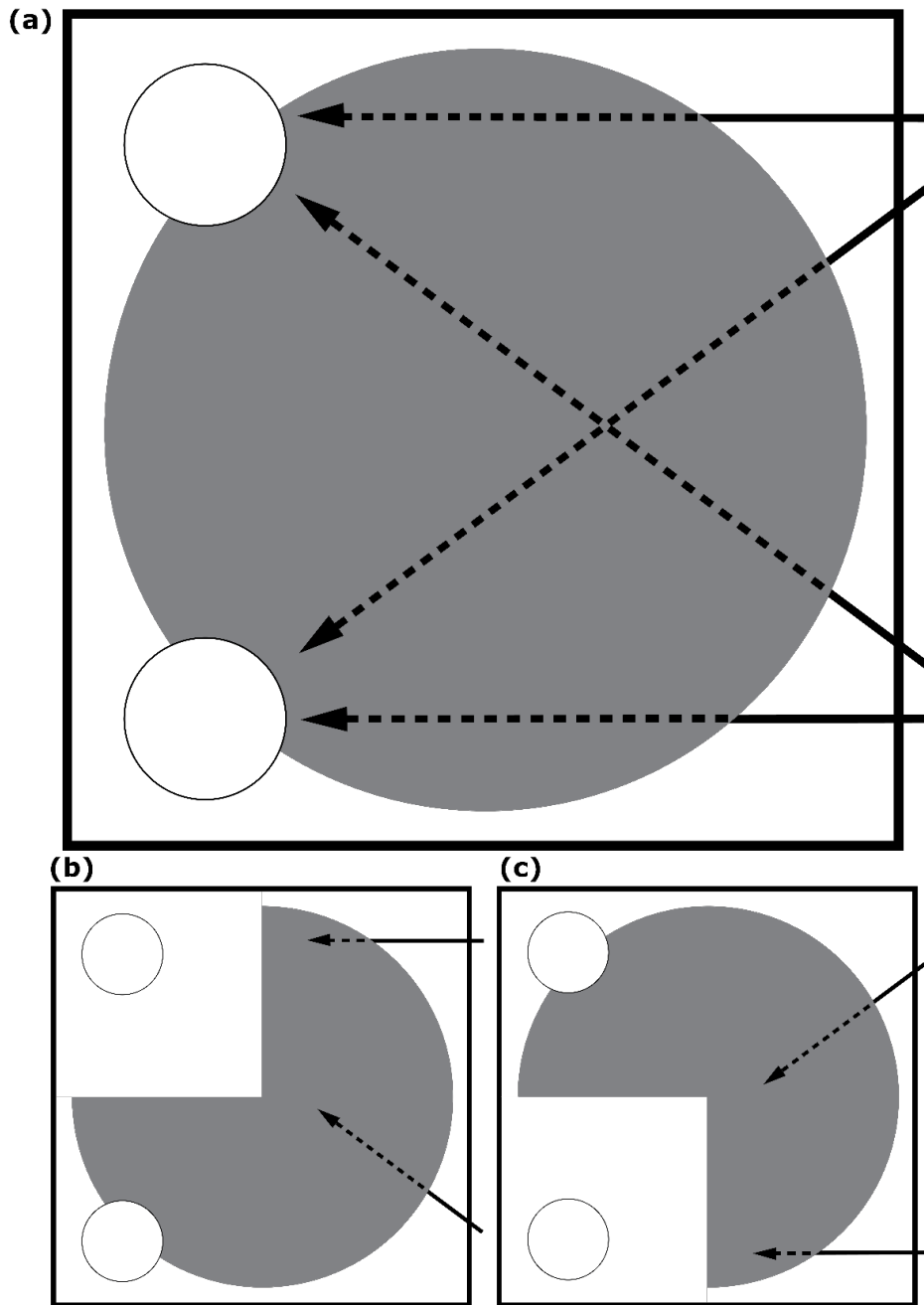


Figure 11. Conditions of Experiment 3. In (a), we show the four trajectory trial types in Experiment 3, from top to bottom: top horizontal, top-bottom diagonal, bottom-top diagonal, bottom horizontal. Parts (b) and (c) illustrate the two possible set-ups for Experiment 3's control conditions, showing the four trajectories: top horizontal, bottom-top diagonal, top-bottom diagonal, bottom horizontal. In all images, the area covered by the occluding screen is shown in grey. Solid lines indicate the visible parts of the trajectory, and dotted lines indicate the trajectories kea were expected to predict, given the hand's initial movement.

Subject	Trajectory Prediction Trials: Performance in First 20 Trials (Number of times black token was chosen)	Control Condition Trials: Performance in First 20 Trials (Number of times black token was chosen)
Blofeld	14/20	18/20
Bruce	17/20	18/20
Loki	16/20	19/20
Moriarty	16/20	19/20
Neo	17/20	17/20
Plankton	13/20	16/20
Taz	10/20	20/20

Table 13. Performance of seven kea in Experiment 3. Subjects highlighted in green showed performance consistent with predicting the end-point of a novel trajectory. Bold performances denote Bayes Factor > 3.

Subjects' performance was not stable over time: only one subject, Blofeld, consistently improved in their performance over the course of 80 trials, whilst it decreased over the course of multiple trials for Loki and Moriarty, who performed no differently from chance in their final block of 20 trials.

We also analysed subject performance in terms of the number of times they selected the window closest to the last location they last saw the experimenter's hand. We found no evidence to suggest that any subjects were using this proximity cue to make their choices, a hypothesis that would have predicted above-chance selection of the top window for both top horizontal and top-bottom trajectories, and selection of the bottom window for bottom horizontal and bottom-top trajectories (two-tailed Bayesian binomial tests, relative preference 0.5, Table 14). However, this analysis revealed a clear bias for the bottom window in one subject, Plankton (two-tailed Bayesian binomial tests, relative preference 0.5, BF > 3).

Trajectory Prediction Trials		
Subject	Proximity Top	Proximity Bottom
Blofeld	16/40	21/40
Bruce	18/40	22/40
Loki	9/40	25/40
Moriarty	14/40	19/40
Neo	16/40	22/40
Plankton	8/40	32/40
Taz	26/40	19/40

Table 14. Kea’s choices of window most closely located to last place hand was seen (use of a proximity cue). No subjects showed performance consistent with using this proximity cue. Bold performances denote Bayes Factor > 3.

5.5. Discussion

Our study shows that kea are capable of simultaneously remembering the identity of two hidden tokens and representing out-of-sight trajectories. In Experiment 1, four out of ten kea tested identified the end location of a preferred token when it was hidden in a hand and then moved behind a screen along a specific trajectory. However, they either selected the wrong hiding location, or performed at chance, when the trajectory was surreptitiously changed behind a screen, suggesting that they were tracking hand trajectories. Experiment 2 showed three of these four kea were not simply following an associative rule where they picked the hand that appeared on the opposite side of the occluder after moving behind it. In Experiment 3, four out of seven kea tested were able to predict the endpoint of four novel trajectories within their first

twenty trials, and then flexibly adjusted their choices when these trajectories became impossible.

Taken together, these results demonstrate that kea have the capacity to represent the trajectories and identities of hidden objects simultaneously, as well as predict the endpoints of novel trajectories. Three subjects performed above chance in crossed and parallel trials for Experiment 1, and in visible and hidden trials for Experiment 2, as would be predicted if they were using a mental representation strategy to solve both tasks. One of these three subjects (Cheeky) was transferred to another facility prior to Experiment 3, but the remaining subjects (Neo and Loki) also passed Experiment 3. Kea's performance across these three experiments suggests that they have the capacity to simulate out-of-sight trajectories, and that the successful subjects did not rely on unintentional experimenter cues, hand identities, or simple associative strategies to succeed in these tasks.

However, the variation between performances in our sample of kea shows that although kea, as a species, have the cognitive capacity to represent object identity and trajectory simultaneously, some do not spontaneously adopt this as a primary strategy in such tasks. Two individuals clearly searched in the last location they saw a hand in Experiment 1, irrespective of its trajectory. One individual used a strategy of searching opposite where they last saw a hand, as evidenced by their failure in Experiment 2. Further research is required to establish exactly under which contexts mental simulation becomes more or less likely to emerge as the preferred strategy in kea, and other nonhuman animals, and to further investigate performance at both the individual and group levels.

There was also variation in performance within individuals over the course of eighty trajectory prediction trials for Experiment 3. A large number of trials may have proven repetitive and therefore frustrating for the kea, leading to reduced performance over time.

Similarly, kea may have stopped attending as closely to the task after experiencing a large number of trials on the same problem. This decline in performance was not observed in the control conditions, possibly due to the easier nature of this task, which may have elicited less frustration when repeated over a large number of trials. It is also possible that kea selected the only likely option in the control condition before the token began on its trajectory, realising that this was the only possible solution to the problem without having to reason about possible trajectories. However, as kea were required to attend to the visible trajectory prior to making a choice in our experiments, they were not given the opportunity to show anticipatory behaviour, that is, select the only possible choice before the trajectory began. Nevertheless, four subjects' above chance performance within their first twenty trials of the trajectory prediction conditions in Experiment 3 supports the claim that kea can make predictions about trajectories' end-destinations, regardless of which strategy kea used to solve the control condition.

Past work on chickens, psittacines and corvids have shown that birds can track the trajectory of a single object (Vallortigara et al., 1998; Freire & Nicol, 1999; Regolin et al., 2005; Hoffmann et al., 2011; Auersperg et al., 2014), but these studies have not simultaneously tested an ability to track hidden moving objects and represent their identities. Our results provide the first evidence that kea can simultaneously represent the identities of two objects and their trajectories. These abilities are essential to understanding more complex causal mechanisms that occur partially or entirely out-of-sight (Wilcox & Baillargeon, 1998; Kushnir et al., 2010). This research therefore opens up the possibility of exploring how much kea, and other animals, understand the complex causal mechanisms of the world around them.

Chapter VI

*Kea show three signatures of domain-general statistical inference*⁵

6.1. Abstract

One key aspect of domain-general thought is the ability to integrate information across different cognitive domains. Here, we tested whether kea can use relative quantities when predicting sampling outcomes, and then integrate both physical information about the presence of a barrier, and social information about the biased sampling of an experimenter, into their predictions. Our results show that kea exhibit three signatures of statistical inference, and therefore can integrate knowledge across different cognitive domains to flexibly adjust their predictions of sampling events. This result provides evidence that true statistical inference is found outside of the great apes, and that aspects of domain-general thinking can convergently evolve in brains with a highly different structure from primates. This has important implications not only for our understanding of how intelligence evolves, but also for research focused on how to create artificial domain-general thought processes.

⁵ Based upon: Bastos, A. P. M. & Taylor, A. H. (2020). Kea show three signatures of domain-general statistical inference. *Nature Communications*, 11, 828. doi:10.1038/s41467-020-14695-1

6.2. Introduction

There is currently great debate on the extent to which both human and nonhuman intelligence is domain-specific (Duchaine, Cosmides, & Tooby, 2001; Shettleworth, 2010, 2012) or domain-general (Horn & Cattell, 1967; Carroll, 1993; Deary, Spinath, & Bates, 2006; Burkart, Schubiger, & van Schaik, 2017): that is, whether subunits of the mind have evolved to solve specific adaptive problems, or whether intelligence evolves more generally, with the same cognitive mechanisms applied flexibly to multiple problems (Cosmides & Tooby, 1987). In humans, one source of evidence for domain-general intelligence, rather than domain-specific intelligence, are correlations between performance at different tasks ('g'; Carroll, 1993; Spearman, 1904). Further evidence for domain-generality in humans comes from our ability to transfer and combine information across different domains (Mithen, 1996; Gentner, Holyoak, & Kokinov, 2001; Carruthers, 2002; Burkart et al., 2017). In animals, while there is some evidence for 'g' (Anderson, 1993; Wass et al., 2012; Shaw, Boogert, Clayton, & Burns, 2015; Shaw & Schmelz, 2017), this remains controversial (Amici, Barney, Johnson, Call, & Aureli, 2012; Shettleworth, 2012; Burkart et al., 2017; van Horik, Langley, Whiteside, Laker, & Madden, 2018; Volter, Tinklenberg, Call, & Seed, 2018), and there is currently little evidence for cross-modular integration of information (Carruthers, 2002; Premack, 2007; Stevens, Kennedy, Morales, & Burks, 2016; Burkart et al., 2017). This has led to claims that such integration is unique to humans (Evans, 2003; Premack, 2007, 2010; Shettleworth, 2012; Spelke, 2013) and dependent on language (Mithen, 1996; Carruthers, 2002).

Reasoning under uncertainty is a central part of human decision-making (Bell, Raiffa, & Tversky, 1988; Koehler & Harvey, 2004). Making inferences about uncertainty involves generating logical predictions about future events based on limited information (Wang, 2007). This ability emerges much earlier in human development than expected for such an advanced form of cognition (Denison, Reed, & Xu, 2012; Denison & Xu, 2014), and this type of

reasoning has a number of key characteristics. First, when observing sampling events with a large number of objects, infants show true statistical inference, using the relative frequency of objects in a population to infer the most likely sampling outcome, rather than using quantity heuristics based on the absolute number of objects (Denison & Xu, 2014). Second, infants can integrate information about physical constraints into their statistical inferences (Téglás, Girotto, Gonzalez, & Bonatti, 2007; Denison & Xu, 2010; Téglás et al., 2011). For example, infants override predictions based purely on relative probabilities when some objects in a population cannot be sampled because they are held back by a physical barrier (Téglás et al., 2007). Third, infants integrate social information about the preferences of a sampler into their statistical inferences, using their knowledge of an individual's bias to again override predictions based purely on relative probabilities. When an agent shows a preference by consistently selecting a minority item from a population, infants integrate this knowledge into their sampling predictions and expect biased sampling in the future (Xu & Denison, 2009; Gweon, Tenenbaum, Schulz, & Carey, 2010; Kushnir, Xu, & Wellman, 2010; Ma & Xu, 2011; Wellman, Kushnir, Xu, & Brink, 2016). These results suggest that infant statistical inference has three signatures: it uses relative frequencies (Signature 1) and is domain-general, as infants can make predictions that integrate relative frequency judgements with information from both the physical domain (Signature 2) and the social domain (Signature 3).

Great apes are the only nonhuman species that have demonstrated true statistical inference, as they use the relative numbers of items within and between populations when predicting sampling events (Rakoczy et al., 2014; Eckert, Rakoczy, Call, Herrmann, & Hanus, 2018a; Eckert, Call, Hermes, Herrmann, & Rakoczy, 2018b), rather than using quantity heuristics based on the absolute number of positive or negative objects. In contrast, capuchins use quantity heuristics based on the absolute frequency of negative items (Tecwyn, Denison, Messer, & Buchsbaum, 2017), and it is not yet clear whether rhesus monkeys, long-tailed

macaques, pigeons, and African grey parrots use relative frequency or the absolute number of either positive or negative items (or events) when predicting sampling outcomes (Clements, Gray, Gross, & Pepperberg, 2018; Roberts, MacDonald, & Lo, 2018; De Petrillo & Rosati, 2019; Placi, Padberg, Rakoczy, & Fischer, 2019). At present, there is no evidence that any nonhuman animal can take physical constraints into account during sampling, but chimpanzees are capable of integrating social information about the preferences of a sampler into statistical inference. When given the choice of two experimenters who had previously both sampled preferred food items from a population, chimpanzees preferred to take a hidden sample from the experimenter who had shown a preference for picking the preferred food item from an unfavourable population (Eckert et al., 2018a).

Birds are an ideal group to test for domain-general statistical inference. This group has shown evidence not only of complex cognition (Taylor, 2014; Güntürkün & Bugnyar, 2016), but also of behaviour suggestive of domain-general intelligence (Bird & Emery, 2009; Laumer, Bugnyar, & Auersperg, 2016). Here, we examined whether the kea, a parrot species endemic to New Zealand, show three signatures of human statistical inference, using comparable tasks to those administered to infants (Téglás et al., 2007; Denison et al., 2012; Denison & Xu, 2014; Wellman et al., 2016) and primates (Rakoczy et al., 2014; Tecwyn et al., 2017; Eckert et al., 2018a; Placi, Eckert, Rakoczy, & Fischer, 2018).

6.3. Methods

6.3.1. Subjects and Apparatus

Our subjects were six male kea at Willowbank Wildlife Reserve (Table 15). Kea were housed in a large outdoor aviary, where food and water were available *ad libitum*. Participation in the study was voluntary and subjects were free to leave mid-session at any point. This research was conducted under ethics approval from The University of Auckland Ethics Committee (reference number 001816).

Subject	Hatch Date (Known or estimated)	Sex	Rearing History
Blofeld	August 2013	Male	Captive bred
Bruce	September 2012	Male	Wild
Loki	August 2014	Male	Captive bred
Neo	September 2012	Male	Captive bred
Plankton	August 2014	Male	Captive bred
Taz	September 2012	Male	Captive bred

Table 15. All subjects were parent reared. Hatch dates were either known (for captive bred subjects) or estimated from veterinarian evaluations (for wild subjects).

Each subject was allocated an individual training platform (42cm x 42cm) within the aviary on which they were tested. Performance in trials was rewarded with soaked Science Hill Diet pellets. A small wooden shelf (60cm x 20cm) with a plexiglass screen (43cm x 29cm) was used to separate subjects from the apparatus and the experimenter during testing. Transparent jars (\varnothing 10.5cm, 16cm tall) were used during training and testing which contained populations of either rewarding (black) or unrewarding (orange) wooden tokens (7cm x 1cm x 1cm). Each jar held a maximum of 120 tokens. When the jars were too large for a population of tokens,

tokens sat on a cardboard platform that was placed inside the jar, to ensure subjects could not see the experimenter's hands during sampling. Semi-circular cardboard lids (\varnothing 11.5cm, 5.5cm tall) were attached to the top of each jar to ensure subjects could not see which tokens were being sampled. Where barriers were used, a blue foam disk (\varnothing 10.5cm, 1cm thick) was added into the jar.

6.3.2. General Testing Procedures

Throughout training and testing, subjects were required to select which of two closed hands contained an out-of-sight rewarding token, while ignoring the hand containing the unrewarding token. The rewarding token would then be exchanged with the experimenter for a food reward. Where subjects attempted to exchange an unrewarding token, this was taken by the experimenter but not rewarded.

Before each experimental session, subjects were given motivation trials, where they had to select and exchange a rewarding (black) token and ignore a nearby unrewarding (orange) token with the experimenter three times in a row, prior to the start of the session. This ensured subjects were eager to work and remembered which of the two tokens they should search for at test. Testing was carried out by three experimenters who were blind to experimental design and hypotheses, wearing mirrored sunglasses. Subjects only proceeded to the next testing condition or experiment upon reaching a criterion of 17/20 correct choices within the same block or completing 240 trials (12 blocks) without reaching criterion. This ensured that subjects were confident in the current task before proceeding to a more demanding one. Throughout testing, hand presentation (parallel or crossed), and location of the rewarding hand at time of choice were all pseudorandomised and counterbalanced within blocks of 20 trials. Throughout training and testing, kea could only see the experimenter's hand disappear behind the cardboard

occluder on the top of the jar. Therefore, subjects were unable to see how far down the populations the experimenter's hand reached, or which token it sampled from the population. In test conditions with very disparate ratios of rewarding to unrewarding tokens, we ensured that at least two tokens from the minority population were fully visible to the subjects in every trial.

6.3.3. Training for Experiment 1

Subjects were trained to attend to and track hand trajectories for a previous study. Subjects were trained specifically for this study on hand-tracking so that they could follow the motion of sampling and make inferences about sampling from token populations in two jars, by selecting a hand that picked a rewarding token from a population of 100% rewarding tokens, over a hand that sampled from a population of 100% unrewarding tokens. In order to allow for a full counterbalancing of trial presentations at test and minimise side biasing, subjects were also taught to simultaneously attend to the side on which jars were placed and whether hands were presented in parallel or crossed over. This was trained over four separate training phases, which are described below.

In Phase 1, subjects observed a sampling motion: the experimenter picked up a single token from a wooden shelf behind a plexiglass barrier and presented it to the kea within one of their fists. They then did the same motion with their other hand. One fist contained the rewarding (black) token, whilst the other contained the unrewarding (orange) token. This ensured the kea could track objects that were picked up and then held, out-of-sight in an experimenter's hand. Hands were presented either in parallel or crossed over each other. Subjects had to pass this stage of training at 17/20 to proceed to the next phase.

In Phase 2, subjects observed sampling from a population of tokens: the experimenter picked up a single token from either a transparent jar with 120 rewarding (black) tokens, or 120 unrewarding (orange) tokens. Kea could not see what the experimenter picked up, as the transparent jars had a cardboard lid which hid the experimenter's hands. Both jars were presented at the same time, though hands took turns sampling. The order of sampling was pseudorandomised. Jar positions remained the same for 10 consecutive trials at a time, and the experimenter's fists could be presented either in parallel or crossed over. This training step taught kea to attend to the contents of each jar and infer what could be sampled from them, i.e., only rewarding tokens from a 100% rewarding population, and only unrewarding tokens from a 100% unrewarding population. Subjects had to pass this stage of training at 17/20 to proceed to the next phase.

Phase 3 involved sampling from the same jars as before, one with 120 rewarding tokens and another with 120 unrewarding tokens, but now the experimenter's fists were only presented in parallel for every trial. The position of jars was pseudorandomised and counterbalanced within blocks of 20 trials. This ensured that subjects attended to the fact that jar sides could be swapped regularly between trials. Only parallel hand motions were used in order to keep the new task demands to a bare minimum and facilitate learning. Subjects had to pass this stage of training at 17/20 to proceed to the next phase.

The final phase of training was identical to the previous stage, but now the experimenter's fists could be presented either in parallel or crossed over. Now, subjects were expected to synthesise all learning from previous steps together in a single demanding task. They had to attend to the contents of each jar, the position of each jar, and the trajectory followed by the experimenter's fists (parallel or crossed over) to succeed in any given trial. Experimenters wore mirrored sunglasses as a control for Clever Hans effects. To ensure that subjects were not attending to specific cues of a particular experimenter, subjects had to pass

this stage of training at 17/20 with two different experimenters. Once subjects passed all four phases of sampling training, they proceeded to Experiment 1.

6.3.4. Procedure for Experiment 1

This experiment investigated whether kea are able to make statistical inferences about two populations of objects using relative frequencies. Over three conditions, we tested whether kea would prefer a sample from a population containing a majority of rewarding tokens, as opposed to a population where they were in the minority, and whether kea rely on (i) relative frequencies, (ii) the absolute number of rewarding tokens, or (iii) the absolute number of unrewarding tokens, when choosing between samples from two populations. Illustrations for the three conditions are provided in Figure 12.

The first condition aimed to test whether kea would prefer a sampled token from a population where there was a higher probability of randomly sampling a rewarding token, as opposed to a population where there was a higher probability of sampling an unrewarding token. Two jars were presented: one contained a 1:5 ratio of rewarding-to-unrewarding (rewarding-to-unrewarding tokens), and the other contained a 5:1 rewarding-to-unrewarding ratio. Both jars contained 120 tokens in total.

The second condition tested whether kea were making their choices based on absolute frequencies or relative frequencies. In order to make this distinction, subjects were presented with two jars containing the same number of rewarding (black) tokens, in differing proportions. One jar had a 1:5 rewarding-to-unrewarding population of 120 tokens, whilst the other had a 5:1 rewarding-to-unrewarding population of 24 tokens. If kea were using the absolute number of rewarding tokens to guide their choices, we predicted they would choose at chance. If, in

contrast, they were taking into account the relative proportion of rewarding-to-unrewarding tokens, we predicted they would choose the jar with only four unrewarding tokens.

In the third condition, we presented subjects with two jars containing the same number of unrewarding tokens: one jar had a 57:63 rewarding-to-unrewarding population (120 tokens total), whilst the other had a 3:63 rewarding-to-unrewarding population (66 tokens total). If kea were simply selecting the jar containing the fewest unrewarding tokens rather than comparing between the frequencies of token populations between jars, they should perform at chance in this condition.

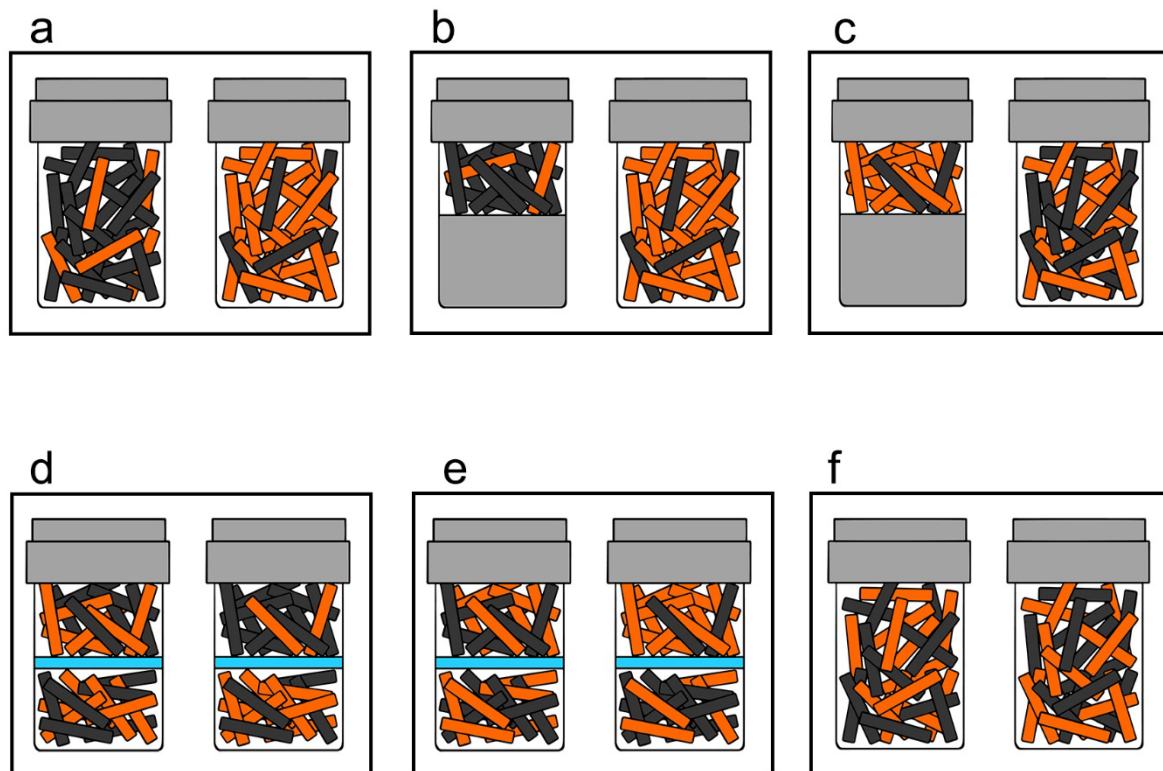


Figure 12. Proportional representation of token populations for Experiments 1-3. In the illustrations, orange rectangles represent the unrewarding tokens, and black rectangles represent the rewarding tokens. A-C show the token frequencies for Condition 1, Condition 2 and Condition 3. D-E shows the token frequencies for Experiment 2, with the blue lines representing a physical barrier. F shows the token frequencies at test for Experiment 3.

6.3.5. Training for Experiment 2

Kea were given two training experiences before being tested for Experiment 2. The first one of these provided an egocentric learning opportunity, and the second provided an allocentric demonstration of the physical constraint imposed by a barrier. In the first training experience, subjects were shown two small transparent jars (\varnothing 6cm, 7.5cm tall), each containing 20 rewarding (black) tokens, where one had a blue foam barrier (1cm thick) at the top and the other did not. The jar with the barrier was therefore impossible to sample from, whereas the rewarding tokens in the open jar could be reached. Jars were presented by an experimenter for 5 seconds at a time, and sides were pseudorandomised. Subjects were allowed to touch either jar. If subjects touched the jar with the barrier, they were allowed to interact with the barrier for 3 seconds, then both jars were removed, and the next trial was started immediately. Subjects learned to sample a black token from the accessible jar and were encouraged to exchange it for a food pellet with the experimenter. If subjects sampled more than one token, they were asked to exchange all their tokens before receiving a single food reward. This ensured that subjects were not motivated to take as many rewarding tokens from the accessible jar as possible within any given trial. Subjects had to touch the accessible jar first, and sample a rewarding token from it, on 17/20 trials in order to proceed to the next training experience.

In the second training experience, subjects watched as an experimenter sampled from two large jars (\varnothing 10.5cm, 16cm tall) of 80 rewarding (black) tokens. One jar's population was made inaccessible by a blue foam barrier (1cm thick) positioned above the tokens, whilst the other did not contain a barrier. The experimenter attempted to sample from the inaccessible jar three times before closing their fist holding nothing, then with their other hand successfully sampled a rewarding token from the other jar. The order in which the two sampling demonstrations occurred was pseudorandomised and counterbalanced, as were the sides of the two jars. Both hands were then presented to the subject either in parallel or crossed over. As in

previous training and testing, subjects indicated their chosen hand by touching the experimenter's fist with their beak. If they found a rewarding (black) token, this was handed over and exchanged for a food reward. If they selected the closed fist, the experimenter opened their hands and showed them an empty palm, then proceeded to the next trial.

6.3.6. Procedure for Experiment 2

This experiment investigated whether kea are able to integrate physical constraints into their sampling inferences. Over two conditions, we presented kea with two populations of tokens which were split in the middle by physical barriers and tested whether kea understood that only the population above the barrier could be sampled from.

In the first condition, both jars each contained 40 rewarding and 40 unrewarding tokens. One jar had a 1:1 rewarding-to-unrewarding population (40 tokens) both above and below the barrier, and the other had a 5:1 rewarding-to-unrewarding population (24 tokens) above the barrier and 5:9 rewarding-to-unrewarding population (56 tokens) below it. This was used to test whether kea were simply attending to which jar had the largest number of rewarding tokens near the top, which should lead to performance at chance, as opposed to comparing between the relative frequencies of tokens for the two accessible populations. Subjects were also expected to perform at chance in this condition if they were comparing between the token frequencies of the two jars without taking the barrier into account, as both jars contained the same absolute number and relative frequencies of rewarding and unrewarding tokens, 1:1 (40 rewarding, 40 unrewarding).

The second condition was identical, but with reversed proportions: one jar had a 1:1 rewarding-to-unrewarding population of 40 tokens above and below the barrier, whilst the other had a 1:5 rewarding-to-unrewarding population (24) tokens above the barrier and the

remaining 9:5 rewarding-to-unrewarding population below it. This condition tested whether kea were selecting the jar with the fewest unrewarding tokens near the top, in which case they should perform at chance, or comparing between the relative frequencies of the two accessible populations in the two jars. Again, both jars contained the same absolute number and relative frequencies of rewarding and unrewarding tokens.

6.3.7. Training for Experiment 3

Two experimenters were randomly assigned and counterbalanced between birds as either biased (hereafter ‘E1’) or unbiased (hereafter ‘E2’). The procedure was based on the study by Eckert and colleagues (2018a) with chimpanzees and required four experience phases.

In the first phase, we ensured that kea could tell the difference between the two experimenters: E1 and E2 stood next to each other and either picked up a food pellet or nothing into their right hand, then closed their fist. E1 and E2 either switched sides or stood on the same side for 5 seconds, before calling the subject’s name in turn and presenting their hands simultaneously for the subject to make a choice. The experimenter’s sides, the order of their actions, whether or not they switched sides (and whether the experimenter that switched sides did so by walking behind or in front of the other), and the order in which the subject’s name was called, were all pseudorandomised and counterbalanced within sessions of 10 trials. Subjects received this training until they achieved a 17/20 criterion.

Following this, subjects were given a preference test. E1 and E2 offered an empty hand to the subject as it held a rewarding token. The subject then had a choice of whom to deliver the token to, in exchange for a reward. Which experimenter placed the token on the platform and the side on which each experimenter stood were pseudorandomised and counterbalanced

within blocks of 20 trials. In order to proceed to the next stage, subjects were required to show no preference for either experimenter, that is, select E1 at between 9/20 and 11/20.

Subjects then observed demonstrations by the two experimenters where they had the opportunity to learn that E2 picked randomly from a population of tokens, whilst E1 acted as a biased sampler. For the demonstration, E1 and E2 stood next to each other, and neither wore mirrored sunglasses so the kea could see their eyes. E2 always had a 10:1 rewarding-to-unrewarding population of 110 tokens, whilst E1 always had a 1:10 rewarding-to-unrewarding population of 110 tokens. Therefore, based on sampling probability alone, E2 was far more likely to sample a rewarding token than E1. During the demonstrations, E1 and E2 took turns sampling, and E2 always tilted their heads back and looked up whilst sampling, whilst E1 lowered their heads close to the jar and looked into it as they made a choice, keeping their hands in the jar for 3 seconds. Both experimenters always sampled a rewarding token, so that they were equally reinforced. After sampling, either both experimenters stood on the same side for 5 seconds, or switched sides, before presenting their closed fists to the subject simultaneously. Which side each experimenter stood on, who sampled first, whether or not they switched sides (and whether they did so by going behind or in front of the other experimenter), were all pseudorandomised and counterbalanced within sessions of 10 trials. In order to proceed to the next experience phase, subjects had to select E1 at 9/20 or above, showing that they had no preference for E2 and were therefore not simply attending to the token populations within jars during demonstrations. All subjects passed this criterion within 20 trials except for Neo, who experienced two blocks (40 trials) of demonstrations.

The final experience phase before test was a memory probe. In this phase, E2 presented each bird with a block of 20 trials where two jars of 120 tokens each contained either 100% rewarding or 100% unrewarding tokens. E2 wore mirrored sunglasses for this phase, and presented their hands in parallel or crossed over, as in previous experiments. This was done by

E2 because they were the unbiased experimenter. We predicted that if greater exposure to one or another person before test could affect test results, then carrying out an extra set of trials with E2 would make the choice of E1 less likely at test. Similarly, an increased number of positive “rewarding token” experiences with E2 should make the choice of E1 less likely at test. Jar sides and hand presentation were counterbalanced and pseudorandomised. This phase ensured that subjects could and would still attend to the contents of jars following the demonstrations and had not simply learned to ignore jar contents during the demonstration phase.

6.3.8. Procedure for Experiment 3

Experiment 3 tested whether kea could take a biased sampler’s preference into account during a sampling event. At test, subjects observed three trials of demonstrations identical to those encountered during training, and then jars were swapped to 1:1 rewarding-to-unrewarding populations of 110 tokens. Based on token probability alone, E1 and E2 were now equally likely to sample a rewarding token. However, E1 and E2 behaved in identical fashion to demonstration trials, suggesting that they were biased and unbiased samplers, respectively. At test, E2 sampled truly randomly, whilst E1 continued to sample only the rewarding token in each trial. We expected that if kea understood that E1 was a biased sampler, they should choose them significantly above chance.

6.3.9. Analyses

All trials were filmed and coded *in situ*. Subject performance was blind coded for 10% of all video data and compared to *in situ* coded data. Inter-observer reliability was perfect (Cohen’s kappa = 1.0).

Performance in the first 20 trials of each condition were analysed at the individual level, using two-tailed Bayesian binomial tests with a test value of 0.5. We used Bayesian correlation tests to investigate average performance across the first 20 trials of each condition over trial number, and average performance on the first 20 trials of each condition across the six experimental conditions. We used default parameters (non-directional correlation, prior width = 1) for all correlation tests. These statistical analyses were carried out in JASP 0.9.2 (Jasp Team, 2019). We followed the convention that a Bayes factor (BF) < 0.33 shows substantial support for the null hypothesis, whilst a BF > 3 shows substantial support for the competing hypothesis (Wagenmakers et al., 2018).

We also analysed first-trial performance at the group level using a Bayesian intercept-only model, using a Bernoulli distribution. We fitted our model to all thirty-six first trial data points, across all individuals and conditions. Intercepts were given weakly-informative Gaussian priors ($M = 0$, $SD = 1$), to reduce overfitting. Reported pMCMC values reflect the probability of performance differing from a 0.5 chance baseline. This analysis was conducted in R 3.4.1 (R Core Team, 2019) using the “brms” package (Bürkner, 2017). We used Stan to run Hamiltonian Monte Carlo estimations (Carpenter et al., 2017).

6.4. Results

6.4.1. Experiment 1

In Experiment 1, we presented six kea with three tasks where they watched sampling events from two populations of mixed objects (rewarding and unrewarding tokens) at different proportions. Over the course of three conditions, we tested whether kea could make sampling predictions using relative rather than absolute quantities. In Condition 1, subjects were presented with two populations, one with 100 rewarding tokens and 20 unrewarding tokens,

the other with these frequencies reversed. Kea observed an experimenter sampling from each of these populations and were then presented with two closed hands. Three of six kea spontaneously showed a preference for the hand that had sampled the population with 100 rewarding tokens within their first 20 trials (Bayesian binomial test, relative preference 0.5, $BF > 3$, Table 16). We then gave all kea experience with this task until they chose the hand sampling the population with 100 rewarding tokens in 17/20 trials, which took kea 120 trials ($s = 61.97$ trials) on average.

Condition 2 tested if kea were using an absolute quantity-based heuristic by selecting the jar with the most rewarding tokens. Here, kea had to choose between two hands that both sampled from a population containing 20 rewarding tokens. However, one population had 100 unrewarding tokens, and the other had 4 unrewarding tokens. Four kea chose the jar with fewer unrewarding tokens above chance within their first 20 trials (Bayesian binomial test, relative preference 0.5, $BF > 3$). Again, we gave subjects experience with this task until they reached a 17/20 criterion, which took 66.67 trials ($s = 41.31$ trials) on average.

Condition 3 controlled for a second absolute quantity-based heuristic, the avoidance of the jar with the most unrewarding tokens, a control which capuchins fail (Tecwyn et al., 2017). Kea observed sampling from one jar which contained 63 unrewarding tokens and 57 rewarding tokens and a second that again contained 63 unrewarding tokens but only 3 rewarding tokens. All six kea chose the jar with 57 rewarding tokens above chance within their first 20 trials (Bayesian binomial test, relative preference 0.5, $BF > 3$), and took 46.67 trials ($s = 27.33$ trials) on average to reach the 17/20 criterion.

The results of Experiment 1 provide conclusive evidence that kea show true statistical inference using the relative frequency of items, rather than using quantity heuristics based on the absolute number of items. Four of our six subjects performed above chance within their

first 20 trials of both Conditions 2 and 3, indicating they had not learnt during their past experience to use a heuristic based on choosing the population with either the most rewarding, or most unrewarding items. Kea therefore did not behave as capuchins do, in using the absolute number of positive or negative tokens within the jars to make decisions (Placi et al., 2018; Tecwyn et al., 2017). Instead, kea mirrored the performance of infants (Denison & Xu, 2014) and chimpanzees (Rakoczy et al., 2014) in using the relative frequency of objects across this experiment.

	Experiment 1			Experiment 2		Experiment 3
	Condition 1	Condition 2	Condition 3	Condition 1	Condition 2	Condition 1
Blofeld	10/20 (BF=0.27)	14/20 (BF = 1.29)	15/20 (BF = 3.22)	18/20 (BF = 262.80)	15/20 (BF = 3.22)	12/20 (BF = 0.40)
Bruce	12/20 (BF = 0.40)	16/20 (BF = 10.31)	16/20 (BF = 10.31)	17/20 (BF = 43.80)	15/20 (BF = 3.22)	17/20 (BF = 43.80)
Loki	15/20 (BF = 3.22)	16/20 (BF = 10.31)	19/20 (BF = 2496.61)	18/20 (BF = 262.80)	19/20 (BF = 2496.61)	11/20 (BF = 0.30)
Neo	14/20 (BF = 1.29)	14/20 (BF = 1.29)	15/20 (BF = 3.22)	15/20 (BF = 3.22)	14/20 (BF = 1.29)	15/20 (BF = 3.22)
Plankton	15/20 (BF = 3.22)	15/20 (BF = 3.22)	16/20 (BF = 10.31)	17/20 (BF = 43.80)	17/20 (BF = 43.80)	10/20 (BF = 0.27)
Taz	15/20 (BF = 3.22)	16/20 (BF = 10.31)	17/20 (BF = 43.80)	17/20 (BF = 43.80)	16/20 (BF = 10.31)	17/20 (BF = 43.80)

Table 16. Number of correct trials performed by each subject ($n = 6$) within the first block of 20 trials for each condition of the three Experiments. In Experiment 3, correct trials constituted trials in which the subject chose the biased sampler, E1. Which of the two experimenters was the biased sampler (E1) was counterbalanced across subjects split into two groups: Neo, Bruce, Blofeld; and Loki, Plankton, Taz. Performance was tested using two-tailed Bayesian binomial tests (test value of 0.5, default Beta prior parameters at 1.0). Green cells indicate values with a Bayes Factor greater than 3.

6.4.2. Experiment 2

Experiment 2 tested whether kea could integrate information about a physical constraint into their prediction of a sampling event, as infants do (Téglás et al., 2007). After giving kea experience of a barrier, we presented them with two tasks where two jars, each with a barrier placed half-way down the jar, were sampled from. Each jar contained identical overall populations of tokens (80 tokens total: 40 rewarding and 40 unrewarding) but the proportions differed above and below the barriers. In Condition 1, one jar contained 20 rewarding and 20 unrewarding tokens above the barrier, and the same below it. The other contained 20 rewarding and 4 unrewarding tokens above the barrier and the remaining 56 tokens below the barrier (20 rewarding, 36 unrewarding). Condition 2 was identical but with the frequencies reversed. In one jar, one population contained 20 rewarding tokens and 20 unrewarding tokens above and below the barrier. The other contained 4 rewarding, and 20 unrewarding tokens above the barrier, and the remaining tokens below it (36 rewarding, 20 unrewarding). If kea could integrate knowledge of the barrier into their relative frequency judgements, we predicted they would choose the jar with 20 rewarding tokens and 4 unrewarding tokens above the barrier in Condition 1 and the jar with equal numbers of rewarding and non-rewarding tokens above the barrier in Condition 2.

As in Experiment 1, kea were given further training until they reached a criterion of 17/20 trials, which took them, on average, 26.67 trials ($s = 16.33$ trials) for Condition 1 and 48 trials ($s = 30.33$ trials) for Condition 2. Five of the six subjects tested performed above chance in the first 20 trials of both conditions of Experiment 2 (Bayesian binomial test, relative preference 0.5, $BF > 3$).

These results not only confirm the results of Experiment 1, in showing kea use the relative frequency of objects to make statistical inferences, but also show that kea can flexibly

integrate physical knowledge into these inferences. When a barrier was placed in the jar, kea used only the relative frequency above the barrier when deciding which hand was more likely to contain a rewarding token.

6.4.3. Experiment 3

Experiment 3 investigated whether kea could integrate social information about sampler biases into their predictions. We closely matched the procedure used in chimpanzees (Eckert et al., 2018a), providing kea with experience of a biased and an unbiased sampler. Kea observed demonstrations of biased and unbiased sampling from the two experimenters. While the biased sampler selected rewarding tokens from a population of 110 tokens containing 10 rewarding and 100 unrewarding tokens, the unbiased sampler selected rewarding tokens from a population of 10 unrewarding and 100 rewarding tokens. Therefore, during demonstrations, both samplers were equally associated with a rewarding sampling outcome: both experimenters always sampled a rewarding token, but whilst the biased sampler did so by looking into a population with a minority of rewarding tokens, the unbiased sampler did so by blindly sampling from a population with a majority of rewarding tokens.

At test, kea observed as the same two samplers picked from populations with an equal number of rewarding and unrewarding tokens (55 rewarding, and 55 unrewarding tokens). If kea continued to use the relative frequencies of the tokens in each jar, we expected them to choose at chance. In contrast, if the kea understood that the biased sampler was indeed biased to choose a rewarding token, while the unbiased sampler had only been choosing rewarding tokens at the same frequency as the biased sampler due to the populations they were sampling from, kea should choose the biased sampler at test. This was because while the unbiased sampler would now be likely to choose a rewarding token half the time, the biased sampler should continue to choose the rewarding token in every trial. In order to succeed at this task,

kea would need to integrate the social knowledge acquired during the demonstration phase into their probabilistic sampling predictions. Three of the six kea chose the biased sampler above chance at test (Table 16). None of these kea had previously shown a preference for either experimenter in the token-exchange task or during the demonstration phase. These results therefore mirror those from infants (Wellman et al., 2016) and chimpanzees (Eckert et al., 2018a) in showing that kea can integrate social information on sampler biases into their statistical inferences.

6.4.4. First Trial Performances

In order to ensure that kea's performance was not merely a consequence of associative learning within the first 20 trials of each condition, we analysed first trial performance across the three experiments. Across all conditions and all experiments, subjects' first trials were correct in 72.22% of trials. Taking into account only the subjects that succeeded within the first 20 trials of each condition, first trial performances were correct in 81.48% of cases. We fitted an intercept-only Bayesian model to our first trial data for all subjects. When compared against a 0.5 baseline probability of success, our model revealed that the median posterior probability of a randomly selected kea succeeding within their first trial, regardless of condition, was 0.70 (pMCMC = 0.005).

6.4.5. Learning Effect Analyses

We ran Bayesian correlation tests of average performance across the first 20 trials of each condition to examine whether performance increased over the course of the first 20 trials. We found no evidence for learning effects ($BF < 3$; results for each condition's analysis are reported in Table 17). Similarly, we examined if performance increased across conditions, but found no

correlation between condition number and average performance within the first 20 trials of each condition ($BF = 0.508$). These results show that subjects' performance did not improve over time within or between test conditions, and therefore kea were not relying on low-level associative strategies to make their sampling predictions.

		r	Bayes Factor
Experiment 1	Condition 1	0.019	0.278
	Condition 2	-0.275	0.527
	Condition 3	-0.157	0.339
Experiment 2	Condition 1	0.335	0.733
	Condition 2	-0.371	0.928
Experiment 3	Condition 1	0.174	0.357

Table 17. Results of correlation Bayesian correlation tests (non-directional correlation, prior width = 1) of trial number and average performance across all subjects, for each of the conditions tested across the three experiments. A Bayes factor (BF) < 0.33 shows substantial support for the null hypothesis, whilst a $BF > 3$ shows substantial support for the competing hypothesis.

6.4.6. Accidental Token Visibility Coding and Simulations

Up to this point our experimental and data processing methods did not account for the fact that tokens may have been visible between the experimenter's fingers in some of the experimental trials. This could have provided kea with additional information when selecting between the jars in the experiment.

To rectify this issue, three independent coders first reviewed the video data files for all trials ($n = 720$) to identify the trials in which both experimenters' hands were visible in the recording ($n = 432$). The video files for these trials were then reviewed using frame-by-frame analysis to identify trials in which token colours might have been visible to the kea. Coders

identified 10 potentially problematic trials, out of the total 432 coded trials, where the subject may have had additional information prior to making their choice. These trials were then either substituted with the next available training trial (for kea that did not reach the 17/20 criteria in their first 20 trials) or removed entirely where no subsequent trials were available (for kea that did reach the 17/20 criteria in their first 20 trials). In the video data for the remaining 288 trials, it was not possible to see both experimenters' hands from an appropriate angle. Therefore, we randomly applied the observed error rate (10/432 trials, 2.31%) to these 288 trials, and so randomly allocated 7 trials out of these 288 to be either substituted by the next available trial within that condition or removed entirely where subsequent trials were not available. The updated version of the original results is presented below as Table 18. The recoding and error simulation did not significantly affect the performance of any kea in any condition over the course of our three experiments.

To check that our initial simulation was not an outlier, and so represents a likely set of substitutions or removals of seven randomly assigned trials, we ran it an additional five times. In four of five of these repetitions, trial removals or substitutions did not affect subjects' performance in any blocks: above chance performances remained above chance, and chance performances remained at chance. In one of the five repetitions, the simulation affected three blocks: one showed a performance change from chance levels to above chance, and two changed from above chance performance to chance. Following the changes, Blofeld showed above performance at 15/20 in Experiment 1's Condition 2, and Plankton performed at 14/20 in the same condition, which, therefore, resulted in no difference to the total number of kea passing this test (4 of 6), and thus does not change our conclusions in any way. One of Neo's correct trials in Experiment 3 was removed without substitution, reducing his score from 15/20 (BF = 3.22) to 14/19 (BF = 2.254) resulting in two, rather than three kea passing this condition. Thus, only in one of the six simulations that we ran in total (one for our actual results, a further

five to check the simulation robustness) did performance significantly change for any kea in any condition. Furthermore, even the one significant change we did observe across our six simulations does not affect our conclusions about kea cognition, because a further two kea passed the condition in question.

	Experiment 1			Experiment 2		Experiment 3
	Condition 1	Condition 2	Condition 3	Condition 1	Condition 2	Condition 1
Blofeld	10/20*** (BF=0.27)	14/20** (BF = 1.29)	15/20** (BF = 3.22)	18/20* (BF = 262.80)	15/20** (BF = 3.22)	12/20*** (BF = 0.40)
Bruce	12/20* (BF = 0.40)	15/20* (BF = 3.22)	16/20* (BF = 10.31)	17/20* (BF = 43.80)	15/20* (BF = 3.22)	17/20*** (BF = 43.80)
Loki	15/20*** (BF = 3.22)	16/20** (BF = 10.31)	19/20*** (BF = 2496.61)	18/20*** (BF = 262.80)	18/19*** (BF = 1379.71)	10/18*** (BF = 0.32)
Neo	14/20* (BF = 1.29)	14/20* (BF = 1.29)	15/20* (BF = 3.22)	15/20* (BF = 3.22)	14/20* (BF = 1.29)	15/20** (BF = 3.22)
Plankton	15/20* (BF = 3.22)	15/20** (BF = 3.22)	16/20** (BF = 10.31)	17/20* (BF = 43.80)	17/20*** (BF = 43.80)	9/19*** (BF = 0.28)
Taz	15/20*** (BF = 3.22)	16/20* (BF = 10.31)	17/20* (BF = 43.80)	15/18* (BF = 16.91)	16/20* (BF = 10.31)	17/20*** (BF = 43.80)

Table 18. Number of correct trials performed by each subject (n = 6) within the first block of 20 trials for each condition of the three Experiments. In Experiment 3, correct trials constituted trials in which the subject chose the biased sampler, E1. Which of the two experimenters was the biased sampler (E1) was counterbalanced across subjects split into two groups: Neo, Bruce, Blofeld; and Loki, Plankton, Taz. Performance was tested using two-tailed Bayesian binomial tests (test value of 0.5, default Beta prior parameters at 1.0). Green cells indicate values with a Bayes Factor greater than 3. The seventeen blocks denoted by a single asterisk had all 20 trials manually verified by an independent coder. Seven further blocks, denoted by two asterisks, had at least 50% of all trials manually verified by an independent coder, and non-verified trials were included in the error simulation. Twelve further blocks, denoted by three asterisks, had all trials included in the error simulation.

6.4.7. Clever Hans Control

In order to ensure that no other potential biases may have affected our data, we ran a control experiment which addresses whether kea may have relied on any unintentional experimenter cues, such as body posture or breathing patterns, to select the hand holding the rewarding black token. We presented all six individuals included in the original study with twenty trials where both jars contained 55 rewarding tokens and 55 unrewarding tokens (50:50 ratio). In each set of trials, the experimenter consistently sampled a rewarding token from one jar, and an unrewarding token from the other. As in all test conditions, the experimenters wore mirrored sunglasses whilst sampling from the two jars. If subjects could use any unintentional experimenter cues or biases to choose the hand containing a rewarding black token, we would expect them to perform above chance across the twenty trials in this condition. On the other hand, if subjects were relying on the proportions of rewarding and unrewarding tokens to make their choices, they should perform at chance. None of the six kea performed above chance level at this control (Table 19) and therefore this additional experiment finds no support for the hypothesis that the kea used any unintentional experimenter cues to guide their choices.

Subject	Performance in Clever Hans Control
Blofeld	12/20 (BF=0.40)
Bruce	13/20 (BF = 0.64)
Loki	9/20 (BF = 0.30)
Neo	8/20 (BF = 0.40)
Plankton	12/20 (BF = 0.40)
Taz	11/20 (BF = 0.30)

Table 19. Number of correct trials performed by each subject ($n = 6$) within the first block of 20 trials for the experiment controlling for experimenter cues or biases. Performance was tested using two-tailed Bayesian binomial tests (test value of 0.5, default Beta prior parameters at 1.0).

6.5. Discussion

Our study shows that kea display three signatures of human statistical inference. Experiment 1 shows that, like infants (Téglás et al., 2007; Denison & Xu, 2010, 2014; Wellman et al., 2016) and chimpanzees (Rakoczy et al., 2014; Eckert et al., 2018a), kea showed evidence of true statistical inference. Kea predicted likely sampling outcomes based on the ratio of objects in the populations being sampled from, rather than using absolute quantity heuristics, such as selecting the population with the greatest number of positive tokens or avoiding the population with the most negative tokens. Experiment 2 shows that kea, like infants (Téglás et al., 2007; Denison & Xu, 2010; Téglás et al., 2011), can integrate knowledge about a physical barrier into their predictions of a sampling outcome, even though the overall population distributions were identical. Finally, Experiment 3 shows that, like infants (Wellman et al., 2016) and chimpanzees (Eckert et al., 2018), kea are capable of integrating social information about biased and unbiased samplers into their predictions. As in past work on chimpanzees (Eckert et al., 2018a), kea took observed information about the biases of experimenters and integrated it into their predictions of what would occur when these experimenters sampled from equally distributed populations of objects. Therefore, just like infants and the great apes, kea made statistical inferences using relative rather than absolute quantities, and then integrated social and physical information into their predictions, using their knowledge of physical barriers and the bias of a sampler, to override predictions based purely on relative probabilities (Denison & Xu, 2010).

To observe these three signatures of domain-general statistical inference in kea is surprising, even given recent developments in avian cognition. Much work in this field over the past fifteen years has focused on corvids, which, as a group, have produced their most impressive problem-solving performances predominantly on domain-specific, ecologically relevant tasks, such as those involving caching or tool use (Dickinson & Clayton, 1998; Emery

& Clayton, 2001; Chappell & Kacelnik, 2002; Dally, Emery, & Clayton, 2006; Alexis, Raby, Clayton, & Dickinson, 2007; Taylor, Hunt, Holzhaider, & Gray, 2007; Taylor, Elliffe, Hunt, & Gray, 2010; Weir & Kacelnik, 2006; Wimpenny, Weir, Clayton, Rutz, & Kacelnik, 2009; von Bayern, Danel, Auersperg, Mioduszevska, & Kacelnik, 2018). Parrots have only recently become the focus of a sustained, global research effort (Auersperg & von Bayern, 2019). Behaviour suggestive of more domain-general processes have emerged in both groups from studies examining the ability of non-tool users to solve tool problems (Bird & Emery, 2009; Auersperg, Gajdon, & Huber, 2010; Auersperg, von Bayern, Gajdon, Huber, & Kacelnik, 2011; Auersperg, Huber, & Gajdon, 2011; Auersperg, Szabo, von Bayern, & Kacelnik, 2012; Auersperg et al., 2016; Laumer et al., 2016). Our results both support these claims and greatly extend them, in showing that the integration of very different types of information – concerning physical barriers and sampling biases – into statistical inferences is possible in at least the mind of one parrot species. Furthermore, both the token transfer behaviours, and the observation of sampling events from a population of objects, had very little ecological relevance to the kea, a species that feeds by extractive foraging in alpine and sub-alpine environments (Greer, Gajdon, & Nelson, 2015). Despite this, kea not only learned to solve these problems, but did so by using domain-general statistical inference, rather than quantity heuristics.

It is important to note, however, that the level of information integration across our experiments was different. Experiment 2 tested if kea could simultaneously integrate information about relative frequency with the presence of a barrier in order to make a judgement, because relying solely on one source of information – either the presence of a barrier, or the overall frequency of the population – would have led to the incorrect choice. In Experiment 3, however, kea needed to override their prior reliance on relative frequency information in favour of information about the sampling bias of an experimenter. Our results therefore suggest that kea can fully integrate physical information with relative frequency by

using both sources of information simultaneously and can integrate social information by overriding relative frequency information with social information. More work is needed to test if kea can also simultaneously consider social and relative frequency information, or can make judgements that simultaneously combine social, physical, and relative frequency information together.

Unlike previous studies in primates (Rakoczy et al., 2014; Tecwyn et al., 2017; Eckert et al., 2018a, b; Placi et al., 2018, 2019), kea were presented with additional training trials before progressing to the next condition. This procedure is commonly used during cognitive experiments on birds (Seed, Tebbich, Emery, & Clayton, 2006; Teschke et al., 2013; Gruber et al., 2019) to ensure subjects are consistent at a specific condition before being given a novel one. It seems unlikely this additional training would have affected the strategy kea were using to make their sampling predictions across our experiments, as this additional experience would have only consolidated whatever strategy kea were using at the time, be it the use of relative frequencies or quantity heuristics. That is, the learning of the kea was unstructured: both quantity heuristics and domain-general statistical inference would have worked to solve several of the early problems presented to kea, yet kea clearly mirrored the type of statistical inferences made by humans and the great apes when solving these problems.

It is currently unclear how infants, apes, and the kea in this study extract statistical information. In particular, as Denison and Xu (2014) note, it is not yet clear if subjects make inferences using discrete or continuous quantity representations, but this does not detract from the results here: either way, subjects were extracting information about the relative frequencies, using it to make predictions about the relative probability of reward (Experiment 1), and then integrating these judgements with both physical (Experiment 2) and social (Experiment 3) information. One key area for future work will be determining exactly how statistical information is extracted and represented by humans, apes, and kea.

Birds last shared a common ancestor with humans at least 312 million years ago (Benton & Donoghue, 2007). This evolutionary distance suggests that domain-general statistical inference may have arisen twice on our planet via convergent evolution. However, further work is required over a wider range of avian and primate species to provide a more accurate evolutionary account of when statistical inference has emerged in different taxa, and so test this hypothesis further.

The statistical inference abilities observed in our study are particularly interesting given that birds have brains with a much smaller absolute size than humans, a very different structure (Güntürkün & Bugnyar, 2016), and much higher neuronal densities (Olkowicz et al., 2016). Our results therefore suggest that (i) aspects of domain-general thought can evolve multiple times, rather than being a one-off or a product of a highly unlikely sequence of evolutionary events, and (ii) that a particular brain architecture, specifically a layered neocortex, is not necessary for this type of higher-order intelligence to evolve. This has important implications not only for our understanding of how intelligence evolves, but also for research focused on creating artificial, domain-general thought processes (artificial general intelligence, AGI), specifically on the degree to which such processes should mirror mammalian cortical processes and structures (Riesenhuber & Poggio, 1999; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007; Hassabis, Kumaran, Summerfield, & Botvinick, 2017; Krizhevsky, Sutskever, & Hinton, 2017).

Chapter VII

*Are parrots naïve realists? Kea behave as if the real and virtual worlds are continuous*⁶

7.1. Abstract

Human psychology and animal cognition have increasingly used virtual stimuli to test cognitive abilities, with the expectation that participants are “naïve realists”, that is, that they perceive virtual environments as both equivalent and continuous with real-life equivalents. However, there have been no attempts to investigate whether nonhuman subjects in fact behave as if physical processes in the virtual and real worlds are continuous. As kea parrots have previously shown the ability to transfer knowledge between real stimuli and both images on paper and touch screens, here we test whether kea behave as naïve realists and so expect physical processes to be continuous between the physical and virtual worlds. We find that, unlike infants, kea do not discriminate between these two contexts, and that they do not exhibit a preference for either. Our findings therefore validate the use of virtual stimuli as a powerful tool for testing the cognition of nonhuman animal species.

⁶ Based upon: Bastos, A. P. M., Wood, P. M., Taylor, A. H. (2021) Are parrots naïve realists? Kea behave as if the real and virtual worlds are continuous. *Biology Letters*.

7.2. Introduction

The use of virtual stimuli has become common in studies of both human psychology and animal cognition (Bohil, Alicea, & Biocca, 2011; Dolins, Klimowicz, Kelley, & Menzel, 2014; Egelkamp & Ross, 2019; Lovato & Waxman, 2016; Naik, Bastien, Navab, & Couzin, 2020; Seitz et al., 2021; Stowers et al., 2017). Subjects are sometimes required to respond to virtual stimuli through real-life choices, for example by selecting between real-life objects or agents after observing relevant videos on a screen (e.g., Krupenye & Hare, 2018; Lucca, Pospisil, & Sommerville, 2018). The validity of both virtual reality (VR) and mixed-context tasks therefore relies on the assumption that individuals are “naïve realists”, interpreting the virtual and real worlds as both equivalent and continuous, and behaving identically in both contexts. However, a recent study with 19-month-old human infants suggests that this is not necessarily the case: infants do not act as naïve realists, as they do not perceive the real and virtual worlds as continuous (Revencu & Csibra, 2020). When observing a seesaw depositing a ball into one of two boxes, infants correctly predicted the location of the ball when all stimuli were virtual or when all stimuli were real, but not when a virtual seesaw appeared to deposit a ball in a real box outside of the screen. This suggests that, at 19 months, infants already perceive a disconnect between events occurring in virtual and real environments, and do not expect physical processes to crossover between the two. Therefore, both experiments in VR environments and those employing a blend of real and virtual stimuli may not accurately assess how infants behave in real situations (Revencu & Csibra, 2020).

To date, there have been no attempts to test whether animals also expect a discontinuity in physical events between the real and virtual worlds, despite the regular use of virtual stimuli and VR in animal testing (Dolins, Klimowicz, Kelley, & Menzel, 2014; Egelkamp & Ross, 2019; Naik, Bastien, Navab, & Couzin, 2020; Seitz et al., 2021; Stowers et al., 2017), and studies showing that they recognise images of real objects on both photographs and screens

(Huskisson, Jacobson, Egelkamp, Ross, & Hopper, 2020; Johnson-Ulrich et al., 2016; Railton, Foster, & Temple, 2014; Spetch & Weisman, 2012; Winkinson, Mueller-Paul, & Huber, 2013). It is therefore unclear if animals are naïve realists and behave equivalently across VR, mixed, and fully real contexts or if, like 19-month-old infants, they perceive a disconnect between events occurring in virtual and real environments.

Kea, *Nestor notabilis*, present an excellent model species to assess whether parrots, and potentially other species, might perceive the real and virtual worlds as continuous, given that they can generalise object discriminations learnt from photographs and touch screens to real objects (Wein et al., 2015) and display ape-like performances in a range of cognitive tasks (Huber & Gajdon, 2006; Auersperg et al., 2009, 2010, 2011; Heaney et al., 2017; Bastos & Taylor, 2020). One study has also shown that kea perform better in associative and reversal learning tasks with real stimuli compared to touchscreen stimuli, raising the possibility that they do not perceive the two as equivalent (O'Hara et al., 2015).

Here, we tested if kea act as naïve realists, and so treat a physical process in the real and virtual worlds as continuous, by adapting the infant study by Revencu & Csibra (2020) for within-subject animal testing. In our first experiment, kea observed scenes where either: (i) a real seesaw deposited a real spherical token into one of two real boxes, (ii) a virtual seesaw deposited a virtual token into one of two virtual boxes, or (iii) a virtual seesaw appeared to deposit a virtual spherical token into one of two real boxes. We then assessed whether kea had an innate preference for real or virtual stimuli by examining if kea chose a box where a real skewer deposited a real spherical token over a real box where a virtual skewer appeared to deposit a virtual token.

7.3. Experiment 1

7.3.1. Subjects and Apparatus

Our subjects were six male kea housed at Willowbank Wildlife Reserve. Subjects were tested on a 76 x 45 x 1.5 cm wooden platform containing a purpose-built 37 x 4 x 29 cm touchscreen case and anti-glare shield. This held an ASUS VivoBook Flip TP401M 14" capacitive touchscreen laptop. A 45 x 0.3 x 30cm transparent acrylic barrier was used during trials to stop subjects approaching the stimuli during manipulations. Across different conditions, the setup was modified to include either a real or virtual seesaw (measuring 30cm wide; the real seesaw had a 13cm handle protruding from the back so it could be surreptitiously tipped by the experimenter), and two virtual or real boxes (measuring 6 x 6 x 6cm; insides of real boxes were padded with felt) with opaque occluders above them. All animations were set over a white background, and when the real seesaw was used, it was placed in front of a white card background (Figure 13).

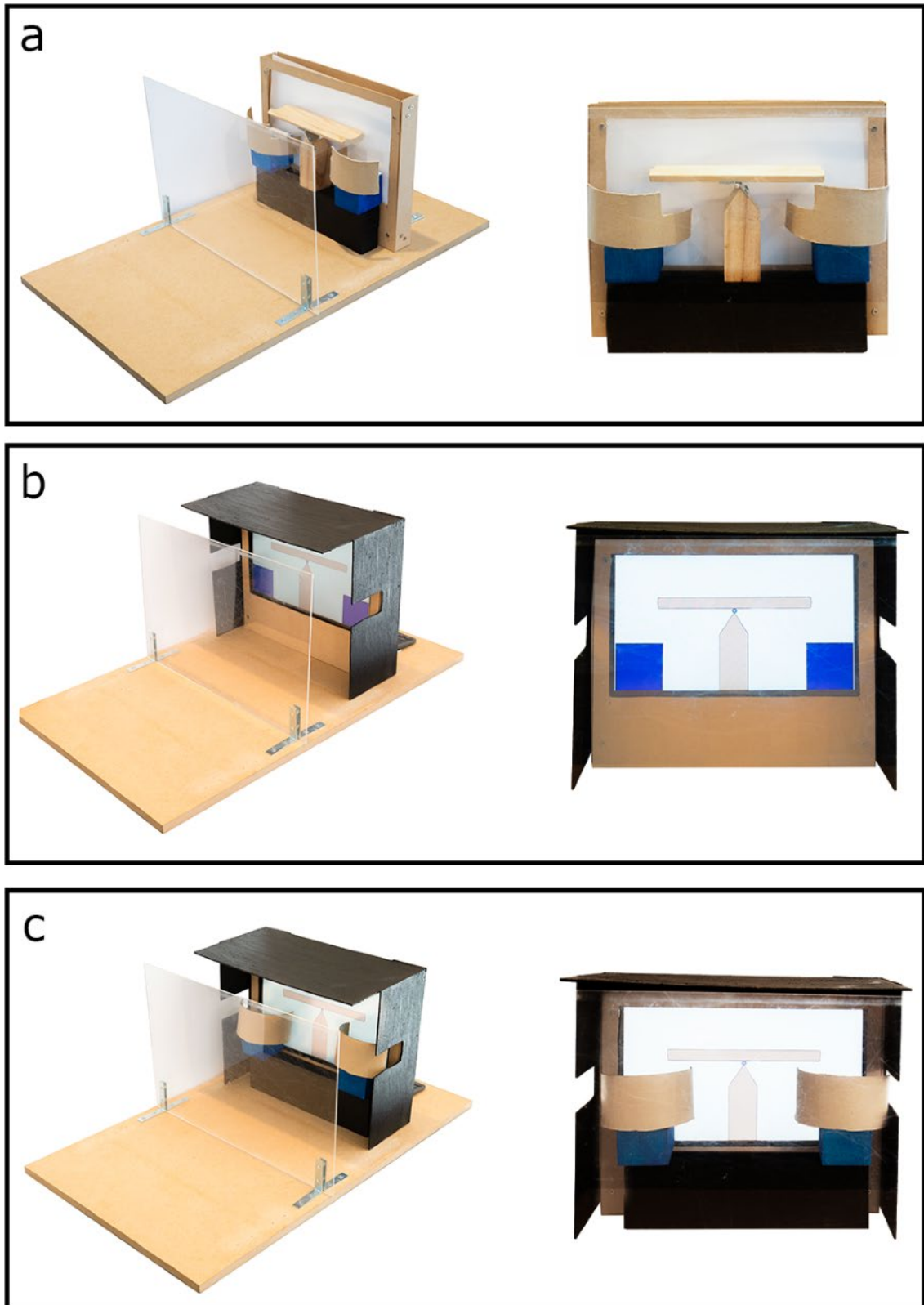


Figure 13. The experimental apparatus used for (a) Reality Condition, (b) Virtual Condition, and (c) Crossover Condition.

7.3.2. Procedure

Subjects received only two experiences with touchscreens prior to this study: they were trained to use their tongues to activate the screen, and they were trained to select a black token over an orange token when both were presented simultaneously on the screen. Subjects received four stages of training for this study:

1. Token Exchange: Kea were required to exchange a spherical black token (\varnothing 2.5 cm) with an experimenter ten times in a row.
2. Real Box Searching Training: Subjects were trained to indicate which of two real boxes contained the spherical token after the experimenter placed it into a box with their hand, then closed both boxes' lids simultaneously. Subjects were rewarded for correctly indicating the location of the token in 17/20 trials, but the contents of the boxes were never revealed.
3. Virtual Box Searching Training: This was identical to Stage 2, but the hand, token, and boxes were virtual. Kea had to select the correct box in 17/20 trials.
4. Seesaw Habituation: Kea observed the experimenter visibly tip the seesaw with their hands to both sides. The session ended when subjects were approached and touched the seesaw, which was not rewarded.

Following this training, kea experienced three conditions analogous to those presented in the infant study (Revcu & Csibra, 2020), in the following order: (i) Reality Condition, where they observed a real seesaw tip to either side, silently depositing the spherical foam token into one of the boxes; (ii) Virtual Condition, where they observed the same scene but with virtual seesaw, boxes, and token; and (iii) Crossover Condition, where they observed a virtual seesaw with a virtual spherical token tip toward one of two real boxes. In all conditions, the top of each box was hidden by an occluder so subjects could not see the ball being deposited

into either box. Subjects were given additional blocks of 20 trials in the Reality and Virtual Conditions until they passed criteria at 17/20, to ensure that they could track the spherical token's location when all stimuli were real and when all stimuli were virtual, so any failures in the Crossover Condition could not be attributed to a poor understanding of the task. The Crossover Condition was presented to subjects until they performed differently from chance (i.e., performed at $< 5/20$ or $> 15/20$). In all trials, subjects indicated their choice by touching a box with their beak or tongue and were rewarded for selecting the correct box. An acrylic barrier ensured that subjects could only approach the boxes when they were allowed to make their choice. Before subjects could select between real boxes, a lid was placed over both boxes and the occluders were removed. In the Virtual Condition, the occluders disappeared before subjects made their choice.

7.3.3. Coding and Analyses

Subjects' responses were coded *in situ* (by an observing experimenter for Reality and Crossover Conditions; automatically by testing software in the Virtual Condition) and recorded on video. Agreement between *in situ* and video coding for 120 trials was perfect (33.3% of all trials; Cohen's kappa = 1.0). Individual performances were analysed with two-tailed Bayesian binomial tests (test value = 0.5, Beta prior widths = 1.0). We considered Bayes Factor values > 3 as evidence for the alternative hypothesis, and values < 0.33 as evidence for the null hypothesis (Wagenmakers et al., 2018). All analyses were carried out in JASP 0.14.1.0 (Jasp Team, 2019).

7.3.4. Results

Three kea performed above chance within their first block of the Reality Condition, and four subjects performed above chance within their first block of the Virtual Condition (Table 20). All kea performed at or above the training criterion of 17/20 in these conditions, taking a maximum of 60 trials over 3 blocks to do so (Reality Condition: $M = 43.3$ trials, $SD = 15.1$ trials; Virtual Condition: $M = 36.7$ trials, $SD = 19.7$ trials). In the Crossover Condition, five subjects performed above chance within their first block, selecting the box the virtual token moved towards in the majority of trials. The only subject that exhibited no significant preference (Taz) demonstrated the same pattern in his subsequent block, selecting that box in 18/20 trials. Therefore, given experience of the same task in both fully virtual and fully real contexts, kea respond to a crossover of virtual and real stimuli as if they are continuous.

	Reality Condition	Virtual Condition	Crossover Condition
Blofeld	18/20 [1]	20/20 [1]	19/20 [1]
Loki	16/20 [2]	16/20 [2]	18/20 [1]
Moriarty	15/20 [2]	17/20 [1]	18/20 [1]
Neo	13/20 [3]	14/20 [3]	16/20 [1]
Plankton	14/20 [2]	17/20 [1]	17/20 [1]
Taz	11/20 [3]	14/20 [2]	14/20 [2]

Table 20. Number of correct choices in the first block of each of the three conditions of Experiment 1 (for the Crossover Condition, the choice of the box where the token would have landed if the virtual stimuli were continuous with the real world as scored as ‘correct’). In square brackets, the number of blocks experienced by subjects until they passed training criteria at 17/20 (Reality and Virtual Conditions) or performed differently from chance (below 5/20 or above 15/20; Crossover Condition). Values with a Bayes Factor > 3 (Bayesian binomial tests, test value = 0.5, default Beta prior parameters = 1.0) are shown in green-shaded cells and in bold.

7.4. Experiment 2

7.4.1. Rationale

In Experiment 2, we directly pitted real and virtual stimuli against each other to distinguish between three hypotheses to explain kea's performances in Experiment 1, which are, in increasing order of complexity: (i) kea's performance in the Crossover Condition is attributable to a learnt associative rule for selecting the box the token moved closest to, irrespective of whether the stimuli were real or virtual; (ii) kea are true naïve realists and believed that the real box in the Crossover Condition would contain the virtual token apparently deposited into it; and (iii) kea do in fact understand that the real and virtual worlds are discontinuous, but this was masked by using the aforementioned associative rule.

7.4.2. Subjects and Apparatus

All kea now observed trials with the same apparatus as before, but where the seesaw scene was substituted for two wooden skewers (one virtual, one real), which took turns depositing spherical tokens into one of two boxes.

7.4.3. Procedure

In Condition 1, kea watched as a real spherical token was deposited into a real box by a real wooden skewer, and a virtual token was deposited into the other real box by a virtual skewer. Given that they now observed a token being deposited into each of the two boxes, we expected kea to choose the box the real token was deposited into if they thought that the real and virtual worlds are discontinuous, but this belief had been masked in Experiment 1 by an associative rule for selecting the box the token had moved towards. That is because tokens were now

deposited into both boxes, making the correct option the box containing a real and retrievable token. Therefore, if kea's naïve realism in Experiment 1 was an artefact of a learnt associative rule, pitting these two options against each other should eliminate the usefulness of this rule. These trials were provided as 20 unrewarded probe trials interspersed among 80 Virtual Box Searching Training trials identical to those received prior to Experiment 1. This ensured that subjects remained motivated to participate in probe test trials despite receiving no reinforcement for their choices.

In Condition 2, kea observed the same demonstrations as before, with one real skewer and one virtual skewer. However, this time, the virtual skewer deposited the token onto the top of the box. The token then rolled along the top of the box and dropped next to it and off the bottom of the screen. If the behaviour of the kea was being driven by an associative rule whereby they simply selected the box the token was seen to move closest to, we now expected kea to perform at chance. If, rather, kea are naïve realists and saw real and virtual stimuli as synonymous, then we expected them to preferentially choose the box containing the real token. The order of events and the side where the real stimulus was presented were counterbalanced and pseudorandomised.

7.4.4. Coding and Analyses

Interobserver agreement between *in situ* coding and video-coding for 40 randomly selected test trials for each experiment was perfect (33.3% of trials; Cohen's kappa = 1.0). Individual performances were analysed with two-tailed Bayesian binomial tests (test value = 0.5, Beta prior widths=1.0). Average performances across the six subjects over the course of their 20 trials were correlated to trial number to test for possible learning effects, using non-directional Bayesian correlations (Beta prior widths = 1.0).

7.4.5. Results

None of the subjects demonstrated a preference for either stimulus type in Condition 1 (Table 21), suggesting that they may have perceived both as equivalent and judged both boxes as equally likely to contain a rewarding spherical token. However, as a majority of subjects preferred the real stimulus on their first trial, we tested for any change in behaviour across trials by conducting a post-hoc Bayesian correlation (non-directional, default stretched beta prior width=1.0) between average performance across subjects and trial number. This did not reveal conclusive evidence for a decrease in preference for the real stimulus over time (Pearson's $r = -0.439$, $BF_{10} = 1.593$, 95% CI [-0.710, 0.013]). In Condition 2, three subjects performed above chance, suggesting that at least these subjects were not using an associative rule to guide their choices throughout Experiments 1 and Condition 1 of Experiment 2.

Subject	Condition 1	Condition 2
Blofeld	9/20	11/20
Loki	10/20	19/20
Moriarty	12/20	14/20
Neo	13/20	15/20
Plankton	11/20	13/20
Taz	9/20	15/20

Table 21. Number of trials where subjects preferred the real stimulus over the virtual stimulus in Conditions 1 and 2 of Experiment 2. Values with a Bayes Factor > 3 (Bayesian binomial tests, test value = 0.5, default Beta prior parameters = 1.0) are shown green-shaded cells.

7.5. Discussion

Unlike 19-month-old infants, kea behaved as naïve realists, predicting that a virtual seesaw would deposit a spherical token in a real box placed in front of the computer screen. In a follow-up experiment, three subjects not only did not display a preference for a real token over a virtual token when both were apparently deposited into two real boxes, but also did not use a simple associative rule to guide their choices, suggesting that subjects did not distinguish real from virtual stimuli in this context.

These results are in clear contrast to those from 19-month-old infants, who differentiate between the real and virtual worlds. In the infant study, participants did not receive extensive training and were each tested in only one of the three conditions (Revencu & Csibra, 2020). In contrast to this between-group structure, the within-individual nature of our tasks gave kea additional experience with virtual and real stimuli acting in a continuous manner. However, Condition 1 and 2 of Experiment 2 provide a clear test of the effect of this experience. If kea actually understood the difference between the real and the continuous worlds, but this understanding was masked by their experience of Experiment 1, we would expect them to show a preference for the real-world stimuli in Condition 1, but they did not. Furthermore, if kea's performances were simply a consequence of following a simple associative rule, then no subjects should have preferred the real stimuli in Condition 2 of Experiment 2, but three kea did so. This provides the first evidence that nonhuman animals might be naïve realists. Future work could examine the performance of animals with different levels of training and testing history with screens to examine if animals ever stop behaving as naïve realists and could also test younger infants to examine if humans ever behave as naïve realists.

These results support the use of tasks mixing real and virtual environments with kea and potentially other avian species, suggesting that their behavioural responses are

interchangeable between these two contexts. Additionally, our results raise the hypothesis that kea are unable to perceive representations as such, and instead understand virtual stimuli as perceptually and functionally identical to real stimuli. Testing if other species also treat the real and virtual worlds as equivalent and continuous offers a way to test the hypothesis that the ability to understand and distinguish between representations and real objects might be a core – and potentially unique – aspect of our cognition (Bulley, Redshaw, & Suddendorf, 2020; Revenu & Csibra, 2020; Suddendorf, Redshaw, & Bulley, *in press*).

Chapter VIII

*Macphail's null hypothesis of vertebrate intelligence: Insights from avian cognition*⁷

8.1. Abstract

Macphail famously criticized two foundational assumptions that underlie the evolutionary approach to comparative psychology: that there are differences in intelligence across species, and that intelligent behavior in animals is based on more than associative learning. Here, we provide evidence from recent work in avian cognition that supports both these assumptions: intelligence across species varies, and animals can perform intelligent behaviours that are not guided solely by associative learning mechanisms. Finally, we reflect on the limitations of comparative psychology that led to Macphail's claims and suggest strategies researchers can use to make more advances in the field.

⁷ Based upon: Bastos, A. P. M. & Taylor, A. H. (2020) Macphail's null hypothesis of vertebrate intelligence: Insights from avian cognition. *Frontiers in Psychology*, 11, 1692. doi:10.3389/fpsyg.2020.01692

8.2. Background

Euan Macphail sparked great controversy in the 1980s, following his synthesis of the current state of comparative psychology. Macphail argued that, given the body of evidence available at the time, there appeared to be no quantitative or qualitative differences in intelligence across species (Macphail, 1982, 1985; Macphail & Bolhuis, 2001), and that seemingly intelligent behavior is underpinned by associative learning (Macphail, 1982, 1985; Macphail & Bolhuis, 2001). In Macphail's general process view, species differences in performance within tasks could be ascribed to contextual variables, rather than to any real differences in their underlying cognition. He directly pitted this view against the widely regarded ecological view (Macphail & Bolhuis, 2001), which takes an evolutionary perspective of cognition, suggesting that species evolve cognitive adaptations to their environment, just as they do physiological adaptations. Here, we will consider two of these lines of Macphail's criticism in light of recent developments in avian cognition: namely, that there are differences in intelligence across species, and that intelligent behavior cannot be explained by associative learning alone.

8.3. Differences in Intelligence Across Species

Macphail's null hypothesis of vertebrate intelligence posits that all animals use the same general mechanisms, to the same level of ability, to solve cognitive tasks. Whilst this may be true in considering some basic processes such as operant conditioning, which appear to be universal across species, this hypothesis fails to consider differences in intelligence at finer, and more ecologically relevant, scales (Shettleworth, 1987).

A strong line of evidence suggesting not only that intelligence is quantitatively different across species, but that these differences emerge as a direct consequence of their ecology, explores the relationship between species' social complexity and cognitive task performance.

The social intelligence hypothesis predicts that complex social environments require better memory and overall cognitive capacity, and so social complexity drives the evolution of intelligence (Dunbar, 1998, 2008). Comparative studies across both parrot and corvid species provide support for this hypothesis. For example, parrots living in complex groups involving fission-fusion dynamics outperform those that form smaller and more stable family groups in string-pulling tasks testing means-end comprehension (Krasheninnikova et al., 2013). Similarly, pinyon jays, which live in large flocks of up to five hundred individuals, outperform western scrub-jays, which form small family groups, in tasks of transitive inference (Bond et al., 2003). Pinyon jays also outperform two less social species, western scrub-jays and Clark's nutcrackers, in both colour and spatial reversal tasks (Bond et al., 2007). Correlations between social complexity and cognitive capacity may be particularly strong in the corvids and parrots, due to the long life expectancy in these species, which might facilitate exposure to a greater number of social partners over time (Boucherie et al., 2019).

There is also good evidence that quantitative cognitive differences between species are driven by their ecological differences in comparative work between caching and non-caching corvids. While caching and non-caching corvids perform similarly in a colour-based task, caching species outperform non-caching species in a spatial task (Olson et al., 1995). Findings from studies such as these suggest that ecology plays an important role in shaping the cognitive abilities of species. Given that comparative psychology has so far been restricted to a minority of species, it seems likely that as a greater number of species are tested, more differences in intelligence are likely to emerge (Elepfandt, 1987; Kamil, 1987; Shultz & Dunbar, 2010; van Horik & Emery, 2011), generating more clear and testable links between differences in ecology and cognitive ability.

Evidence for convergent evolution in the cognitive abilities of great apes, corvids, and parrots also suggests that quantitative differences in intelligence do exist across species, and

that these differences relate to their ecology (Emery, 2004; Clayton, 2012; van Horik et al., 2012; Güntürkün & Bugnyar, 2016; Auersperg & von Bayern, 2019). One clear prediction that the contextual variable argument makes is that differences in methodology should become more difficult to control for and, therefore, have a greater impact on task performance, the further apart two species are phylogenetically. This is because species that are more similar are more likely to share more of the same perceptual abilities and biases than those that are more distantly related. Thus, if species do not differ in intelligence, as Macphail claims, we should expect problem solving performances to differ more as phylogenetic distance increases, due to contextual variables becoming harder to control.

However, the great apes, parrots, and corvids, despite being evolutionarily distant taxa, converge in several of their cognitive abilities (Emery, 2004, 2006; Emery & Clayton, 2004; Seed et al., 2009; Clayton, 2012; van Horik et al., 2012; Güntürkün & Bugnyar, 2016; Auersperg & von Bayern, 2019). The Piagetian framework for object permanence describes different stages of development for this ability, which requires an individual to understand that an object continues to exist when hidden within a container (Piaget, 1954). Its final stage requires an understanding of invisible displacement: that is, tracking a container which presumably contains the hidden object as it moves behind a series of screens or occluders, and guessing where it may have been left once the container is shown to be empty. The great apes (de Blois et al., 1998; Call, 2001; Collier-Baker & Suddendorf, 2006; Collier-Baker et al., 2006; Mallavarapu, 2009), corvids (Pollok et al., 2000; Zucca et al., 2007; Hoffmann et al., 2011; Ujfalussy, Miklósi, & Bugnyar, 2013), and parrots (Pepperberg & Funk, 1990; Pollok et al., 2000) succeed at the final stage of object permanence, even though other species of both birds and mammals do not. Four species of lemurs (Deppe et al., 2009; Mallavarapu, 2009) succeed only at visible displacement tasks, where the reward can be seen as it moves between two or more occluders. Several other mammals also fail to understand invisible displacement

tasks (for a review, see Jaakkola, 2014), even though they understand visible displacements, suggesting that contextual variables are not to blame. Similarly, ring doves can successfully retrieve a hidden reward, but fail to track its displacement within a container (Dumas & Wilkie, 1995). Given that parrots, corvids, and the great apes show similar performance whilst more closely related mammalian and avian species fail, it seems likely that stage 6 object permanence – the ability to understand invisible displacements – emerged convergently in the great apes and these two avian taxa and represents a real quantitative difference in cognitive ability across species.

A similar convergence in cognitive ability appears in the object transposition task. In this task, a reward is hidden under one of two cups, and their positions are changed. In children, the ability to solve the transposition task emerges later than the ability to solve invisible displacement tasks (Sophian & Sage, 1983; Sophian, 1984; Barth & Call, 2006), suggesting that this is a more challenging type of problem. A large number of mammals either fail to solve transposition tasks or may use associative strategies to guide their choices, including cats (Doré et al., 1996), dogs (Doré et al., 1996; Rooijakkers et al., 2009; Fiset & Plourde, 2013), wolves (Fiset & Plourde, 2013), wild boars (Albiach-Serrano et al., 2012), pigs (Albiach-Serrano et al., 2012), goats (Nawroth et al., 2015), dolphins (Jaakkola et al., 2010), and bears (Hartmann et al., 2017). Despite this selection of species including herbivores, omnivores, and carnivores, as well both domesticated and wild animals, only parrots (Pepperberg et al., 1997; Auersperg et al., 2014) and primates (Beran & Minahan, 2000; Call, 2001, 2003; Beran et al., 2005; Barth & Call, 2006; Rooijakkers et al., 2009) have been conclusively shown to succeed at object transposition tasks. Rather than relying on associative learning strategies, these two taxa appear able to represent the change to the objects' spatial locations.

Another example is the ability to reason through inference by exclusion (Güntürkün & Bugnyar, 2016). In tests of inference by exclusion, subjects must infer that one of two stimuli

contains or is associated with a reward, after a demonstration that the other stimulus is not. Where two cups are used, for example, they must reason that if the reward is not hidden in the cup shown to be empty, then it must be in the other one. Several species of corvids (Schloegl et al., 2009; Mikolasch et al., 2012; Shaw et al., 2013; Jelbert, Taylor, & Gray, 2015b), parrots (Schloegl et al., 2009; Mikolasch et al., 2011; Pepperberg et al., 2013; O'Hara et al., 2015, 2016; Bastos & Taylor, 2019; Subias et al., 2019), and apes (Call, 2004, 2006; Hill et al., 2011) readily reason in this way. The ability to reason by exclusion is present in some New World monkeys (Sabbatini & Visalberghi, 2008; Marsh et al., 2015; Takahashi et al., 2015). Some capuchin monkeys are capable of this form of inference, whilst squirrel monkeys fail at both auditory and visual versions of the task (Marsh et al., 2015). This ability seems to be absent from other mammalian species including rats, golden hamsters, and tree shrews (Takahashi et al., 2015). This pattern suggests that the ability to reason through inference by exclusion varies quantitatively across species and has emerged convergently in the primate and avian lineages.

The ability to use probabilistic information to make predictions about uncertain events also appears to have evolved convergently in the great apes and parrots (Rakoczy et al., 2014; Bastos & Taylor, 2020). When choosing between two hidden samples taken from two mixed populations of rewarding and unrewarding objects, capuchin monkeys appear to use a heuristic strategy of simply avoiding the sample from the population with the greatest absolute number of unrewarding objects (Tecwyn et al., 2017). On the other hand, both the great apes and kea make their choices by relying on probabilistic information, by comparing the ratios of objects within and between the two populations (Rakoczy et al., 2014; Bastos & Taylor, 2020). This ability, known as true statistical inference, has so far not been conclusively shown outside of these two taxa, as other studies on primates and birds have not been able to exclude the absolute number heuristic as a potential strategy (Clements et al., 2018; Roberts et al., 2018; De Petrillo & Rosati, 2019; Placì et al., 2019).

Macphail went further than suggesting there are no quantitative differences in intelligence between species, claiming there are also no qualitative differences (Macphail, 1982, 1985; Macphail & Bolhuis, 2001). Research in this area has rather focused on whether there are types of thought that are unique to humans (Penn et al., 2008), rather than whether different animal species might think in qualitatively different ways. At present, therefore, it is not clear whether this hypothesis has been tested sufficiently to make conclusions either way. One route to testing this hypothesis further is focusing more on testing whether there are differences in the information processing biases, errors and limits of species showing similar levels of performance at different behavioural tasks (Taylor, 2014; Taylor, Bastos, Brown & Allen, *under review*).

8.4. Intelligent Behaviour Beyond Pure Association

Another of Macphail's claims is that all intelligent behavior can be explained by associative learning alone (Macphail, 1982, 1985; Macphail & Bolhuis, 2001). Critics of Macphail have highlighted that a purely associative view of cognition is insufficient to explain the intelligent behaviours observed in vertebrates (Shettleworth, 1987), including birds. There are certainly areas of the literature on avian cognition where there is great debate as to whether the problem-solving performances of birds can be explained by associative learning alone. For example, there is currently debate surrounding the role of associative learning and more complex cognition in research on planning in ravens (Redshaw et al., 2017; Lind, 2018; Dickerson et al., 2018; Hampton, 2019), stone-dropping in corvids (Taylor & Gray, 2009; Cheke et al., 2011; Taylor et al., 2011; Jelbert et al., 2014; Logan et al., 2014; Ghirlanda & Lind, 2017; Hennefield et al., 2018, 2019), and string-pulling in a wide variety of birds (e.g., Taylor et al., 2010b, 2012; for a review of the species tested on string pulling, see Jacobs & Osvath, 2015). However,

several lines of evidence indicate the presence of specific cognitive mechanisms other than associative learning in birds.

First, work on the innate cognitive capacities of birds has shown that prior experience is not required for complex problem solving to emerge. Without any prior experience, chicks can solve several problems in the physical realm, including detouring around a barrier by moving away from a desired object (Regolin et al., 1995), mentally representing the possible location of a hidden object when choosing between two different screens (Vallortigara et al., 1998; Chiandetti & Vallortigara, 2011), and recognizing partially hidden objects by representing their complete outline (Regolin & Vallortigara, 1995; Regolin et al., 2004). Research on imprinted ducklings has also revealed an innate ability to distinguish between the abstract concepts of “same” and “different”: when imprinted on two identical objects, ducklings preferred to approach pairs of identical objects rather than pairs of different objects, even though the objects in either case were different from those they were originally imprinted on (Martinho & Kacelnik, 2016). Given that these studies used inexperienced chicks and ducklings, this line of work strongly suggests that intelligence operates on more cognitive processes than associative learning alone.

Work in chicks also offers further support for an innate approximate number system. Inexperienced chicks can distinguish between both small quantities from one to four (Rugani et al., 2013a) and larger quantities between five and ten (Rugani et al., 2013b). This capacity develops in birds into a numerical ability of surprising complexity. A seminal study in pigeons trained subjects to select images including one, two, or three shapes in ascending order, after which pigeons were asked to order sets with numerosities between one and nine (Scarf et al., 2011). Pigeons succeeded in this task despite never having received training on stimuli including between four and nine shapes, suggesting that they represent one through nine on an ordinal scale.

Work on the social cognition of birds has also found clear evidence of birds performing beyond the predictions of associative learning. In a recent prosociality experiment, African gray parrots readily transferred tokens through a window to a conspecific who could exchange them for a food reward, when they could not exchange them themselves (Brucks & von Bayern, 2020). The study's control conditions suggest that this response was not driven by associative learning alone, as token transfers occurred significantly less often when their partner was unable to exchange tokens, or when the partner was absent. Similarly, an associative account would suggest that their tendency to transfer tokens would increase over time, but most subjects acted prosocially in their first trial.

Caching studies provide evidence that birds can flexibly use information learnt in an egocentric manner to make allocentric predictions about the behavior of conspecifics in their environment. For example, Western scrub-jays, which pilfer other individual's caches, strategically relocate their caches (Emery & Clayton, 2001; Dally et al., 2005, 2006) in response to novel cues of a conspecific's presence, so as to reduce the likelihood of their caches being stolen in the future. In order to do this, individuals must have pilfered others' caches before, but need not have observed a pilfering event by another individual (Emery & Clayton, 2001), suggesting that they can project their own experience onto others. An associative learning explanation fails to acknowledge how they might shift between these egocentric and allocentric perspectives. A more recent study on ravens shows these birds will re-cache food when they believe they are being watched, and not as a learned response to a conspecific's gaze (Bugnyar et al., 2016). Ravens were similarly fast to cache when they heard sound recordings of a conspecific in a nearby compartment with a peephole, which could have granted the conspecific visual access to the cache, and when a conspecific was fully visible in the nearby compartment. In contrast, ravens cached slower and made more improvements to their caches in a control non-observed condition where they could hear a conspecific in a nearby

compartment, but this conspecific was neither visible nor had a peephole available to look through. Ravens, therefore, appeared to flexibly use their egocentric experiences, in this case looking through a peephole at the caching chamber, to predict that another individual at the peephole would be able to see them caching.

Work on tool use in birds has produced a number of intriguing findings, suggesting that birds are capable of sophisticated technical intelligence (Weir et al., 2002; Taylor et al., 2007, 2010a; Tebbich et al., 2007; Bird & Emery, 2009; von Bayern, Heathcote, Rutz, & Kacelnik, 2009; von Bayern, Danel, Auersperg, Mioduszevska, & Kacelnik, 2018; Wimpenny et al., 2009; Auersperg et al., 2010, 2011b, 2012b; Teschke & Tebbich, 2011; St Clair & Rutz, 2013; Laumer et al., 2016; Jelbert et al., 2018, 2019; Fayet et al., 2020). While some of these studies suggest that birds might be capable of mental trial and error during tool use, conclusive evidence that birds can mentally represent tool problems only emerged recently from a study on New Caledonian crows (Gruber et al., 2019). This showed that these birds can pre-plan a sequence of behaviors up to three steps ahead, taking an available tool to the correct apparatus (the sub-goal) in order to retrieve another tool, which only then could be used to obtain a food reward (the overall goal). New Caledonian crows correctly planned and executed this sequence of behaviors even though all components of the sequence were out-of-sight of each other. This, therefore, required the crows to mentally represent the location and identity of the correct out-of-sight sub-goal and then use this representation to form a plan to solve the problem without error. Clear evidence of future-directed thought also comes from work on caching corvids. Western scrub-jays can anticipate their future needs, storing food that is unlikely to be available the following morning in a particular location (Raby et al., 2007; Cheke & Clayton, 2011), regardless of their current satiation state (Correia et al., 2007).

Evidence for the use of mental representations during tool manufacture has also emerged recently. After learning to insert a tool of a particular size into a vending machine,

New Caledonian crows, when given a sheet of paper, were able manufacture tools of the correct size to insert into the machine (Jelbert et al., 2018). This was despite no tool template being available at the time of manufacture for use as a reference. Instead, the crows had to rely solely on their mental representation of the tool's size. Additionally, crows were not rewarded at test for making tools of the correct size. Instead, half of all tools made were rewarded irrespective of size, meaning there was no differential reinforcement for making the correct size tool at test. This has since also been shown in Goffin's cockatoos (Laumer, Jelbert, Taylor, Rössler, & Auersperg, 2021).

Recently, work has also begun to show that birds can solve problems that require domain-general intelligence, rather than problems involving domain-specific, ecologically relevant behaviors such as tool use and caching. Initial evidence that birds might have more domain-general cognitive processes comes from studies examining the ability of non-tool users to solve tool problems (Bird & Emery, 2009; Auersperg et al., 2010, 2011a, b, 2012b, 2016; Laumer et al., 2016). More recently, a study in kea showed that they can not only make accurate probabilistic comparisons between the two sampling events, as described above, but also integrate information across different domains (Bastos & Taylor, 2020). In one experiment, the two jars contained a physical barrier, and the otherwise identical populations of tokens were unevenly distributed above and below these barriers. Kea considered the physical constraint imposed by the barriers, adjusting their predictions of the likely sampling outcomes from the two jars. Another experiment in this study provided the kea with social information on sampling biases: one human demonstrator showed they had a preference for rewarding tokens by taking them from a jar rewarding tokens were in the minority, while the other demonstrated they were an unbiased blind sampler by taking rewarding tokens from a jar where such tokens were in the majority. When both these demonstrators sampled from jars with an even split of rewarding and unrewarding tokens, kea preferentially selected the samples from the biased

demonstrator. These results showed that kea integrated either social or physical information into their probabilistic predictions, performing comparably to human infants (Téglás et al., 2007; Xu & Denison, 2009; Denison & Xu, 2010, 2014; Denison et al., 2012) and chimpanzees (Eckert et al., 2018a, b; Rakoczy et al., 2014), and outperforming monkeys (Tecwyn et al., 2017).

8.5. Echoes of Macphail’s Criticisms in the 21st Century

Despite recent research not finding support for several of Macphail’s claims, it is important to consider why Macphail may have raised these points in the first place, and why they are relevant today. The reasoning behind Macphail’s null hypothesis for differences in intelligence across vertebrates highlights a flaw that has pervaded comparative psychology for many years: it is often impossible to tell why animals fail at a task. Differences in apparatus, methodology, motivation, and other contextual factors may affect species’ performance in cognitive tasks. As highlighted by Macphail, failure at a task might be a true reflection of the species’ ability, or the result may be caused by some contextual variable in that task. Researchers may attempt to resolve this in two ways: either by presenting an identical task across species, or by modifying some contextual variables in the task so it is better suited to a particular species. However, these two solutions are equally problematic.

When contextual variables are changed to suit a particular species, this makes it even more difficult to establish the reason for a species’ failure at the task (e.g., Caldwell & Whiten, 2002; Schloegl et al., 2009; Liedtke et al., 2011; Auersperg et al., 2012a; Krasheninnikova et al., 2019; Farrar et al., 2020). Small changes in contextual variables may affect how a species interprets a task and therefore affect their performance, and this is likely to make it difficult to compare performances in a task across multiple species. One clear example of this comes from

work on the trap-tube, a problem where an animal must extract food from a tube with a tool while avoiding a trap set into the lower surface of the tube. Apes' performance at this task changed dramatically once subjects were allowed to pull food with a tool towards themselves, rather than push food away (Mulcahy & Call, 2006), with learning speed increasing greatly and subjects passing the key "inverted-tube" control, where the tube was turned upside down and therefore the trap was rendered irrelevant. This example highlights how small changes to a task can affect animal performance greatly and offers a cautionary reminder of how hard it can be to interpret failure by a species at a cognitive task.

Even presenting an identical task to two very different species may lead to false positives, or false negatives, when the two species interpret the same task differently. This has been highlighted in a number of studies where animals failed at tasks involving a human demonstrator but could have performed better had that contextual variable been changed (e.g., Erdőhegyi et al., 2007; Mikolasch et al., 2012; Shaw et al., 2013; Nawroth et al., 2014; Jelbert et al., 2016). Given that failure at a task does not necessarily represent a species' true cognitive abilities, negative results often become ambiguous and difficult to interpret, contributing to a "file-drawer effect" and publication bias (Fanelli, 2012; Farrar & Ostojic, 2019).

These issues in comparative psychology are highlighted in a landmark study by MacLean and colleagues (MacLean et al., 2014), which presented two identical tasks across thirty-six species to measure self-control: an A-not-B task, where a reward was visibly moved between two locations after being previously rewarded in only one of them, and a cylinder task, where an opaque cylinder containing food was presented and then substituted for a translucent cylinder. According to the authors, greater self-control should enable species to successfully switch search locations in the A-not-B task and avoid reaching directly for the food in the cylinder task. However, the study failed to consider how different species may perceive these tasks (Jelbert et al., 2016; Kabadayi et al., 2017): for example, that birds may perform poorly

in the A-not-B task due to a poor innate understanding of human hands (Jelbert et al., 2016), rather than an inability to exert self-control. In support of this critique, New Caledonian crows performed poorly at this task without experience of tracking hands, but after hand-tracking training actually performed comparably to the great apes in the same task (Jelbert et al., 2016).

Macphail's view suggests that errors such as these could be ruled out by exhaustively varying perceptual task features and other contextual variables to ensure that they are not responsible for subjects' failures, but in real terms this is often impossible (Kamil, 1987). One potential solution to this problem is to present pre-test baselines to different species (e.g., Jelbert et al., 2016). These baselines would comprise simple tasks that the animal would be expected to succeed at, given that the testing methods – or contextual variables – were appropriate. Success at such baselines could act as a checkpoint prior to test, ensuring that all species in the experiment understood the basic requirements of the test. Provided that a species succeeded at these baseline tasks, it would be possible to more confidently attribute failure at test to a lack of understanding, rather than other aspects of the task. For example, in the New Caledonian crow A-not-B study mentioned above, subjects first experienced hand-tracking training, watching the experimenter's hand bait a container among multiple other hand movements involving several cups and their lids (Jelbert et al., 2016). Had the subjects not first experienced this baseline training, it would not have been possible to determine if failure at the subsequent A-not-B task was due to a lack of experience with tracking human hands or reflected a failure to inhibit a response to investigate a previously rewarded container. Similarly, as highlighted earlier, the performance of various mammal species that pass visible displacement tasks, but fail invisible displacement tasks, provides stronger evidence for this failure being due to cognition rather than contextual variables, because the visible displacement task acts as a baseline test for the more complex invisible displacement task.

Another criticism of MacLean et al. (2014) is a lack of clarity on exactly which cognitive mechanisms were being tested (Beran, 2015; Brucks et al., 2017). It is unclear whether the self-control measures discussed in the study might reflect a single cognitive process, or a combination of several mechanisms (Beran, 2015). Self-control has been used as a term to describe either the ability to incur a cost in order to obtain a more valuable reward instead of a less-costly, lower-value reward (Beran, 2015), or the ability to inhibit a response (MacLean et al., 2014). The two interpretations of this ability are not necessarily underpinned by the same cognitive process. Work on dogs shows that even inhibition alone is not consistent across different tasks, suggesting that different tests for the same ability are not actually tapping into the same cognitive mechanisms (Brucks et al., 2017). Similarly, a recent study in pheasant chicks showed that comparisons across multiple tasks might not accurately reflect cognitive ability (van Horik and Madden, 2016). In the pheasant study, two hundred chicks experienced three foraging tasks, meant to assess whether individual variation in performance was robust and driven by real differences in cognitive ability. The study failed to find any consistent differences in problem solving ability between individuals across the three tasks, suggesting that motivation, and not cognitive capacity, was the main driver for these differences.

One way to help move past these issues would be to focus more on exploring how animals succeed at some tasks, and how they fail at others, rather than whether they simply fail or succeed at certain problems. The signature-testing approach, and research focused on cognitive processes rather than task performance, are a viable strategy for this (Kacelnik, 2009; Taylor et al., *under review*; Taylor & Gray, 2009; Seed et al., 2012; Fletcher & Carruthers, 2013; Taylor, 2014). A process-driven approach allows researchers to generate specific hypotheses about which errors, biases, limits, and specific patterns of performance identify particular cognitive mechanisms, and design experimental tasks that tease these potential processes apart. This is analogous to the strong inference approach (Platt, 1964), which aims

to successively rule out alternative hypotheses through the design of experiments that specifically test these hypotheses with clear predicted outcomes for each alternative explanation. Researchers can triangulate several of these signatures within or between tasks (Heyes, 1993; Taylor, 2014), to pinpoint exactly which of several qualitative forms of intelligence different species are capable of. This approach could provide a more powerful system through which we can better address Macphail's null hypothesis, particularly in terms of qualitative differences in intelligence. Several of the studies discussed earlier provide clear examples of behavioural signatures that constrain the possible cognitive mechanisms an animal can be using to solve a problem. For example, the presence of a distance effect bias in pigeons' numerical discriminations, where pigeons are more accurate and quicker to discriminate number pairs when the numerical distance between them is greater, provides a clear behavioural signature that numbers are represented on an ordinal scale (Scarf et al., 2011). Similarly, the ability of crows to solve problems without mistake when downstream aspects of the problem are out-of-sight, shows they are not limited by having problems out-of-sight and so provides a clear signature for pre-planning, as decisions have to be made using mental representations of the problem (Gruber et al., 2019).

Finally, a Bayesian framework may provide useful tools in interpreting animal performances from a statistical viewpoint. Given that research questions and methods are appropriately framed, the Bayesian framework can distinguish between a lack of power in the data and direct support for the null hypothesis (Wagenmakers, 2007; Stevens, 2017; Wagenmakers et al., 2018). In the frequentist framework, these two forms of non-significance are often confounded. This leaves researchers with inconclusive data which often ends up unpublished (the file-drawer effect: e.g., Fanelli, 2012; Farrar et al., 2020). In contrast, Bayesian analyses can be much more informative than their frequentist counterparts when animals fail at an experimental task. The Bayesian framework enables researchers to provide

claims both for and against the existence of particular cognitive capacities in their target species, rather than it being unclear whether negatives are due to low sample size or a true failure at a task. While clearly this framework does not resolve all of the issues surrounding the interpretation of ‘evidence of absence’ in comparative psychology, it does offer a route toward bringing more quantitative and qualitative differences in intelligence to light in the literature (Stevens, 2017).

8.6. Discussion

Macphail’s support of a null hypothesis for no quantitative differences in intelligence across species, and his claim that all intelligent behavior is association-based, fall short in the light of recent research in avian cognition. Avian cognition provides clear evidence for robust differences in intelligence among avian species, as well as between birds and other taxa, and for problem solving that extends beyond simple associative learning.

However, Macphail’s criticisms of comparative psychology are relevant to this day and can inspire researchers to make more advances in this field. Thirty-five years ago, Macphail highlighted the difficulty in establishing whether animals fail at a task because they cannot understand it, or because their performance was affected by variations in methodology (Macphail, 1982, 1985; Macphail & Bolhuis, 2001). Today, much of the field still grapples with this distinction. Researchers often cannot tell why subjects fail at some tasks but not others, and comparative psychology suffers from widespread publication bias (Farrar et al., 2020).

Macphail also highlighted that some of the preconceptions of the field at the time had not been appropriately tested. One of these was the belief that intelligence varies predictably across species, with humans showing the greatest intelligence, followed by their closest

relatives (Jensen, 1980). According to this view, one might expect an inverse correlation in intelligence with evolutionary distance from humans and other primates. Macphail argued that such a *scala naturae* assumption might be erroneous (Macphail, 1985), so helping move the field past this early anthropocentric attitude and towards the present day, where researchers focus on testing intelligence across a phylogenetically broad array of animal species, albeit often still with tests that have been designed for human intelligence.

In sum, while some of Macphail's claims do not hold up to the current body of evidence, a number of his criticisms of the field of comparative psychology still hold in the present day. We suggest three strategies researchers can use to combat these issues: (i) using baseline tasks to ensure that contextual variables cannot explain subjects' failure at test (Jelbert et al., 2016); (ii) focusing on a signature-testing, process-driven approach, that specifically seeks to pinpoint the cognitive mechanisms that animals rely on to solve problems (Kacelnik, 2009; Taylor, 2014; Taylor et al., *under review*; Taylor & Gray, 2009; Seed et al., 2012; Carruthers & Fletcher, 2013); and (iii) taking advantage of the Bayesian framework to distinguish between support for the null hypothesis and a lack of statistical power (Wagenmakers, 2007; Stevens, 2017; Wagenmakers et al., 2018). Put together, these three strategies can help researchers identify both quantitative and qualitative differences in intelligence between species, learn from animals' successes as well as their failures, and triangulate evidence for complex cognition that is not rooted exclusively in associative learning.

Chapter IX

Concluding Remarks

9.1. Summary of the Main Findings

In this thesis, I presented the findings of some of my work investigating different aspects of kea's numerical, physical, and social cognition, broadening the existing knowledge of the behaviour and cognition of this species, and contextualised my findings within comparative cognition more broadly. As outlined in **Chapter 1**, kea are a particularly interesting model species for the study of avian cognition, as they fulfil many of the evolutionary and ecological hypotheses that predict the evolution of large brain sizes and complex cognition in vertebrates. Kea inhabit a harsh and changeable environment where food is scarce and seasonal and that therefore require an ability to find and extract a wide range of food items (Diamond & Bond, 1999; Greer et al., 2015; Young et al., 2012), and they experience rich social lives where they interact with a large number of conspecifics (Diamond & Bond, 1999; Schwing et al., 2016) over a long lifespan (Young et al., 2012). Their suitability as a study species is also aided by their unusual neophilia and exploratory behaviour (Diamond & Bond, 1999), which may encourage the emergence of new object-driven behaviours (Auersperg et al., 2014; Gajdon et al., 2014; Goodman et al., 2018).

These innovative behaviours can include tool use, as I reported in **Chapter 2**. In this chapter, I developed a rigorous approach to reporting anecdotal behaviour in a single individual, through repeated observations and successive hypothesis-testing. In this case, I demonstrated that Bruce, a captive kea missing his upper mandible, was able to innovate self-care tooling by using small pebbles for preening. As hypothesised, the majority of the instances where he manipulated a pebble co-occurred with preening, which he performed repeatedly and specifically with pebbles. When Bruce dropped a pebble during preening, he often retrieved it

or replaced it with another pebble. Bruce also preferred pebbles of a particular size which he could use to preen, and this action that was never performed by individuals with typical beaks. Finally, Bruce interacted with smaller pebbles than those selected by other individuals, despite being able to comfortably hold larger objects. These results demonstrate that rare behaviours, such as tooling innovations, can be verified robustly through repeated observations, and provide the first evidence of self-care tooling with a pebble tool in a bird. Furthermore, this work supports the hypothesis that tooling can be innovated by individuals when necessary.

The study on Bruce's pebble tooling highlighted that tool use in parrots is remarkably underreported in the scientific literature, despite ample video evidence of tool use behaviours in parrots. Therefore, in **Chapter 3** I set out to investigate the prevalence of tool use behaviours across all parrot species. In this chapter, I demonstrated how crowdsourced data from online visual media platforms can be used to identify the presence of rare behaviours across a broad range of species and trace their likely ancestry using Bayesian phylogenetic ancestral trait reconstruction. These results indicate not only that tool use is not particularly rare in parrots, but also that one clade – the Psittaculidae – may have lost the ability to innovate or perform tool use behaviours. Importantly, the strong phylogenetic signal for tool use in parrots suggests that the crowdsourced behavioural observations in this study likely comprise examples of intentional tool use, rather than accidental combinations of two behaviours (namely, holding an object and attempting to scratch oneself), which should not be predicted by phylogeny.

Given that kea are part of the most basal clade in the parrot order, and our evidence of widespread tool use outside of the Psittaculidae, these results also suggest that the capacity for tool use is ancestral to all parrots. However, it is important to highlight that kea have been shown to use three different forms of tool use, both through spontaneous behavioural innovation and following training (Auersperg et al., 2010, 2011a, b; Bastos, Horváth, Webb, Wood, & Taylor, 2021a; Goodman et al., 2018), but the variety and flexibility of tool use in

less well-studied species remains to be investigated. Future research should investigate a wide range of parrot species to establish the extent to which tool use varies within species, in terms of the breadth and depth of different species' tool use.

Previous work on kea does suggest that they might perform better than most other birds at other physical cognition tasks, such as string-pulling (Werdenich & Huber, 2006). In **Chapter 4**, I investigated this claim by presenting kea with the loose-string connectivity task both before and after experience of the perceptual-motor feedback loop with a vertical string. The goal of this chapter was two-fold. First, it directly tested the insight, planning, and means-end understanding hypotheses, all of which predict success at this task even without feedback experience. Second, if kea failed the loose-string connectivity task, it tested whether feedback facilitates an understanding of string properties, which in turn enables successful performance following vertical string-pulling experience. I found that kea failed the loose-string task both before and after perceptual-motor feedback loop experience with a vertical string, suggesting that string-pulling behaviour in this species is not explainable by the insight, planning, or means-end understanding hypotheses, and that feedback experience alone cannot predict success at this task. This is curious given that kea have previously succeeded at other tests of connectivity (Auersperg et al., 2009), and successfully made choices between non-string stimuli using the same apparatus during training with the same apparatus. It therefore appears that success at the loose-string connectivity task is modulated by more than an ability to identify connectedness and experience of string-pulling alone. More work is required both on kea and other bird species to establish the cognitive mechanisms underlying string-pulling behaviour, and what experiences and contextual factors enable success at the loose-string connectivity task.

In **Chapter 5**, I investigated another aspect of physical cognition not previously studied in kea: the ability to mentally represent hidden objects and their trajectories. In this chapter, I

showed that kea can not only remember the locations of two tokens of different colours and values, but also continuously track their locations as they move while remaining hidden. I also demonstrated that kea can predict the likely end point of a hidden token's trajectory by observing the start of the trajectory before it is obscured by an occluder. This chapter therefore provides the first evidence that birds can mentally represent the hidden identities and trajectories of two objects simultaneously and demonstrates that kea also mentally represent likely hidden trajectories so as to predict their end point.

In **Chapter 6**, I used the abilities kea demonstrated in the previous chapter to test whether these birds are capable of making statistical inferences. In this study, kea were always presented with combinations of two different jars, both of which contained a mixed population of rewarding black and unrewarding orange tokens, and then observed as experimenters obtained hidden samples from each jar. Their task was to select the hand which they believed was most likely to contain a rewarding black token. Over the course of three experiments, I demonstrated that kea not only make true probabilistic predictions based on the proportions of black and orange tokens in the jars (not simply selecting the jar with the highest number of black tokens or avoiding the jar with the most orange tokens), but also combine physical and social bias information into their predictions. When a physical barrier was placed halfway down both jars, with distinct proportions of black and orange tokens above and below the barrier, kea considered only the reachable top half of the jars when making their probabilistic predictions. Even more surprisingly, after observing that one experimenter exhibited a bias for selecting black tokens but was equally reinforced to an unbiased experimenter, kea remembered this bias and chose the biased person at test. This chapter therefore provides the first direct evidence for domain-general intelligence in a bird, given that kea integrated information from different cognitive domains to make a single judgement, rivalling the performances of human infants and chimpanzees. One aspect that remains to be explored, both

in kea but also in infants and apes, is whether this domain-general integration occurs upstream of the decision point, with information from different domains being combined simultaneously, or in tandem, such that a physical or social considerations must override purely probabilistic judgements in an all-or-nothing fashion.

Chapter 7 investigated whether kea would expect physical processes on a screen to be continuous with real-life, acting as naïve realists. Despite the increasing use of touchscreens to test kea in cognitive studies (Wein et al., 2015; O’Hara et al., 2015; Laschober et al., 2021), we still have very little understanding of how this species perceives virtual stimuli and environments. Avian visual systems are very different from our own: parrots have lower visual acuity, better colour vision, faster visual temporal resolution (measured as the critical flicker fusion frequency, or CFF), and an equivalent mechanism for depth perception only at close distances (reviewed in Nelson & Fijn, 2013). Additionally, different methods of presentation matter. As an example, chickens do not respond appropriately to video stimuli presented on cathode-ray tube (CRT) monitors (D’Eath & Dawkins 1996; Patterson-Kane et al. 1997), which have lower flicker-rates than chickens’ CCF requires (Lisney et al. 2011) but have no issue with modern LCD screens (Nelson, Wilson, & Evans, 2008; Smith et al. 2009). Therefore, it can be difficult to tell whether their responses in touchscreen-based tasks is comparable to how they might respond to real stimuli, and to determine the extent to which they believe virtual stimuli to be real, provided that they are presented in a manner appropriate to the avian visual system. To test this, I adapted a procedure used in infants (Revenu & Csibra, 2020) for within-individual testing to establish whether kea would expect a physical process occurring on-screen to be continuous with the real-world stimuli off-screen. I found that, unlike 19-month-old infants, kea behaved as true naïve realists, expecting physical processes that took place on-screen to continue into real-life. This suggests that naïve kea respond to virtual stimuli as if

they are real, as they see the real and screen worlds as equivalent and continuous with each other.

Finally, in **Chapter 8** I revisited a long-standing debate in comparative psychology through the lens of avian cognition, including some of my own work with kea which is described in the preceding chapters. In the 1980's, Euan Macphail suggested that intelligence does not vary quantitatively or qualitatively across species, but rather that species perform differently in cognitive tasks because they perceive the contextual variables of these tasks differently (Macphail, 1982, 1985; Macphail & Bolhuis, 2001). I discussed how contextual variables are still an important challenge in comparing the cognitive abilities of different species, but also described the recent evidence pointing to quantitative differences in intelligence across avian species. Qualitative differences were more difficult to identify, possibly due to the current theoretical and methodological limitations of the field. I therefore outlined three important strategies that animal cognition scientists can use to move forward from the contextual issue that shadows the field and more robustly identify both quantitative and qualitative differences in cognition between species, namely: (i) using baseline tasks to demonstrate task comprehension, (ii) using the signature-testing approach to identify the cognitive mechanisms responsible for task failures and successes, and (iii) utilising a Bayesian framework to make stronger inferences about experimental results.

As well as presenting new insights into the cognition of kea, over the course of this thesis I also showcased how several different approaches can inform the study of animal cognition: **Chapter 2** employed an observational approach to report on a single-individual case study of tooling, **Chapter 3** used both crowdsourcing and phylogenetic analyses to contextualise rare tool use behaviours from an evolutionary perspective, **Chapters 4, 5 and 7** established whether kea were capable of performing in different tasks using binary success-testing, and **Chapter 6** used the signature-testing approach (Taylor, 2014; Taylor et al., *under*

review) to identify the similarities between kea's and humans' abilities to perform domain-general statistical inferences. In most chapters, kea's performances were also directly compared to those of either human infants or other nonhuman species, leveraging a comparative framework to place kea's cognitive abilities in a broader context.

9.2. Back to the Wild

It is estimated that the total kea population comprises between 1,000 and 5,000 individuals (Anderson, 1986), and current modelling suggests that the effective population size comprises fewer than 1,000 individuals (Dusseix & Robertson, 2017). Kea population sizes are difficult to estimate due to the remoteness and inaccessibility of alpine habitats (Dusseix & Robertson, 2017), but declining numbers (Elliot & Kemp, 2004) have encouraged their recent reclassification as an endangered species (BirdLife International, 2017).

The challenges of protecting kea in the wild can benefit greatly from research on kea behaviour and intelligence, and to some extent this work is already having real impacts on kea conservation. For example, pest traps – which target invasive predators but are sometimes deactivated by kea – are painted in dark colours because kea are more attracted to light coloured objects (Kea Conservation Trust, 2019). However, there are still many unexplored possibilities for applying behavioural or cognitive research to wild kea conservation. In this section I will suggest how, beyond providing insights into the evolution of avian intelligence, some of the findings of this thesis and other research can be harnessed to support kea conservation efforts.

Perhaps the greatest threat to kea's survival in the wild is conflict and competition with invasive predators (Innes et al., 2010). Rats, stoats, ferrets, feral cats, and possums introduced to New Zealand by early settlers have caused enormous damage to native bird species (Innes et al., 2010), which evolved in the absence of mammalian land predators (Gibbs, 2006;

Tennyson, 2010) and are therefore poorly adapted to defend themselves against them. For burrow-nesting kea, invasive predators pose the greatest danger to eggs and young chicks (BirdLife, 2017; Kemp et al., 2018). Additionally, attacks by mammalian predators also disproportionately affect adult females, as they undertake most of the egg incubation and protection of fledglings in the nest (Kemp et al., 2018). Therefore, not only do invasive predators contribute to very high mortality of kea younger than one year old (Bond & Diamond, 1992; Jackson, 1969; Kea Conservation Trust, 2017), but they are also the leading cause for a worsening population imbalance of ever-diminishing numbers of females (Kea Conservation Trust, 2017), reducing the effective population size of kea in the wild (Dusseix & Robertson, 2017) and therefore making mate selection and courtship even more difficult for the species.

Unfortunately, important conservation measures implemented to protect kea and other New Zealand native species from invasive predators are sometimes hindered or counteracted by kea's exploratory and neophilic behaviour. For example, pest control traps targeting invasive predators can lead to accidental injuries of kea (Kea Conservation Trust, 2017, 2019). Research investigating what features make objects most or least interesting to kea can help inform the design of pest traps. Similarly, our ongoing research suggests that kea might be especially interested in some materials over others, preferring to interact with wood and soft plastics over hard metals, such as stainless steel and aluminium (*unpublished data*). This can inform the best materials to use in the construction of pest traps in order to deter kea.

Additionally, wild kea sometimes deactivate pest traps by using stick tools (Goodman et al., 2018), reducing the effectiveness of this method of pest control. Given our recent finding that an individual kea innovated a completely novel form of tool use never before seen in parrots (namely, using a pebble for preening; Bastos et al., 2021a, **Chapter 2**), it is possible that kea might innovate other forms of tool use to deactivate or damage pest traps in the future. In addition to their flexible and sophisticated technical intelligence (Huber & Gajdon, 2006;

Bastos & Taylor, 2020; Bastos et al., 2021a), kea's ability to combine multiple sources of information in a domain-general manner (Bastos & Taylor, 2020, **Chapter 6**) also raises issues in regard to trap design. It is clearly possible for kea to combine different sources of information to aid their problem solving, which also needs to be taken into consideration when designing pest traps. Future research investigating the limits of kea's innovation ability and their ability to flexibly combine different sources of information could provide valuable information for the design of traps that kea cannot damage or deactivate in both the short and long-term.

Invasive predators are also targeted by 1080 pellets dispersed through aerial drops (Elliott et al., 2016; Kemp et al., 2018). However, due to their neophilic tendencies, a minority of kea interact with and ingest these pellets, leading to the unintentional by-kill of kea (Veltman & Westbrooke, 2011). Several projects are currently investigating how to best target invasive predators without this unwanted side-effect. For example, in one recent study, researchers investigated whether kea would learn a conditioned aversion to anthraquinone (Nichols et al., 2020), a bird repellent that causes gut irritation when ingested. They presented pellets identical to those used for predator poisoning, but which contained anthraquinone rather than the 1080 poison. This study shows that kea readily learn to avoid the aversive bait, suggesting that in the future, aerial drops of 1080 could be preceded by presentations of non-lethal repellent baits, reducing kea by-kill and more effectively targeting invasive predators. Ongoing research is also investigating the duration of this learned aversion, which could inform the precise timing for anthraquinone aerial drops preceding real 1080 drops (*unpublished data*) for maximal benefit to wild kea populations.

Further research should investigate the extent to which conditioned aversion might be socially learnt in kea. This work would build on the first studies of wild kea behaviour, which concerned the social learning of problem-solving behavioural strategies, such as bin opening (Gajdon et al., 2006), opening artificial fruit puzzle-boxes (Hubert et al., 2001), and pushing

baited tubes off diagonal poles (Gajdon et al., 2004). These studies found little evidence to suggest that kea learn from observing others in wild settings, despite their ability to do so in captivity (Gajdon et al., 2011). Even though wild kea do not necessarily learn problem-solving foraging strategies through observation, it is possible that they could acquire social information about food quality, given that this occurs in other species: for example, great tits can socially learn the appetitive value of prey items by observing conspecifics' reactions to eating them (Thorogood, Kokko & Mappes, 2017).

Additionally, the ability of kea to rapidly solve problems, and associate stimuli with both rewarding and unrewarding outcomes (Laschober et al., 2021) could be used to reduce kea by-kill from 1080 aerial drops. It could be possible to manipulate the foraging locations of wild kea at specific moments in time by presenting specific signals about different foraging locations. For example, dispensers for both food and anthraquinone baits could be placed at two different locations, with each dispenser containing a signal visible to a flying kea but not to invasive mammalian predators. Each signal could indicate either the presence of food or the presence of repellent anthraquinone baits. Wild kea could therefore learn these different signals and use them to inform their daily foraging habits, whilst introduced predators would be unable to learn these signal patterns. Running such an experiment over two broad geographical locations – for example lowland versus highland areas – could enable a system whereby kea populations can be encouraged to move between two areas. Immediately preceding 1080 bait drops, signals indicating the presence of repellent baits could be positioned at the targeted geographical locations, discouraging kea from foraging in these areas. Simultaneously, food signals could be placed in other areas to encourage kea foraging in safe locations. If wild kea can be directed to forage away from areas targeted for 1080 bait drops, whilst no information is provided to ground-dwelling mammalian species, then this strategy could reduce accidental kea by-kill from bait ingestion whilst still ensuring the effectiveness of 1080 poisoning for

targeted invasive predators. Clearly, research is required to test the potential of this management strategy, particularly in terms of the distances required between feeders to ensure that kea move sufficiently between locations.

Another important threat to wild kea consists of dangerous and potentially deadly interactions with humans and human-made objects (Kea Conservation Trust, 2017). Due to their inquisitive and neophilic nature, and their lack of evolved responses to natural predators, wild kea will readily interact with, and show little or no fear of, humans and human-made objects. Recent projects have constructed roadside “kea gyms” to keep kea off busy roads and reduce the risk of traffic accidents and their interaction with humans (Roy, 2018). Again, initiatives such as these could greatly benefit from research investigating the types of materials or objects kea are most interested in.

Wild kea also readily interact with and eat objects and foods offered to them by humans. This can lead to consumption of unhealthy or toxic substances, such as chocolate (Gartrell & Reid, 2007), which is deadly to parrots (Cole & Murray 2005; Dumonceaux & Harrison 1994; Gartrell & Reid, 2007; Labonde 1995). At present, people visiting kea-populated parts of New Zealand feed wild kea, despite being explicitly discouraged from interacting with the animals and from feeding them (Orams, 2002). Our recent work shows that kea judge the probability of different events by comparing the relative frequencies of positive and negative outcomes (Bastos & Taylor, 2020). This raises one interesting potential manipulation for reducing kea’s interest in and approach behaviours towards humans. At present, most interactions with humans might result in kea obtaining food, assuming that humans who are most likely to ignore the requests against interacting with kea are also the most likely to ignore requests against feeding them. Therefore, kea may perceive that, although they do not interact with humans very often, any interactions that do take place are quite likely to yield food and are therefore likely to be rewarding. In contrast, if people were encouraged to interact with kea, but not to

feed them, kea perception of their interactions with humans might shift. Now, out of many interactions with humans, only a minor fraction would result in food. Future research could investigate if shifting human attitudes towards kea, by encouraging interactions but discouraging feeding specifically, might actually reduce the perceived value of human interactions to kea and therefore discourage kea from seeking out these interactions. Furthermore, repeated unrewarding interactions with humans could potentially reduce the novelty value of humans and human-made objects, so reducing kea's neophilic responses to humans in their environment. However, this is a high-risk strategy which should first be tested in captive settings and trialled with a small wild population, to ensure that additional interaction with humans does not unintentionally cause the opposite effect: that kea learn that humans are not threatening, and therefore become even more interested in and likely to approach humans and human-made objects.

Kea that are fed by humans also learn to regularly visit human settlements and display problematic behaviours, such as stealing food and opening rubbish bins (Gajdon et al., 2006). Here again, kea's ability to judge relative frequencies could be exploited to reduce problematic behaviour. At present, if kea open a bin, there is a high probability of that bin containing food. Given that kea can judge probabilities, this behaviour may be assessed by kea as being more likely to yield food than more natural foraging activities. In contrast, if real bins were intermixed with decoy empty bins in kea-populated human settlements, kea would regularly open a bin, only to find it empty. These experiences, if they occurred often enough, could lead kea to judge other foraging contexts as more likely to yield food than bin-opening behaviour, and so reduce this behaviour in the population.

Additionally, given that kea display a poor understanding of string connectivity (Bastos et al., 2021b, **Chapter 4**) and that most birds may require a perceptual-motor feedback loop to learn string-pulling (Taylor et al., 2010, 2012), another possible solution to problematic bin-

opening behaviours might be the use of rope- or chain-based pulley systems for opening outdoors bins in kea-populated areas. Provided that the rope or chain must be pulled in a counter-intuitive direction (for example, down and away from the bin lid) in order to open the bin, this should hinder kea's bin opening behaviour. However, as kea might still learn this associatively without perceptual-motor feedback over an extended period of time, these should still require considerable force to open to ensure that they are also physically challenging for the birds. The combination of difficult-to-open bins and decoy bins offers a promising management strategy to change kea behaviour in human settlements.

Human settlements pose further risks to kea beyond food. A considerable issue is their interactions with lead, which is found in the fixtures of old buildings, car wheel weights, and bullets (Jarrett 1998; McLelland et al. 2010; Reid et al., 2011). Even in small amounts, lead can suppress immunity, impair development and cognitive function, impair motor ability (leading to being run over by cars, for example), reduce appetite, and ultimately, lead to death, either directly or indirectly (De Francisco, Ruiz Troya, & Agüera, 2003). Several organisations are actively working towards substituting the lead in their environments with safer alternatives (Kea Conservation Trust, 2017). Additionally, wildlife hospitals throughout the South Island of New Zealand regularly treat lead poisoning in wild kea with chelation, whereby injected Edetate Calcium Disodium (Calcium EDTA) binds with the toxic lead and enables its excretion through urine (Flora & Pachauri, 2010).

Research on kea cognition and behaviour offers multiple strategies that can be employed to keep kea away from human settlements. As well as provisioning kea with “gyms” in natural areas which might distract them away from human buildings, work revealing that kea perceive real and virtual stimuli as equivalent and continuous, responding to virtual stimuli as they would to real world stimuli (Wein et al., 2015; Bastos et al., 2021c, **Chapter 7**), could also be used for this purpose. Screens featuring either attractants (placed away from human

establishments) or repellents (placed on human establishments) could be used to displace and relocate kea. For example, displaying imagery of conspecifics or interesting novel objects in natural environments could help attract kea away from human settings. It is possible that printed images might also achieve the same effect, given that kea easily transfer concepts learnt about real objects to photographs of these objects (O’Hara et al., 2015; Wein et al., 2015), although fixed and unchanging images will likely result in faster habituation than a larger selection of stimuli presented on video screens, which would likely be more beneficial given kea’s high neophilia. Provided that the screens are robustly cased in acrylic boxes or similarly resistant materials and attached securely in place, this could provide a long-term strategy for managing the locations of kea populations that overlap with human settlements. If implemented, careful consideration is needed to ensure that screens are placed such that they are visible to wild kea, but minimally disruptive to people, given that many common areas of overlap between kea and human populations are in national parks (e.g., Arthur’s Pass, see: Bond & Diamond, 1992) and therefore popular tourist destinations.

Further research to determine if kea detect and respond appropriately to the distress of conspecifics could also be useful for deterring kea from human settlements and roads. For example, screens placed near roads could feature realistic animated stimuli of conspecifics being injured by human-made objects such as cars and trucks and subsequently exhibiting signs of distress. If kea can socially learn from video-based presentations such as these, this strategy has the potential to reduce the risk of road accidents involving kea. Again, more research is needed to explore the viability of this manipulation.

Playback calls could also be used to manage kea distribution in the wild. Schwing and colleagues (2017) demonstrated that a particular vocalisation in this species, the warble call, elicits play and therefore creates positive emotional contagion between individuals. Upon hearing warble calls, kea engage in either solitary or social play, and produce this vocalisation

themselves. The same study also identified alarm calls in kea, such as the trill call. Therefore, kea management strategies could involve playing warble calls to attract kea to suitable natural areas, and trill calls to repel them from human establishments. Again, this strategy is most likely to be successful if used sparingly and cycling through various different calls, to avoid habituation to artificial playbacks. One potential solution would be to vary the relative frequency of different call types, relying on kea's ability to judge relative frequencies. Kea would therefore hear calls from human settlements that suggest a relatively threatening environment (relatively few warble and screech calls, and a high frequency of trill calls) and those from wilderness areas that suggest an appealing environment (relatively few trill and screech calls, and a high frequency of warble calls). This would better mirror natural call rates, by shifting the relative frequency of different call types, rather than simply making different call types present or absent in different areas, which might be more easily detected as artificial stimuli. A manipulation of this sort could provide appropriate social information about different locations to guide kea behaviour, while also reducing the likelihood of habituation to playbacks. Additionally, conspecific call playbacks could also be used to displace kea from areas designated for 1080 drops, helping reduce accidental by-kill of kea from poisonous bait ingestion. Future studies could determine the potential effects of this approach on wild kea social interactions and its effectiveness in shifting kea's movement and foraging patterns.

Finally, kea's neophilia and intelligence also make them difficult to breed for release. Not only are they a challenging species to keep in captivity due to their requirement for considerable and constant cognitive enrichment (Lopez et al., 2016), but also their interactions with humans in captive settings would likely encourage problematic behaviours upon release in the wild, such as foraging in bins (Gajdon et al., 2006), consuming human food (Kea Conservation Trust, 2017), and destroying property (Kea Conservation Trust, 2017). Put together, these two factors offer a significant compounded challenge to captive breeding of kea

for reintroduction: their requirement for enrichment is likely to entail extensive interaction with human caretakers, but extensive interaction might make individuals unsuitable for release. Here, kea's tendency to interpret real and virtual stimuli as continuous and equivalent (Wein et al., 2015; O'Hara et al., 2015; Bastos et al., 2021c) provides a clear potential advantage: if captive kea that are bred for release can have their cognitive enrichment needs met almost exclusively through screen-based presentations, in the absence of humans, then they might remain suitable for reintroduction to wild settings.

9.3. Final Remarks

Over the course of this thesis, I presented five empirical chapters testing different aspects of kea cognition, demonstrating previously untapped abilities in this species. I found that kea can innovate tool use when necessary, that they fail to perceive the difference between connected and unconnected strings even following perceptual-motor feedback experience with vertical string, that they mentally represent the identities and trajectories of objects, that they can use this representational capacity to make statistical inferences, and that they do not distinguish between real-world and virtual contexts. In a chapter using crowdsourced data, I expanded on my discovery of self-care tool use in kea by providing a phylogenetic overview of tool use in parrots, demonstrating that kea are not unique in their ability to innovate self-care tool use, but also investigating the ancestral state of this behaviour at different points in the parrot phylogeny. Finally, in the last two chapters, I contextualised my findings on kea behaviour and cognition within an ongoing debate in the field of animal cognition and outlined how some of the current knowledge of kea behaviour and cognition can be used to kea's advantage in helping conserve this species in the wild. The findings reported in this thesis therefore contribute not

only to our understanding of avian intelligence, but also have real-world applications for conserving this iconic New Zealand species in the wild.

References

- Albiach-Serrano, A., Bräuer, J., Cacchione, T., Zickert, N., & Amici, F. (2012). The effect of domestication and ontogeny in swine cognition (*Sus scrofa scrofa* and *S. s. domestica*). *Applied Animal Behaviour Science*, *141*(1-2), 25-35. doi:10.1016/j.applanim.2012.07.005
- Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Søvik, E., & Chittka, L. (2016). Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLoS Biology*, *14*(10), e1002564. doi:10.1371/journal.pbio.1002564
- Allen, C. (2017). On (not) defining cognition. *Synthese*, *194*(11), 4233-4249. doi:10.1007/s11229-017-1454-4.
- Amici F., Barney B., Johnson V. E., Call J., & Aureli, F. (2012) A Modular Mind? A Test Using Individual Data from Seven Primate Species. *PLoS One*, *7*(12), e51918. doi:10.1371/journal.pone.0051918
- Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? *Neuroscience Letters*, *153*(1), 98-102.
- Anderson, P. (2003). A bird in the house: An anthropological perspective on companion parrots. *Society & Animals*, *11*(4), 393-418. doi:10.1163/156853003322796109
- Anderson, R. (1986). Keas for keeps. *Forest and Bird*, *17*, 2-5.
- Ashton, B. J., Kennedy, P., & Radford, A. N. (2020). Interactions with conspecific outsiders as drivers of cognitive evolution. *Nature Communications*, *11*(1), 4937. doi:10.1038/s41467-020-18780-3
- Auersperg, A. M. I., Borasinski, S., Laumer, I., & Kacelnik, A. (2016). Goffin's cockatoos make the same tool type from different materials. *Biology Letters*, *12*(11), 20160689. doi:10.1098/rsbl.2016.0689

- Auersperg, A. M. I., Gajdon, G. K., & Huber, L. (2009). Kea (*Nestor notabilis*) consider spatial relationships between objects in the support problem. *Biology Letters*, 5(4), 455-458. doi:10.1098/rsbl.2009.0114
- Auersperg, A. M. I., Gajdon, G. K., & Huber, L. (2010). Kea, *Nestor notabilis*, produce dynamic relationships between objects in a second-order tool use task. *Animal Behaviour*, 80(5), 783-789. doi:10.1016/j.anbehav.2010.08.007
- Auersperg, A. M. I., Gajdon, G. K., & von Bayern, A. M. P. (2012). A new approach to comparing problem solving, flexibility and innovation. *Communicative & Integrative Biology*, 5(2), 140-145. doi:10.4161/cib.18787
- Auersperg, A. M. I., Huber, L., & Gajdon, G. K. (2011a). Navigating a tool end in a specific direction: Stick-tool use in kea (*Nestor notabilis*). *Biology Letters*, 7(6), 825-828. doi:10.1098/rsbl.2011.0388
- Auersperg, A. M. I., Schwing, R., Mioduszewska, B., O'Hara, M., & Huber, L. (2020). Do puffins use tools? *Proceedings of the National Academy of Sciences*, 117(22), 11859. doi:10.1073/pnas.2001988117
- Auersperg, A. M. I., Szabo, B., von Bayern, A. M. P., & Bugnyar, T. (2014). Object permanence in the Goffin cockatoo (*Cacatua goffini*). *Journal of Comparative Psychology*, 128(1), 88-98. doi:10.1037/a0033272
- Auersperg, A. M. I., Szabo, B., von Bayern, A. M. P., & Kacelnik, A. (2012). Spontaneous innovation in tool manufacture and use in a Goffin's cockatoo. *Current Biology*, 22(21), R903-R904. doi:10.1016/j.cub.2012.09.002
- Auersperg, A. M. I., van Horik, J. O., Bugnyar, T., Kacelnik, A., Emery, N. J., & von Bayern, A. M. P. (2014). Combinatory actions during object play in Psittaciformes (*Diopsittaca nobilis*, *Pionites melanocephala*, *Cacatua goffini*) and corvids (*Corvus corax*, *C. monedula*, *C. moneduloides*). *Journal of Comparative Psychology*, 129(1), 62-71. doi:10.1037/a0038314

- Auersperg, A. M. I., & von Bayern, A. M. P. (2019). Who's a clever bird - now? A brief history of parrot cognition. *Behaviour*, *156*(5-8), 391-407. doi:10.1163/1568539X-00003550
- Auersperg, A. M. I., von Bayern, A. M. P., Gajdon, G. K., Huber, L., & Kacelnik, A. (2011b). Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLoS One*, *6*(6), e20231. doi:10.1371/journal.pone.0020231
- Auersperg, A. M. I., von Bayern, A. M. P., Weber, S., Szabadvari, A., Bugnyar, T., & Kacelnik, A. (2014). Social transmission of tool use and tool manufacture in goffin cockatoos (*Cacatua goffini*). *Proceedings of the Royal Society B: Biological Sciences*, *281*(1793), 20140972. doi:10.1098/rspb.2014.0972
- Aydinonat, D., Penn, D. J., Smith, S., Moodley, Y., Hoelzl, F., Knauer, F., & Schwarzenberger, F. (2014). Social isolation shortens telomeres in African grey parrots (*Psittacus erithacus erithacus*). *PLoS One*, *9*(4), e93839. doi:10.1371/journal.pone.0093839
- Bagotskaya, M., Smirnova, A., & Zorina, Z. (2012). Corvidae can understand logical structure in baited string-pulling tasks. *Neuroscience and Behavioral Physiology*, *42*(1), 36-42. doi:10.1007/s11055-011-9529-z
- Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, *23*(1), 21-41. doi:10.1016/0010-0277(86)90052-1
- Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. *Visual perception and cognition in infancy* (pp. 273-324). Psychology Press. doi:10.4324/9780203763032-12
- Baillargeon, R. (2004). Infants' reasoning about hidden objects: evidence for event-general and event-specific expectations. *Developmental Science*, *7*(4), 391-414. doi:10.1111/j.1467-7687.2004.00357.x

- 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*). *Journal of Experimental Psychology: Animal Behavior Processes*, 32(3), 239-252. doi:10.1037/0097-7403.32.3.239
- Barton, R., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology*, 24(20), 2440-2444. doi:10.1016/j.cub.2014.08.056
- Bastos, A. P. M., Horváth, K., Webb, J., Wood, P. M., & Taylor, A. H. (2021a) Self-care tooling innovation in a disabled kea (*Nestor notabilis*). *Scientific Reports*, 11, 18035. doi:10.1038/s41598-021-97086-w.
- Bastos, A. P. M., & Taylor, A. H. (2019). Kea (*Nestor notabilis*) represent object trajectory and identity. *Scientific Reports*, 9(1), 19759. doi:10.1038/s41598-019-56380-4
- Bastos, A. P. M., & Taylor, A. H. (2020). Kea show three signatures of domain-general statistical inference. *Nature Communications*, 11, 828. doi:10.1038/s41467-020-14695-1
- Bastos, A. P. M. & Taylor, A. H. (2020) Macphail's null hypothesis of vertebrate intelligence: Insights from avian cognition. *Frontiers in Psychology*, 11, 1692. doi:10.3389/fpsyg.2020.01692
- Bastos, A. P. M., Wood, P. M., Taylor, A. H. (2021b) Kea (*Nestor notabilis*) fail a loose-string connectivity task. *Scientific Reports*, 11, 15492. doi:10.1038/s41598-021-94879-x
- Bastos, A. P. M., Wood, P. M., & Taylor, A. H. (2021c) Are parrots naïve realists? Kea behave as if the real and virtual worlds are continuous. *Biology Letters*.
- Bauchot, R., Bauchot, M. L., Platel, R., & Ridet, J. M. (1977). Brains of Hawaiian tropical fishes; brain size and evolution. *Copeia*, 1977(1), 42-46. doi:10.2307/1443502
- Bell, D. E., Raiffa, H. & Tversky, A. (1988). *Decision Making*. New York, NY: Cambridge University Press.

- Bentley-Condit, V., & Smith, E. O. (2010). Animal tool use: Current definitions and an updated comprehensive catalog. *Behaviour*, *147*(2), 185-32A. doi:10.1163/000579509X12512865686555
- Benton, M. J. & Donoghue, P. C. J. (2007). Paleontological Evidence to Date the Tree of Life. *Molecular Biology and Evolution*, *24*(1), 26-53. doi:10.1093/molbev/msl150
- Beran, M. J. (2015). The comparative science of “self-control”: What are we talking about? *Frontiers in Psychology*, *6*, 51. doi:10.3389/fpsyg.2015.00051
- Beran, M. J., Beran, M. M., & Menzel, C. R. (2005). Chimpanzees (*Pan troglodytes*) use markers to monitor the movement of a hidden item. *Primates*, *46*(4), 255-259. doi:10.1007/s10329-005-0150-z
- Beran, M. J., & Minahan, M. F. (2000). Monitoring spatial transpositions by bonobos (*Pan paniscus*) and chimpanzees (*P. troglodytes*). *Comparative Psychology*, *13*(1), 1-15.
- Bercovitz, A. B. (1987). *Avian sex identification techniques*. In E. W. Burr (Ed.), *Companion bird medicine* (pp. 197-203). Ames: Iowa State University Press.
- Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive non tool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370-10375. doi:10.1073/pnas.0901008106
- Bird, C. D., & Emery, N. J. (2009). Rooks use stones to raise the water level to reach a floating worm. *Current Biology*, *19*(16), 1410-1414. doi:10.1016/j.cub.2009.07.033
- BirdLife International. (2017). *Nestor notabilis*. the IUCN red list of threatened species 2017: E.T22684831A119243358. Retrieved from <https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T22684831A119243358.en>
- Boeckle, M., Schiestl, M., Frohnwieser, A., Gruber, R., Miller, R., Suddendorf, T., ... Clayton, N. S. (2020). New Caledonian crows plan for specific future tool use. *Proceedings of the Royal Society B*, *287*(1938), 20201490. doi:10.1098/rspb.2020.1490

- Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752-762. doi:10.1038/nrn3122
- Bond, A. B., & Diamond, J. (1992). Population estimates of Kea in Arthur's Pass National Park. *Notornis*, *39*(3), 151-160.
- Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour*, *65*(3), 479-487. doi:10.1006/anbe.2003.2101
- Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus*, *Nucifraga columbiana*, *Aphelocoma californica*). *Journal of Comparative Psychology*, *121*(4), 372-379. doi:10.1037/0735-7036.121.4.372
- Boucherie, P. H., Loretto, M. C., Massen, J. J. M., & Bugnyar, T. (2019). What constitutes “social complexity” and “social intelligence” in birds? Lessons from ravens. *Behavioral Ecology and Sociobiology*, *73*(1), 1-14. doi:10.1007/s00265-018-2607-2
- Borsari, A., & Ottoni, E. (2005). Preliminary observations of tool use in captive hyacinth macaws (*Anodorhynchus hyacinthinus*). *Animal Cognition*, *8*(1), 48-52. doi:10.1007/s10071-004-0221-3
- Bräuer, J. & Call, J. (2011) The magic cup: Great apes and domestic dogs (*Canis familiaris*) individuate objects according to their properties. *Journal of Comparative Psychology*, *125*(3), 353-361. doi:10.1037/a0023009
- Brosnan, S. F. (2011). A hypothesis of the co-evolution of cooperation and responses to inequity. *Frontiers in Neuroscience*, *5*, 43. doi:10.3389/fnins.2011.00043
- Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), 1251776. doi:10.1126/science.1251776

- Brucks, D., Marshall-Pescini, S., Wallis, L. J., Huber, L., & Range, F. (2017). Measures of dogs' inhibitory control abilities do not correlate across tasks. *Frontiers in Psychology*, 8, 849. doi:10.3389/fpsyg.2017.00849
- Brucks, D., & von Bayern, A. M. P. (2020). Parrots voluntarily help each other to obtain food rewards. *Current Biology*, 30(2), 292-297.e5. doi:10.1016/j.cub.2019.11.030
- Brunton-Martin, A. L., Nichols, M., & Gaskett, A. C. (2021). Assessing kea perception of cereal baits using modelling of spectral reflectance. *New Zealand Journal of Ecology*, 45(1), 3421.
- Buckner, C. (2015). A property cluster theory of cognition. *Philosophical Psychology*, 28(3), 307-336. doi:10.1080/09515089.2013.843274
- Bulley, A., Redshaw, J., & Suddendorf, T. (2020). *The future-directed functions of the imagination: From prediction to metaforesight*. In A. Abraham (Eds.) *The Cambridge Handbook of the Imagination* (pp. 425-442). Cambridge, UK: Cambridge University Press.
- Burghardt, G. M. (2006). *The genesis of animal play*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/3229.001.0001
- Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, 7(1), 10506. doi:10.1038/ncomms10506
- Burkart, J., Schubiger, M., & Van Schaik, C. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, 40, E195. doi:10.1017/S0140525X16000959
- Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1-28. doi:10.18637/jss.v080.i01
- Byrne, R. W. (1997). The technical intelligence hypothesis: An additional evolutionary stimulus to intelligence? In R. W. Byrne, & A. Whiten (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 289-311). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511525636.012

- Byrne, R. W., & Whiten, A. (1997). Machiavellian intelligence. In R. W. Byrne, & A. Whiten (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 1-23). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511525636.002
- Cabrera-Alvarez, M. J., & Clayton, N. S. (2020). Neural processes underlying tool use in humans, macaques, and corvids. *Frontiers in Psychology, 11*, 560669. doi:10.3389/fpsyg.2020.560669
- Caldwell, C. A., & Whiten, A. (2002). Evolutionary perspectives on imitation: is a comparative psychology of social learning possible? *Animal Cognition, 5*, 193-208. doi:10.1007/s10071-002-0151-x
- Call, J. (2001). Object permanence in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and children (*Homo sapiens*). *Journal of Comparative Psychology, 115*(2), 159-171. doi:10.1037/0735-7036.115.2.159
- Call, J. (2003). Spatial rotations and transpositions in orangutans (*Pongo pygmaeus*) and chimpanzees (*Pan troglodytes*). *Primates, 44*, 347-357. doi:10.1007/s10329-003-0048-6
- Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118*(2), 232-241. doi:10.1037/0735-7036.118.2.232
- Call, J. (2006). Inferences by exclusion in the great apes: the effect of age and species. *Animal Cognition, 9*(4), 393-403. doi:10.1007/s10071-006-0037-4
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan: A Probabilistic Programming Language*. *Journal of Statistical Software, 76*(1), 1-32. <http://dx.doi.org/10.18637/jss.v076.i01>
- Carrete, M., & Tella, J. (2008). Wild-bird trade and exotic invasions: A new link of conservation concern? *Frontiers in Ecology and the Environment, 6*(4), 207-211. doi:10.1890/070075
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

- Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, 25(6), 657-674. <http://doi.org/10.1017/S0140525X02000122>
- Chappell, J. (2017). Bird brains: Does absolute size matter? *Learning & Behavior*, 45(1), 1-2. doi:10.3758/s13420-016-0247-9
- Chappell, J., Kacelnik, A. (2002). Tool selectivity in a non-primate, the New Caledonian crow (*Corvus moneduloides*). *Animal Cognition*, 5, 71-78. doi:10.1007/s10071-002-0130-2
- Cheke, L. G., Bird, C. D., & Clayton, N. S. (2011). Tool-use and instrumental learning in the Eurasian jay (*Garrulus glandarius*). *Animal Cognition*, 14, 441-455. doi:10.1007/s10071-011-0379-4
- Cheke, L. G., & Clayton, N. S. (2011). Eurasian jays (*Garrulus glandarius*) overcome their current desires to anticipate two distinct future needs and plan for them appropriately. *Biology Letters*, 8(2), 171-175. doi:10.1098/rsbl.2011.0909
- Cheney, D. L., & Seyfarth, R. M. (1999). Recognition of other individuals' social relationships by female baboons. *Animal Behaviour*, 58(1), 67-75. doi:10.1006/anbe.1999.1131
- Cheney, D. L., & Seyfarth, R. M. (2005). Social complexity and the information acquired during eavesdropping by primates and other animals. In P. K. McGregor (Ed.), *Animal communication networks* (pp. 583-603). Cambridge, England: Cambridge University Press.
- Chiandetti, C., & Vallortigara, G. (2011). Intuitive physical reasoning about occluded objects by inexperienced chicks. *Proceedings of the Royal Society B: Biological Sciences*, 278(1718), 2621-2627. doi:10.1098/rspb.2010.2381
- Cicchetti, D. (1994) Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology. *Psychological Assessment*, 6(4), 284. doi:10.1037/1040-3590.6.4.284
- Clarke, C. (1970). Observations on population movements and food of the kea (*Nestor notabilis*). *Notornis*, 17(2), 105-114.

- Clayton, D. H., Moyer, B. R., Bush, S. E., Jones, T. G., Gardiner, D. W., Rhodes. Barry, B., & Goller, F. (2005). Adaptive significance of avian beak morphology for ectoparasite control. *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 811-817. doi:10.1098/rspb.2004.3036
- Clayton, N. S. (2012). Corvid cognition: Feathered apes. *Nature*, 484, 453-454. doi:10.1038/484453a
- Clayton, N. S. & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395, 272-274. doi:10.1038/26216
- Clements, K. A., Gray, S. L., Gross, B., & Pepperberg, I. M. (2018). Initial evidence for probabilistic reasoning in a grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 132(2), 166-177. doi:10.1037/com0000106
- Cole, G., & Murray, M. (2005). Suspected chocolate toxicosis in an African Grey parrot (*Psittacus erithacus*). *Proceedings of the Annual Conference of the Association of Avian Veterinarians*, Monterey CA, USA, pp 339-40.
- Collier-Baker, E., Davis, J. M., Nielsen, M., & Suddendorf, T. (2006). Do chimpanzees (*Pan troglodytes*) understand single invisible displacement? *Animal Cognition*, 9, 55-61. doi:10.1007/s10071-005-0004-5
- Collier-Baker, E., & Suddendorf, T. (2006). Do chimpanzees (*Pan troglodytes*) and 2-year-old children (*Homo sapiens*) understand double invisible displacement? *Journal of Comparative Psychology*, 120(2), 89-97. doi:10.1037/0735-7036.120.2.89
- Colombo, M., & Scarf, D. (2020). Are there differences in “Intelligence” between nonhuman species? the role of contextual variables. *Frontiers in Psychology*, 11, 2072. doi:10.3389/fpsyg.2020.02072
- Correia, S. P. C., Dickinson, A., & Clayton, N. (2007). Western scrub-jays anticipate future needs independently of their current motivational state. *Current Biology*, 17(10), 856-861. doi:10.1016/j.cub.2007.03.063

- Cosmides, L. & Tooby, J. (1987). *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, MA: MIT Press.
- Custance, D., Whiten, A., & Fredman, T. (1999). Social learning of an artificial fruit task in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 113(1), 13-23. doi:10.1037/0735-7036.113.1.13
- Dally, J. M., Emery, N. J., & Clayton, N. S. (2005). Cache protection strategies by western scrub-jays, *Aphelocoma californica*: Implications for social cognition. *Animal Behaviour*, 70(6), 1251-1263. doi:10.1016/j.anbehav.2005.02.009
- Dally, J. M., Emery, N. J., & Clayton, N. S. (2006). Food-caching western scrub-jays keep track of who was watching when. *Science*, 312(5780), 1662-1665. doi:10.1126/science.1126539
- Danel, S., von Bayern, A. M. P., & Osiurak, F. (2019). Ground-hornbills (*Bucorvus*) show means-end understanding in a horizontal two-string discrimination task. *Journal of Ethology*, 37(1), 117-122. doi:10.1007/s10164-018-0565-9
- Davis, W. E. (2008). Atlas of bird distribution in New Zealand 1999-2004. *The Auk*, 125(3), 753-754. doi:10.1525/auk.2008.4708.3
- Dawkins, M. S. (1993). *Through our eyes only?* Oxford: Freeman.
- Deary, I. J., Spinath, F. M. & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics*, 14, 690-700. doi:10.1038/sj.ejhg.5201588
- D'eath, R. B., & Dawkins, M. S. (1996). Laying hens do not discriminate between video images of conspecifics. *Animal Behaviour*, 52(5), 903-912. doi:10.1006/anbe.1996.0238
- de Blois, S. T., Novak, M. A., & Bond, M. (1998). Object permanence in orangutans (*Pongo pygmaeus*) and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, 112(2), 137-152. doi:10.1037/0735-7036.112.2.137

- Dechaume-Moncharmont, F. (2020). Touchy matter: The delicate balance between Morgan's canon and open-minded description of advanced cognitive skills in the animal. *Peer Community in Ecology*, 100042. doi:10.24072/pci.ecology.100042
- De Francisco, N., Ruiz Troya, J. D., & Agüera, E. I. (2003). Lead and lead toxicity in domestic and free living birds. *Avian Pathology*, 32(1), 3-13. doi:10.1080/0307945021000070660
- De Petrillo, F., & Rosati, A. G. (2019). Rhesus macaques use probabilities to predict future events. *Evolution and Human Behavior*, 40(5), 436-446. doi:10.1016/j.evolhumbehav.2019.05.006
- Denison, S., Reed, C., & Xu, F. (2012). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, 49(2), 243-249. doi:10.1037/a0028278
- Denison, S., & Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science*, 34(5), 885-908. doi:10.1111/j.1551-6709.2010.01111.x
- Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, 130(3), 335-347. doi:10.1016/j.cognition.2013.12.001
- Deppe, A. M., Wright, P. C., & Szelistowski, W. A. (2009). Object permanence in lemurs. *Animal Cognition*, 12(2), 381-388. doi:10.1007/s10071-008-0197-5
- DeWeerd, S. (2019). How to map the brain. *Nature*, 571(7766), S6-S8. doi:10.1038/d41586-019-02208-0
- Diamond, J., & Bond, A. (1999). *Kea, bird of paradox: The evolution and behavior of a New Zealand parrot*. Berkely, CA: University of California Press.
- Diamond, J., & Bond, A. (2003). A comparative analysis of social play in birds. *Behaviour*, 140(8-9), 1091-1115. doi:10.1163/156853903322589650
- Diamond, J., & Bond, A. (2004). Social play in kaka (*Nestor meridionalis*) with comparisons to kea (*Nestor notabilis*). *Behaviour*, 141(7), 777-798. doi:10.1163/1568539042265680

- Diamond, J., Eason, D., Reid, C., & Bond, A. B. (2006). Social play in kakapo (*Strigops habroptilus*) with comparisons to kea (*Nestor notabilis*) and kaka (*Nestor meridionalis*). *Behaviour*, *143*(11), 1397-1423. doi:10.1163/156853906778987551
- Dickerson, K. L., Ainge, J. A., & Seed, A. M. (2018). The role of association in pre-schoolers' solutions to "Spoon tests" of future planning. *Current Biology*, *28*(14), 2309-2313.e2. doi:10.1016/j.cub.2018.05.052
- Dolins, F. L., Klimowicz, C., Kelley, J., & Menzel, C. R. (2014). Using virtual reality to investigate comparative spatial cognitive abilities in chimpanzees and humans. *American Journal of Primatology*, *76*(5), 496-513. doi:10.1002/ajp.22252
- Doré, F. Y., Fiset, S., Goulet, S., Dumas, M. C., & Gagnon, S. (1996). Search behavior in cats and dogs: Interspecific differences in working memory and spatial cognition. *Animal Learning & Behavior*, *24*(2), 142-149. doi:10.3758/BF03198962
- Drews, C. (2001). Wild animals and other pets kept in Costa Rican households: Incidence, species and numbers. *Society & Animals*, *9*(2), 107-126. doi:10.1163/156853001753639233
- Duchaine, B., Cosmides, L. & Tooby, J. (2001). Evolutionary psychology and the brain. *Current Opinion in Neurobiology*, *11*(2), 225-230. doi:10.1016/s0959-4388(00)00201-4
- Dugas-Ford, J., Rowell, J. J., & Ragsdale, C. W. (2012). Cell-type homologies and the origins of the neocortex. *Proceedings of the National Academy of Sciences*, *109*(42), 16974-16979.
- Dumas, C., & Wilkie, D. M. (1995). Object permanence in ring doves (*Streptopelia risoria*). *Journal of Comparative Psychology*, *109*(2), 142-150. doi:10.1037/0735-7036.109.2.142
- Dumonceaux, G., & Harrison, G. J. (1994). Toxins. In: Ritchie, B. W., Harrison, G. J., Harrison, L. R. (Eds). *Avian Medicine: Principles and Applications* (pp. 1034-44). Lake Worth, FL: Wingers Publishing.
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178-190. doi:10.1002/(SICI)1520-6505(1998)6:53.0.CO;2-8

- Dunbar, R. I. M. (2008). Cognitive constraints on the structure and dynamics of social networks. *Group Dynamics: Theory, Research, and Practice*, 12(1), 7-16. doi:10.1037/1089-2699.12.1.7
- Duncan, R. P., Lefebvre, L., Cassey, P., Sol, D., Orians, G. H., & Blackburn, T. M. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences*, 102(15), 5460-5465. doi:10.1073/pnas.0408145102
- Eckert, J., Rakoczy, H., Call, J., Herrmann, E., & Hanus, D. (2018a). Chimpanzees consider humans' psychological states when drawing statistical inferences. *Current Biology*, 28(12), 1959-1963.e3. doi:10.1016/j.cub.2018.04.077
- Eckert, J., Call, J., Hermes, J., Herrmann, E., & Rakoczy, H. (2018b). Intuitive statistical inferences in chimpanzees and humans follow Weber's law. *Cognition*, 180, 99-107. doi:10.1016/j.cognition.2018.07.004
- Edinger, L. (1899). *The anatomy of the central nervous system of man and of vertebrates in general*. Philadelphia, PA: The F. A. Davis Company.
- Egelkamp, C. L., & Ross, S. R. (2019). A review of zoo-based cognitive research using touchscreen interfaces. *Zoo Biology*, 38(2), 220-235. doi:10.1002/zoo.21458
- Elepfandt, A. (1987). Comparative cognition: Inadequate approach, precipitate conclusions. *Behavioral and Brain Sciences*, 10(4), 661-662. doi:10.1017/S0140525X00055059
- Elliott, G. P., & Kemp, J. R. (1999). *Conservation ecology of kea (Nestor notabilis)*. WWF/NZ final report. New Zealand: World Wildlife Fund for Nature.
- Elliott, G., & Kemp, J. (2004) Effect of hunting and predation on kea, and a method of monitoring kea populations: Results of kea research on the St. Arnaud range. *DOC Science Internal Series*, 181, 1-7.
- Elliott, G., & Kemp, J. (2016). Large-scale pest control in New Zealand beech forests. *Ecological Management & Restoration*, 17(3), 200-209. doi:10.1111/emr.12227

- Emery, N. J. (2004). Are corvids ‘feathered apes’? cognitive evolution in crows, jays, rooks, and jackdaws. In S. Watanabe (Ed.), *Comparative Analysis of Minds* (pp. 181-213). Tokyo, Japan: Keio University Press.
- Emery, N. J. (2006). Cognitive ornithology: The evolution of avian intelligence. *Philosophical Transactions of the Royal Society B*, *361*(1465), 23-43. doi:10.1098/rstb.2005.1736
- Emery, N. J., & Clayton, N. S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature*, *414*, 443-446. doi:10.1038/35106560
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, *306*(5703), 1903-1907. doi:10.1126/science.1098410
- Emery, N. J. & Clayton, N. S. (2005). Evolution of the avian brain and intelligence. *Current Biology*, *15*(23), R946-R950. doi:10.1016/j.cub.2005.11.029
- Emery, & Clayton, N. S. (2009). Tool use and physical cognition in birds and mammals. *Current Opinion in Neurobiology*, *19*(1), 27-33. doi:10.1016/j.conb.2009.02.003
- Engebretson, M. (2006). The welfare and suitability of parrots as companion animals: A review. *Animal Welfare*, *15*(3), 263-276.
- Erdőhegyi, Á, Topál, J., Virányi, Z., & Miklósi, Á. (2007). Dog-logic: Inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour*, *74*(4), 725-737. doi:10.1016/j.anbehav.2007.03.004
- Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454-459. doi:10.1016/j.tics.2003.08.012
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891-904. doi:10.1007/s11192-011-0494-7
- Farrar, B. (2020). *Evidence of tool use in a seabird?* Unpublished manuscript. doi:10.31234/OSF.IO/463HK

- Farrar, B. G., Boeckle, M., & Clayton, N. (2020). Replications in comparative cognition: What should we expect and how can we improve? *Animal Behavior and Cognition*, 7(1), 1-22. doi:10.26451/abc.07.01.02.2020
- Farrar, B. G., & Ostojić, L. (2019). The illusion of science in comparative cognition. Unpublished manuscript. doi:10.31234/osf.io/hduyx
- Fayet, A. L., Hansen, E. S., & Biro, D. (2020). Evidence of tool use in a seabird. *Proceedings of the National Academy of Sciences of the United States of America*, 117(3), 1277-1279. doi:10.1073/pnas.1918060117
- Finn, J. K., Tregenza, T., & Norman, M. D. (2009). Defensive tool use in a coconut-carrying octopus. *Current Biology*, 19(23), R1069-R1070. doi:10.1016/j.cub.2009.10.052
- Fiset, S., & Plourde, V. (2013). Object permanence in domestic dogs (*Canis lupus familiaris*) and gray wolves (*Canis lupus*). *Journal of Comparative Psychology*, 127(2), 115-127. doi:10.1037/a0030595
- Flammer, K., Whitt-Smith, D., & Papich, M. (2001). Plasma concentrations of doxycycline in selected psittacine birds when administered in water for potential treatment of *Chlamydophila psittaci* infection. *Journal of Avian Medicine and Surgery*, 15(4), 276-282. doi:10.1647/1082-6742(2001)015[0276:PCODIS]2.0.CO;2
- Fletcher, L., & Carruthers, P. (2013). Behavior-reading versus mentalizing in animals. In J. Metcalfe & H. Terrace (Eds.), *Agency and Joint Attention* (pp. 82-99). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199988341.003.0005
- Flora, S. J. S., & Pachauri, V. (2010). Chelation in metal intoxication. *International Journal of Environmental Research and Public Health*, 7(7), 2745-2788. doi:10.3390/ijerph7072745
- Fontanari, L., Rugani, R., Regolin, L. & Vallortigara, G. (2011). Object individuation in 3-day-old chicks: use of property and spatiotemporal information. *Developmental Science* 14(5), 1235-1244. doi:10.1111/j.1467-7687.2011.01074.x

- Fragaszy, D. M., & Mangalam, M. (2018). Tooling. *Advances in the Study of Behavior*, 50, 177-241.
doi:10.1016/bs.asb.2018.01.001
- Fragaszy, D. M., & Mangalam, M. (2020). Folk physics in the twenty-first century: Understanding tooling as embodied. *Animal Behavior and Cognition*, 7(3), 457-473.
doi:10.26451/abc.07.03.12.2020
- Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325-1330. doi:10.1111/2041-210X.12584
- Freire, R. & Nicol, C. J. (1999). Effect of experience of occlusion events on the domestic chick's strategy for locating a concealed imprinting object. *Animal Behaviour* 58(3), 593-599.
doi:10.1006/anbe.1999.1162
- Funk, M. S. (1996). Development of object permanence in the New Zealand parakeet (*Cyanoramphus auriceps*). *Learning & Behavior* 24, 375-383. doi:10.3758/BF03199009
- Gajdon, G. K., Amann, L., & Huber, L. (2011). Keas rely on social information in a tool use task but abandon it in favour of overt exploration. *Interaction Studies*, 12(2), 304-323.
doi:10.1075/is.12.2.06gaj
- Gajdon, G. K., Fijn, N., & Huber, L. (2004). Testing social learning in a wild mountain parrot, the kea (*Nestor notabilis*). *Animal Learning & Behavior*, 32(1), 62-71. doi:10.3758/BF03196007
- Gajdon, G. K., Fijn, N., & Huber, L. (2006). Limited spread of innovation in a wild parrot, the kea (*Nestor notabilis*). *Animal Cognition*, 9(3), 173-181. doi:10.1007/s10071-006-0018-7
- Gajdon, G. K., Lichtnegger, M., & Huber, L. (2014). What a parrot's mind adds to play: The urge to produce novelty fosters tool use acquisition in kea. *Open Journal of Animal Sciences*, 4(2), 51-58. doi:10.4236/ojas.2014.42008

- Gajdon, G. K., Ortner, T. M., Wolf, C. C., & Huber, L. (2013). How to solve a mechanical problem: The relevance of visible and unobservable functionality for kea. *Animal Cognition*, *16*(3), 483-492. doi:10.1007/s10071-012-0588-5
- Gamer, M., Lemon, J., & Singh, P. (2012). *Irr: Various Coefficients of Interrater Reliability and Agreement*. Available online: cran.r-project.org/web/packages/irr
- Gartrell, B. D., & Reid, C. (2007). Death by chocolate: A fatal problem for an inquisitive wild parrot. *New Zealand Veterinary Journal*, *55*(3), 149-151. doi:10.1080/00480169.2007.36759
- Gartrell, B. D., Reid, C., McInnes, K., & McLelland, J. M. (2012). Anthropogenic lead (Pb) exposure in populations of a wild parrot (kea *Nestor notabilis*). *New Zealand Journal of Ecology*, *36*(1), 56-63.
- Gaycken, J., Picken, D. J., Pike, T. W., Burman, O. H. P., & Wilkinson, A. (2019). Mechanisms underlying string-pulling behaviour in green-winged macaws. *Behaviour*, *2019*(5-8), 619-631. doi:10.1163/1568539X-00003520
- Gentner, D., Holyoak, K. J. & Kokinov, B. N. (2001). *The Analogical Mind*. Cambridge, MA: MIT Press.
- Ghirlanda, S., & Lind, J. (2017). “Aesop’s fable” experiments demonstrate trial-and-error learning in birds, but no causal understanding. *Animal Behaviour*, *123*, 239-247. doi:10.1016/j.anbehav.2016.10.029
- Gibbs, G. (2006). *Ghosts of Gondwana*. Nelson, New Zealand: Craig Potton Publishing.
- Gill, F., Donsker, D., & Rasmussen, P. (2020). *IOC world bird list 10.1*. World Bird Names International Ornithologists Union. doi:10.14344/ioc.ml.10.1
- Giraldeau, L., & Caraco, T. (2000). *Social foraging theory*. Princeton, NJ: Princeton University Press.
- Global GfK survey: Pet ownership*. (2016). <http://www.gfk.com/global-studies/global-studies-pet-ownership>

- Goodman, M., Hayward, T., & Hunt, G. R. (2018). Habitual tool use innovated by free-living New Zealand kea. *Scientific Reports*, 8(1), 13935-12. doi:10.1038/s41598-018-32363-9
- Greer, A. L., Gajdon, G. K. & Nelson, X. J. (2015). Intraspecific variation in the foraging ecology of kea, the world's only mountain- and rainforest-dwelling parrot. *New Zealand Journal of Ecology*, 39(2), 254-261.
- Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R. D., Clayton, N. S., & Taylor, A. H. (2019). New Caledonian Crows use mental representations to solve metatool problems. *Current Biology*, 29(4), 686-692.e3. doi:10.1016/j.cub.2019.01.008
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, 20(4), 291-303. doi:10.1016/j.tics.2016.02.001
- Gweon, H., Tenenbaum, J. B., Schulz, L. E. & Carey, S. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9066-9071. doi:10.1073/pnas.1003095107
- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutorials in Quantitative. *Methods for Psychology*, 8(1), 23-34. doi:10.20982/tqmp.08.1.p023
- Hampton, R. (2019). Parallel overinterpretation of behavior of apes and corvids. *Learning & Behavior*, 47, 105-106. doi:10.3758/s13420-018-0330-5
- Hartmann, D., Davila-Ross, M., Wong, S. T., Call, J., & Scheumann, M. (2017). Spatial transposition tasks in Indian sloth bears (*Melursus ursinus*) and Bornean sun bears (*Helarctos malayanus euryspilus*). *Journal of Comparative Psychology*, 131(4), 290-303. doi:10.1037/com0000077
- Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. (2017). Neuroscience-Inspired artificial intelligence. *Neuron*, 95(2), 245-258. doi:10.1016/j.neuron.2017.06.011
- Heaney, M., Bastos, A. P. M., Gray, R. D., & Taylor, A. H. (2019). Are kea prosocial? *Ethology*, 126(2), 176-184. doi:10.1111/eth.12944

- Heaney, M., Gray, R. D., & Taylor, A. H. (2017a). Kea show no evidence of inequity aversion. *Royal Society Open Science*, 4, 160461. doi:10.1098/rsos.160461
- Heaney, M., Gray, R. D., & Taylor, A. H. (2017b). Keas perform similarly to chimpanzees and elephants when solving collaborative tasks. *PLoS One*, 12(2), e0169799. doi:10.1371/journal.pone.0169799
- Heinsohn, R., Zdenek, C. N., Cunningham, R. B., Endler, J. A., & Langmore, N. E. (2017). Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. *Science Advances*, 3(6), e1602399. doi:10.1126/sciadv.1602399
- Heinrich, B. (1995). An experimental investigation of insight in common ravens (*Corvus corax*). *The Auk*, 112(4), 994-1003. doi:10.2307/4089030
- Heinrich, B., & Bugnyar, T. (2005). Testing problem solving in ravens: String-pulling to reach food. *Ethology*, 111(10), 962-976. doi:10.1111/j.1439-0310.2005.01133.x
- Hennefield, L., Hwang, H. G., & Povinelli, D. J. (2019). Going meta: Retelling the scientific retelling of Aesop's the crow and the pitcher. *Journal of Folklore Research*, 56(2-3), 45-70. doi:10.2979/jfolkrese.56.2_3.04
- Hennefield, L., Hwang, H. G., Weston, S. J., & Povinelli, D. J. (2018). Meta-analytic techniques reveal that corvid causal reasoning in the Aesop's fable paradigm is driven by trial-and-error learning. *Animal Cognition*, 21, 735-748. doi:10.1007/s10071-018-1206-y
- Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, 109(1), 10661-10668. doi:10.1073/pnas.1201895109
- Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science*, 12(2), 141-147. doi:10.1111/1467-9280.00324

- Hespos, S. J., Ferry, A. L., Anderson, E. M., Hollenbeck, E. N., & Rips, L. J. (2016). Five-month-old infants have general knowledge of how nonsolid substances behave and interact. *Psychological Science*, 27(2), 244-256. doi:10.1177/0956797615617897
- Heyes, C. M. (1993). Anecdotes, training, trapping and triangulating: Do animals attribute mental states? *Animal Behaviour*, 46(1), 177-188. doi:10.1006/anbe.1993.1173
- Higgins, P. J. (2002). *Handbook of Australian, New Zealand & Antarctic birds* (1st Ed.). Melbourne, Australia: Oxford University Press.
- Hill, A., Collier-Baker, E., & Suddendorf, T. (2011). Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. *Journal of Comparative Psychology*, 125(1), 91-103. doi:10.1037/a0020867
- Hirata, S. (2003). Cooperation in chimpanzees. *Hattatsu*, 95, 103-111.
- Hirata, S., & Fuwa, K. (2007). Chimpanzees (*Pan troglodytes*) learn to act with other individuals in a cooperative task. *Primates*, 48(1), 13-21. doi:10.1007/s10329-006-0022-1
- Hoffmann, A., Rüttler, V., & Nieder, A. (2011). Ontogeny of object permanence and object tracking in the carrion crow, *Corvus corone*. *Animal Behaviour*, 82(2), 359-367. doi:10.1016/j.anbehav.2011.05.012
- Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26(2), 107-129. doi:10.1016/0001-6918(67)90011-X
- Hoyo, J. d., & Bierregaard, R. (2011). *Handbook of the birds of the world*. Barcelona, Spain: Lynx Editions.
- Huber, L. & Gajdon, G. K. (2006). Technical intelligence in animals: the kea model. *Animal Cognition* 9, 295-305. doi:10.1007/s10071-006-0033-8

- Huber, L., Rechberger, S., & Taborsky, M. (2001). Social learning affects object exploration and manipulation in keas, *Nestor notabilis*. *Animal Behaviour*, 62(5), 945-954. doi:10.1006/anbe.2001.1822
- Humphrey, N. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303-317). Cambridge, UK: Cambridge University Press.
- Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows. *Nature*, 379(6562), 249-251. doi:10.1038/379249a0
- Hunt, G. R., & Gray, R. D. (2002). Species-wide manufacture of stick-type tools by New Caledonian crows. *Emu*, 102, 349-353.
- Hunt, G. R., Gray, R. D., & Taylor, A. H. (2013). Why is tool use rare in animals? *Tool use in animals* (pp. 89-118). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511894800.007
- Huskinson, S. M., Jacobson, S. L., Egelkamp, C. L., Ross, S. R., & Hopper, L. M. (2020). Using a touchscreen paradigm to evaluate food preferences and response to novel photographic stimuli of food in three primate species (*Gorilla gorilla gorilla*, *Pan troglodytes*, and *Macaca fuscata*). *International Journal of Primatology*, 41(1), 5-23. doi:10.1007/s10764-020-00131-0
- Innes, J., Kelly, D., Overton, J. M., & Gillies, C. (2010). Predation and other factors currently limiting New Zealand forest birds. *New Zealand Journal of Ecology*, 34(1), 86-114.
- Iwaniuk, A. N., Dean, K. M., & Nelson, J. E. (2004). Interspecific allometry of the brain and brain regions in parrots (Psittaciformes): Comparisons with other birds and primates. *Brain, Behavior and Evolution*, 65(1), 40-59. doi:10.1159/000081110
- Iwaniuk, A. N., Lefebvre, L., & Wylie, D. R. (2009). The comparative approach and brain-behaviour relationships. *Canadian Journal of Experimental Psychology*, 63(2), 150-159. doi:10.1037/a0015678

- Jaakkola, K. (2014). Do animals understand invisible displacement? A critical review. *Journal of Comparative Psychology*, *128*(3), 225-239. doi:10.1037/a0035675
- Jaakkola, K., Guarino, E., Rodriguez, M., Erb, L., & Trone, M. (2010). What do dolphins (*Tursiops truncatus*) understand about hidden objects? *Animal Cognition*, *13*, 103-120. doi:10.1007/s10071-009-0250-z
- Jackson, J. R. (1969). What do keas die of? *Notornis*, *16*(1), 33-44.
- Jacobs, I. F., & Osvath, M. (2015). The string-pulling paradigm in comparative psychology. *Journal of Comparative Psychology*, *129*(2), 89-120. doi:10.1037/a0038746
- Janzen, M. J., Janzen, D. H., & Pond, C. M. (1976). Tool using by the African grey parrot (*Psittacus erithacus*). *Biotropica*, *8*(1), 70.
- Jarrett, M. (1998). Hazards to kea (*Nestor notabilis*) at rubbish dumps. [Master's thesis, Lincoln University]. Research@Lincoln. <https://researcharchive.lincoln.ac.nz/handle/10182/2405>
- Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Medina, L., ... Avian Brain Nomenclature Consortium. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151-159. doi:10.1038/nrn1606
- Jasp Team (2020). JASP (Version 0.13.1) [Computer Software].
- Jelbert, S. A., Hosking, R. J., Taylor, A. H., & Gray, R. D. (2018). Mental template matching is a potential cultural transmission mechanism for New Caledonian crow tool manufacturing traditions. *Scientific Reports*, *8*, 8956. doi:10.1038/s41598-018-27405-1
- Jelbert, S. A., Miller, R., Schiestl, M., Boeckle, M., Cheke, L. G., Gray, R. D., Taylor, A. H., & Clayton, N. S. (2019). New Caledonian crows infer the weight of objects from observing their movements in a breeze. *Proceedings of Royal Society B: Biological Sciences*, *286*(1894), 20182332. doi:10.1098/rspb.2018.2332

- Jelbert, S. A., Taylor, A. H., Cheke, L. G., Clayton, N. S., & Gray, R. D. (2014). Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PLoS One*, *9*(3), e92895. doi:10.1371/journal.pone.0092895
- Jelbert, S.A., Taylor, A.H. & Gray, R.D. (2015a). Investigating animal cognition with the Aesop's Fable paradigm: current understanding and future directions. *Communicative & Integrative Biology*, *8*(4), e1035846. doi:10.1080/19420889.2015.1035846
- Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2015b). Reasoning by exclusion in New Caledonian crows (*Corvus moneduloides*) cannot be explained by avoidance of empty containers. *Journal of Comparative Psychology*, *129*(3), 283-290. doi:10.1037/a0039313
- Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, *12*(2), 20150871. doi:10.1098/rsbl.2015.0871
- Jensen, A. R. (1980). *Bias in mental testing*. London, UK: Methuen.
- Jerison, H. J. (1973). *Evolution of the brain and intelligence*. New York, NY: Academic Press.
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444-448. doi:10.1038/nature11631
- Johnson-Ulrich, Z., Vonk, J., Humbyrd, M., Crowley, M., Wojtkowski, E., Yates, F., & Allard, S. (2016). Picture object recognition in an American black bear (*Ursus americanus*). *Animal Cognition*, *19*(6), 1237-1242. doi:10.1007/s10071-016-1011-4
- Kabadayi, C., Krasheninnikova, A., O'Neill, L., van de Weijer, J., Osvath, M., & von Bayern, A. M. P. (2017). Are parrots poor at motor self-regulation or is the cylinder task poor at measuring it? *Animal Cognition*, *20*, 1137-1146. doi:10.1007/s10071-017-1131-5
- Kacelnik, A. (2003). The evolution of patience. In G. Loewenstein, D. Read & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 115-138). New York, NY: Russell Sage Foundation.

- Kacelnik, A. (2009). Tools for thought or thoughts for tools? *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), 10071-10072. doi:10.1073/pnas.0904735106
- Kacelnik, A., Chappell, J., Weir, A. A. S., & Kenward, B. (2006). Cognitive adaptations for tool-related behaviour in New Caledonian crows. *Comparative Cognition: Experimental Explorations of Animal Intelligence* (515-528). Oxford, UK: Oxford University Press.
- Kamil, A. C. (1987). A synthetic approach to the study of animal intelligence. Nebraska Symposium on Motivation. *Nebraska Symposium on Motivation*, 35, 257-308.
- Karten, H. J. (1969). The organization the avian telencephalon and some speculations on the phylogeny of the amniote telencephalon. *Annals of the New York Academy of Sciences*, 167(1), 164-179. doi:10.1111/j.1749-6632.1969.tb20442.x
- Karten, H. J. (1997). Evolutionary developmental biology meets the brain: The origins of mammalian cortex. *Proceedings of the National Academy of Sciences*, 94(7), 2800-2804. doi:10.1073/pnas.94.7.2800
- Karten, H. (2013). Neocortical evolution: Neuronal circuits arise independently of lamination. *Current Biology*, 23(1), R12-R15. doi:10.1016/j.cub.2012.11.013
- Kea Conservation Trust. (2019). *Safe ground-based pest control in kea habitat*. <https://www.keaconservation.co.nz/wp-content/uploads/2019/01/SOP-safe-Pest-Control-in-kea-habitat-v8-updated-Jan2019.pdf>
- Kea Conservation Trust. (2017) *A strategic plan for kea conservation - Strategic Objective 2: National Kea Management Plan*. <https://www.keaconservation.co.nz/wp-content/uploads/2017/05/Strategic-objective-2-DRAFT-April2017.pdf>
- Kellman, P. J. & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, 15(4), 483-524. doi:10.1016/0010-0285(83)90017-8

- Kelly, D., McCarthy, E., Menzel, K., & Engebretson, M. (2014). *How many captive birds: Are population studies giving us a clear picture?* <http://www.avianwelfare.org/issues/overview.htm>
- Kemp, J., Mosen, C., Elliott, G., & Hunter, C. (2018). Effects of the aerial application of 1080 to control pest mammals on kea reproductive success. *New Zealand Journal of Ecology*, *42*(2), 158-168. doi:10.20417/nzj ecol.42.28
- Klump, B. C., Martin, J. M., Wild, S., Horsch, J. K., Major, R. E., & Aplin, L. M. (2021). Innovation and geographic spread of a complex foraging culture in an urban parrot. *Science*, *373*(6553), 456. doi:10.1126/science.abe7808
- Koehler, D. J. & Harvey, N. (2004). *Blackwell Handbook of Judgment and Decision Making*. Malden, MA: Blackwell Publishing.
- Koo T. K., & Li, M. Y. (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012
- Knauff, M., & Wolf, A. G. (2010). Complex cognition: The science of human reasoning, problem-solving, and decision-making. *Cognitive Processing*, *11*(2), 99-102. doi:10.1007/s10339-010-0362-z
- Krasheninnikova, A. (2013). Patterned-string tasks: Relation between fine motor skills and visual-spatial abilities in parrots. *PLoS One*, *8*(12), e85499. doi:10.1371/journal.pone.0085499
- Krasheninnikova, A., Berardi, R., Lind, M., O'Neill, L., & von Bayern, A. M. P. (2019). Primate cognition test battery in parrots. *Behaviour*, *2019*(5-8), 721-761. doi:10.1163/1568539X-0003549
- Krasheninnikova, A., Bräger, S., & Wanker, R. (2013). Means-end comprehension in four parrot species: Explained by social complexity. *Animal Cognition*, *16*(5), 755-764. doi:10.1007/s10071-013-0609-z

- Krause, J., & Ruxton, G. D. (2008). *Living in groups* (Reprinted ed.). Oxford, UK: Oxford University Press.
- Krizhevsky, A., Sutskever, I. & Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. doi:10.1145/3065386
- Krueger, K., Esch, L., & Byrne, R. (2019). Animal behaviour in a human world: A crowdsourcing study on horses that open door and gate mechanisms. *PLoS One*, 14(6), e0218954. doi:10.1371/journal.pone.0218954
- Kummer, H. (1971). *Primate societies: Group techniques of ecological adaptation*. London, UK: Routledge. doi:10.4324/9781315127415
- Kushnir, T., Gopnik, A., Lucas, C. & Schulz, L. (2010). Inferring hidden causal structure. *Cognitive Science*, 34, 148-160. doi:10.1111/j.1551-6709.2009.01072.x
- Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young Children Use Statistical Sampling to Infer the Preferences of Other People. *Psychological Science*, 21(8), 1134-1140. doi:10.1177/0956797610376652
- Krupenye, C., Hare, B. (2018). Bonobos prefer individuals that hinder others over those that help. *Current Biology*, 28, 280-286.e5. doi:10.1016/j.cub.2017.11.061
- Labonde, J. Household poisonings in cage birds. (1995). In J. D. Bonagura, J. Bonagura, & R. W. Kirk (Eds). *Kirk's Current Veterinary Therapy XII* (pp. 1299-303). Philadelphia, PA: Saunders.
- Lambert, M. L., Jacobs, I., Osvath, M., & von Bayern, A. P. (2019). Birds of a feather? Parrot and corvid cognition compared. *Behaviour*, 2019(5-8), 505-594. doi:10.1163/1568539X-00003527
- Lambert, M. L., Schiestl, M., Schwing, R., Taylor, A. H., Gajdon, G. K., Slocombe, K. E., & Seed, A. M. (2017). Function and flexibility of object exploration in kea and New Caledonian crows. *Royal Society Open Science*, 4(9), 170652. doi:10.1098/rsos.170652

- Lambert, M. L., Seed, A. M., & Slocombe, K. E. (2015). A novel form of spontaneous tool use displayed by several captive greater vasa parrots (*Coracopsis vasa*). *Biology Letters*, *11*(12), 20150861. doi:10.1098/rsbl.2015.0861
- Laschober, M., Mundry, R., Huber, L., & Schwing, R. (2021). Kea (*Nestor notabilis*) show flexibility and individuality in within-session reversal learning tasks. *Animal Cognition*. doi:10.1007/s10071-021-01524-1
- Lasiewski, R. C., & Dawson, W. R. (1967). A re-examination of the relation between standard metabolic rate and body weight in birds. *The Condor*, *69*(1), 13-23. doi:10.2307/1366368
- Laumer, I. B., Bugnyar, T., & Auersperg, A. M. I. (2016). Flexible decision-making relative to reward quality and tool functionality in Goffin cockatoos (*Cacatua goffiniana*). *Scientific Reports*, *6*, 28380. doi:10.1038/srep28380
- Laumer, I. B., Jelbert, S. A., Taylor, A. H., Rössler, T., Auersperg, A. M. I. (2021). Object manufacture based on a memorized template: Goffin's cockatoos attend to different model features. *Animal Cognition*, *24*, 457-470. doi:10.1007/s10071-020-01435-7
- Lefebvre, L., Nicolakakis, N., & Boire, D. (2002). Tools and brains in birds. *Behaviour*, *139*(7), 939-973. doi:10.1163/156853902320387918
- Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, *63*(4), 233-246. doi:10.1159/000076784
- Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, *53*(3), 549-560. doi:10.1006/anbe.1996.0330
- Li, L., & Jiang, Z. (2014). International trade of CITES listed bird species in China. *PLoS One*, *9*(2), e85012. doi:10.1371/journal.pone.0085012
- Liedtke, J., Werdenich, D., Gajdon, G., Huber, L., & Wanker, R. (2011). Big brains are not enough: Performance of three parrot species in the trap-tube paradigm. *Animal Cognition*, *14*, 143-149. doi:10.1007/s10071-010-0347-4

- Lind, J. (2018). What can associative learning do for planning? *Royal Society Open Science*, 5(11), 180778. doi:10.1098/rsos.180778
- Lisney, T. J., Rubene, D., Rózsa, J., Løvlie, H., Håstad, O., & Ödeen, A. (2011). Behavioural assessment of flicker fusion frequency in chicken *Gallus gallus domesticus*. *Vision Research*, 51(12), 1324-1332. doi:10.1016/j.visres.2011.04.009
- Logan, C. J. (2016). How far will a behaviourally flexible invasive bird go to innovate? *Royal Society Open Science*, 3(6), 160247. doi:10.1098/rsos.160247
- Logan, C. J., Jelbert, S. A., Breen, A. J., Gray, R. D., & Taylor, A. H. (2014). Modifications to the Aesop's fable paradigm change New Caledonian crow performances. *PLoS One*, 9(7), e103049. doi:10.1371/journal.pone.0103049
- Lopez, R. R., Kennedy, A., Mendl, M., Dow, S., & Held, S. (2016). Keeping kea entertained: A survey of current enrichment practices. *UK and Ireland Regional Environmental Enrichment Conference*, 6, 16-37.
- Lovato, S. B., & Waxman, S. R. (2016). Young children learning from touch screens: Taking a wider view. *Frontiers in Psychology*, 7, 1078. doi:10.3389/fpsyg.2016.01078
- Lucca, K., Pospisil, J., & Sommerville, J. A. (2018). Fairness informs social decision making in infancy. *PLoS One*, 13(2), e0192848. doi:10.1371/journal.pone.0192848
- Ma, L. & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition*, 120(3), 403-411. doi:10.1016/j.cognition.2011.02.003
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), e2140-e2148. doi:10.1073/pnas.1323533111
- Macphail, E. M. (1982). *Brain and intelligence in vertebrates*. Oxford, UK: Clarendon Press.

- Macphail, E. M. (1985). Vertebrate intelligence: The null hypothesis. *Philosophical Transactions of the Royal Society B*, 308(1135), 37-51. doi:10.1098/rstb.1985.0008
- Macphail, E. M., & Bolhuis, J. J. (2001). The evolution of intelligence: Adaptive specializations versus general process. *Biological Reviews*, 76(3), 341-364. doi:10.1017/S146479310100570X
- Mallavarapu, S. (2009). Object permanence in orangutans, gorillas, and black-and-white ruffed lemurs [Unpublished PhD Thesis]. Georgia Institute of Technology.
- Marsh, H. L., Vining, A. Q., Levendoski, E. K., & Judge, P. G. (2015). Inference by exclusion in lion-tailed macaques (*Macaca silenus*), a hamadryas baboon (*Papio hamadryas*), capuchins (*Sapajus apella*), and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, 129(3), 256-267. doi:10.1037/a0039316
- Martinho III, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept of “same or different”. *Science* 353(6296), 286-288. doi:10.1126/science.aaf4247
- Mason, G. J. (2010). Species differences in responses to captivity: Stress, welfare and the comparative method. *Trends in Ecology & Evolution*, 25(12), 713-721. doi:10.1016/j.tree.2010.08.011
- Mazenganya, P., Bhagwandin, A., Manger, P. R., & Ihunwo, A. O. (2018). Putative adult neurogenesis in old world parrots: The Congo African grey parrot (*Psittacus erithacus*) and Timneh grey parrot (*Psittacus timneh*). *Frontiers in Neuroanatomy*, 12, 7. doi:10.3389/fnana.2018.00007
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. 2nd Edition. CRC Press.
- McGrew, W. C. (2013). Is primate tool use special? chimpanzee and New Caledonian crow compared. *Philosophical Transactions. Biological Sciences*, 368(1630), 20120422. doi:10.1098/rstb.2012.0422
- McLelland J. M., Reid C., McInnes K., Roe W. D., & Gartrell, B. D. (2010). Evidence of lead exposure in a free-ranging population of kea (*Nestor notabilis*). *Journal of Wildlife Diseases*, 46, 532-540.

- McWilliams, S. R., & Karasov, W. H. (2014). Spare capacity and phenotypic flexibility in the digestive system of a migratory bird: Defining the limits of animal design. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1783), 20140308. doi:10.1098/rspb.2014.0308
- Meehan, C. L., Garner, J. P., & Mench, J. A. (2003). Isosexual pair housing improves the welfare of young amazon parrots. *Applied Animal Behaviour Science*, *81*(1), 73-88. doi:10.1016/S0168-1591(02)00238-1
- Melis, A. P., Hare, B., & Tomasello, M. (2006a). Chimpanzees recruit the best collaborators. *Science*, *311*(5765), 1297-1300. doi:10.1126/science.1123007
- Melis, A. P., Hare, B., & Tomasello, M. (2006b). Engineering cooperation in chimpanzees: Tolerance constraints on cooperation. *Animal Behaviour*, *72*(2), 275-286. doi:10.1016/j.anbehav.2005.09.018
- Mennill, D. J., Ratcliffe, L. M., & Boag, P. T. (2002). Female eavesdropping on male song contests in songbirds. *Science*, *296*(5569), 873. doi:10.1126/science.296.5569.873
- Mikolasch, S., Kotrschal, K., & Schloegl, C. (2011). African grey parrots (*Psittacus erithacus*) use inference by exclusion to find hidden food. *Biology Letters*, *7*(6), 875-877. doi:10.1098/rsbl.2011.0500
- Mikolasch, S., Kotrschal, K., & Schloegl, C. (2012). Is caching the key to exclusion in corvids? The case of carrion crows (*Corvus corone corone*). *Animal Cognition*, *15*(1), 73-82. doi:10.1007/s10071-011-0434-1
- Miller, R., Gruber, R., Frohnwieser, A., Schiestl, M., Jelbert, S. A., Gray, R. D., ... Clayton, N. S. (2020). Decision-making flexibility in New Caledonian crows, young children and adult humans in a multi-dimensional tool-use task. *PLoS One*, *15*(3), e0219874. doi:10.1371/journal.pone.0219874
- Millikan, R. G. (2017). *Beyond concepts: Unicepts, language, and natural information* (1st Ed.). Oxford, UK: Oxford University Press. doi:10.1093/oso/9780198717195.001.0001

- Mithen, S. J. (1996). *The Prehistory of the Mind*. London, UK: Thames & Hudson, London.
- Milton, K. (1981). Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. *American Anthropologist*, 83(3), 534-548.
doi:10.1525/aa.1981.83.3.02a00020
- Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, 21(2), 204-218.
doi:10.1037/h0032198
- Miyata, H., Gajdon, G. K., Huber, L., & Fujita, K. (2011). How do keas (*Nestor notabilis*) solve artificial-fruit problems with multiple locks? *Animal Cognition*, 14(1), 45-58.
doi:10.1007/s10071-010-0342-9
- Moore, M. K., Borton, R. & Darby, B. L. (1978). Visual tracking in young infants: Evidence for object identity or object permanence? *Journal of Experimental Child Psychology* 25(2), 183-198.
doi:10.1016/0022-0965(78)90076-0
- Morozov, N. (2015). Why do birds practice anting? *Biology Bulletin Reviews*, 5(4), 353-365.
doi:10.1134/S2079086415040076
- Mulcahy, N. J., & Call, J. (2006). How great apes perform on a modified trap-tube task. *Animal Cognition*, 9, 193-199. doi:10.1007/s10071-006-0019-6
- Nagy, K. A. (1987). Field metabolic rate and food requirement scaling in mammals and birds. *Ecological Monographs*, 57(2), 111-128. doi:10.2307/1942620
- Naik, H., Bastien, R., Navab, N., & Couzin, I. D. (2020). Animals in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(5), 2073-2083.
doi:10.1109/TVCG.2020.2973063
- Natale, F., Antinucci, F., Spinozzi, G. & Potí, P. (1986). Stage 6 Object Concept in Nonhuman Primate Cognition: A comparison between gorilla (*Gorilla gorilla gorilla*) and Japanese macaque (*Macaca fuscata*). *Journal of Comparative Psychology*, 100(4), 335-339.

- Nawroth, C., von Borell, E., & Langbein, J. (2014). Exclusion performance in dwarf goats (*Capra aegagrus hircus*) and sheep (*Ovis orientalis aries*). *PLoS One*, *9*(4), e93534. doi:10.1371/journal.pone.0093534
- Nawroth, C., von Borell, E., & Langbein, J. (2015). Object permanence in the dwarf goat (*Capra aegagrus hircus*): Perseveration errors and the tracking of complex movements of hidden objects. *Applied Animal Behaviour Science*, *167*, 20-26. doi:10.1016/j.applanim.2015.03.010
- Nelson, X. J., & Fijn, N. (2013). The use of visual media as a tool for investigating animal behaviour. *Animal Behaviour*, *85*(3), 525-536. doi:10.1016/j.anbehav.2012.12.009
- Nelson, X. J., Wilson, D. R., & Evans, C. S. (2008). Behavioral syndromes in stable social groups: An artifact of external constraints? *Ethology*, *114*(12), 1154-1165. doi:10.1111/j.1439-0310.2008.01568.x
- Nichols, M., Bell, P., Mulgan, N., & Taylor, A. (2020). Conditioned aversion in kea to cereal bait: A captive study using anthraquinone. *Applied Animal Behaviour Science*, *230*, 105077. doi:10.1016/j.applanim.2020.105077
- O'Hara, M., Auersperg, A. M. I., Bugnyar, T., & Huber, L. (2015). Inference by exclusion in Goffin cockatoos (*Cacatua goffini*). *PLoS One*, *10*(8), e0134894. doi:10.1371/journal.pone.0134894
- O'Hara, M., Gajdon, G. K., & Huber, L. (2012). Kea logics: How these birds solve difficult problems and outsmart researchers. In S. Watanabe (Ed.), *Logic and sensibility* (1st Ed., pp. 23-38). Tokyo, Japan: Keio University Press.
- O'Hara, M., Huber, L., & Gajdon, G. K. (2015). The advantage of objects over images in discrimination and reversal learning by kea, *Nestor notabilis*. *Animal Behaviour*, *101*, 51-60. doi:10.1016/j.anbehav.2014.12.022
- O'Hara, M., Mioduszewska, B., Mundry, R., Yohanna, Haryoko, T., Rachmatika, R., ... Auersperg, A. M. I. (2021). Wild goffin's cockatoos flexibly manufacture and use tool sets. *Current Biology*, *31*, 1-9. doi:10.1016/j.cub.2021.08.009

- O'Hara, M., Schwing, R., Federspiel, I., Gajdon, G. K., & Huber, L. (2016). Reasoning by exclusion in the kea (*Nestor notabilis*). *Animal Cognition*, *19*, 965-975. doi:10.1007/s10071-016-0998-x
- Obozova, T. A., & Zorina, Z. A. (2013). Do great grey owls comprehend means-end relationships? *International Journal of Comparative Psychology*, *26*, 187-201.
- Obozova, T., Bagotskaya, M., Smirnova, A., & Zorina, Z. (2014). A comparative assessment of birds' ability to solve string-pulling tasks. *Biology Bulletin*, *41*(7), 565-574. doi:10.1134/S1062359014070073
- Olah, G., Butchart, S., Symes, A., Guzmán, I., Cunningham, R., Brightsmith, D., & Heinsohn, R. (2016). Ecological and socio-economic factors affecting extinction risk in parrots. *Biodiversity and Conservation*, *25*(2), 205-223. doi:10.1007/s10531-015-1036-z
- Olkowicz, S., Kocourek, M., Lučan, R., Porteš, M., Tecumseh Fitch, W., Herculano-Houzel, S., & Nemeč, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), 7255-7260. doi:10.1073/pnas.1517131113
- Olson, D. J., Kamil, A. C., Balda, R. P., & Nims, P. J. (1995). Performance of four seed-caching corvid species in operant tests of nonspatial and spatial memory. *Journal of Comparative Psychology*, *109*(2), 173-181. doi:10.1037/0735-7036.109.2.173
- Orams, M. B. (2002). Feeding wildlife as a tourism attraction: A review of issues and impacts. *Tourism Management*, *23*(3), 281-293. doi:10.1016/S0261-5177(01)00080-2
- Orr-Walker, T., Adams, N. J., Roberts, L. G., Kemp, J. R., & Spurr, E. B. (2012). Effectiveness of the bird repellents anthraquinone and d-pulegone on an endemic New Zealand parrot, the kea (*Nestor notabilis*). *Applied Animal Behaviour Science*, *137*(1-2), 80-85. doi:10.1016/j.applanim.2012.01.002

- Osvath, M., Kabadayi, C., & Jacobs, I. (2014). Independent evolution of similar complex cognitive skills: The importance of embodied degrees of freedom. *Animal Behavior and Cognition*, 1(3), 249-264. doi:10.12966/abc.08.03.2014
- Pagel, M., et al. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5), 673-684. doi: 10.1080/10635150490522232
- Park, P. J., & Bell, M. A. (2010). Variation of telencephalon morphology of the threespine stickleback (*Gasterosteus aculeatus*) in relation to inferred ecology. *Journal of Evolutionary Biology*, 23(6), 1261-1277. doi:10.1111/j.1420-9101.2010.01987.x
- Patterson-Kane, E., Nicol, C. J., Foster, T. M., & Temple, W. (1997). Limited perception of video images by domestic hens. *Animal Behaviour*, 53(5), 951-963. doi:10.1006/anbe.1996.0385
- Pellis, S. M. (1981). A description of social play by the Australian magpie, *Gymnorhina tibicen*, based on Eshkol-Wachman notation. *Bird Behaviour*, 3(3), 61-79. doi:10.3727/015613881791560685
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(2), 109-130. doi:10.1017/S0140525X08003543
- Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., ... Harmon, L. J. (2014). Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216-2218. doi:10.1093/bioinformatics/btu181
- Pepperberg, I. M. & Kozak, F. A. (1986). Object permanence in the African Grey parrot (*Psittacus erithacus*). *Animal Learning & Behavior*, 14, 322-330. doi:10.3758/BF03200074
- Pepperberg, I. M. (1999). *The Alex Studies*. Cambridge, MA: Harvard Univ. Press.
- Pepperberg, I. M. (2007). Grey parrots do not always 'parrot': The roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences*, 29(1), 1-13. doi:10.1016/j.langsci.2005.12.002

- Pepperberg, I. M. (2015). Reply to Jaakkola (2014): “Do animals understand invisible displacement? A critical review”. *Journal of Comparative Psychology*, *129*(2), 198-201. doi:10.1037/a0038319
- Pepperberg, I. M., & Funk, M. S. (1990). Object permanence in four species of psittacine birds: An African Grey parrot (*Psittacus erithacus*), an Illiger mini macaw (*Ara maracana*), a parakeet (*Melopsittacus undulatus*), and a cockatiel (*Nymphicus hollandicus*). *Learning & Behavior*, *18*, 97. doi:10.3758/BF03205244
- Pepperberg, I. M., Willner, M. R., & Gravitz, L. B. (1997). Development of Piagetian object permanence in a grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, *111*(1), 63-75. doi:10.1037/0735-7036.111.1.63
- Pepperberg, I. M., Koepke, A., Livingston, P., Girard, M., & Hartsfield, L. A. (2013). Reasoning by inference: Further studies on exclusion in grey parrots (*Psittacus erithacus*). *Journal of Comparative Psychology*, *127*(3), 272-281. doi:10.1037/a0031641
- Phillips, W. & Santos, L. R. (2007). Evidence for kind representations in the absence of language: Experiments with rhesus monkeys (*Macaca mulatta*). *Cognition*, *102*(3), 455-463. doi:10.1016/j.cognition.2006.01.009
- Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books. doi:10.1037/11168-000
- Plaçi, S., Eckert, J., Rakoczy, H. & Fischer, J. (2018). Long-tailed macaques (*Macaca fascicularis*) can use simple heuristics but fail at drawing statistical inferences from populations to samples. *Royal Society Open Science*, *5*(9), 181025. doi:10.1098/rsos.181025
- Plaçi, S., Padberg, M., Rakoczy, H., & Fischer, J. (2019). Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty. *Scientific Reports*, *9*, 12107. doi:10.1038/s41598-019-48543-0
- Platt, J. R. (1964). Strong inference. *Science*, *146*(3642), 347-353. doi:10.1126/science.146.3642.347

- Pollok, B., Prior, H., & Güntürkün, O. (2000). Development of object permanence in food-storing magpies (*Pica pica*). *Journal of Comparative Psychology*, *114*(2), 148-157. doi:10.1037/0735-7036.114.2.148
- Povinelli, D. (2003). *Folk physics for apes*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198572190.001.0001
- Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(35), 13861-13867. doi:10.1073/pnas.0706147104
- Premack, D. (2010). Why Humans Are Unique: Three Theories. *Perspectives on Psychological Science*, *5*(1), 22-32. doi:10.1177/1745691609356782
- Price, E. R., Brun, A., Caviedes-Vidal, E., & Karasov, W. H. (2015). Digestive adaptations of aerial lifestyles. *Physiology*, *30*(1), 69-78. doi:10.1152/physiol.00020.2014
- Provine, R. R. (1992). Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bulletin of the Psychonomic Society*, *30*(1), 1-4. doi:10.3758/BF03330380
- Raby, C. R., Alexis, D. M., & Dickinson, A., Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, *445*(7130), 919-921. doi:10.1038/nature05575
- Railton, R., Foster, T. M., & Temple, W. (2014). Object/picture recognition in hens. *Behavioural Processes*, *104*, 53-64. doi:10.1016/j.beproc.2014.01.013
- Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition*, *131*(1), 60-68. doi:10.1016/j.cognition.2013.12.011
- Ramsey, G., Bastian, M. L., & van Schaik, C. (2007). Animal innovation defined and operationalized. *The Behavioral and Brain Sciences*, *30*(4), 393-407. doi:10.1017/S0140525X07002373
- Range, F., Horn, L., Bugnyar, T., Gajdon, G., & Huber, L. (2009). Social attention in keas, dogs, and human children. *Animal Cognition*, *12*(1), 181-192. doi:10.1007/s10071-008-0181-0

- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.
- Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), 4436-4441. doi:10.1073/pnas.062041299
- Redshaw, J., Taylor, A. H., & Suddendorf, T. (2017). Flexible planning in ravens? *Trends in Cognitive Sciences*, 21(11), 821-822. doi:10.1016/j.tics.2017.09.001
- Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, 57(7), 971-976. doi:10.3758/BF03205456
- Regolin, L., Vallortigara, G., & Zanforlin, M. (1995). Object and spatial representations in detour problems by chicks. *Animal Behaviour*, 49(1), 195-199. doi:10.1016/0003-3472(95)80167-7
- Regolin, L., Marconato, F., & Vallortigara, G. (2004). Hemispheric differences in the recognition of partly occluded objects by newly hatched domestic chicks (*Gallus gallus*). *Animal Cognition*, 7(3), 162-170. doi:10.1007/s10071-004-0208-0
- Regolin, L., Garzotto, B., Rugani, R., Pagni, P. & Vallortigara, G. (2005). Working memory in the chick: parallel and lateralized mechanisms for encoding of object- and position-specific information. *Behavioural Brain Research* 157(1), 1-9. doi:10.1016/j.bbr.2004.06.012
- Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., ... Jarvis, E. D. (2004). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *Journal of Comparative Neurology*, 473(3), 377-414. doi:10.1002/cne.20118
- Revenu, B., & Csibra, G. (2020) For 19-month-olds, what happens on-screen stays on-screen. *Proceedings from 42nd annual meeting of the Cognitive Science Society*, 508-513.
- Riesenhuber, M., Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025. doi:10.1038/14819

- Roberts, W., MacDonald, H., & Lo, K. (2018). Pigeons play the percentages: computation of probability in a bird. *Animal Cognition*, *21*, 575-581. doi:10.1007/s10071-018-1192-0
- Rooijackers, E. F., Kaminski, J., & Call, J. (2009). Comparing dogs and great apes in their ability to visually track object transpositions. *Animal Cognition*, *12*, 789-796. doi:10.1007/s10071-009-0238-8
- Rosati, A. G. (2017). Foraging cognition: Reviving the ecological intelligence hypothesis. *Trends in Cognitive Sciences*, *21*(9), 691-702. doi:10.1016/j.tics.2017.05.011
- Roy, E. A. (2018) *Beak fitness: New Zealand develops roadside gym for endangered keas*. <https://www.theguardian.com/environment/2018/jan/11/gym-for-birds-new-zealand-develops-exercise-playground-for-endangered-kea>
- Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition*, *16*, 557-564. doi:10.1007/s10071-012-0593-8
- Rugani, R., Vallortigara, G., & Regolin, L. (2013). Numerical abstraction in young domestic chicks (*Gallus gallus*). *PLoS One*, *8*(6), e65262. doi:10.1371/journal.pone.0065262
- Rutz, C., Klump, B. C., Komarczyk, L., Leighton, R., Kramer, J., Wischniewski, S., ... Masuda, B. M. (2016). Discovery of species-wide tool use in the Hawaiian crow. *Nature*, *537*(7620), 403-407. doi:10.1038/nature19103
- Sabbatini, G., & Visalberghi, E. (2008). Inferences about the location of food in capuchin monkeys (*Cebus apella*) in two sensory modalities. *Journal of Comparative Psychology*, *122*(2), 156-166. doi:10.1037/0735-7036.122.2.156
- Safi, K., & Dechmann, D. K. N. (2005). Adaptation of brain regions to habitat complexity: A comparative analysis in bats (Chiroptera). *Proceedings of the Royal Society B: Biological Sciences*, *272*(1559), 179-186. doi:10.1098/rspb.2004.2924

- Sándor, K., & Miklósi, Á. (2020). How to report anecdotal observations? A new approach based on a lesson from "puffin tool use". *Frontiers in Psychology, 11*, 555487. doi:10.3389/fpsyg.2020.555487
- Scarf, D., Hayne, H., & Colombo, M. (2011). Pigeons on par with primates in numerical competence. *Science, 334*(6063), 1664. doi:10.1126/science.1213357
- Schloegl, C., Dierks, A., Gajdon, G. K., Huber, L., Kotrschal, K., & Bugnyar, T. (2009). What you see is what you get? Exclusion performances in ravens and keas. *PLoS One, 4*(8), e6368. doi:10.1371/journal.pone.0006368
- Schmidt, G. F., & Cook, R. G. (2006). Mind the gap: Means-end discrimination by pigeons. *Animal Behaviour, 71*(3), 599-608. doi:10.1016/j.anbehav.2005.06.010
- Schuck-Paim, C., Alonso, W. J., & Ottoni, E. B. (2008). Cognition in an ever-changing world: Climatic variability is associated with brain size in neotropical parrots. *Brain, Behavior and Evolution, 71*(3), 200-215. doi:10.1159/000119710
- Schuck-Paim, C., Borsari, A., & Ottoni, E. (2009). Means to an end: Neotropical parrots manage to pull strings to meet their goals. *Animal Cognition, 12*(2), 287-301. doi:10.1007/s10071-008-0190-z
- Schwing, R., Jocteur, E., Wein, A., Noë, R., & Massen, J. J. M. (2016). Kea cooperate better with sharing affiliates. *Animal Cognition, 19*(6), 1093-1102. doi:10.1007/s10071-016-1017-y
- Schwing, Nelson, X. J., Wein, A., & Parsons, S. (2017). Positive emotional contagion in a New Zealand parrot. *Current Biology, 27*(6), R213-R214. doi:10.1016/j.cub.2017.02.020
- Schwing, R., Weber, S., & Bugnyar, T. (2017). Kea (*Nestor notabilis*) decide early when to wait in food exchange task. *Journal of Comparative Psychology, 131*(4), 269-276. doi:10.1037/com0000086
- Schwing, R., Weiss, F., Tichy, A., & Gajdon, G. (2019). Testing the causal understanding of water displacement by kea (*Nestor notabilis*). *Behaviour, 156*(5-8), 447-478. doi:10.1163/1568539X-00003536

- Schwing, R., Reuillon, L., Conrad, M., Noë, R., Huber, L., & Massen, J. (2020). Paying attention pays off: Kea improve in loose-string cooperation by attending to partner. *Ethology*, *126*(2), 246-256. doi:10.1111/eth.12984
- Schwing, R., Meaux, E., Piseddu, A., Huber, L., & Noë, R. (2021). Kea, *Nestor notabilis*, achieve cooperation in dyads, triads, and tetrads when dominants show restraint. *Learning & Behavior*, *49*(1), 36-53. doi:10.3758/s13420-021-00462-9
- Seed, A., & Byrne, R. (2010). Animal tool-use. *Current Biology*, *20*(23), R1032-R1039. doi:10.1016/j.cub.2010.09.042
- Seed, A. M., Emery, N., & Clayton, N. (2009). Intelligence in corvids and apes: A case of convergent evolution? *Ethology*, *115*(5), 401-420. doi:10.1111/j.1439-0310.2009.01644.x
- Seed, A. M., Seddon, E., Greene, B., & Call, J. (2012). Chimpanzee ‘folk physics’: Bringing failures into focus. *Philosophical Transactions of the Royal Society B*, *367*(1603), 2743-2752. doi:10.1098/rstb.2012.0222
- Seed, A. M., Tebbich, S., Emery, N. J. & Clayton, N. S. (2006). Investigating physical cognition in rooks, *Corvus frugilegus*. *Current Biology*, *16*(7), 697-701. doi:10.1016/j.cub.2006.02.066
- Seitz, B. M., McCune, K., MacPherson, M., Bergeron, L., Blaisdell, A. P., & Logan, C. J. (2021). Using touchscreen equipped operant chambers to study animal cognition: Benefits, limitations, and advice. *PLoS One*, *16*(2), e0246446. doi:10.1371/journal.pone.0246446
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust Object Recognition with Cortex-Like Mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(3), 411-426. doi:10.1109/TPAMI.2007.56
- Shaw, R. C., Plotnik, J. M., & Clayton, N. S. (2013). Exclusion in corvids: The performance of food-caching Eurasian jays (*Garrulus glandarius*). *Journal of Comparative Psychology*, *127*(4), 428-435. doi:10.1037/a0032010

- Shaw, R. C., Boogert, N. J., Clayton, N. S. & Burns, K. C. (2015). Wild psychometrics: evidence for 'general' cognitive performance in wild New Zealand robins, *Petroica longipes*. *Animal Behaviour*, *109*, 101-111. doi:10.1016/j.anbehav.2015.08.001
- Shaw, R.C., Schmelz, M. (2017). Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. *Animal Cognition*, *20*, 1003-1018. doi:10.1007/s10071-017-1135-1
- Shettleworth, S. J. (1987). Intelligence: More than a matter of associations. *Behavioral and Brain Sciences*, *10*(4), 679. doi:10.1017/S0140525X00055266
- Shettleworth, S. J. (2010). *Cognition, Evolution, and Behavior*. New York, NY: Oxford University Press.
- Shettleworth, S. J. (2012). Modularity, comparative cognition and human uniqueness. *Philosophical Transactions of the Royal Society B*, *367*(1603), 2794-2802.
- Shettleworth, S. J. (2012). Do animals have insight, and what is insight anyway? *Canadian Journal of Experimental Psychology*, *66*(4), 217-226. doi:10.1037/a0030674
- Shultz, S., & Dunbar, R. I. M. (2010). Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *Journal of Comparative Psychology*, *124*(3), 252-260. doi:10.1037/a0018894
- Shumaker, R. W., Walkup, K. R., Beck, B. B., & Burghardt, G. M. (2011). *Animal tool behavior: The use and manufacture of tools by animals*. Baltimore, MD: Johns Hopkins University Press.
- Silk, J. B. (1999). Male bonnet macaques use information about third-party rank relationships to recruit allies. *Animal Behaviour*, *58*(1), 45-51. doi:10.1006/anbe.1999.1129
- Silva, T., Guzmán, A., Urantówka, A. D., & Mackiewicz, P. (2017). A new parrot taxon from the Yucatán peninsula, Mexico – its position within genus *Amazona* based on morphology and molecular phylogeny. *PeerJ*, *5*, e3475. doi:10.7717/peerj.3475

- Smith, C. L., Van Dyk, D. A., Taylor, P. W., & Evans, C. S. (2009). On the function of an enigmatic ornament: Wattles increase the conspicuousness of visual displays in male fowl. *Animal Behaviour*, 78(6), 1433-1440. doi:10.1016/j.anbehav.2009.07.034
- Sol, D. (2009). Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters*, 5(1), 130-133. doi:10.1098/rsbl.2008.0621
- Sophian, C. (1984). Spatial transpositions and the early development of search. *Developmental Psychology*, 20(1), 21-28. doi:10.1037/0012-1649.20.1.21
- Sophian, C., & Sage, S. (1983). Developments in infants' search for displaced objects. *Journal of Experimental Child Psychology*, 35(1), 143-160. doi:10.1016/0022-0965(83)90075-9
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201-292. doi:10.2307/1412107
- Spelke, E. S. (1991). Physical knowledge in infancy: Reflections on Piaget's theory. *The Epigenesis of Mind* (pp. 145-182). Hove, UK: Psychology Press. doi:10.4324/9781315807805-14
- Spelke, E. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.) *Language in Mind: Advances in the Study of Language and Thought* (pp. 277-311). Cambridge, MA: MIT Press.
- Spetch, M. L., & Weisman, R. G. (2012). Birds' perception of depth and objects in pictures. In O. F. Lazareva, T. Shimizu, & E. A. Wasserman (Eds.), *How animals see the world: Comparative behavior, biology, and evolution of vision* (pp. 217-231). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195334654.003.0013
- Spinozzi, G., & Potí, P. (1993). Piagetian stage 5 in two infant chimpanzees (*Pan troglodytes*): The development of permanence of objects and the spatialization of causality. *International Journal of Primatology*, 14(6), 905-917. doi:10.1007/BF02220259

- Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., & Güntürkün, O. (2020). A cortex-like canonical circuit in the avian forebrain. *Science*, 369(6511), eabc5534. doi:10.1126/science.abc5534
- St Clair, J. J. H., & Rutz, C. (2013). New Caledonian crows attend to multiple functional properties of complex tools. *Philosophical Transactions of the Royal Society B*, 368(1630), 20120415. doi:10.1098/rstb.2012.0415
- Stevens, J. R., Kennedy, B. A., Morales, D., & Burks, M. (2016). The domain specificity of intertemporal choice in pinyon jays. *Psychonomic Bulletin & Review*, 23, 915-921. doi:10.3758/s13423-015-0973-6
- Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology*, 8, 862. doi:10.3389/fpsyg.2017.00862
- Stowers, J. R., Hofbauer, M., Bastien, R., Griessner, J., Higgins, P., Farooqui, S., ... Straw, A. D. (2017). Virtual reality for freely moving animals. *Nature Methods*, 14(10), 995-1002. doi:10.1038/nmeth.4399
- Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, Mass: Sinauer Assoc.
- Su, S., Cassey, P., Vall-Llosera, M., & Blackburn, T. M. (2015). Going cheap: Determinants of bird price in the Taiwanese pet market. *PLoS One*, 10(5), e0127482. doi:10.1371/journal.pone.0127482
- Subias, L., Griffin, A. S., & Guez, D. (2019). Inference by exclusion in the red-tailed black cockatoo (*Calyptorhynchus banksii*). *Integrative Zoology*, 14(2), 193-203. doi:10.1111/1749-4877.12299
- Suddendorf, T. (2013). *The Gap*. New York: Basic Books.
- Suddendorf, T., Redshaw, J., Bulley, A. (in press) *The Invention of Tomorrow: How Foresight Conquered the World*. New York, NY: Basic Books.

- Suh, A., Paus, M., Kiefmann, M., Churakov, G., Franke, F. A., Brosius, J., ... Schmitz, J. (2011). Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nature Communications*, 2(1), 443. doi:10.1038/ncomms1448
- Takahashi, M., Ueno, Y., & Fujita, K. (2015). Inference in a social context: A comparative study of capuchin monkeys (*Cebus apella*), tree shrews (*Tupaia belangeri*), hamsters (*Mesocricetus auratus*), and rats (*Rattus norvegicus*). *Journal of Comparative Psychology*, 129(4), 402-411. doi:10.1037/a0039732
- Taylor, A. H. (2014). Corvid cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(3), 361-372. doi:10.1002/wcs.1286
- Taylor, A. H., Bastos, A. P. M., Brown, R., & Allen, C. (*under review*) The signature-testing approach to mapping biological and artificial intelligences. *Trends in Cognitive Sciences*.
- Taylor, A. H., Elliffe, D. M., Hunt, G. R., Emery, N. J., Clayton, N. S., & Gray, R. D. (2011). New Caledonian crows learn the functional properties of novel tool types. *PLoS One*, 6(12), e26887. doi:10.1371/journal.pone.0026887
- Taylor, A. H., Elliffe, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural innovation in New Caledonian crows. *Proceedings of the Royal Society B: Biological Sciences*, 277(1694), 2637-2643. doi:10.1098/rspb.2010.0285
- Taylor, A. H., & Gray, R. D. (2009). Animal cognition: Aesop's fable flies from fiction to fact. *Current Biology*, 19(17), R731-R732. doi:10.1016/j.cub.2009.07.055
- Taylor, A. H., & Gray, R. D. (2014). Is there a link between the crafting of tools and the evolution of cognition? *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(6), 693-703. doi:10.1002/wcs.1322
- Taylor, A. H., Hunt, G. R., & Gray, R. D. (2012). Context-dependent tool use in New Caledonian crows. *Biology Letters*, 8(2), 205-207. doi:10.1098/rsbl.2011.0782

- Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by New Caledonian crows. *Current Biology*, *17*(17), 1504-1507. doi:10.1016/j.cub.2007.07.057
- Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do New Caledonian crows solve physical problems through causal reasoning? *Proceedings of the Royal Society B: Biological Sciences*, *276*(1655), 247-254. doi:10.1098/rspb.2008.1107
- Taylor, A. H., Knaebe, B., & Gray, R. D. (2012). An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4977-4981. doi:10.1098/rspb.2012.1998
- Taylor, A. H., Medina, F. S., Holzhaider, J. C., Hearne, L. J., Hunt, G. R., & Gray, R. D. (2010). An investigation into the cognition behind spontaneous string pulling in New Caledonian crows. *PLoS One*, *5*(2), e9345. doi:10.1371/journal.pone.0009345
- Tebbich, S., Seed, A. M., Emery, N. J., & Clayton, N. S. (2007). Non-tool-using rooks, *Corvus frugilegus*, solve the trap-tube problem. *Animal Cognition*, *10*, 225-231. doi:10.1007/s10071-006-0061-4
- Tebbich, S., Taborsky, M., & Winkler, H. (1996). Social manipulation causes cooperation in keas. *Animal Behaviour*, *52*(1), 1-10. doi:10.1006/anbe.1996.0147
- Tecwyn, E., Denison, S., Messer, E., & Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition*, *20*(2), 243-256. doi:10.1007/s10071-016-1043-9
- Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(48), 19156-19159. doi:10.1073/pnas.0700271104
- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054-1059. doi:10.1126/science.1196404

- Tennyson, A. J. D. (2010). The origin and history of New Zealand's terrestrial vertebrates. *New Zealand Journal of Ecology*, 34(1), 6-27.
- Teschke, I., Cartmill, E. A., Stankewitz, S., & Tebbich, S. (2011). Sometimes tool use is not the key: No evidence for cognitive adaptive specializations in tool-using woodpecker finches. *Animal Behaviour*, 82(5), 945-956. doi:10.1016/j.anbehav.2011.07.032
- Teschke, I., & Tebbich, S. (2011). Physical cognition and tool-use: Performance of Darwin's finches in the two-trap tube task. *Animal Cognition*, 14, 555-563. doi:10.1007/s10071-011-0390-9
- Teschke, I., Wascher, C. A. F., Scriba, M. F., von Bayern, A. M. P., Huml, V., Siemers, B., & Tebbich, S. (2013) Did tool-use evolve with enhanced physical cognitive abilities? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1630), 20120418. doi:10.1098/rstb.2012.0418
- Thorogood, R., Kokko, H., & Mappes, J. (2018). Social transmission of avoidance among predators facilitates the spread of novel prey. *Nature Ecology & Evolution*, 2(2), 254-261. doi:10.1038/s41559-017-0418-x
- Torres Ortiz, S., Maxwell, A., Krasheninnikova, A., Wahlberg, M., & Larsen, O. N. (2019). Problem solving capabilities of peach-fronted conures (*Eupsittula aurea*) studied with the string-pulling test. *Behaviour*, 2019(5-8), 815-846. doi:10.1163/1568539X-00003539
- Ujfalussy, D., Miklósi, Á, & Bugnyar, T. (2013). Ontogeny of object permanence in a non-storing corvid species, the jackdaw (*Corvus monedula*). *Animal Cognition*, 16, 405-416. doi:10.1007/s10071-012-0581-z
- Vallortigara, G., Regolin, L., Rigoni, M., & Zanforlin, M. (1998). Delayed search for a concealed imprinted object in the domestic chick. *Animal Cognition*, 1, 17-24. doi:10.1007/s100710050003
- van Horik, J. O., & Emery, N. J. (2011). Evolution of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 621-633. doi:10.1002/wcs.144

- van Horik, J. O., Clayton, N. S., Emery, N. J. (2012). Convergent evolution of cognition in corvids, apes and other animals. *The oxford handbook of comparative evolutionary psychology*. Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199738182.013.0005
- van Horik, J. O., & Madden, J. R. (2016). A problem with problem solving: Motivational traits, but not cognition, predict success on novel operant foraging tasks. *Animal Behaviour*, *114*, 189-198. doi:10.1016/j.anbehav.2016.02.006
- van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R. & Madden, J. R. (2018). Intra-individual variation in performance on novel variants of similar tasks influences single factor explanations of general cognitive processes. *Royal Society Open Science*, *5*(7), 171919. doi:10.1098/rsos.171919
- Veltman, C. J., & Westbrooke, I. (2011). Forest bird mortality and baiting practices in New Zealand aerial 1080 operations from 1986 to 2009. *New Zealand Journal of Ecology*, *35*(1), 21-29.
- Vergara-Tabares, D. L., Cordier, J. M., Landi, M. A., Olah, G., & Nori, J. (2020). Global trends of habitat destruction and consequences for parrot conservation. *Global Change Biology*, *26*(8), 4251-4262. doi:10.1111/gcb.15135
- Vincze, O. (2016). Light enough to travel or wise enough to stay? Brain size evolution and migratory behavior in birds. *Evolution*, *70*(9), 2123-2133. doi:10.1111/evo.13012
- Volter, C. J., Tinklenberg, B., Call, J. & Seed, A. M. (2018). Comparative psychometrics: establishing what differs is central to understanding what evolves. *Philosophical Transactions of the Royal Society B*, *373*(1576), 20170283. doi:10.1098/rstb.2017.0283
- von Bayern, A. M. P., Danel, S., Auersperg, A. M. I., Mioduszewska, B., & Kacelnik, A. (2018). Compound tool construction by New Caledonian crows. *Scientific Reports*, *8*, 15676-8. doi:10.1038/s41598-018-33458-z

- von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C., & Kacelnik, A. (2009). The role of experience in problem solving and innovative tool use in crows. *Current Biology*, *19*(22), 1965-1968. doi:10.1016/j.cub.2009.10.037
- von Bayern, A. M. P., Jacobs, I., & Osvath, M. (2020). Tool-using puffins prick the puzzle of cognitive evolution. *Proceedings of the National Academy of Sciences*, *117*(6), 2737-2739. doi:10.1073/pnas.1922117117
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi:10.3758/BF03194105
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Morey, R. D. (2018). Bayesian inference for psychology. part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35-57. doi:10.3758/s13423-017-1343-3
- Wakonig, B., Auersperg, A. M. I., & O'Hara, M. (2021). String-pulling in the Goffin's cockatoo (*Cacatua goffiniana*). *Learning & Behavior*, *49*(1), 124-136. doi:10.3758/s13420-020-00454-1
- Wang, L., Luo, Y., Wang, X., Maierdiyali, A., Chang, H., & Li, Z. (2019). Azure-winged magpies solve string-pulling tasks by partial understanding of the physical cognition. *Current Zoology*, *65*(4), 385-392. doi:10.1093/cz/zoy070
- Wang, Y. (2007). The Cognitive Processes of Formal Inferences. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *1*(4), 75-86. doi:10.4018/jcini.2007100106
- Wass, C., Denman-Brice, A., Rios, C., Light, K. R., Kolata, S., Smith, A. M., & Matzel, L. D. (2012). Covariation of learning and "reasoning" abilities in mice: Evolutionary conservation of the operations of intelligence. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 109-124. doi:10.1037/a0027355

- Wedel, M. J. (2019). Postcranial skeletal pneumaticity in sauropods and its implications for mass estimates. *The sauropods: Evolution and paleobiology* (pp. 201-228). Berkeley, CA: University of California Press. doi:10.1525/9780520932333-010
- Wein, A., Gajdon, G. K., Schwing, R., & Wright, J. (2015). Picture-object recognition in kea (*Nestor notabilis*). *Ethology*, *121*(11), 1059-1070. doi:10.1111/eth.12423
- Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, *297*(5583), 981. doi:10.1126/science.1073433
- Weir, A. A. S., Kacelnik, A. (2006). A New Caledonian crow (*Corvus moneduloides*) creatively re-designs tools by bending or unbending aluminium strips. *Animal Cognition*, *9*, 317. doi:10.1007/s10071-006-0052-5
- Wellman, H.M., Kushnir, T., Xu, F. & Brink, K.A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy*, *21*(5), 668-676. doi:10.1111/infa.12131
- Werdenich, D., & Huber, L. (2006). A case of quick problem solving in birds: String pulling in keas, *Nestor notabilis*. *Animal Behaviour*, *71*(4), 855-863. doi:10.1016/j.anbehav.2005.06.018
- Weser, C., & Ross, J. (2013). The effect of colour on bait consumption of kea (*Nestor notabilis*): Implications for deterring birds from toxic baits. *New Zealand Journal of Zoology*, *40*(2), 137-144. doi:10.1080/03014223.2012.710639
- Whiten, A., Custance, D. M., Gomez, J., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *110*(1), 3-14. doi:10.1037/0735-7036.110.1.3
- Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., ... Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682-685. doi:10.1038/21415
- Wilcox, T. & Baillargeon, R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology*, *37*(2), 97-155. doi:10.1006/cogp.1998.0690

- Wilkinson, A., Mueller-Paul, J., & Huber, L. (2013). Picture-object recognition in the tortoise *Chelonoidis carbonaria*. *Animal Cognition*, *16*(1), 99-107. doi:10.1007/s10071-012-0555-1
- Wimpenny, J. H., Weir, A. A. S., Clayton, L., Rutz, C., & Kacelnik, A. (2009). Cognitive processes associated with sequential tool use in New Caledonian crows. *PLoS One*, *4*(8), e6471. doi:10.1371/journal.pone.0006471
- Winkler, H., Leisler, B., & Bernroider, G. (2004). Ecological constraints on the evolution of avian brains. *Journal Für Ornithologie*, *145*(3), 238-244. doi:10.1007/s10336-004-0040-y
- Winkler, S. L., & Bryant, G. A. (2021). Play vocalisations and human laughter: A comparative review. *Bioacoustics*, *30*(5), 499-526. doi:10.1080/09524622.2021.1905065
- Wright, T. F., Schirtzinger, E. E., Matsumoto, T., Eberhard, J. R., Graves, G. R., Sanchez, J. J., ... Fleischer, R. C. (2008). A multilocus molecular phylogeny of the parrots (Psittaciformes): Support for a Gondwanan origin during the cretaceous. *Molecular Biology and Evolution*, *25*(10), 2141-2156. doi:10.1093/molbev/msn160
- Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*(1), 97-104. doi:10.1016/j.cognition.2009.04.006
- Young, A. M., Hobson, E. A., Lackey, L. B., & Wright, T. F. (2012). Survival on the ark: Life-history trends in captive parrots. *Animal Conservation*, *15*(1), 28-43. doi:10.1111/j.1469-1795.2011.00477.x
- Young, L. M., Kelly, D., & Nelson, X. J. (2012). Alpine flora may depend on declining frugivorous parrot for seed dispersal. *Biological Conservation*, *147*(1), 133-142. doi:10.1016/j.biocon.2011.12.023
- Zelenkov, N. V. (2016). The first fossil parrot (Aves, Psittaciformes) from Siberia and its implications for the historical biogeography of Psittaciformes. *Biology Letters*, *12*(10), 20160717. doi:10.1098/rsbl.2016.0717

Zentall, T. R. (2021). Macphail (1987) revisited: Pigeons have much cognitive behavior in common with humans. *Frontiers in Psychology, 11*, 618636. doi:10.3389/fpsyg.2020.618636

Zucca, P., Milos, N., & Vallortigara, G. (2007). Piagetian object permanence and its development in Eurasian jays (*Garrulus glandarius*). *Animal Cognition, 10*, 243-258. doi:10.1007/s10071-006-0063-2