

How to succeed in human modified environments

[Logan CJ](#)^{1,2}, [Shaw R](#)³, [Lukas D](#)¹, [McCune KB](#)^{1,2}

2022-08-25. Open... access [code](#)

Affiliations: 1) Max Planck Institute for Evolutionary Anthropology, Germany; 2) University of California Santa Barbara, USA; 3) Victoria University of Wellington, New Zealand.
Corresponding authors: corina_logan@eva.mpg.de, rachael.shaw@vuw.ac.nz, kelseybmccune@gmail.com

ABSTRACT

Human modifications of environments are increasing, causing global changes that other species must adjust to or suffer from. Behavioral flexibility (hereafter ‘flexibility’) could be key to coping with rapid change. Behavioral research can contribute to conservation by determining which behaviors can predict the ability to adjust to human modified environments and whether these can be manipulated. When research that manipulates behavior in a conservation context occurs, it primarily trains a specific behavior to improve individual success in the wild. However, training a domain general cognitive ability, such as flexibility, has the potential to change a whole suite of behaviors, which could have a larger impact on influencing success in adjusting to human modified environments. This project asks whether flexibility can be increased by experimentally increasing environmental heterogeneity and whether such an increase can help species succeed in human modified environments. We explore whether it is possible to take insights from highly divergent species and apply them to address critical conservation challenges. This pushes the limits in terms of understanding how conserved these abilities may be and to what extent they can be shaped by the environment. We aim to 1) conduct flexibility interventions in flexible species that are successful in human modified environments (great-tailed grackles and California scrub-jays or blue jays) to understand how flexibility relates to success; and 2) implement these interventions in two vulnerable species (toutouwai and Florida scrub-jays) to determine whether flexibility as a generalizable cognitive ability can be trained and whether such training improves success in human-modified environments. This research will significantly advance our understanding of the causes and consequences of flexibility, linking behavior to environmental change, cognition, and success in human modified environments through a comparative and global framework. This registered report launches our reproducible research program, ManyIndividuals (<https://github.com/ManyIndividuals/ManyIndividuals>), which is a global network of researchers with field sites investigating hypotheses that involve generalizing across many individuals.

REGISTERED REPORT DETAILS

- **Level of bias = 6:** This registered report was written (Jul 2021-May 2022), and revised after two rounds of peer review at Peer Community in Registered Reports (Jul and Aug 2022) prior to collecting any data.
- **Programmatic registered report:** Three Stage 2 articles will result from this one Stage 1 registered report: one for toutouwai, one for grackles, and one for jays.
- **Deviations from the Stage 1 registered report:** [to be filled in as needed after data collection begins]

INTRODUCTION

Human modified environments are increasing (Goldewijk, 2001; X. Liu et al., 2020; Wu et al., 2011), causing global changes that other species must adjust to or suffer from (Alberti, 2015; Chejanovski et al., 2017; Ciani, 1986; Federspiel et al., 2017). Behavioral flexibility (hereafter ‘flexibility’) could be key for adjusting to such change: individuals interact with their environment through behavior, making it crucial to an ecologically valid understanding of how species adjust to environmental changes (Lee & Thornton, 2021). One of the top priorities for behavioral research to maximize conservation progress is to determine which cognitive abilities and behaviors can predict the ability to adjust to human modified environments and whether these can be manipulated (Moseby et al., 2016). The rare research that manipulates behavior in a conservation context usually focuses on training specific behaviors (for example, predator recognition through predator exposure) to improve individual success in the wild (Jolly et al., 2018; Moseby et al., 2012; Ross et al., 2019; West et al., 2018; see review in Tetzlaff et al., 2019). However, training a general cognitive ability, such as flexibility – the ability to rapidly adapt behavior to changes through learning throughout the lifetime (see the theory behind this definition in Mikhalevich et al., 2017) – has the potential to change a whole suite of behaviors and more broadly influence success in adjusting to human modified environments. Recent evidence supports this hypothesis: as far as we are aware, we were the first to show that flexibility can be manipulated using serial reversal learning of color preferences, and that the manipulated individuals were more flexible in a new context (locus switching on a puzzlebox) as well as being more innovative (solved more loci on a puzzlebox) (C. Logan et al., 2022).

Environments where informational cues about resources vary in a heterogeneous (but non-random) way across space and time are hypothesized to open a pathway for species to functionally detect and react to such cues via flexibility (Mikhalevich et al., 2017). Human modified environments likely provide a different set of informational cues that vary heterogeneously across space and time, and the species that are successful in such environments are likely those who are able to detect and track such cues. Because heterogeneous environments are hypothesized to select for flexibility (Wright et al., 2010), we expect that experimentally manipulating environments to be more heterogeneous will result in an increase in flexibility in individuals, which will then increase their success in such environments (Figure 1). Success can relate to any number of variables regarding the usage of and investment in resources and response to threats, from improved foraging efficiency to increased dispersal and survival within human modified environments, to placing nests in more protective locations. Whether a measure of success is predicted to relate to flexibility depends on what is already known about the particular population and their particular environment.



Figure 1. The theory behind this research illustrated by a directed acyclic graph (DAG), which is a theoretical model of the causal relationships among the key variables in our investigation. Based on the theoretical background provided by Mikhalevich et al. (2017), we assume that more heterogeneity causes more flexibility, which then causes more success in human modified environments.

This investigation asks whether flexibility can be increased by experimentally increasing environmental heterogeneity (via serial reversal learning) and whether such an increase can help species succeed in human modified environments. We explore whether it is possible to take insights from highly divergent species and apply them to address critical conservation challenges. Serial reversal learning tasks have been performed with a wide diversity of species (birds: Bond et al., 2007; bumblebees: Strang & Sherry, 2014; stingrays: Daniel & Schluessel, 2020). There is variation across individuals and species in their performance, however almost all previous studies show that individuals improve their flexibility if the reversal intervention is given multiple times in sequence (rats: Mackintosh et al., 1968; guppies: Lucon-Xiccato & Bisazza, 2014; poison frogs: Y. Liu et al., 2016). We aim to conduct a flexibility intervention in flexible species that are successful in human modified environments (great-tailed grackles and California scrub-jays or blue jays) to understand how flexibility relates to success, and implement these interventions in two vulnerable species (toutouwai and Florida scrub-jays) to determine whether flexibility as a generalizable cognitive ability can be trained and whether such training improves success in human modified environments (Figure 2).

While we do not examine the potential spread of the post-manipulation success behaviors from manipulated individuals to individuals that are not involved in our studies, we acknowledge that this is a possibility worthy of future investigation. Manipulating the flexibility of a few individuals could have population-level effects because significant research on social information use in birds (e.g., Valente et al., 2021) demonstrates the potential for the manipulated behavior to disseminate to conspecifics (for example, if manipulated individuals are faster at locating new resources, which could attract the attention of conspecifics, or if unmanipulated individuals copy the manipulated individuals' nesting or foraging locations). In the event that social learning is not used by a given population to spread the behaviors of manipulated individuals, investing in the training of specific individuals to increase their success in the wild could still have conservation impacts. In some cases, it is possible to train many individuals in a population or a species because there are not many individuals left [Greggor 2021pre]. It is also possible to train all individuals involved in a conservation management event such as a translocation [Greggor 2021pre]. Therefore, there can still be significant population consequences even if each individual needs to be trained to achieve the goal.

This comparative approach will ultimately reveal how conserved these abilities may be and to what extent they can be shaped by the environment. To increase the generalizability of the conclusions from the ManyIndividuals project, we here also provide multiple methodological options that other researchers can use to test these questions in additional species. The results will substantially advance our understanding of the causes and consequences of flexibility,

linking behavior to environmental change, cognition, and success in human modified environments through a comparative and global framework.

RESEARCH QUESTIONS

Can behavioral flexibility in individuals be increased by increasing environmental heterogeneity? If so, does increased flexibility help individuals succeed in human modified environments?

Prediction 1: Flexibility can be increased in individuals and such an increase **improves the likelihood of success in human modified environments**. This would indicate that the abilities involved in tracking changing resources in the environment are the same as or related to the abilities involved in succeeding in human modified environments. It would also indicate that flexibility is trainable and that such training could be a useful conservation tool for threatened and endangered species.

Prediction 1 alternative 1: Flexibility can be increased in individuals, but such an increase **does not improve the likelihood of success** in human modified environments. This would indicate that species associated with human modified environments form this association for reasons other than their flexibility, and that threatened species are likely not very successful in human modified environments for reasons unrelated to their ability to change their behavior with changing circumstances. An alternative could be that the changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success [changes in grackles were still present for four weeks after the manipulation and longer time periods were not attempted so the threshold is unknown @logan2022flexmanip].

Prediction 1 alternative 2: Flexibility can be increased in some populations, but not others. This would indicate that **flexibility manipulations may not work for all populations**, and that the effectiveness of such experiments should first be tested in the population of interest before including such an intervention in a conservation plan. If flexibility is not manipulatable in threatened populations, this would indicate that they are likely not very successful in human modified environments because of their inability to change their behavior with changing circumstances, and that flexibility is not trainable. If flexibility is not manipulatable in populations that are successful in human modified environments, this could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary [@wright2010behavioral]. In populations where flexibility is not manipulatable, this would indicate that the abilities involved in tracking changing resources in the environment are independent of the abilities involved in succeeding in human modified environments.

Population-specific background and tailored research questions

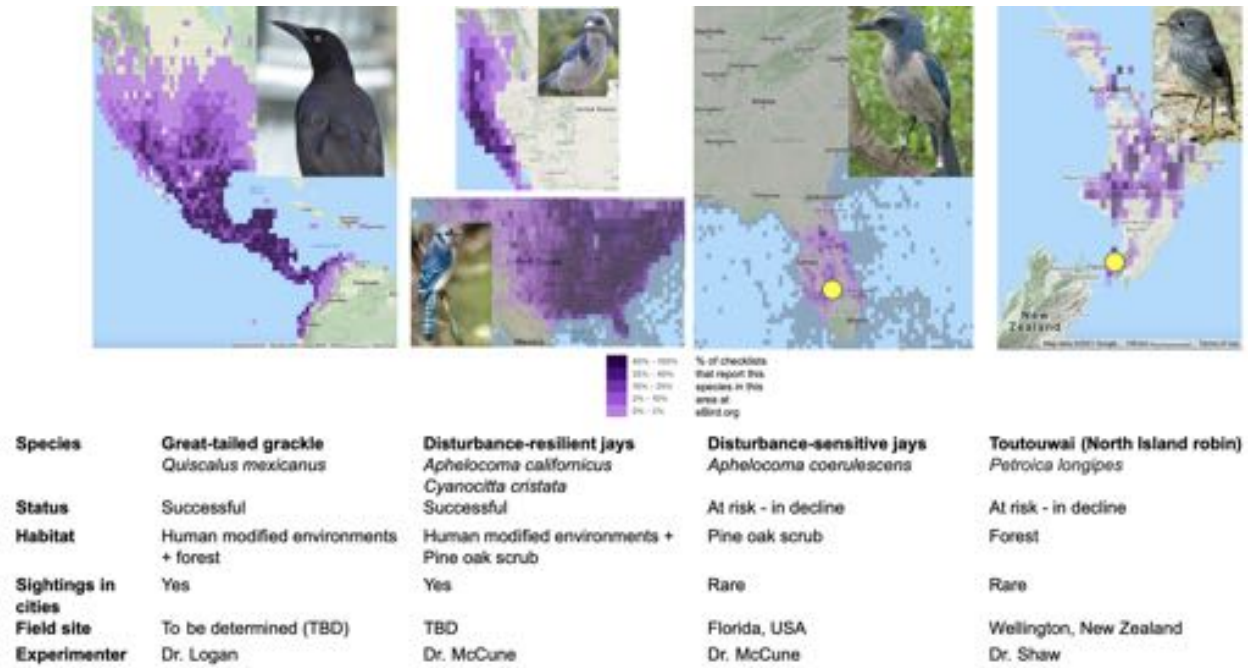


Figure 2. Comparing the species involved in this investigation relative to their geographic range and association with human modified habitats. The yellow dots represent field site locations. Photo credit: grackle and CASJ, Corina Logan; blue jay, Rhododendrites; FLSJ, VvAndromedavV; toutouwai, Rachael Shaw.

Great-tailed grackles (*Quiscalus mexicanus*)

Background

Great-tailed grackles are flexible (Logan, 2016; Logan, MacPherson, et al., 2019), highly associated with human modified environments (Johnson & Peer, 2001), and have been rapidly expanding their geographic range across North America over the past 140 years (Wehtje, 2003). They are social and polygamous, and eat a diversity of human foods as well as foraging on insects and on substrates for other natural food items (Johnson & Peer, 2001). Males tend to be the socially dominant sex (Johnson et al., 2000) and also the sex that disperses away from their natal area (Sevchik et al., 2019). Rodrigo et al. (2021) found that more grackles are present and more foraging events occur during garbage pick ups when garbage tends to spill out of the bags, thus increasing food availability. Attending to garbage trucks, potentially across space and time, is an example of how flexibility can help individuals meet foraging needs in the context of changing environmental cues. Great-tailed grackle behavioral flexibility is manipulatable using serial reversal learning, and this manipulation improves their flexibility in a

new context as well as their innovativeness (Logan, MacPherson, et al., 2019), which shows that training a general cognitive ability can affect more behaviors than the behavior that was trained and potentially make them more successful in human modified environments. We are in the middle of long-term data collection on grackles to answer questions about how flexibility relates to exploration (McCune KB et al., 2019), the predictability of their space use (McCune KB et al., 2020) and their foraging breadth (Logan, Lukas, et al., 2019).

Research questions

- **G.Q1: Do flexibility manipulated individuals differ in the proportion of time spent at cafes and garbage dumpsters when food is present?** We will investigate this question by tracking their presence at cafes and dumpsters when food is present versus when it is absent before and after manipulating their flexibility using serial reversal learning in the wild (Figure 3). Table 1 summarizes our predictions, analysis plans, interpretations for the various directions the results could go, and the hypotheses that could be contradicted given the various outcomes.
- **G.Q2: Does manipulating behavioral flexibility alter the number of microhabitats used?** We will investigate this question by tracking their presence in a variety of microhabitats before and after manipulating their flexibility using serial reversal learning in the wild. We only count that a microhabitat was used if the individual had at least 5% of their data points there. This prevents a microhabitat from being counted even if an individual was simply moving through it, and therefore not necessarily using it.
- **G.Q3: Does manipulating flexibility alter the number of different food items taken by grackles?** We will investigate this question by tracking the various food items they take before and after manipulating their flexibility using serial reversal learning in the wild.

Note: we may not have the time or personnel to collect data for G.Q2 and G.Q3, however we will attempt to answer these questions if possible.

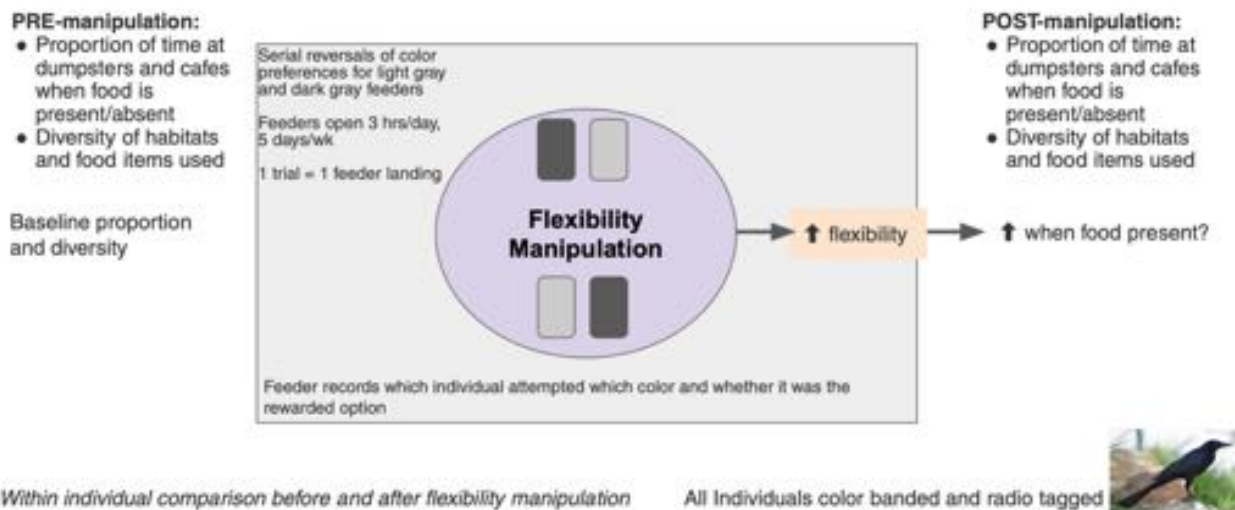


Figure 3. The reversal learning experiment in a group context (Design 2) tailored to the great-tailed grackle research question.

Table 1. Study design for the great-tailed grackle research. References that were not already cited in the introduction: Duckworth (2009), Dingemans & Wolf (2013), Grinnell (2017); Peterson (2011).

Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
<p>1. Do flexibility manipul ed individuals differ in the proportion of time spent at cafes and garbage dumpsters when food is present?</p>	<p>Prediction 1.1: There is an increase in the proportion of time spent at cafes and dumpsters when food is present after their flexibility has been manipulated relative to before the manipulation.</p>	<p>Simulations using bespoke Bayesian models showed a high likelihood of detecting differences with a sample size of 20 when the change in proportion between the before and after conditions is at least 0.1 (see Analysis Plan)</p>	<p>Bayesian model: Response at cafes and dumpster s when food present / total duration at cafes and dumpsters <u>Explanatory Condition (before/after)</u> <u>Random ID</u> (see Analysis Plan)</p>	<p>Contrasts will determine whether the before and after conditions differed from each other. We will conclude there is a difference if the confidence interval does not cross zero. False positives: the power analyses suggest that false positives are unlikely even with small sample sizes. Accordingly, we will interpret any contrast that does not cross zero as</p>	<p>The increase in proportion of time spent indicates that flexibility is involved in this foraging behavior. Environmental cues could signal the lack of food availability (e.g., reduced human presence, no visible food, no garbage trucks or garbage spilled on the ground). Flexibility is the functional tracking of resources across time and space: one knows a variety of places to obtain food and prefers to forage at those with higher pay offs. If no food is available at a usual food location, then the more functional choice is to go somewhere else that does have food.</p>	<p>Individual differences in the behavior are under selection (Duckworth 2009) and/or variation in behavior is caused by specific genetic variants (Dingemans & Wolf 2013), therefore behavior is not manipulatable within a short time period</p>
	<p>Prediction 1.2: There is no difference in the proportion of time spent in the presence of food at cafes and dumpsters between the before and after conditions.</p>				<p>This could indicate that the flexibility manipulation did not manipulate an ability that is linked with foraging at cafes and dumpsters. Alternatively, it could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary, or that changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success.</p>	<p>Flexibility facilitates adapting to environmental change (see Introduction)</p>

<p>Prediction 1.3: There is a decrease in the proportion of time spent in the presence of food at cafes and dumpsters after the manipulation relative to before.</p>		<p>indicating an effect. False negatives: the power analyses suggest that, especially with small sample sizes, we will not have sufficient power to exclude the possibility that an effect exists even though our model does not indicate an effect (contrast crosses zero).</p>	<p>This could indicate that the flexibility manipulation changed their behavior such that they spend more time seeking food at other or new locations.</p>	<p>NA</p>
<p>Prediction 1.4: If predictions 1.2 or 1.3 are supported, this could be due to the differences in the regularity of food availability between cafes (more predictable, less environmental heterogeneity) and dumpsters (less predictable, more environmental heterogeneity). We predict that the flexibility manipulation will increase the proportion of time spent in the presence of food at dumpsters more than at cafes.</p>		<p>Run the above model on only the cafe data and then only the dumpster data (see Analysis Plan)</p>	<p>Garbage pick up is generally once per week and lids can be sporadically left open. Cafes are generally open daily and people eat lunch outside daily during good weather. Because flexibility is linked with environmental heterogeneity, we will see a larger increase at dumpsters because of their increased environmental heterogeneity relative to cafes. It is possible that dumpster lid openings are not predictable, thus making it a random occurrence and there should be no link with flexibility.</p>	<p>If cafes and dumpsters differ: See P 1.1 If no difference: Environmental heterogeneity drives flexibility, therefore differences in the predictability of different types of food sources should coincide with differences in flexibility (Mikhalovich et al. 2017)</p>
<p>2. Habitat: Does manipulating behavior flexibility alter the number of microhabitats used?</p>	<p>Prediction 2.1: Flexibility can be increased and such an increase alters daily habitat use to include more variety of habitats</p>	<p>Simulation s using bespoke Bayesian models showed a high likelihood of detecting differences with a sample size of 20 when mean difference in the proportion of</p>	<p>Bayesian model: Response : Number of microhabitats used per individual Explanatory Condition (before/after) Random Condition IID (see Analysis)</p>	<p>(1) Habitat preferences and the foraging niche are fixed within species because each species evolves within a specific ecological niche (Grinnell 1917; Peterson et al. 2011). Behavioral flexibility facilitates the use of novel habitats and invasion success through dietary generalism (Sol et al. 2002)</p>
<p>Prediction 2.2: Flexibility can be increased and such an increase alters daily habitat use to decrease the variety of habitats used</p>	<p>Prediction 2.3: Flexibility can be increased but has no effect on the variety of habitats used</p>	<p>Increasing behavioral flexibility potentially leads to increased foraging breadth or use of resources within one habitat, rather than leading to sampling across habitat types.</p>	<p>This suggests that the cognitive ability behavioral flexibility may not generalize to all domains (e.g. may relate to foraging but not habitat use). Alternatively, it could indicate</p>	<p>(2) Flexibility facilitates adapting to environmental change (see Introduction) (3) Behavioral flexibility is a</p>

		<p>microhabitats used was at least 0.1 (standard deviation= 0.1) or 0.15 (SD=0.2) (see Analysis/Plan)</p>	<p>Plan)</p>		<p>that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary, or that changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success.</p>	<p>general cognitive ability (see Introduction)</p>
<p>3. Foraging: Does manipulating flexibility alter the number of different food items taken by grackles?</p>	<p>Prediction 3.1: Flexibility can be increased and such an increase alters daily foraging breadth to include more variety of food items</p> <p>Prediction 3.2: Flexibility can be increased and such an increase alters daily foraging breadth to decrease the variety of food items taken</p>	<p>Simulations using bespoke Bayesian models showed a high likelihood of detecting differences with a sample size of 20 when mean differences in the number of foods taken were at least 1 (and a standard deviation of 2) (see Analysis/Plan)</p>	<p>Bayesian model: Response - Number of foods taken per individual</p> <p>Explanatory: Condition (before/after)</p> <p>Random: Condition IID</p> <p>(see Analysis/Plan)</p>		<p>Increasing behavioral flexibility, with a serial reversal learning manipulation, increases the likelihood the individual will sample new food sources while foraging.</p> <p>Increasing behavioral flexibility potentially leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat.</p>	<p>(1)</p> <p>Behavioral flexibility facilitates the use of novel foods through sampling new foods and foraging strategies (Sol et al. 2002)</p>
	<p>Prediction 3.3: Flexibility can be increased but has no effect on foraging breadth</p>				<p>This suggests that the cognitive ability behavioral flexibility may not generalize to all domains (e.g. may relate to habitat use but not foraging breadth). Alternatively, it could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary, or that changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success.</p>	<p>(2), (3)</p>

Jays (*Aphelocoma californica* and/or *A. coerulescens* and/or *Cyanocitta cristata*)

We are not yet sure which jay species will be feasible to test as part of this investigation. We aim to test a disturbance resilience species like the California scrub-jay or blue jay. We also aim to test the disturbance sensitive Florida scrub-jay, but we are still waiting for permits on the Florida scrub-jay and blue jay. Regardless of the species that end up in the investigation, we will use the same hypotheses and predictions below.

Background

Jay species exhibit a diversity of social systems and success in colonizing suburban and urban areas. California scrub-jays (*Aphelocoma californica*, hereafter “CASJ”) and blue jays (*Cyanocitta cristata*, hereafter “BLJA”) are singular, monogamous breeders that are increasing in abundance, expanding their range sizes, and highly successful in natural, suburban, and urban areas (Blair, 1996; Curry et al., 2017). We therefore consider these “disturbance-resilient” (DR) jay species. In contrast, the Florida scrub-jay (*A. coerulescens*; hereafter “FLSJ”) is a “disturbance-sensitive” (DS) jay species that is threatened, endemic, and range-restricted to xeric oak scrub habitat in Florida (Woolfenden & Fitzpatrick, 1996).

These species forage primarily on mast (acorns, hazelnuts, etc.) that they cache throughout their territory, which makes it available to eat year-round. They are also opportunistic omnivores and specifically need high-fat and high-protein arthropods to feed to nestlings and fledglings (Curry et al., 2017). Nesting and foraging substrates can be drastically different in human modified environments compared to natural areas ([e.g. predominance of non-native vegetation](#); Tuomainen & Candolin, 2011), and it is unknown whether suburban and urban jays are able to persist in these environments through behavioral adjustments. The DS jay species, the FLSJ, can persist in suburban habitats after conversion from xeric oak scrub, however suburban populations of FLSJ steadily decline (Bowman pers. comm.). This is potentially due to the presence of suboptimal habitat resulting from fire suppression (Woolfenden & Fitzpatrick, 1996), higher rates of brood reduction through nestling starvation (Shawkey et al., 2004), and the lack of nutritionally complete prey items (Shawkey et al., 2004) in suburban habitats. It is possible that behavioral flexibility in habitat use and foraging breadth underlies the ability of some FLSJ to persist in human-dominated areas.

We aim to compare behavioral flexibility within species, between suburban and natural populations to determine whether variation in flexibility relates to variation in presence in these habitats. Subsequently, we will compare flexibility between DS and DR jay species to determine whether this trait is related to the greater success of DR jay species, like the CASJ and BLJA, in human-dominated areas. Lastly, we will test whether manipulating flexibility increases the foraging and microhabitat breadth of jays in human modified environments. Manipulating the flexibility of a subset of individuals has the potential to affect the population because previous research demonstrates that both species have the capacity to use foraging information discovered by others (social learning) to flexibly change their behavior (K. B. McCune, 2018; Midford et al., 2000).

While we have begun to acquire permits to work with these species, the processing time for permits to research threatened and endangered species can be very long and unpredictable. Therefore, while we aim to collect data on the FLSJ during the 5-year timeline of this project, it is possible this species will have to be omitted if the permits are not approved in time. If this

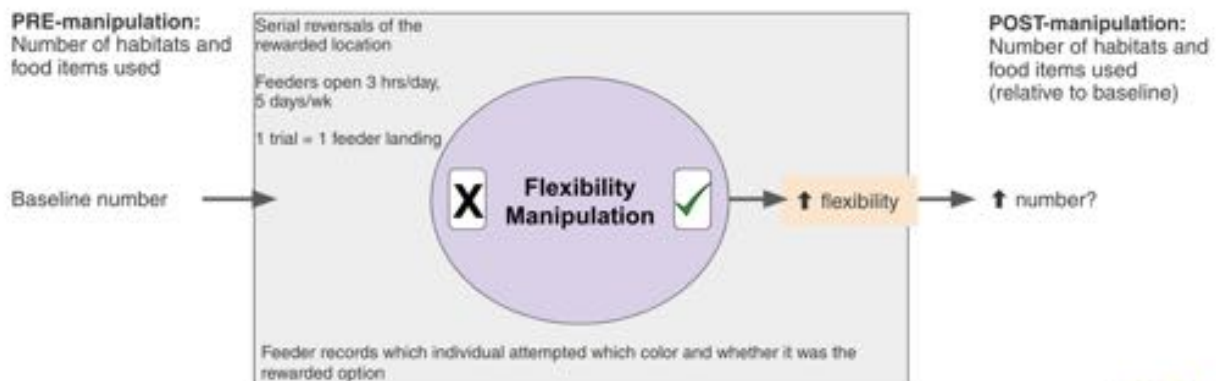
occurs, we will not be able to compare CASJ/BLJA and FLSJ (Question 2 in Table 2), but we will still be able to compare behavioral flexibility of CASJ/BLJA populations in human modified and natural areas, and assess whether the behavioral flexibility manipulation affects success in human modified environments.

Research questions

For all research questions, Table 2 summarizes our predictions, analysis plans, interpretations for the various directions the results could go, and the hypotheses that could be contradicted given the various outcomes.

- **J.Q1: Do jay populations in human modified areas differ in baseline behavioral flexibility compared to populations in natural areas?** We will investigate this question by comparing performance on serial reversal learning in the wild between jays in natural areas and jays in human modified areas.
- **J.Q2: Are disturbance-resilient (DR) jays more behaviorally flexible than disturbance-sensitive (DS) jays?** We will investigate this question by comparing performance on serial reversal learning in the wild ~~between~~ between DR and DS jay species.
- **J.Q3: Does manipulating behavioral flexibility alter the number of microhabitats used?** We will investigate this question by tracking their presence in a variety of microhabitats before and after manipulating their flexibility using serial reversal learning in the wild. We only count that a microhabitat was used if the individual had at least 5% of their data points there. This prevents a microhabitat from being counted even if an individual was simply moving through it, and therefore not necessarily using it.
- **J.Q4: Does manipulating flexibility alter the number of different food items taken by jays?** We will investigate this question by tracking the various food items they take before and after manipulating their flexibility using serial reversal learning in the wild.

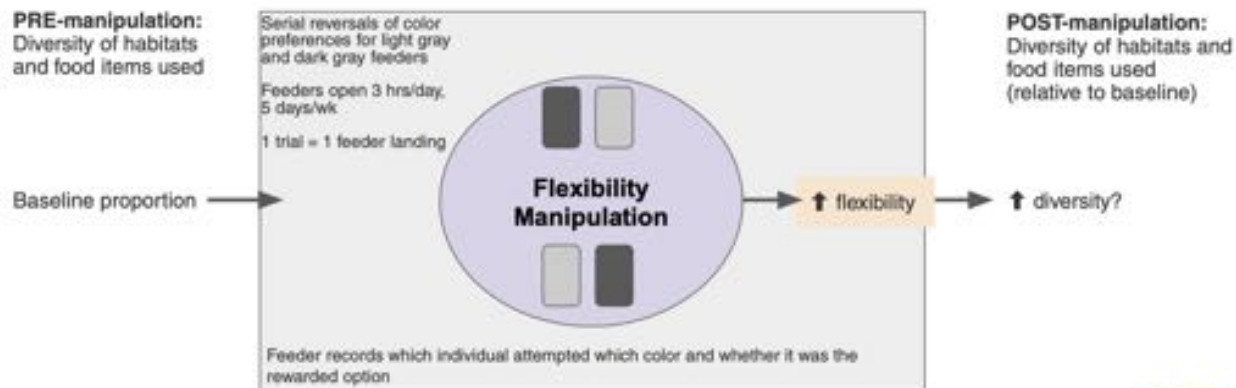
Do flexibility manipulated individuals differ in how they interact with their environment?



Within individual comparison before and after flexibility manipulation

All individuals color banded and radio tagged





Within individual comparison before and after flexibility manipulation

All individuals color banded and radio tagged

Figure 4. The reversal learning experiment in a group context (Design 2) tailored to the jay research questions. The white rectangles represent feeder locations, the feeder with the X is in the unrewarded location while the feeder with the green check is the rewarded location.

Table 2. Study design for the jay research. References that were not already cited in the introduction: Galbraith et al. (2015); Lapedra et al. (2017); Rice et al. (2003); Emery & Clayton (2004); Sol et al. (2002).

Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding test sensitivity to confirm/disconfirm hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes					
1. Do jay populations in human modified areas differ in baseline behavioral flexibility compared to populations in natural areas?	<p>Prediction 1.1: Suburban jays are more flexible than jays in natural areas</p> <p>Prediction 1.2: Suburban jays are less flexible than jays in natural areas</p> <p>Prediction 1.3: There is no difference in flexibility between suburban jays and jays in natural areas</p>	<p>Simulations using bespoke Bayesian models in Logan et al. (2021) showed a high likelihood of detecting differences with a sample size of 15 when mean differences in phi were at least 0.01 and lambda at least 3 (see <i>Analysis Plan</i>)</p>	<p>Bayesian model: <u>Response:</u> phi and lambda</p> <p><u>Explanator Y:</u> Habitat (suburban/natural)</p> <p>(see <i>Analysis Plan</i>)</p>	<p>Contrasts will determine whether the before and after conditions differed from each other. We will conclude there is a difference if the confidence interval does not cross zero.</p> <p>False positives: the power analyses suggest that false positives are unlikely even with small sample sizes. Accordingly, we will interpret any contrast that does not cross zero as indicating an effect.</p> <p>False negatives: the power analyses suggest that, especially with small sample sizes, we will not have</p>	<p>This implies that flexibility is related to the ability to occupy human modified environments where spatial and temporal heterogeneity of resources is high.</p> <p>This implies that human modification of the environment has led to less spatial and temporal heterogeneity of resources. For example, the prevalence of bird feeders in suburban areas leads to consistently available food.</p> <p>This implies that additional behavioral (e.g. boldness, Lapedra et al., 2017) or genetic traits may facilitate success in human modified environments.</p>	<p>Selection for exploitation of supplementary food (Galbraith et al. 2015) where individuals learn to depend on anthropogenic food sources and are less likely to flexibly sample alternative resources</p> <p>(1) Flexibility facilitates adapting to environmental change (see Introduction)</p>					
							<p>Prediction 2.1: DR jays are more flexible than DS jays</p>	<p>Bayesian model: <u>Response:</u> phi and lambda</p>	<p>False negatives: the power analyses suggest that, especially with small sample sizes, we will not have</p>	<p>This difference may explain the range expansion and greater success of DR jays in human modified environments.</p>	<p>Range expansion instead relates to ecological niche</p>

behaviorally flexible than disturbance sensitive (DS) jays) ?			<p>Explainer Y: Species (see Analysis Plan)</p>	<p>sufficient power to exclude the possibility that an effect exists even though our model does not indicate an effect (contrast crosses zero).</p>	<p>This implies that flexibility is not related to success in human modified environments and that flexibility may instead be related to a different, unknown social or environmental characteristic. For example, the cooperative breeding system of the DS jay species, the Florida scrub-jay, may favor increased flexibility for responding to group mates' behavior.</p>	<p>differentiation (Rice et al. 2003) where DR jays evolved to occupy a niche that more closely resembles human modified environments than DS jays</p>
	<p>Prediction 2.2: DR jays are less flexible than DS jays</p>					
<p>3. Habitat: Does manipulating behavioral flexibility alter the number of microhabitats used?</p>	<p>Prediction 3.1: Flexibility can be increased and such an increase alters daily habitat use to include more variety of habitats</p>	<p>Simulations using bespoke Bayesian models showed a high likelihood of detecting differences with a sample size of 20 when mean difference in the proportion of microhabitats used was at least 0.1 (standard)</p>	<p>Bayesian model: Response: Number of microhabitats used per individual Explainer Y: Condition (before/after) Random: Condition ID (see</p>	<p>"</p>	<p>Increasing behavioral flexibility, with a serial reversal learning manipulation, increases the likelihood the individual will sample new areas while foraging.</p>	<p>(2) Habitat preferences and the foraging niche are fixed within species because each species evolves within a specific ecological niche (Grinnell 1917; Peterson et al. 2011).</p>

<p>Prediction 3.2: Flexibility can be increased and such an increase alters daily habitat use to decrease the variety of habitats used</p>	<p>deviation=0.11 or 0.15 (SD=0.2) <i>(see Analysis Plan)</i></p>	<p><i>Analysis Plan)</i></p>	<p>Increasing behavioral flexibility potentially leads to increased foraging breadth or use of resources within one habitat, rather than leading to sampling across habitat types.</p>	<p>Behavioral flexibility facilitates the use of novel habitats and invasion success through dietary generalism (Sol et al. 2002)</p>	
<p>Prediction 3.3: Flexibility can be increased but has no effect on the variety of habitats used</p>				<p>This suggests that the cognitive ability behavioral flexibility may not generalize to all domains (e.g. may relate to foraging but not habitat use). Alternatively, it could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary, or that changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success.</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see Introduction)</p>
<p>4. Foraging: Does manipulating flexibility alter the number of different food items taken by jays?</p>	<p>Prediction 4.1: Flexibility can be increased and such an increase alters daily foraging breadth to include more variety of food items</p>	<p>Simulations using bespoke Bayesian models showed a high likelihood of detecting differences with a sample size of 20 when mean differences in the number of foods taken were at least 1 (and a standard deviation of 2) <i>(see Analysis Plan)</i></p>	<p>Bayesian model: <u>Response</u>: Number of foods taken per individual <u>Explanator</u> Y: Condition (before/after)</p>	<p>Increasing behavioral flexibility, with a serial reversal learning manipulation, increases the likelihood the individual will sample new food sources while foraging.</p>	<p>(2)</p>
<p>Prediction 4.2: Flexibility can be increased and such an increase alters daily foraging breadth to decrease the variety of food items taken</p>		<p><u>Random</u>: Condition ID <i>(see Analysis Plan)</i></p>	<p>Increasing behavioral flexibility potentially leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat.</p>	<p>Behavioral flexibility facilitates the use of novel foods through sampling new foods and foraging strategies (Sol et al. 2002)</p>	
<p>Prediction 4.3: Flexibility can be</p>			<p>This suggests that the cognitive ability behavioral flexibility may not generalize to all domains (e.g.</p>	<p>(1) (3)</p>	

	increased but has no effect on foraging breadth				may relate to habitat use but not foraging breadth). Alternatively, it could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary, or that changes induced by the increase in flexibility do not persist for sufficiently long times to make a difference on the subsequent likelihood of success.	
--	--	--	--	--	---	--

Table 2. Study design for the jay research. References that were not already cited in the introduction: Galbraith et al. (2015);

Lapedra et al. (2017); Rice et al. (2003); Emery & Clayton (2004); Sol et al. (2002); Grinnell (2017); Peterson (2011).

Question	Hypothesis	Sampling plan	Analysis plan	Statistical test	Interpretation given different outcomes	Theory that would be driven strong by the outcomes
1. Do JPV populations in human modified areas differ in baseline behavioral flexibility compared to populations in natural areas?	<p>Prediction 1.1: Suburban jays are more flexible than jays in natural areas.</p> <p>Prediction 1.2: Suburban jays are less flexible than jays in natural areas.</p> <p>Prediction 1.3: There is no difference in flexibility between suburban jays and jays in natural areas.</p>	<p>Randomized using multiple flycatcher trappers in 10 urban and 10 rural areas of (2017) showed a high baseline of flexibility.</p> <p>Observation with a sample size of 15 urban and 15 rural jays in 10 areas of each (2017) and results of base 1.</p> <p>(see Analysis Plan)</p>	<p>Bayesian model (MCMC) fit and model comparison.</p> <p>Model selection (Bayesian model)</p> <p>(see Analysis Plan)</p>	<p>Correlations will indicate whether the behavior and other variables affect each other.</p> <p>There is a difference in the correlation between the variables.</p> <p>There is a difference in the correlation between the variables.</p> <p>There is a difference in the correlation between the variables.</p>	<p>The model that flexibility is related to the ability to occupy human modified environments when spatial and temporal heterogeneity of resources is high.</p> <p>The model that human modification of the environment has led to more spatial and temporal heterogeneity of resources. For example, the presence of bird feeders in suburban areas leads to consistently available food.</p> <p>The model that additional resources (e.g., birdseed, garbage) at a site or greater leads to higher success in human modified environments.</p> <p>The difference only explains the large expansion and greater success of CDE jays in human modified environments.</p>	<p>Support for expansion of rudimentary tool (CDE) at a site, where individuals learn to depend on anthropogenic food resources and are less likely to build simple alternative resources.</p> <p>(1) flexibility facilitates adapting to environmental change (see introduction)</p> <p>(2) The urban Star Jay (Sjostedt et al. 2017) shows more anthropogenic pressures select for flexible individuals.</p> <p>Large expansion toward urban to ecological niche differentiation (Rice et al. 2020), where CDE jays evolved to occupy a niche that more closely resembles human modified environments than CDE jays.</p>
2. Are distributions amongst CDE jays more substantially flexible than distributions amongst CDE jays 1?	<p>Prediction 2.1: CDE jays are more flexible than CDE jays.</p> <p>Prediction 2.2: CDE jays are less flexible than CDE jays.</p> <p>Prediction 2.3: CDE jays and CDE jays are equally flexible.</p>	<p>Randomized using multiple flycatcher trappers in 10 urban and 10 rural areas of (2017) showed a high baseline of flexibility.</p> <p>Observation with a sample size of 15 urban and 15 rural jays in 10 areas of each (2017) and results of base 1.</p> <p>(see Analysis Plan)</p>	<p>Bayesian model (MCMC) fit and model comparison.</p> <p>Model selection (Bayesian model)</p> <p>(see Analysis Plan)</p>	<p>Correlations will indicate whether the behavior and other variables affect each other.</p> <p>There is a difference in the correlation between the variables.</p> <p>There is a difference in the correlation between the variables.</p> <p>There is a difference in the correlation between the variables.</p>	<p>The model that flexibility is not related to success in human modified environments and that flexibility may instead be related to a different, unknown social or environmental characteristic. For example, the cooperation flexibility of the CDE jays species, the Florida scrub jay, may have increased flexibility for responding to great winter weather.</p> <p>The model that flexibility is not related to success in human modified environments and the level of flexibility is primarily an evolutionary constraint that limits a novel human ancestor (Emery & Clayton 2004).</p>	<p>(1)</p> <p>(2)</p>

Table 2. Continued

<p>1. Market: Does manipulating behavioral flexibility alter the number of individuals used?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily habitat use to include more variety of habitats</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of individuals used per individual</p>		<p>increasing behavioral flexibility with a novel resource starting population, increases the likelihood the individual will sample new areas within foraging</p>	<p>(2) Habitat preferences and the foraging niche are fixed within species because each species evolves within a specific ecological niche (Stenseth 1972, Pevzner et al. 2017)</p>
<p>2. Foraging: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>Behavioral flexibility reinforces the use of novel habitats through sampling new foods and foraging strategies (Jain et al. 2022)</p>
<p>3. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily habitat use to include more variety of habitats used</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of individuals used per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>
<p>4. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>
<p>5. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>
<p>6. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>
<p>7. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>
<p>8. Market: Does manipulating behavioral flexibility alter the number of different food items taken by lags?</p>	<p>Population 1, 2, 7 flexibility can be increased and each an increase alters daily foraging breadth to include more variety of food items taken</p>	<p>Individuals using multiple habitats increase. However, a high standard of flexibility</p>	<p>Statistical: Number of food items per individual</p>		<p>increasing behavioral flexibility generally leads to increased use of the same type of food items across habitat types, rather than leading to sampling more food items within one habitat</p>	<p>(1) (3) Behavioral flexibility is a general cognitive ability (see introduction)</p>

Toutouwai (North Island robin, *Petroica longipes*)

Background

Toutouwai are a small insectivorous passerine species that is endemic to the North Island of New Zealand. Prior to the arrival of European settlers, they were abundant and widespread. However, European colonization brought the introduction of mammalian predators, which had devastating consequences for New Zealand's avifauna. As a result, the national conservation status of the toutouwai is now at risk and in decline (Robertson et al., 2016). In response to avifauna declines in New Zealand, many offshore islands have now been established as predator free sanctuaries. Threatened endemic birds have also been reintroduced into 'mainland island' sanctuaries; areas on New Zealand's North and South Islands that are designed to eliminate the threat posed by non-native mammalian predators (Saunders & Norton, 2001). Yet, despite predator control and fences protecting sanctuary populations, these vulnerable species still come into contact with invasive mammalian predators beyond mainland sanctuary boundaries.

At Zealandia, a 225 hectare predator-proof fenced sanctuary located in central Wellington, less than 20 years have passed since the introduction of toutouwai (Miskelly et al., 2005). This species now thrives inside the sanctuary with some individuals surviving up to 14 years, but is struggling to gain a foothold in the surrounding reserves. Toutouwai may undergo juvenile dispersal from 5 weeks post-fledging onwards, traveling anywhere from a few meters to several kilometers from their natal territory (Richard & Armstrong, 2010). Many juvenile toutouwai are sighted establishing territories outside of the sanctuary each year, yet few persist for more than a few weeks (Shaw & Harvey, 2018). One potential reason for the failure to persist, is that toutouwai tend to forage on the ground and are thus at high risk from invasive mammalian predators, which they fail to recognize. It is possible that the more flexible individuals that disperse outside the sanctuary might forage at a diversity of heights and/or more readily learn to recognize novel predators, and therefore have a higher likelihood of surviving post-dispersal, but currently this hypothesis is untested.

Previous research on the cognitive abilities of toutouwai reveals that this species can retain a learned skill for several months in the wild, without reinforcement (Shaw & Harvey, 2020). Moreover, their bold and curious nature means that they will interact with novel objects and readily engage in a suite of cognitive tests (Shaw et al., 2015). As such, they are an ideal species in which to examine whether flexibility manipulations might influence the dispersal decisions made by juveniles, or enhance the survival of juveniles that attempt to establish beyond the sanctuary fence.

Research questions

For all research questions, Table 3 summarizes our predictions, analysis plans, interpretations for the various directions the results could go, and the hypotheses that could be contradicted given the various outcomes.

- **T.Q1: Does a flexibility manipulation alter dispersal timing and distance, as well as the likelihood that juvenile toutouwai will disperse beyond the protection of Zealandia's fence and attempt to establish in the adjacent urban reserves?** We will investigate this question by measuring their age at dispersal, dispersal distance, and habitat dispersed to after either manipulating their flexibility using serial reversal learning in the wild (manipulated group) or not manipulating their flexibility by giving them only

- one reversal rather than serial reversals (control group).
- **T.Q2: Compared to control individuals, are flexibility manipulated individuals more likely to survive their first 16 weeks post-fledging, particularly if they disperse into the urban reserves outside the sanctuary fence?** We will investigate this question by tracking their dispersal destination after either manipulating their flexibility using serial reversal learning in the wild (manipulated group) or not manipulating their flexibility by giving them only one reversal rather than serial reversals (control group).

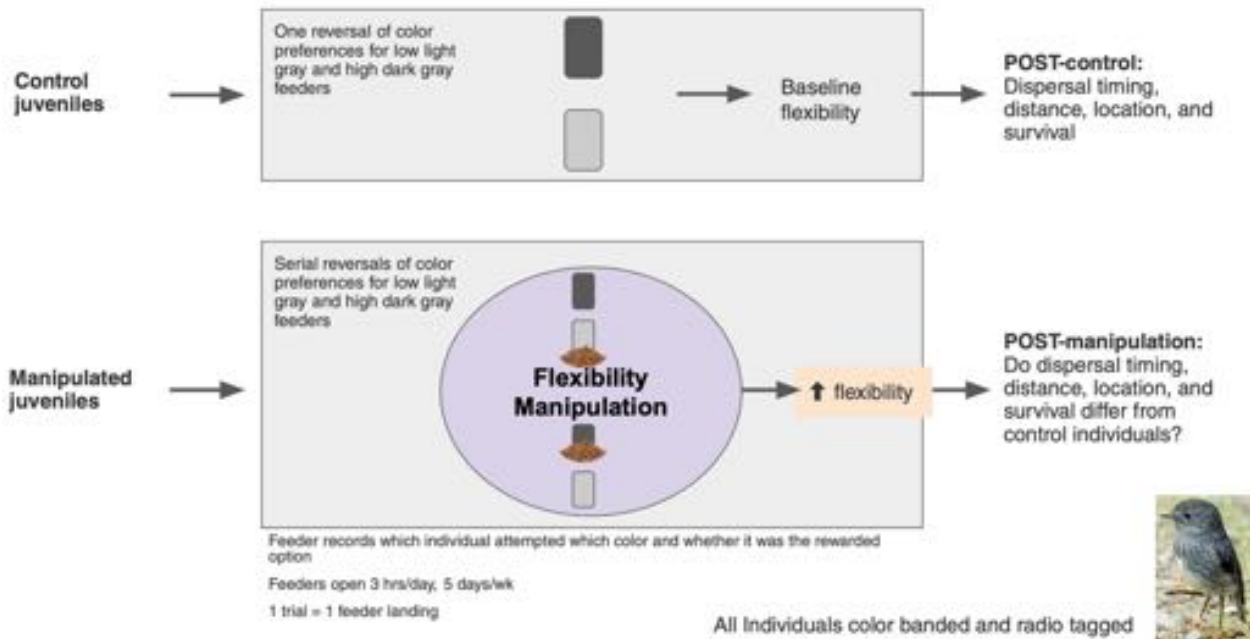


Figure 5. The reversal learning experiment in a group context (Design 2) tailored to the toutouwai research questions.

Table 3. Study design for the toutouwai research. References that were not already cited in the introduction: Pasinelli et al. (2004); Howard (1960).

Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
1. Does a flexibility manipulation alter dispersal timing and distance, as well as the likelihood that juvenile toutouwai will disperse beyond the protection of Zealandia's fence and attempt to establish in the adjacent urban area?	Prediction 1.1: Flexibility manipulated individuals disperse further, are more likely to go into suburban habitats and leave the natal territory earlier relative to control individuals.	Simulations using bespoke Bayesian models showed a high likelihood of detecting differences between conditions with a minimum sample size of:	Bayesian models all contain the <u>explanatory</u> variable: Condition	This implies that flexibility is related to the ability to disperse into human modified environments, potentially because individuals are quick to update their behavioral response to novel foods and habitats encountered outside of the natal territory.	(1) Dispersal patterns are innate (Howard 1960; Pasinelli et al. 2004), where individuals inherit the distance, duration, and movement pattern of dispersal from parents.	
						<p>Prediction 1.2: Flexibility manipulated individuals disperse closer, are less likely to move into suburban habitats and leave the natal territory later than control individuals</p> <p>Prediction 1.3: Flexibility manipulated individuals do not alter the pattern and/or timing of dispersal relative to control individuals</p> <p>Prediction 1.4: Flexibility cannot be increased in individuals using serial</p>
					<p>This implies that another trait (e.g. genetics; Pasinelli et al. 2004) governs dispersal motivation and distance.</p>	(2) Flexibility facilitates adapting to environmental change (see Introduction)
					<p>This indicates either flexibility is not manipulatable or the effects of the manipulation did not last long enough to be detected. Other</p>	(2) (3) The urban filter (Lapiedra et al. 2017), where novel anthropogenic

	reversal learning (i.e., few or none pass the serial reversal criterion)	40 when the difference is at least 0.3 (see Analysis Plan)		experiments would need to be conducted to determine whether flexibility, or behavior in general, is manipulatable at all in this population	pressures select for flexible individuals
--	--	---	--	---	---

Table 3 continued

2. Compared to control individuals, are flexibility manipulated individuals more likely to survive their first 16 weeks post-fledging, particularly if they disperse into the urban area outside the sanctuary fence?	<p>Prediction 2.1: Flexibility manipulated individuals have a higher likelihood of survival in general and in human modified environments relative to control individuals</p>	<p>Simulations using bespoke Bayesian models showed a high likelihood of detecting differences with a minimum sample size of 30 when the change in proportion between the control and manipulated conditions is at least 0.3</p>	<p>Bayesian model: Condition Response</p> <p>Survival at 112 days post fledging</p> <p><u>Explanatory</u>: Condition</p> <p><u>Habitat</u>: Survival at 112 days post fledging</p>	<p>Contrasts will determine whether the before and after conditions differed from each other. We will conclude there is a difference if the confidence interval does not cross zero.</p> <p>False positives: the power analyses suggest that false positives are unlikely even with small sample sizes. Accordingly, we will interpret any contrast that does not cross zero as</p>	<p>This would indicate that the abilities involved in tracking changing resources in the manipulation task are the same as or related to the abilities involved in succeeding in human modified environments. It would also indicate that flexibility is trainable and that such training could be a useful conservation tool for threatened and endangered species.</p>	<p>Adaptation to environmental change occurs through genetic variation and not behavior (Barton & Partridge 2000), therefore behavior is not manipulatable within a short time period</p>
	<p>Prediction 2.2: Flexibility manipulated individuals do not have a higher likelihood of survival in general and in human modified environments relative to control individuals</p>	<p>Differences between survival of control and manipulated individuals in natural vs. urban habitats will not be detectable</p>	<p><u>Explanatory</u>: Condition + Dispersal habitat (see Analysis Plan)</p>	<p>This would indicate that species associated with human modified environments form this association for reasons other than their flexibility, and that threatened species are likely not very successful in human modified environments for reasons unrelated to their ability to change their behavior with changing circumstances.</p>	<p>(2) (3)</p>	<p>This would indicate that flexibility manipulations may not work for all populations, and that the effectiveness of such experiments should first be tested in the population of interest before including such an intervention in a conservation plan. If flexibility is not manipulatable in threatened</p>
<p>Prediction 2.3: Flexibility cannot be increased in individuals using serial reversal learning (i.e., few or none pass the serial reversal criterion)</p>						

		(see Analysis Plan)	<p>indicating an effect.</p> <p>False negatives: the power analyses suggest that, especially with small sample sizes, we will not have sufficient power to exclude the possibility that an effect exists even though our model does not indicate an effect (contrast crosses zero).</p>	<p>populations, this would indicate that they are likely not very successful in human modified environments because of their inability to change their behavior with changing circumstances, and that flexibility is not trainable. If flexibility is manipulatable in populations that are successful in human modified environments, this could indicate that they might have used flexibility in the past when originally forming the association, but the need to maintain flexibility in their repertoire is no longer necessary. In populations where flexibility is not manipulatable, this would indicate that the abilities involved in tracking changing resources in the environment are independent of the abilities involved in succeeding in human modified environments.</p>	
--	--	---------------------	---	---	--

METHODS

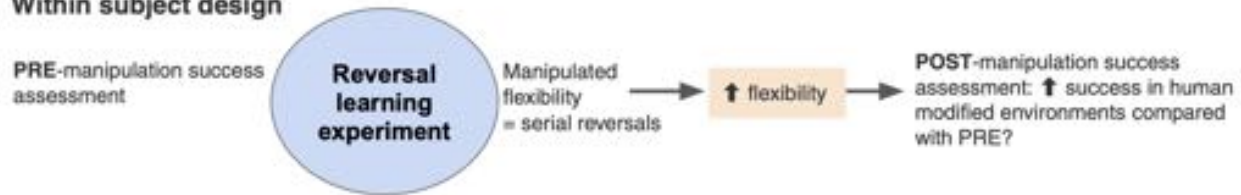
Our approach involves individuals participating in a serial reversal learning task and measuring success in natural behavior in the wild. In the following, we outline the procedure we are planning to use with wild birds: the serial reversal learning will involve a feeder setup (see design 2 below), and the success measures will be matched to the relevant behavior in the respective population. The framework we provide is general to facilitate its adaptation and replication to other populations (i.e., another researcher could adapt our hypotheses, methods, and analysis plans to their system). We start each section of the methods and analysis plans with general considerations (e.g., minimum sample size), before providing the specific details for each of the systems we plan to study.

We present one experimental design for the flexibility intervention (Figure 6) that can be conducted in two ways: in visual isolation (design 1) and in a group context (design 2). Experimenters can decide which (or both) they want to conduct in their population. Only one experiment must be conducted per population to be able to test these hypotheses. Conducting more than one of these experiments per population is acceptable, but not necessary. Before we present the designs, we first validated the reversal passing criteria and made them generalizable to a variety of species (see the next two sections). Depending on the response variable, there is the option to conduct a within- or between-subjects design:

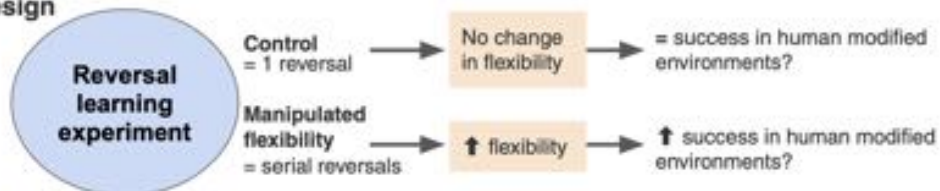
1. **Within-subjects:** run the manipulation on all individuals and compare pre- and post-manipulation success measures
2. **Between-subjects:** with manipulated and control groups and compare post-manipulation success measures. In this case, 50% of the individuals will be assigned to the control condition and 50% to the flexibility manipulation condition. Assignment to condition will be random (using the random number generator random.org).

Template reversal learning experiment: do manipulated individuals do more/less of X in human modified environments?

Within subject design



Between subjects design



- Individuals are identifiable and tracked throughout

Figure 6. Template design for the reversal learning experiment.

Determining when to switch each individual to the next reversal: reversal passing criterion

Different criteria exist to decide whether an individual has learned an association between the presence of a reward and some other feature (e.g., color or shape). The two main two criteria used are to switch an individual after it either has chosen 10 out of 12 choices correct (e.g., Shaw et al., 2015) or 17 out of 20 choices correct (e.g., Logan, 2016). The criteria are further modified depending on whether choices are assessed continuously or grouped in predetermined blocks.

Here, we assess whether achieving 10 correct choices out of the last 12 continuously counted choices can be used as a reliable reversal passing criterion. To determine reliability and suitability, we investigated five questions (see below) by generalizing previously simulated reversal learning data from Logan CJ et al. (2020), based on data from great-tailed grackles. We simulated the choices individuals with different learning rates (ϕ) and rates of deviating from learned associations (λ) would make in the initial discrimination and in the first reversal. Grackles are fast to reverse preferences compared with many other species (Logan, 2016), therefore we generalized the simulations to other species by setting the parameters that guide performance (ϕ and λ) to lead to slower performances.

The findings from these simulated data indicate that deciding that an individual has passed the reversal when they choose 10 out of the last 12 consecutive trials correctly is functional and reliable because of the following:

1) individuals will be finished after fewer trials than with other criteria

With the 10 out of 12 criterion, individuals pass the reversal 8 trials faster (median) than with the

17 out of 20 criterion. This means that, for most individuals, the two rules are equally effective because they will pass both in the same amount of trials (i.e., the individual who met the 17/20 criterion in 50 trials would have met the 10/12 criterion in 42 trials), but because the 10 out of 12 criterion is restricted to 12 trials instead of 20, individuals need 8 fewer trials to meet the passing criterion. No individual needs more trials with the 10 out of 12 criterion. When trials are grouped into blocks of 10 such that they could only pass on trial 20, 30, etc., individuals need a median of 5 more trials compared to when choices are assessed continuously.

2) classification of individuals using the 10/12 criterion is less noisy because there is less of a chance for individuals to approach the criterion and not pass or never pass

The average improvement in the number of trials individuals need to reach the respective criterion is larger than the median of 8 trials. This occurs because there are no individuals who are faster with the 17 out of 20 criterion, and because there is a subset of individuals who need considerably fewer trials with the 10/12 criterion (Figure 7). Individuals who require a larger number of trials (>100) to pass almost never occur with the 10/12 criterion, whereas they are more common with the 17/20. With more trials, there is a higher chance that an individual will deviate from their preference by chance. This is also reflected in that 65 of the 626 simulated individuals never reached the 17/20 criterion within the maximum 300 trials, whereas there were only 4 individuals with the 10/12 criterion. Accordingly, an additional benefit of choosing the 10/12 criterion is that it is more likely that data for all individuals, even those who are slow to learn an association, can be collected.

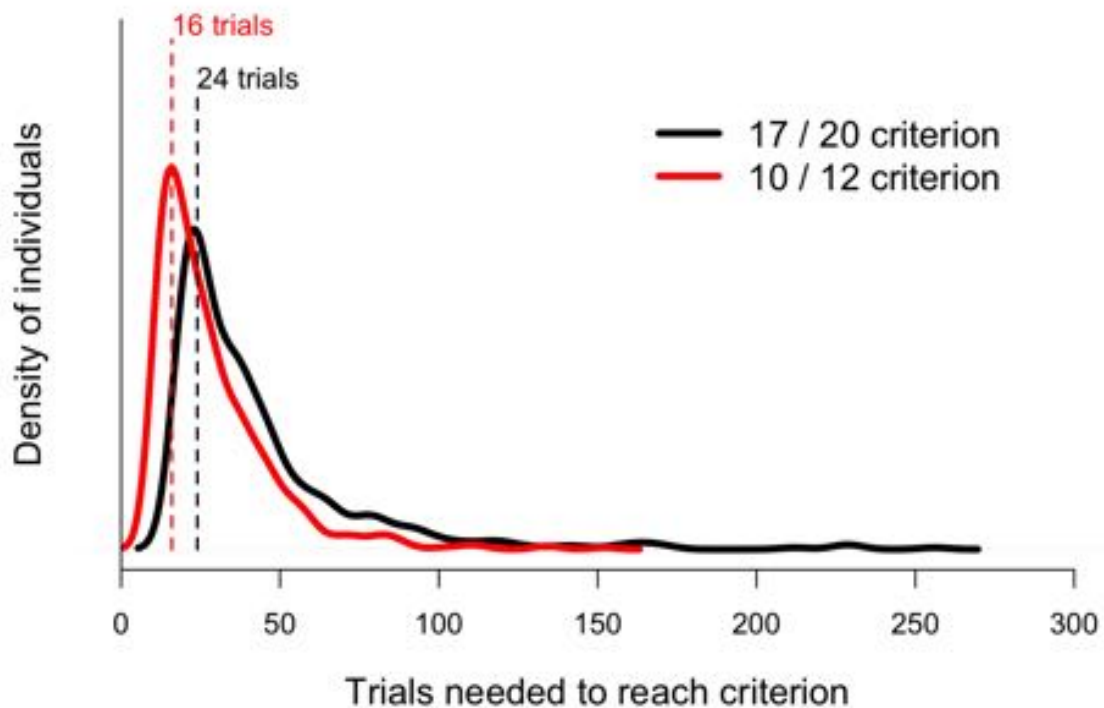


Figure 7. There is less variation with the 10/12 reversal learning passing criterion and it requires fewer trials to reach than the 17/20 passing criterion. The lines represent the densities of

individuals (estimated with smoothing which means values can go down to zero) across the 626 simulated individuals that needed a certain number of trials to either reach the 10/12 (red) or the 17/20 (black) criterion. With the 10/12 criterion, most individuals need 8 fewer trials (indicated by the lines showing the mode of the two density distributions). In addition, there are only very few individuals who need 100 or more trials with the 10/12 criterion, while there are several individuals that needed such large numbers with the 17/20 criterion.

3) variation among individuals with the 10/12 criterion is still present and similar to the variation detected with other criteria

As described in point 1, when changing the criterion from 17/20 to 10/12, most individuals need 8 fewer trials. This also means that the differences among individuals, which might contain relevant information about variation among them, is preserved. When transforming performance with the two criteria to ranks, individuals are sorted essentially in the same order independent of which criterion is used. This is shown in Figure 7: most points are shifted up by exactly 8 trials.

4) individuals can be assumed to have reliably learned the association using the 10/12 criterion

Based on the two reversal passing criteria (10/12 and 17/20), we can extract the attractions that simulated individuals have formed toward both the rewarded and the unrewarded option at the point at which they meet each of these criteria. Comparing the two attractions (to the rewarded and unrewarded options), we can determine whether individuals are likely to have learned an association or not. Independent of the criterion, individuals generally formed a preference for the rewarded option: 89% of individuals favor the rewarded option between 2.5 and 14 times more than the unrewarded option. With both criteria, individuals always have a stronger attraction to the rewarded than the unrewarded option. The smallest difference between the attraction scores to the rewarded and unrewarded options we observe at the point of passing is the same with both criteria. With the 10/12 criterion, individuals would in the next trial, on average, choose the rewarded option with a probability of 76% (3 times more likely to choose rewarded over unrewarded option), whereas this is 84% with the 17/20 criterion (5 times more likely).

5) the learned association means that individuals who move to the next reversal are unlikely to solve the reversed association by chance

As expected, based on the relative attraction scores at the end of the previous reversal, most individuals are unlikely to choose the now rewarded option. We expect that, on average, individuals will choose the newly rewarded option in 4 or fewer trials out of the first 12 trials (red line in Figure 8). This is a lower number of trials compared to individuals who have no association with either option (gray line in Figure 8), and a slightly higher number compared to individuals who use the 17/20 criterion (black line in Figure 8). The probability that an individual would, after a reversal, immediately choose the rewarded option 10 times during the first 12 trials (and pass) by chance is 0.001. However, even such rare individuals will have actually reversed their preference during their first 12 trials because they update their attractions on every trial.

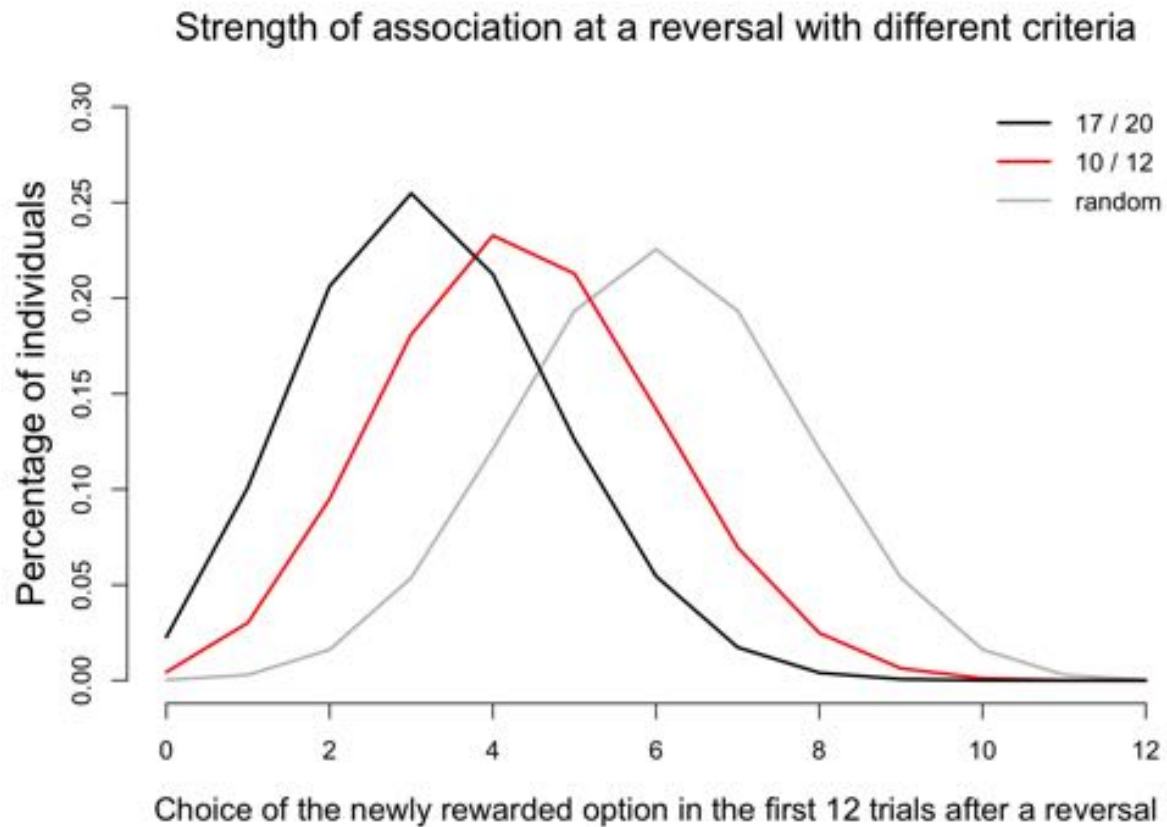


Figure 8. Individuals form strong enough preferences using the 10/12 passing criterion as indicated by the fact that they are unlikely to pass in the first 12 trials of their next reversal (red line). These individuals would take longer to switch their preference than individuals who have no preference (gray line), and they would be slightly faster at switching their preference than individuals who formed their previous association using the 17/20 criterion (black line).

Code

Determining after which reversal an individual has completed the experiment: serial reversal passing criterion

Data from previous serial reversal experiments suggests that individuals who go through multiple reversals will end up with a performance that is similar to the individuals who needed the fewest trials on the first reversal (C. Logan et al., 2022; Lucon-Xiccato & Bisazza, 2014). This suggests that the manipulation changes individuals within their natural range of variation rather than pushing them to new limits. This means that we can use the performance of the fastest individuals in the first reversal to set the criterion for passing the serial reversal experiment. Accordingly, we can only set the serial reversal passing criterion after the data from the first reversal begins to become available. Some species might already have data from

previous studies on reversal learning, however it is important to set the passing criterion for this experiment using ~~the~~ this particular setup. Therefore, the criterion must be established from scratch for each species using this setup.

The serial reversal passing criterion: reach the reversal passing criterion (10 out of 12 trials correct) in X trials or fewer in two consecutive reversals.

X = the number of trials required that marks the fastest 20% of individuals in the first reversal. For example, if you test 20 individuals, the number of trials for the 4th fastest individual will be the criterion. For 10 individuals, use the number of trials for the 2nd fastest individual. The fastest 20% was validated using the grackle data (C. Logan et al., 2022): it aligns with the one sigma rule from a normal distribution, indicating the percentage of individuals who are faster than the mean number of trials minus one standard deviation. If more than 20% of individuals reach this number of trials in their first reversal (because there might be a tie), choose the next fastest number of trials to pass. Particularly near the beginning of the experiment, it will be important to set the passing criterion to a lower number to ensure that individuals will be overtrained rather than undertrained.

As the data for additional individuals becomes available, this number can change accordingly. If the number changes across the experiment, we will check whether any currently participating individuals would have already passed according to this criterion and end their experiment.

Individuals need to meet this criterion in two consecutive reversals to pass the serial reversal experiment to ensure that their behavior is consistent and that their speedy performance did not occur by chance. Previous serial reversal experiments show that reversal performance plateaus after a certain number of reversals (e.g., 6-8 reversals in great-tailed grackles C. Logan et al., 2022). If individuals show no consistent improvement after 12 reversals and have not yet met the serial reversal passing criterion, they will be excluded from the experiment. We will plan to start with many more individuals than the minimum sample size to allow for potential drop outs.

We do not expect that the serial reversal manipulation will introduce new negative effects because the passing criterion is set such that the manipulated individuals are only as fast as the fastest 20% of tested individuals. This means that we are not introducing an unnatural amount of flexibility because we are not making any individuals more flexible than what already exists in their population.

There will be individual variation in terms of baseline flexibility before the manipulation such that the flexibility training might influence individuals differently. For example, individuals who are already flexible before the manipulation will not benefit much from the manipulation, while the less flexible individuals will benefit more. Individuals who are already flexible and pass the serial reversals in fewer reversals will still meet the experiment's passing criterion and be considered to have completed the manipulation, even if they did not improve. Baseline flexibility differences could also be reflected in their pre-manipulation success measures (i.e., individuals with high baseline flexibility might already be successful before the manipulation) if these success measures relate to flexibility. Our statistical models account for these baseline individual differences in success as they might relate to performance on the flexibility manipulation because they include an interaction between the intercept (the value at which individuals start) and slope (by how much they change).

Planned Sample

For each population, depending on the response variable, we ran separate power analyses to

determine the planned sample size (see Analysis Plan). For each population, we will aim to reach the minimum sample size required to detect the expected effects of the intervention on the response variables. However, given the difficulties of working with wild individuals, there might be instances where we might not reach a particular target. In such a case, we will interpret the result in light of the power that the particular sample size provides, as indicated by the power analyses. The minimum sample size depends on whether the intervention is performed as a within-subject design (higher power), measuring the response for the same individuals before and after the intervention, or whether it is performed as a between-subjects design, where half the individuals are randomly assigned to the intervention group. In addition, it will depend on whether the response variable has a binary or a continuous outcome (higher power), and in the latter case whether the measure is open-ended (lower power) and therefore individuals will show a large range of values (e.g., dispersal distance).

We will stop collecting data for the flexibility manipulation experiment when a buffer above the minimum sample size is reached or when the season in which the minimum sample size is reached comes to an end, or when the minimum sample sizes for the success measures have been reached. When conducting the manipulation experiment, it is important to aim to test more than the minimum number of individuals because some might not have data in the post-intervention stage.

DESIGN 1 - Reversal learning experiment in visual isolation

Although we do not use this design in our planned studies, we present this design as an option for researchers interested in using the ManyIndividuals framework. Half of the individuals (manipulated group) undergo serial reversal learning until they meet the passing criterion, while the other half (control group) receive only one reversal (Figure 6). A reversal of where the reward is placed represents environmental heterogeneity, and those individuals who have enough experience in this heterogeneous environment through multiple reversals are able to learn to improve their flexibility (the number of trials it takes to change a preference). This flexibility intervention has been shown to work in great-tailed grackles where 8 out of 9 individuals in the serial reversal group achieved the passing criterion (passing 2 consecutive reversals in 50 trials or less) within their given time frame (Logan, MacPherson, et al., 2019). Individuals are presented with two options that differ in color, shape, or in some other way, with one option being the rewarded option. The first rewarded option (i.e., color, shape, etc.) in reversal learning is counterbalanced across individuals at each site. The rewarded option location is pseudorandomized for side. Pseudorandomization consists of alternating location of the rewarded option for the first two trials of a session and then keeping the same color on the same side for at most two consecutive trials thereafter. A list of all [88 unique trial sequences](#) for a 10-trial session, following the pseudorandomization rules, will be generated in advance for experimenters to use during testing (e.g., a randomized trial sequence might look like: LRLRLRLRLR, where L and R refer to the location, left or right, of the rewarded tube). Randomized trial sequences will be assigned randomly to any given 10-trial session using a random number generator (random.org) to generate a number from 1-88. The individual is only allowed one choice per trial and the option on the left is always placed first, with the experimenter always turning to the right when setting up and taking down each trial (if a live experimenter is involved). Once a preference for the rewarded option is reached (10/12 trials correct), the reward is then always placed in the previously non-rewarded option until a preference is reached (using the same criterion). After the control group's first reversal, they

receive a similar amount of experience as the manipulated group, but without the functionality: two apparatuses are still used in every trial, but they are the same color/shape/etc. and both contain food. As with the manipulated group, only one choice, the first choice, is allowed per trial. Choices are scored as: 1=chose the correct option (even if they do not eat the food), 0=chose the incorrect option, and -1=did not make a choice.

Passing criteria:

- **Habituation:** leave one color/shape/etc. apparatus that is not used in the experiment in the enclosure overnight and feed the individual off of it until they readily approach it for food.
- **Training to look for non-visible food:** obtain the food from the habituation apparatus (only one presented per trial) in 5 consecutive trials when the food is hidden inside the apparatus (and not visible to the individual unless they approach the apparatus and choose to look inside). Choices are scored as: 1=ate first from inside the apparatus, 0=ate food from around the apparatus but not inside it, and -1=did not eat any food.
- **Training to eliminate any color/shape/etc. preference:** choose one color/shape, of the two possible colored/shaped apparatuses, first 8 or fewer times out of 10 trials, which indicates no preference. Use both experimental apparatuses (e.g., two colors or shapes) per trial with the food openings taped over so the individual cannot look inside the tubes. Place both apparatuses in the test area at the same time and place food on the outside of both apparatuses at the same time (if you need to spend more time on one apparatus because the food falls off or something, make mirror movements on the other apparatus at the same time so both apparatuses get the same amount of attention and in the same way). Choices are scored as: 1=ate first from the color/shape that will be the rewarded option, 0=ate first from the color/shape that will be the non-rewarded option, and -1=did not eat any food.
- **Reversal (including initial discrimination):** using the same apparatuses as above but with the openings untaped, must obtain the food from the rewarded color/shape on at least 10 of the latest 12 trials, calculated in 1-trial windows (i.e., the individual can pass on trial 21, 35, 44, etc.).
- **Serial reversal manipulation group:** across reversals, must achieve the reversal criterion for two consecutive reversals in 50 trials or less. This passing criterion was generated using great-tailed grackles (Logan, MacPherson, et al., 2019) and might need to be adjusted depending on the population.
- **Control group:** receives as many trials with the control apparatuses (two containers of identical color/shape, both rewarded) as the average number of trials that manipulated individuals require to pass serial reversals. If this is unknown at the beginning of the experiment, test a manipulated individual first and match the control individuals to this number until an average can be obtained. The average can continue to be updated as more manipulated individuals complete testing.

Protocols and data sheet templates

[Protocol](#) for reversal learning of a color preference used by Logan, MacPherson, et al. (2019) for great-tailed grackles.

See the [data sheet templates](#) in Logan, MacPherson, et al. (2019).

Interobserver reliability

We conduct interobserver reliability by having hypothesis-blind video coders code all of the videos for 20% of the individuals in the experiment: [instructions for video coders](#). Live coder data is then compared with video coder data by analyzing whether the individual made the correct choice (1) or not (0) or whether they did not make a choice (-1) using Cohen's unweighted kappa (Landis & Koch (1977), with the psych package in R Revelle (2017)). This measure indicates how replicable the experiment and the coding instructions are. Before a video coder is approved to begin coding, they must first pass training where they code a different set of videos and reach an unweighted kappa of 0.89 or above.

Code

DESIGN 2 - Reversal learning experiment in a group context

Feeders are set up in the field, counterbalanced for color/location/etc, and will be available for opening when the experiment is being conducted (Figure 7). The feeders will be fitted with technology to automatically record which individual visits which feeder and when (unless a particular population is easy to track visually without such automated technology). Individuals are habituated to the open feeders filled with food until at least half of the minimum sample size has visited at least 1 feeder. All individuals start on the same rewarded option for the initial discrimination to improve motivation for participating in the experiment (e.g., if the first rewarded option was counterbalanced across individuals, subjects might be slower to learn their rewarded option if they use social information about which option is rewarded). If the subject visits the rewarded feeder, the feeder will automatically deliver a small amount of food, and then close and reset more food in preparation for the next opening. If the subject visits the non-rewarded feeder, the presence data will be recorded, but the feeder will not open. All feeders will contain one type of high value food. Ideally, feeders will be programmed to automatically switch which feeder type is rewarded as soon as an individual passes criterion in the middle of a test session. If automation is not possible, then the data sheets will be checked at the end of each test session to determine which individuals have passed criterion and their rewarded feeder type will be changed in the next test session.

Individuals in the control condition (if there is one), will receive 1 reversal and, after they pass criterion on reversal 1, feeders of both options (previously rewarded and previously non-rewarded options) will open for these individuals. This will help keep the whole group interested in visiting the feeders while the individuals in the manipulated group complete their serial reversals. Data are collected on success measures (see below) either before and after the flexibility intervention or only after the intervention (depending on feasibility and how the study design needs to be tailored for each population).

Passing criteria:

- **Feeder habituation:** all feeders at all locations will have food and be open for several hours daily or until at least half of the minimum sample size in each condition (control and manipulated) have visited at least one of the feeders.
- **Reversal passing criterion:** an individual is considered to have a preference when they choose 10 of the most recent 12 trials (choices) correct (the rewarded option). This criterion applies to the initial discrimination, and to each reversal.
- **Manipulation passing criterion:** pass two consecutive reversals in X trials or less (see Serial reversal passing criterion above).

Protocols

Food in the feeders: All feeders are opaque and always have food in them to eliminate the confound due to olfactory differences between the feeders that could be introduced if only the rewarded feeders have food in them. If a feeder needs to be refilled, refill all feeders consecutively in the same time period and refill each for the same amount of time even if that feeder does not need much or any food (in these cases, pretend to fill the feeder as you normally would). This eliminates confounds from cues provided by a differential amount of attention experimenters give the feeders depending on which needs refilling.

Reliability of feeder data

During the experiment, we will determine the reliability of the automated feeder data. We will verify that the feeders are correctly detecting tags by using a (motion-activated, Go-Pro, or similar) camera focused on 20% of the feeder perches. Hypothesis-blind video coders will then code who landed on the perch: individual ID, time of day, and whether it ate food. We will then compare the automated feeder data sheets with the video coder's data sheets by calculating the Cohen's unweighted kappa for individual ID and the intra-class correlation coefficient (ICC) on the time of day (kappa: Landis & Koch (1977), with the psych package in R Revelle (2017); ICC: Hutcheon et al. (2010), with the irr package in R: Gamer et al. (2012)). Cohen's kappa is used when the distance between measurements is not quantifiable numerically, and the ICC is used for continuous variables with equal distances between units.

Assessment of the likelihood of success in human modified environments with regard to the flexibility manipulation

After the manipulation, compare control individuals with individuals in the flexibility manipulation and/or pre-manipulation and post-manipulation measures on the same individuals using one or more of the following success measures. Choose population-relevant success variables that are predicted to be the most likely to be used in human modified environments. This list is not exhaustive - it serves as a place to generate ideas about what the best measures could be in a given population. Observational methods to collect these data may vary among populations and we describe below the methods that we will use (i.e., focal follows and all occurrences sampling).

- **Fitness variables:** nest success, number of offspring who survived to independence or adulthood, longevity, etc.
- **Foraging variables:** diet breadth, number of foraging techniques used, etc.
- **Movement variables:** predictability of movement behavior [e.g., step length and turning angles; see McCune KB et al. (2020)], ability to disperse from a lower risk environment (e.g. a sanctuary, or largely intact natural habitat) to a higher risk, more heavily human modified environment (assess success/survival after dispersal if possible), etc.
- **Habitat use variables:** foraging substrate (ground, bushes, trees, human modified substrates, human-provided supplemental food), nesting substrate (high or low, tree, bush or reeds), etc.

Begin collecting post-manipulation data on an individual as soon as it passes the manipulation because it is unknown for how long any potential effects of the manipulation will last. For the more social species and for populations who participate in the experiment in groups, there is the potential that individuals who were not in the manipulation condition or who have not yet passed the manipulation condition will learn about post-manipulation success behaviors from the manipulated individuals who passed the experiment. We are ultimately interested in determining whether we can change success behaviors as a result of the flexibility manipulation. If part of this change is the result of social learning from some of the manipulated individuals, it will still result in a change even if we do not quantify what percentage of the mechanism comes from individual learning after the manipulation or social learning after the manipulation. If there is a change in success measures between before and after the manipulation, the manipulation will have been the cause in either case.

If other researchers are interested in beginning to quantify whether social learning is involved in the spread of success behaviors, then they could collect data on post-manipulation success measures from the individuals who did not pass the manipulation and compare them with individuals who passed the manipulation. If there is no difference between both groups' post-manipulation success measures, this indicates that social learning was involved in the spread of success behaviors. They could then use this data in a future registered report examining the role of social learning in the spread of success behaviors. If researchers are not planning on a social learning component in future research and/or do not have the time or resources to collect more data, they can refrain from collecting post-manipulation data on success measures on the individuals who did not pass the manipulation criterion.

Observational methods

We will use slightly different observational methods for grackles and jays to collect data on foraging and microhabitat use. However, the categorization of food and habitat types will be the same:

Microhabitat types in the suburban habitat (<100m from human structure) include: vertical human structure (e.g. building, bench), native vegetation, non-native vegetation, grass, impervious surface, and dirt. In the natural habitat (>100m from human structure), microhabitat types include all previous categories, but not human structure or impervious surface. All categories can be further defined by whether the subject was high (>3m) or low (<3m) (for example, grass and impervious surfaces can occur above 3m if grass is on the roof of a building, and if an individual is walking on the impervious surface of an upper floor parking garage).

Food types are broken down into plant (seed, fruit, human-provided, or unknown plant) and animal (insect larva, adult insect, amphibian, reptile, mammal, bird, egg, human-provided, or unknown animal). “Human-provided” indicates any food item that was acquired from a store at some point and is left out by humans. For example, sunflower seeds would be considered human-provided if they are in the form of bird seed or a human snack. Sunflower seeds would only be counted in the “seed” category if the bird is seen eating it from a plant. Data will be collected on the four plant subcategories and nine animal subcategories and used in the analyses.

During follows, we will record each microhabitat the individual is present in and all food items consumed. Before data analysis, to ensure that we are only including the microhabitats individuals use (rather than just pass through), we will filter the data to

include only microhabitats that account for at least 5% of their data points. Although we may not observe every possible microhabitat or food item the individual may use, by equally sampling before and after the manipulation we can detect changes in habitat use.

Additionally, for all observational methods we use binoculars so that we can always attempt to stay far enough from the focal individual that our presence is not affecting their behavior. Because that distance can be different for each individual and species, we hesitate to give a specific number. However, if at any point the focal individual shows it is affected by our presence by looking directly at the researcher, alarm calling or startling, then we end the focal immediately, drop all data from that follow and attempt the follow again the next day.

We will attempt to balance follows for each observational method in the morning and afternoon for all individuals in the study. In this way, we will collect a random sample of data from active and inactive time periods for all individuals. Additionally, we will use automated tracking technology on most of the species, which records daily movement patterns. This will tell us whether there is temporal or spatial variation in bird behavior.

Focal follows

We will collect data on foraging and microhabitat use of grackles during 10 minute focal follows and our minimum sample size will be 20 individuals. We will do 4 focal follows per individual before the flexibility manipulation and 4 focal follows per individual after the manipulation (8 total 10-min follows for 80 minutes of follow per bird). To minimize the temporal and spatial autocorrelation of behavior, we believe it is better to do the shorter follows of 10 minutes and space sequential follows apart by at least 1 week.

All occurrences sampling

We will collect data on the foraging and microhabitat use of jays during spatial movement tracking (the latter as part of another investigation). We will follow jays for 60 minutes and record at 1-minute intervals the spatial location (GPS coordinates), any food items consumed, the microhabitat the jay is present in, and breeding behaviors if it is the breeding season. We will do 4 tracks per individual before, and 4 tracks per individual after the flexibility manipulation.

Open data

The data will be published in the Knowledge Network for Biocomplexity's data repository.

Great-tailed grackles

- Grackle [protocol](#) and [data sheet templates](#)
- [Protocol for applying radio tags and conducting GPS tracks](#) from McCune KB et al. (2020)

Planned sample

We will catch at least 20 grackles using walk in traps, mist nets, and bow nets; collect their biometric data, blood, and feathers; apply colored leg bands for individual identification, one band will have a PIT tag that will interface with the automated feeders; attach a radio tag using a leg loop harness; and release them at their point of capture. We will collect data on pre- and post-manipulation success measures and conduct the experiment within the non-breeding season to control for potential temporal differences in the environment and behavior.

Before and after manipulation success measures

The proportion of time spent at dumpsters and outdoor cafes when food is present will be collected using automation from Cellular Tracking Technologies. Individuals will wear radio tags whose signals are detected (within approximately 75m) by nodes that are placed at approximately 200m intervals. Before and after tracking data will be collected on a minimum of 20 individuals with a minimum of 4 separate visit events per individual pre-manipulation and a minimum of 4 separate visit events per individual post-manipulation. This data will allow us to detect pre- vs post-manipulation within-individual differences in visits to dumpsters and cafes to determine whether individuals change the proportion of time they spend at these locations when food is present.

A determination will be made daily about whether food was present at a location and in which particular time period. We will visit the dumpsters each morning to record whether the lids were open and there was garbage inside or garbage on the ground around the dumpsters (1=food present) or whether the lids were closed and there was no garbage on the ground (0=food absent). We will initially visit cafes during lunch hours to determine when people generally eat outside and then set the default food presence period for each cafe to the widest time period we observed for that cafe (e.g., 12-2p = food present, before 12p and after 2p = food absent).

We will track baseline behavior and changes after the flexibility manipulation in habitat use and foraging using 10 minute focal follows (Altmann, 1974). We document all occurrences of microhabitats used and foraging of the focal individual (see Observational methods, above). These data will allow us to detect any pre- and post-manipulation within-individual differences in diversity of habitat use and foraging breadth.

Flexibility manipulation (Design 2 reversal learning)

We will set up 4 feeders (2 dark gray and 2 light gray) at one location in a particular spatial arrangement: one dark gray and one light gray feeder will be oriented in the same way 1-2m apart, and the other dark gray and light gray feeders will be facing them in a mirrored position, but 5-10m away. The feeders will be available for opening for 3 hours per day, 5 days per week. The feeders will be fitted with RFID readers to automatically record which individual (fitted with a PIT tag attached to a leg band) visits which feeder and when. If the subject visits the feeder that has the rewarded color, the feeder will automatically deliver a small amount of food, and then close and reset more food in preparation for the next opening. If the subject visits the feeder that has the non-rewarded color, the presence data will be recorded, but the feeder will not open. All

feeders will contain one type of high value food (e.g., goldfish crackers, cheetos, cheez-its, or peanuts).

Feeder habituation: All feeders will have food available and be open for 3 hours per day on 5 consecutive days or until at least 20 banded grackles have visited at least 1 feeder.

The experiment: Each time a PIT tagged subject lands on a feeder it is recorded as the color choice and counts as one trial. The experiment will begin with the dark gray feeder being the rewarded feeder for all individuals.

Jays

- Jay [protocol](#) and [data sheet templates](#)
- [Protocol for applying radio tags and conducting GPS tracks](#) from McCune KB et al. (2020)
- Jay [processing protocol](#)

Planned sample

We will catch up to 60 jays per year using walk in traps, mist nets, and bow nets; collect their biometric data, and a blood sample; apply colored leg bands; attach a radio tag using a leg loop harness; and release them at their point of capture. We will collect data on pre- and post-manipulation success measures and conduct the experiment within the non-breeding season to control for potential temporal differences in the environment and behavior.

To determine whether the flexibility manipulation has influenced the ability of jays to persist in human modified environments, we will catch half of the jays in areas with access to human-supplemented food (i.e. private property, a university campus, parks adjacent to neighborhoods with feeders) and the other half in natural areas (wildlife management areas, reserves).

Before and after manipulation success measures

We will track baseline behavior and changes after the flexibility manipulation via spatial movement tracking that lasts for 60 min, noting the GPS location and the jay's behavior at 1 minute intervals. During tracks, we document all occurrences of microhabitats used within the territory (see Observational methods, above), foraging (see Observational methods, above), and breeding behaviors of the focal individual if it is the breeding season (Altmann, 1974). The minimum sample size will be 20 individuals per species with a minimum of 4 tracks per individual (at least 2 per month) pre-manipulation and a minimum of 4 tracks per individual (at least 2 per month) post-manipulation (at least 320 tracks in total). These data will allow us to detect any pre- and post-manipulation within-individual differences in space use, diversity of habitat use, and foraging breadth.

Flexibility manipulation (Design 2 reversal learning)

We will set up feeding stations at a minimum of 4 study sites, each containing multiple jay territories, spaced at least 2km apart: 2 sites in natural habitat, 2 in human-modified habitat. If

individuals are dispersed across multiple areas, we will attempt to add an equal number of sites in natural and human-modified habitats. Gravity feeders will be set up in territories in each of the sites in a similar spatial arrangement as in the grackle experiment (above).

Feeder habituation

All feeders at all locations will have food and be open for 2-hour sessions between 8a-3p on a minimum of 4 consecutive days or until at least 1 banded jay per territory per site (natural/human modified) has visited at least 3 feeders.

The experiment

Prior to the flexibility manipulation, we will collect the 4 minimum tracking sessions to determine the baseline values for space use, nest success (if appropriate), microhabitat use and foraging breadth. Afterwards, we will set up feeders to initiate the flexibility manipulation phase. Once jays are habituated to the feeders we will manipulate behavioral flexibility using serial reversals of the open feeder location or color. Only one feeder will be opened at a consistent location (or that is a consistent color) within territories across days and the manipulation treatment will consist of 30-min sessions per day per territory, up to 4 days per week, where each visit by the focal jay to a feeder is considered a trial. Jays pass a given reversal when they correctly choose the rewarded feeder in at least 10 trials out of the most recent 12 trials. Serial reversals will continue until jays pass two consecutive reversals in X trials or less (see Serial reversal passing criterion above). At this criterion, the jays will be considered to have increased their behavioral flexibility. After the manipulation is complete in each territory, we will again conduct the tracking sessions to measure space use, nest success (if appropriate), microhabitat use and foraging breadth.

Toutouwai

- Toutouwai [protocol](#) and [data sheet templates](#)

Planned sample

In the Zealandia toutouwai population the breeding season runs from October to February. Pairs typically produce 1-3 fledglings per nesting attempt and can nest up to 3 times per season. Each season, around 30 pairs nest in the long term study area (Figure 9), so we will aim to catch up to 30 fledgling toutouwai a year (one per pair, 60 in total) using a drop trap. We will collect their biometric data and a feather sample for DNA sexing. On each bird we will apply 2 colored leg bands for individual ID (one to each leg), a uniquely numbered metal BP sized band (supplied by the New Zealand National Bird Banding Scheme) and a single RFID tag attached to a leg band (on the opposite leg to the metal band). We will also attach a radio tag using a leg loop harness that degrades over time. Birds will be released at their point of capture. We will conduct the experiment in the breeding season and collect data on post-manipulation success measures in the breeding season as well as the non-breeding season. In this case, we do not need to control for potential temporal differences in the environment and behavior because we are measuring juvenile dispersal behavior.

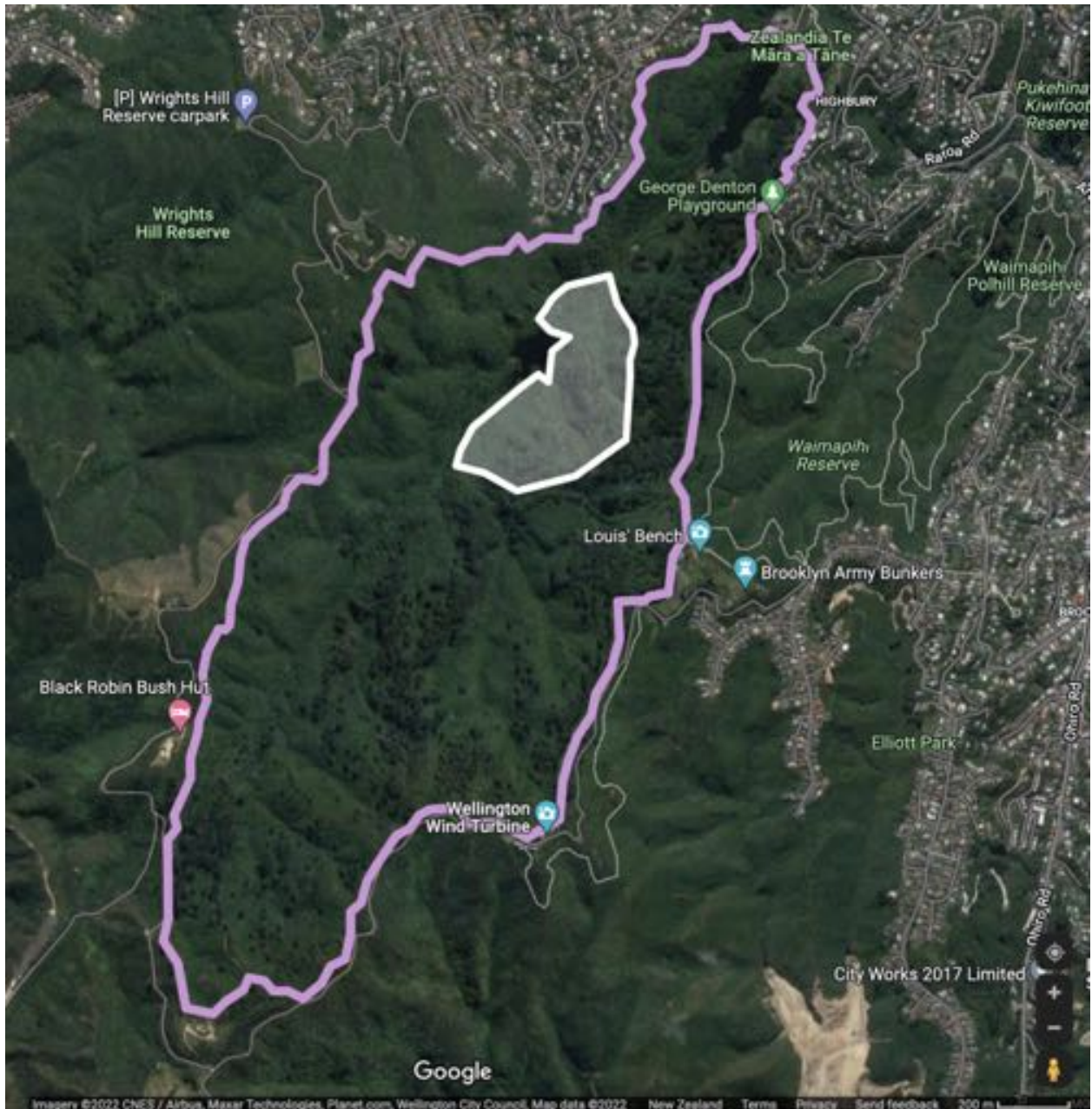


Figure 9. The toutouwai study area shown in white, is situated centrally within Zealandia Ecosanctuary, outlined in purple.

Flexibility manipulation (Design 2 reversal learning)

Juvenile toutouwai typically begin to forage independently on their caregiver parent's territory from 3-4 weeks post-fledging. We will set up two experimental feeders on the caregiver parent's territory when the juvenile toutouwai is at least 3 weeks post-fledging and has been captured, radio tagged and fitted with an RFID tag. One juvenile per territory will be randomly selected to participate in the experiment. Half of all juveniles will be assigned to a manipulated condition

and half will be assigned to a control condition. The condition (control or manipulation) will be randomly assigned to each bird using random.org. On each territory the two RFID activated feeders will be spaced a maximum of 4 m apart and within sight of each other, with one feeder mounted within 20 cm of the ground, and the other mounted at 2 m height. Feeders on the ground will be light grey in colour and feeders mounted at height will be dark grey, to make them both spatially and visually distinct. To ensure that the juvenile belonging to the territory is the only bird that can access the feeders, each feeder will only open when the individual fitted with the RFID tag coded to the feeder approaches.

Feeder habituation

Both feeders on a territory will be available to the RFID tagged individual(s) from 8 am - 3 pm for 5 consecutive days, or until the juvenile has visited both feeders. When a juvenile lands on the perch of an available feeder, the RFID tag will be read and the feeder will automatically open to dispense a live mealworm reward.

The experiment

Juveniles in both the manipulation and control conditions will initially undergo one reversal learning procedure consisting of a discrimination learning phase and a reversal learning phase. During the initial discrimination learning phase, only one feeder will be available (hereafter called the rewarded feeder) and the location of this rewarded feeder (high or low) will be assigned randomly, with half of the birds receiving the low and half receiving the high feeder as their rewarded feeder. The rewarded feeder will be available for 2 hours each day, or until a maximum of 100 trials have been completed (whichever occurs first). The rewarded feeder will automatically open when the target bird (fitted with the correct RFID tag) lands on the feeder perch. However, when the target bird lands on the perch of either feeder (rewarded or unrewarded), the visit will be logged and counted as a single trial. To pass the discrimination phase, the bird must visit the rewarded feeder in 10 out of 12 consecutive trials. Once this criteria is reached, the juvenile toutouwai will be given a reversal phase, where the previously unrewarded feeder is now rewarded. To pass this reversal phase, the bird must again achieve the criterion of 10 out of 12 consecutive trials visiting the correct feeder. Birds in the manipulation group will then receive serial reversals until the point at which they are switching feeder preferences and pass the serial reversal criterion (see Methods). For juveniles in the control group, after the initial single reversal both feeders will remain available and rewarded for two hours each day (or until food is depleted) for a minimum of 8 days.

Post-manipulation success measures

We will radio track fledglings for 12 weeks following the end of the flexibility manipulation (i.e., until they are 112 days post-fledging). Tag signals will either be detected with a hand-held antenna, or by nodes that are placed at approximately 200m intervals across the 25 ha robin study area and around the sanctuary perimeter. Each fledgling will be located once per week over this time period and their location will be GPS marked to reconstruct dispersal tracks. By the end of 12 weeks, fledglings will be approximately 4 months post-fledging, at which point they are likely to be attempting to establish their own territories. The measures used to investigate the effect of the manipulation versus control treatment will include the final dispersal location (inside or outside the sanctuary), the total dispersal distance between the natal territory and final location, the age at which dispersal is first detected (defined as when the bird has left the caregiver parent's territory for a minimum of 3 days), and their survival status at 16 weeks post-fledging (alive/dead).

ANALYSIS PLAN

We run analyses in R [current version 4.1.2; R Core Team (2017)] using the following R packages: rethinking (McElreath, 2020a), rstan (Stan Development Team, 2020), cmdstanr (Gabry & Češnovar, 2022), knitr (Xie, 2018), and irr (Gamer et al., 2012).

Can flexibility be increased to help individuals succeed in human modified environments?

DESIGN 1 - Reversal learning experiment in visual isolation

Can flexibility be increased? If most individuals in the flexibility manipulation (serial reversals) passed the passing criterion, then the answer is yes.

Do the flexibility manipulated (serial reversal) individuals have better success in human modified environments?

Response variable: success variable (e.g., predictability of movement behavior, number of different food items taken, etc.)

Explanatory variable:

- Condition (control, manipulated)

OR

- Condition (pre-manipulation, post-manipulation)

Random variable: ID (when response variable has multiple data points per individual)

DESIGN 2 - Reversal learning experiment in a group context

Same questions, response/explanatory/random variables as in Design 1.

Great-tailed grackles

G.Q1 Do flexibility manipulated individuals differ in the proportion of time spent at cafes and garbage dumpsters when food is present?

The model

Bayesian model for a binomial distribution:

$R_i \sim \text{Binomial}(t_i, p_i)$

R_i is the duration spent at cafes and dumpsters when food rewards were present and t_i is the

total duration spent at cafes and dumpsters either before or after (i) the flexibility manipulation. A binomial distribution was used because the response variable is a proportion (McElreath, 2020b).

$$\text{logit}(p_i) = \gamma[\text{condition}] + \alpha[\text{ind}],[\text{condition}]$$

$\gamma[\text{condition}]$ is the average log-odds for each condition (before/after) and $\alpha[\text{ind}],[\text{condition}]$ is the effect for each individual in each condition. The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in the proportion of time spent in the presence of a reward in the before vs. after conditions (Figure 10). We simulated the proportions of times that different sample sizes of individuals would spend in the presence of the reward before and after the flexibility manipulation. We analyzed these simulated data with the model we will use to analyze the actual data, estimating the change in the proportion of time spent in the presence of the reward between the before and after conditions. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the change in proportion of time spent in the presence of a reward before the flexibility manipulation is similar to after. If the ratio is close to zero, the model assumes individuals have changed their behavior. For changes smaller than 0.1, models are likely to assume that no changes occurred even with large sample sizes. If the change in the proportion of time an individual spends at a location when food is present before the flexibility manipulation vs. after is 0.1, on average 93% of the posterior of the model based on a sample size of 20 individuals will be larger than zero. This means that the model is quite certain there is a difference that is larger than zero. In addition, none of the models for a sample size of 20 at the mean change of 0.1 have a ratio larger than 0.3, meaning that the risk of having a false negative is unlikely.

In general, with sample sizes at or above 20 and mean changes in the proportion of time spent in the presence of a reward is 0.1 or larger, then it is highly likely that the model will indicate that individuals have changed their behavior. Mean changes below 0.1 can still be detected, however there is a higher risk that there will be a false negative. If the change in the proportion of time an individual spends at a location when food is present before the flexibility manipulation vs. after is 0.025, on average 61% of the posterior of the model based on a sample size of 20 individuals will be larger than zero. In addition, only 20% of the models for a sample size of 20 at the mean change of 0.025 have a ratio larger than 0.3, meaning that the risk of having a false negative is still low.

With small mean changes in the response variable, some individuals might not increase or even decrease their response after the manipulation because there is variation around the mean change in individual responses. With small sample sizes, there is a risk that only individuals who did not clearly increase their response will be studied, whereas larger sample sizes are more likely to include a wider spectrum of individuals.

To estimate the risk of detecting false positives, we set the mean change in the proportion of time spent in the presence of the reward to zero so there was no change between the before and after conditions. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 20, 43% have a ratio smaller than 0.3,

meaning that the risk of having a false positive is high. The risk would be lower if the variation among individuals was lower than what we assumed (across all models, the standard deviation of the mean change in proportion was 0.1, which is a conservative estimate).

Code
Code

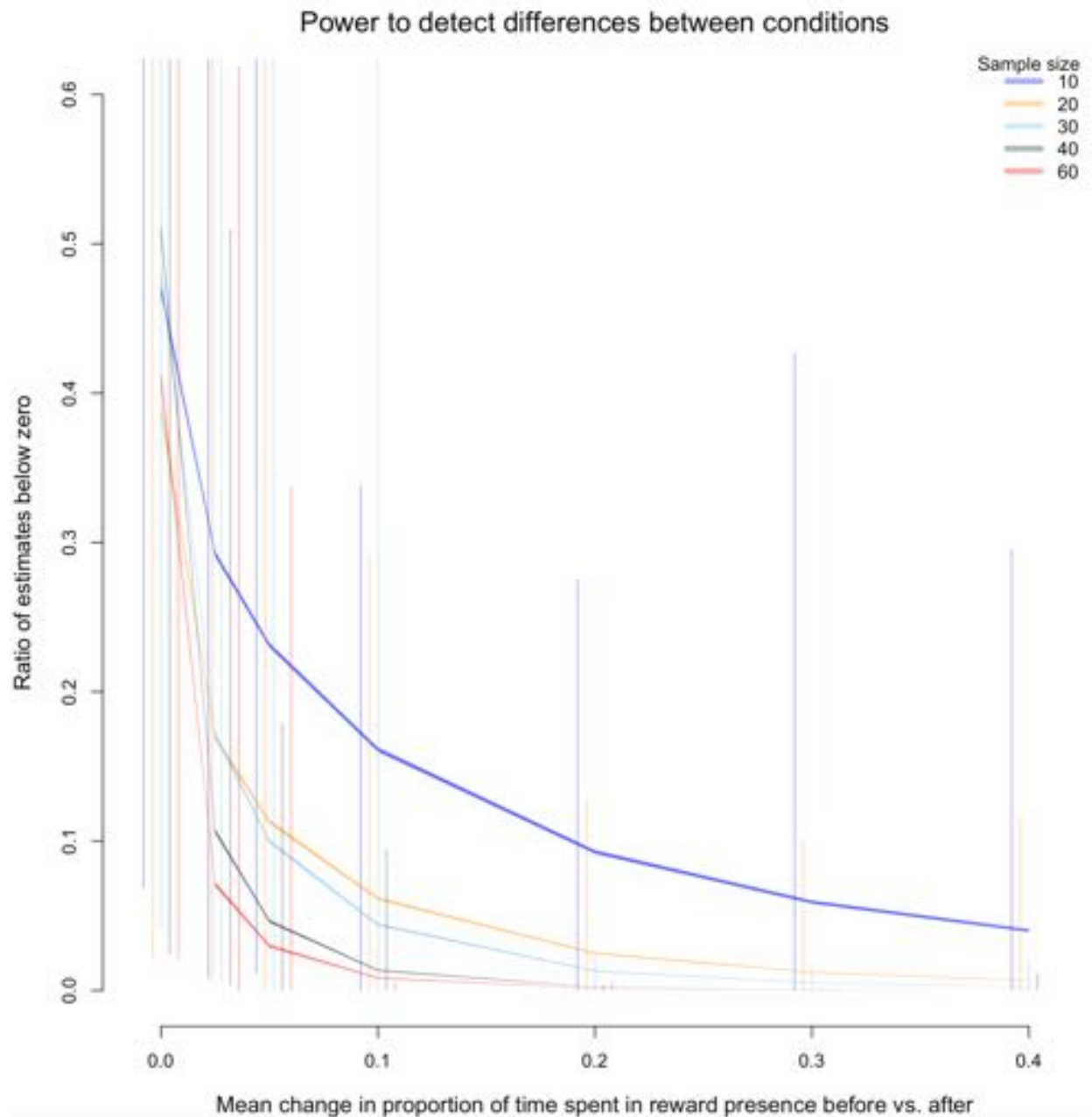


Figure 10. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different

potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other. Across all models, the standard deviation of the mean change in proportion was 0.1 (a conservative estimate).

Code

Run this model on the actual data

Run the code below to determine whether there were differences between the before and after conditions in their proportion of time spent at locations when food was present.

Code

G.Q2 More flexible = use more microhabitats?

The model

Bayesian model with a normal distribution:

$\text{habitatuse} \sim \alpha[\text{ind}] + \beta[\text{ind}] * \text{before}$

habitatuse is the response variable: the total number of different microhabitats used per individual. There will be one intercept, α , and one slope β per individual, which will be estimated for the two conditions, before (and after) the manipulation. ID is nested within condition as a random effect because there is more than one data point per individual: each individual has a data point in the before condition and in the after condition. A normal distribution was used because the response variable is a sum without an expected skew to the curve (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis

See the jay Q3 microhabitat model for the power analysis.

Run this model on the actual data

Run the code below to determine whether there were differences between the before and after conditions in the proportion of microhabitats used. Only count that a microhabitat was used if the individual had at least 5% of their data points there. This prevents a microhabitat from being counted even if an individual was simply moving through it, and therefore not necessarily using it.

Code

G.Q3 More flexible = more food types?

The model

Bayesian model with a normal distribution:

$y \sim \alpha[\text{ind}] + \beta[\text{ind}] * \text{before}$

y is the response variable: the total number of different food types taken per individual. There will be one intercept, α , and one slope β per individual, which will be estimated for the two conditions, before (and after) the manipulation. ID is nested within condition as a random effect because there is more than one data point per individual: each individual has a data point in the

before condition and in the after condition. A normal distribution was used because the response variable is a sum without an expected skew to the curve (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis

See the jay Q4 foraging model for the power analysis.

Run this model on the actual data

Run the code below to determine whether there were differences between the before and after conditions in the number of food types taken.

Code

Jays

J.Q1 Do jay populations in human modified areas differ in baseline behavioral flexibility compared to populations in natural areas?

The model

We used the reversal learning Bayesian model in Logan CJ et al. (2020) to simulate and analyze population differences in reversal learning, and calculate our ability to detect differences between populations. The model “accounts for every choice made in the reversal learning experiment and updates the probability of choosing either option after the choice was made depending on whether that choice contained a food reward or not. It does this by updating three main components for each choice: an attraction score, a learning rate (ϕ), and a rate of deviating from learned attractions (λ)” (Logan CJ et al., 2020).

Equation 1 (attraction and ϕ):

$$A_{i,j,t+1} = (1-\phi_j)A_{i,j,t} + \phi_j \pi_{i,j,t}$$

Equation 1 “tells us how attractions to different behavioral options $A_{i,j,t+1}$ (i.e., how preferable option i is to the bird j at time $t+1$) change over time as a function of previous attractions $A_{i,j,t}$ and recently experienced payoffs $\pi_{i,j,t}$ (i.e., whether they received a reward in a given trial or not). Attraction scores thus reflect the accumulated learning history up to this point. The (bird-specific) parameter ϕ_j describes the weight of recent experience. The higher the value of ϕ_j , the faster the bird updates their attraction. It thus can be interpreted as the *learning or updating rate of an individual*. A value of $\phi_j=0.04$, for example, means that receiving a single reward for one of the two options will shift preferences by 0.02 from initial 0.5-0.5 attractions, a value of $\phi_j=0.06$ will shift preferences by 0.03 and so on” (Blaisdell et al., 2021).

Equation 2 (λ):

$$P(i)_{t+1} = \frac{\exp(\lambda_j A_{i,j,t})}{\sum_{m=1}^2 \exp(\lambda_j A_{m,j,t})}$$

Equation 2 “expresses the probability an individual j chooses option i in the next round, $t+1$, based on the latent attractions. The parameter λ_j represents the *rate of deviating from learned attractions* of an individual (also called inverse temperature). It controls how sensitive choices are to differences in attraction scores. As λ_j gets larger, choices become more deterministic, as it gets smaller, choices become more exploratory (random choice if $\lambda_j=0$). For instance, if an

individual has a 0.6-0.4 preference for option A, a value of $\lambda_j=3$ means they choose A 65% of the time, a value of $\lambda_j=10$ means they choose A 88% of the time and a value of $\lambda_j=0.5$ means they choose A only 53% of the time” (Blaisdell et al., 2021). We used the ϕ_j and λ_j values as the response variable in the Bayesian model to examine whether there were differences in flexibility between the habitats: $y \sim \alpha[\text{habitat}]$

y is the response variable (ϕ_j and λ_j , which are extracted from the correct and incorrect choices in the serial reversals). There is one intercept, α , per habitat (suburban or natural) and we will estimate the habitat’s average and standard deviation of the response variable.

Power analysis

Simulations using bespoke Bayesian models in Logan CJ et al. (2020) (the same model structure we use here) showed a high likelihood of detecting differences with a minimum sample size of 15 when mean differences in phi were at least 0.01 and mean differences in lambda at least 3.

Run this model on the actual data

Run the code below to determine whether there were differences between the two habitats in their phi and lambda flexibility measures.

Code

J.Q2 Are disturbance-resilient jays more flexible than disturbance-resistant jays?

The model

Same as in J.Q1 above.

Power analysis

Same as in J.Q1 above.

Run this model on the actual data

Run the code below to determine whether there were differences between the species in their phi and lambda flexibility measures.

Code

J.Q3 More flexible = use more microhabitats?

The model

Bayesian model with a normal distribution:

$\text{habitatuse} \sim \alpha[\text{ind}] + \beta[\text{ind}]*\text{before}$

habitatuse is the response variable: the total number of different microhabitats used per individual. There will be one intercept, α , and one slope β per individual, which will be estimated for the two conditions, before (and after) the manipulation. ID is nested within condition as a random effect because there is more than one data point per individual: each individual has a data point in the before condition and in the after condition. A normal distribution was used because the response variable is a sum without an expected skew to the curve (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in the proportion of different microhabitats used per individual in the before vs. after conditions (Figure 11). We simulated the proportion of habitats used for different sample sizes of individuals before and after the flexibility manipulation. We analyzed these simulated data with the model we will use to analyze the actual data, estimating the change in the proportion of habitats used between the before and after conditions. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the change in the proportion of habitats used before the flexibility manipulation is similar to after. If the ratio is close to zero, the model assumes individuals have changed their behavior. For changes in the proportion of habitats used smaller than 0.15 (standard deviation=0.2) or 0.1 (SD=0.1), models are likely to assume that no changes occurred even with large sample sizes. If the change in the proportion of habitats used before the flexibility manipulation vs. after is 0.15 with a standard deviation of 0.2, on average 94% of the posterior of the model based on a sample size of 20 individuals will be larger than zero (93% with a standard deviation of 0.1). This means that the model is quite certain there is a difference that is larger than zero. In addition, only four of the 30 models for a sample size of 20 at the mean change of 0.15 have a ratio larger than 0.3, meaning that the risk of having a false negative is not very likely.

In general, with sample sizes at or above 20 and mean changes in the proportion of habitats used at 0.1 or larger, it is highly likely that the model will indicate that individuals have changed their behavior. Mean changes below 0.1 can still be detected, however there is a higher risk that there will be a false negative and this risk is independent of sample size.

With small mean changes in the response variable, some individuals might not increase or even decrease their response after the manipulation because there is variation around the mean change in individual responses. With small sample sizes, there is a risk that only individuals who did not clearly increase their response will be studied, whereas larger sample sizes are more likely to include a wider spectrum of individuals.

To estimate the risk of detecting false positives, we set the mean change in the proportion of habitats used to zero so there was no change between the before and after conditions. As expected, the average ratio of estimates below zero is close to but below 0.5 and independent of sample size. The estimates went generally below 0.5 because the maximum number of habitats used was set to 10 and we had a condition where individuals before the manipulation used a mean of 7 habitats. Accordingly, if individuals randomly either increase or decrease their number of habitats used, decreases will be more severe because individuals can only increase by 3 habitats, but potentially decrease by 6 habitats. With a sample size of 20, 27% have a ratio smaller than 0.3, meaning that the risk of having a false positive is high. The risk would be lower if the variation among individuals was lower than what we assumed (the standard deviation of the mean change in number of habitats was 0.2, which is a conservative estimate).

Code
Code

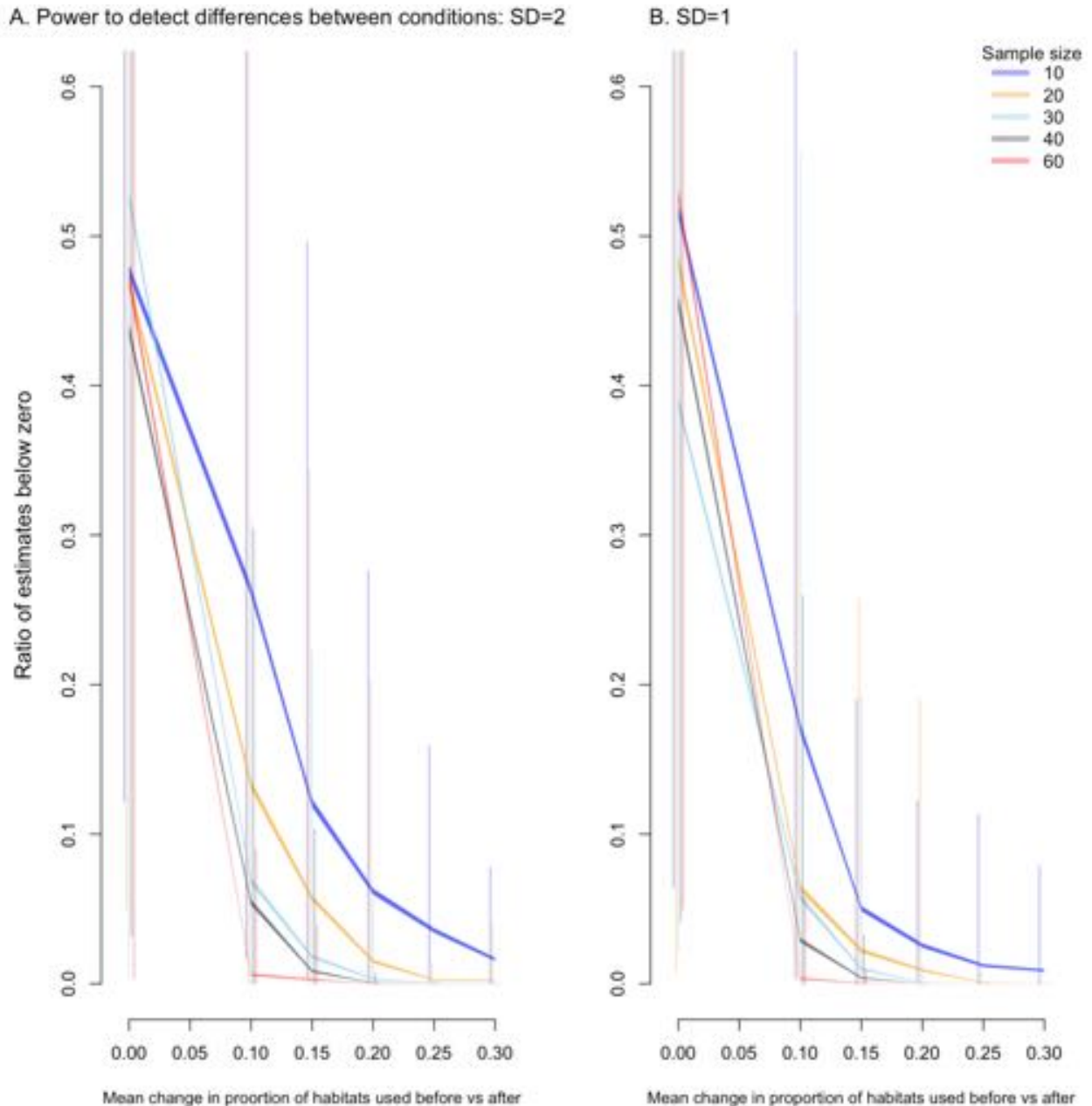


Figure 11. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. Across all models, the standard deviation of the mean change in proportion of habitats used was 0.2 (A) or 0.1 (B). A mean change in proportion of habitats of 0.3 is associated with a difference of 3 habitats (when the maximum number of habitats is 10). The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

Run this model on the actual data

Run the code below to determine whether there were differences between the before and after conditions in the proportion of microhabitats used. Only count that a microhabitat was used if the individual had at least 5% of their data points there. This prevents a microhabitat from being counted even if an individual was simply moving through it, and therefore not necessarily using it.

Code

J.Q4 More flexible = more food types?

The model

Bayesian model with a normal distribution:

$$y \sim \alpha[\text{ind}] + \beta[\text{ind}] * \text{before}$$

y is the response variable: the total number of different food types taken per individual. There will be one intercept, α , and one slope β per individual, which will be estimated for the two conditions, before (and after) the manipulation. ID is nested within condition as a random effect because there is more than one data point per individual: each individual has a data point in the before condition and in the after condition. A normal distribution was used because the response variable is a sum without an expected skew to the curve (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in the total number of different food types taken per individual in the before vs. after conditions (Figure 12). We simulated the number of food types taken for different sample sizes of individuals before and after the flexibility manipulation. We analyzed these simulated data with the model we will use to analyze the actual data, estimating the change in the number of food types taken between the before and after conditions. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the change in the number of food types taken before the flexibility manipulation is similar to after. If the ratio is close to zero, the model assumes individuals have changed their behavior. For changes in the number of food types taken smaller than 1, models are likely to assume that no changes occurred even with large sample sizes. If the change in the number of food types taken before the flexibility manipulation vs. after is 1 with a standard deviation of 2, on average 99.96% of the posterior of the model based on a sample size of 20 individuals will be larger than zero. This means that the model is quite certain there is a difference that is larger than zero. In addition, none of the 30 models for a sample size of 20 at the mean change of 1 have a ratio larger than 0.3, meaning that the risk of having a false negative is not very likely.

In general, with sample sizes at or above 20 and mean changes in the number of food types taken at 1 or larger, it is likely that the model will indicate that individuals have changed their behavior. Mean changes below 1 can still be detected, however there is a higher risk that there will be a false negative and this risk is independent of sample size. For example, 17% of the models for a sample size of 20 at the mean change of 0.5 have a ratio larger than 0.3, meaning

there is a risk of having a false negative.

With small mean changes in the response variable, some individuals might not increase or even decrease their response after the manipulation because there is variation around the mean change in individual responses. With small sample sizes, there is a risk that only individuals who did not clearly increase their response will be studied, whereas larger sample sizes are more likely to include a wider spectrum of individuals.

To estimate the risk of detecting false positives, we set the mean change in the number of food types taken to zero so there was no change between the before and after conditions. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 20, 30% have a ratio smaller than 0.3, meaning that the risk of having a false positive is high. The risk would be lower if the variation among individuals was lower than what we assumed (the standard deviation of the mean change in number of foods was 2, which is a conservative estimate).

Code
Code

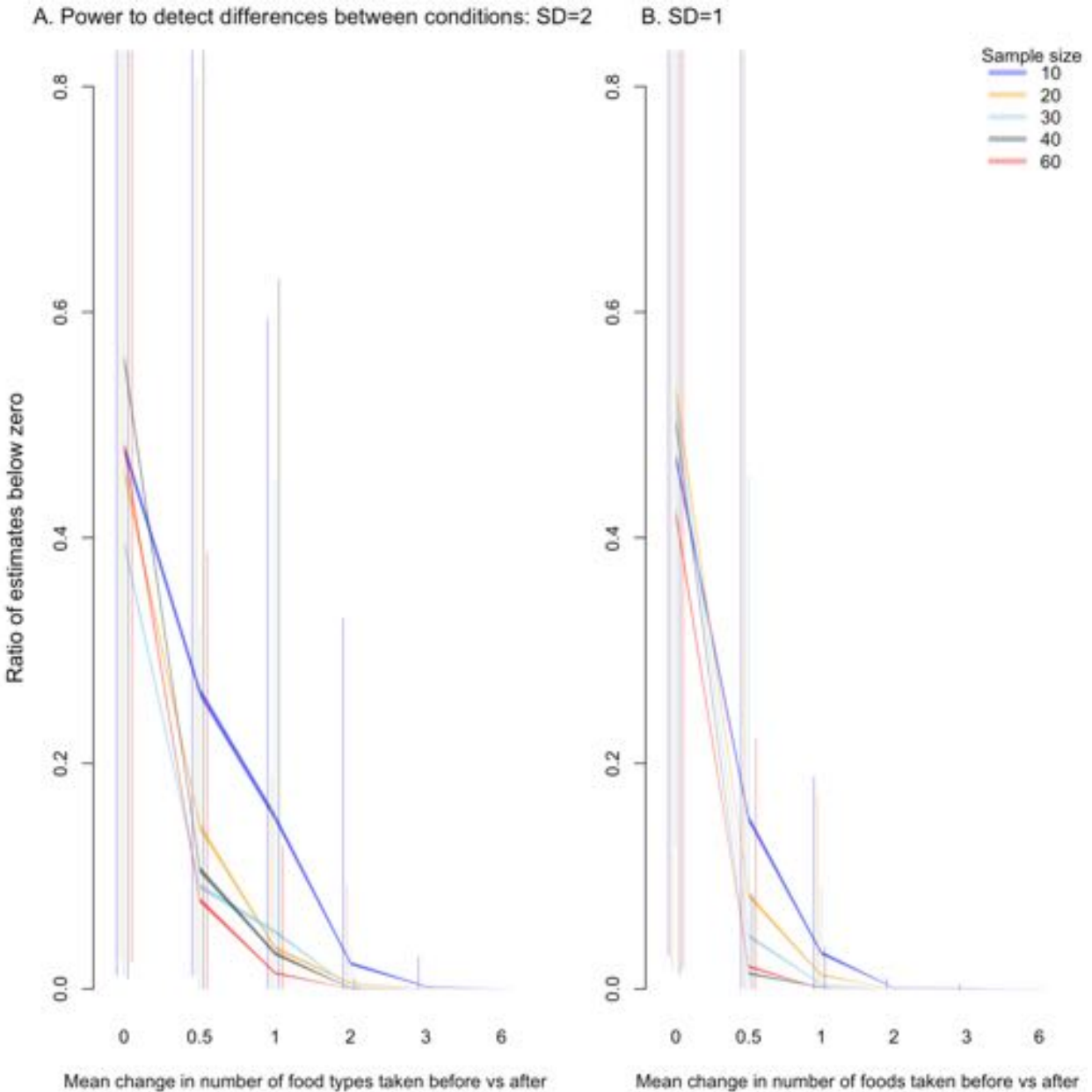


Figure 12. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. Across all models, the standard deviation of the mean change in number of food types taken was 2 (A) or 1 (B), and the number of food types taken before the flexibility manipulation was 6.5. A mean change in proportion of habitats of 0.3 is associated with a difference of 3 habitats (when the maximum number of habitats is 10). The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

Run this model on the actual data

Run the code below to determine whether there were differences between the before and after conditions in the number of food types taken.

Code

Toutouwai

T.Q1 Does a flexibility manipulation alter dispersal timing (1.1) and distance (1.2), as well as the likelihood that juvenile toutouwai will disperse beyond the protection of Zealandia's fence and attempt to establish in the adjacent urban area (1.3)?

The model: dispersal timing (1.1)

Bayesian model for a zero-inflated poisson distribution:

$D_i \sim \text{Zero-inflated Poisson}(\pi_i, \lambda_{d_i})$

D_i is how many days before the cut-off of observations at 112 days individuals dispersed (for example, if a bird dispersed 50 days after fledging, D_i is $112-50=72$); the subtraction covers cases where the age at dispersal would be very late such that some birds might disperse after day 112 post-fledging. The latter individuals will be included with the non-dispersers as having a zero value. λ_{d_i} is the mean of the Poisson distribution describing the average day (and variance) for individuals who have dispersed. Dispersal is defined as the time point at which the juvenile is first detected leaving the natal territory on three consecutive days. A zero inflated Poisson distribution was used because a small percentage of individuals do not disperse in their first season (zero values) and individuals who do disperse have a response variable that is a count number of days post-fledging do so within a relatively short period, which is captured by the Poisson distribution which counts the number of days (McElreath, 2020b). These patterns are described by (Richard, 2007), which shows that only a small minority of individuals does not disperse, and that dispersers were observed to leave their natal territory on average 47.5 days after fledging, with a range between 32-72 days. We assume that both the likelihood to disperse and the day at which dispersers leave might differ between individuals in the manipulated and the control conditions:

$$\log(\pi_i) = \gamma_p[\text{condition}] + \alpha_p[\text{parentid}], [\text{condition}]$$

The two $\gamma[\text{condition}]$ estimates reflect the difference between control and manipulated birds for the probability to not disperse (γ_p) and when to disperse (γ_l). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis (1.1)

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in age at which individuals disperse after fledging (Figure 13). We simulated the age at which individuals would disperse in the control condition (with averages of 36, 48, or 72 days after fledging), and assumed that manipulated individuals might disperse sooner. We analyzed these simulated data with the model we will use to analyze the actual data, estimating the contrast in average time at which individuals disperse. The model also accounts for differences in the likelihood to not disperse at all between control and manipulated birds, but we did not examine this here in the simulation given that the number of individuals

who will not disperse is likely to be very low. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero, indicating that there would be a clear difference between control and manipulated birds.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the age at which individuals disperse is similar for control and manipulated birds. If the ratio is close to zero, the model assumes individuals have changed their behavior and disperse earlier when they were manipulated. We find that for changes smaller than 1 standard deviation (which equals roughly 7 days), models are likely to assume that there are no differences between control and manipulated individuals even with large sample sizes. If the contrast in time of dispersal of manipulated compared to control individuals is 1 standard deviation, on average 92% of the posterior of the model based on a sample size of 20 individuals will be larger than zero. This means that the model is quite certain there is a contrast that is larger than zero.

In general, with sample sizes at or above 20 and mean contrasts between manipulated and control individuals of 1 standard deviation or higher, it is highly likely that the model will indicate that manipulated individuals have changed their behavior. Mean contrasts of less than 1 standard deviation can still be detected, however there is a higher risk that there will be a false negative (27% risk of a false negative with a sample size of 20 when the contrast is 0.25 standard deviations = ~2 days earlier dispersal).

To estimate the risk of detecting false positives, we set the mean change in age of dispersal for manipulated individuals compared to control individuals to zero so there was no change. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 20, <7% of the simulations have a ratio smaller than 0.1, meaning that the risk of having a false positive is low.

Code
Code

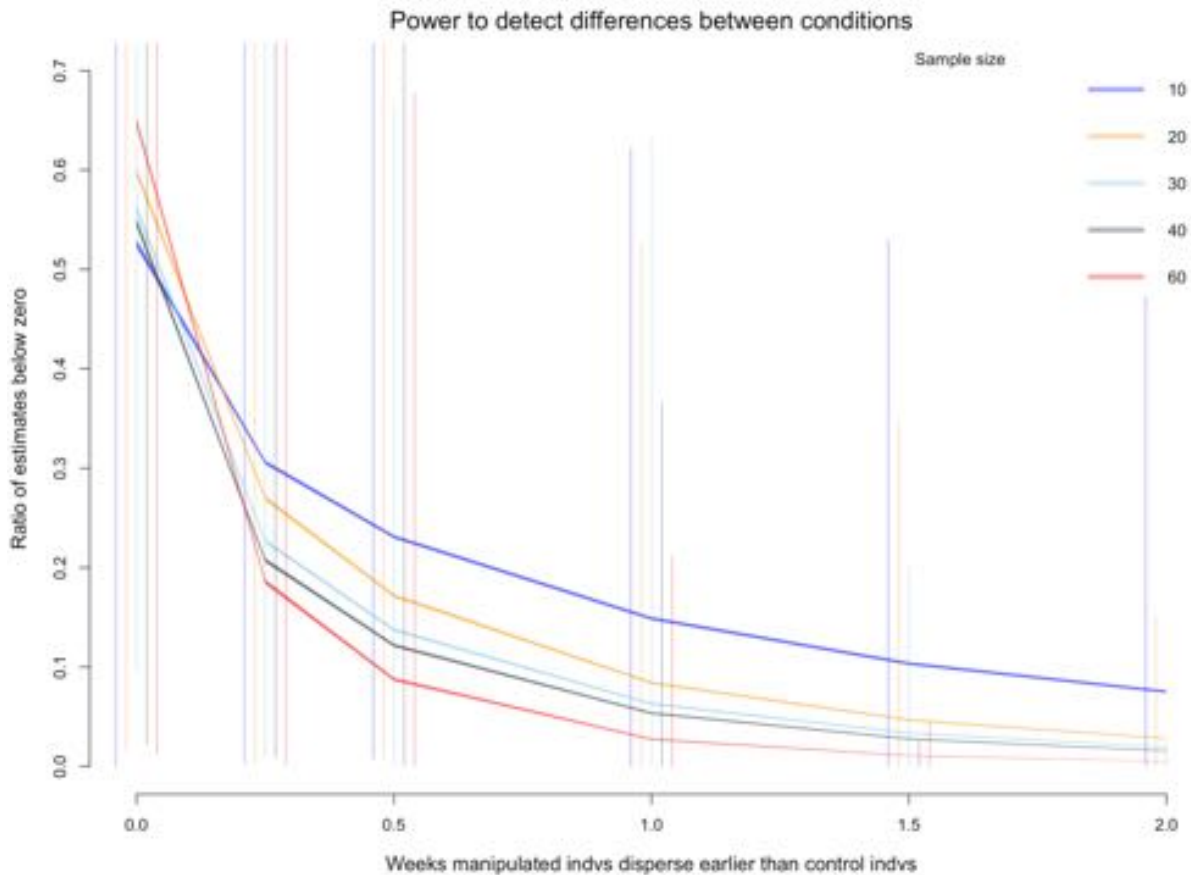


Figure 13. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

Run this model on the actual data (1.1)

Run the code below to determine whether the manipulated individuals differ from the control individuals in their age of dispersal.

Code

The model: dispersal distance (1.2)

Bayesian model for a gamma-poisson distribution:

$D_i \sim \text{Gamma-Poisson}(\lambda_{d_i}, \phi_i)$

D_i is the distance in meters that individuals are found away from their parents' territory on day 112 after they fledged (measured as a straight line between the center of the natal territory and final location the bird is detected at or the point of death/disappearance), $\lambda_{i|c}$ is the mean of the poisson describing the distance individuals who have dispersed are found on average away from their parents' territory, and ϕ_i controls the variance. A gamma-poisson distribution was used because the dispersal distance is a non-negative count, almost all individuals disperse, but the distribution is skewed. Most individuals disperse not very far, but a small proportion can disperse for large distances. These patterns are described by (Richard, 2007), which shows that the average dispersal distance $\sim 1,000\text{m}$ and the maximum dispersal distance is $\sim 10,000\text{m}$. We assume that distance dispersers move might differ between individuals in the manipulated and the control conditions:

$$\log(\lambda_{i|c}) = \gamma[\text{condition}]$$

The $\gamma[\text{condition}]$ reflects the average distance individuals disperse for each condition (control/manipulated) to disperse. The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis (1.2)

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in distance individuals disperse (Figure 14). We simulated the distance individuals would disperse in the control condition (with averages of 500m, 1000m, or 1500m), and assumed that manipulated individuals might disperse up to 1000m farther. We analyzed these simulated data with the model we will use to analyze the actual data, estimating the contrast in average distance individuals disperse. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero, indicating that there would be a clear difference between control and manipulated birds.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the distance individuals disperse is similar for control and manipulated birds. If the ratio is close to zero, the model assumes individuals have changed their behavior and disperse earlier when they were manipulated. We find that for changes smaller than 300m, models are likely to assume that there are no differences between control and manipulated individuals even with large sample sizes. Also for the larger increases in dispersal distance, we would need sample sizes of at least 40 (for a 1000m increase) or 60 individuals (for a 300m increase) to reliably detect differences between control and manipulated individuals. If the contrast in dispersal distance of manipulated compared to control individuals is 500m, on average 93% of the posterior of the model based on a sample size of 60 individuals will be larger than zero (81% with a sample size of 20). This means that the model is quite certain there is a contrast that is larger than zero.

In general, with sample sizes at or above 40 and mean contrasts between manipulated and control individuals of 300m or more, it is likely that the model will indicate that manipulated individuals have changed their behavior. Mean contrasts of less than 300m can still be detected, however there is a high risk that there will be a false negative (67% risk of a false negative with a sample size of 40 when the contrast is 200m).

To estimate the risk of detecting false positives, we set the mean change in age of dispersal for manipulated individuals compared to control individuals to zero so there was no change. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 40, <7% of the simulations have a ratio smaller than 0.1, meaning that the risk of having a false positive is low.

Code
Code

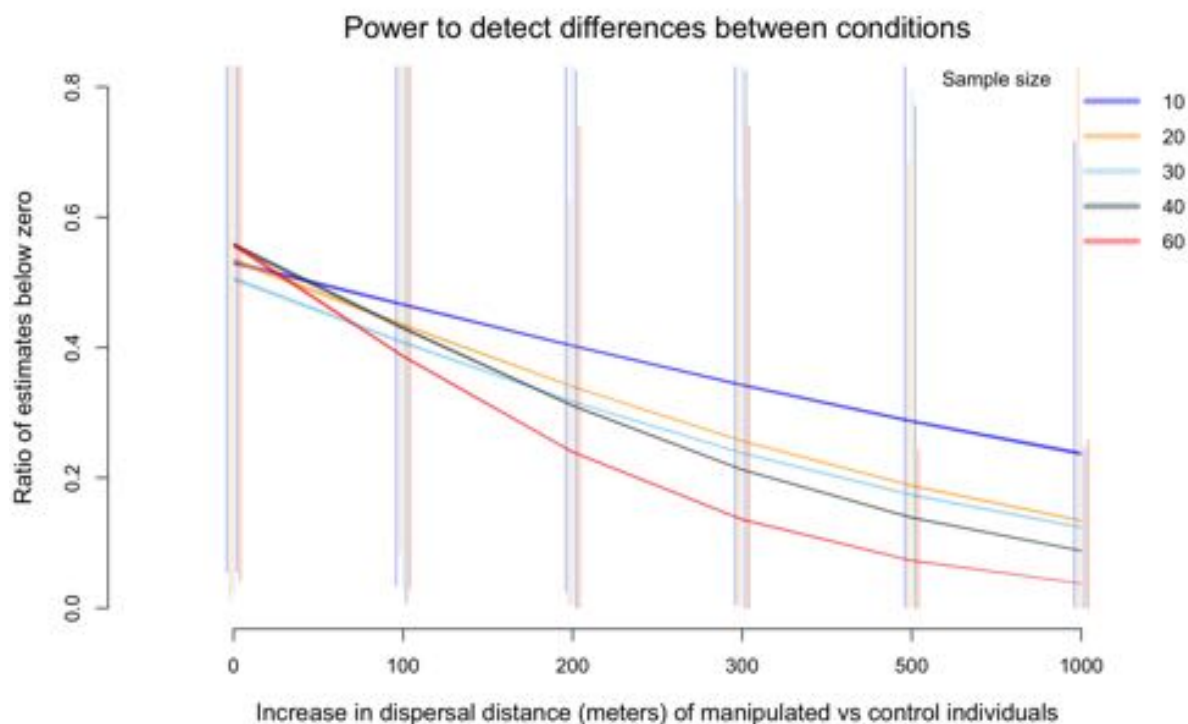


Figure 14 Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

Run this model on the actual data (1.2)

Run the code below to determine whether there were differences between the control and manipulated conditions in the distance of dispersal.

Code

The model: dispersal location (1.3)

```
location ~ dbinom(1,p)
```

```
logit(p) <- g[condition]
```

The response variable is dispersal location (0=inside Zealandia, 1=outside Zealandia) on day

112 post-fledging. $g[\text{condition}]$ is the average log odds for each condition (control or flexibility manipulated). A binomial distribution was used because the response variable is binomial (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Power analysis (1.3)

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in the dispersal location in the control vs. manipulated conditions (Figure 15). We analyzed simulated data with the model we will use to analyze the actual data, estimating the change in dispersal location between the conditions. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the change in the number of food types taken before the flexibility manipulation is similar to after. If the ratio is close to zero, the model assumes individuals have changed their behavior. For differences in the dispersal location at or larger than 0.3, models are likely to assume that no changes occurred even with large sample sizes. If the difference in dispersal location between control and manipulated conditions is 0.3, on average 92% of the posterior of the model based on a sample size of 40 individuals will be larger than zero. This means that the model is quite certain there is a difference that is larger than zero.

In general, with sample sizes at or above 40 and mean differences in the dispersal location of 0.3 or larger, it is likely that the model will indicate that the flexibility experiment influenced their behavior in a way that affected their dispersal location. Mean changes below 0.3 can still be detected, however there is a higher risk that there will be a false negative. For example, 17% of the models for a sample size of 40 at the mean change of 0.2 have a ratio larger than 0.3, meaning there is a large risk of having a false negative.

With small mean changes in the response variable, some individuals might not increase or even decrease their response after the manipulation because there is variation around the mean change in individual responses. With small sample sizes, there is a risk that only individuals who did not clearly increase their response will be studied, whereas larger sample sizes are more likely to include a wider spectrum of individuals.

To estimate the risk of detecting false positives, we set the mean change in survival to zero so there was no change between the conditions. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 40, 60% have a ratio smaller than 0.3, meaning that the risk of having a false positive is high.

Code

Code

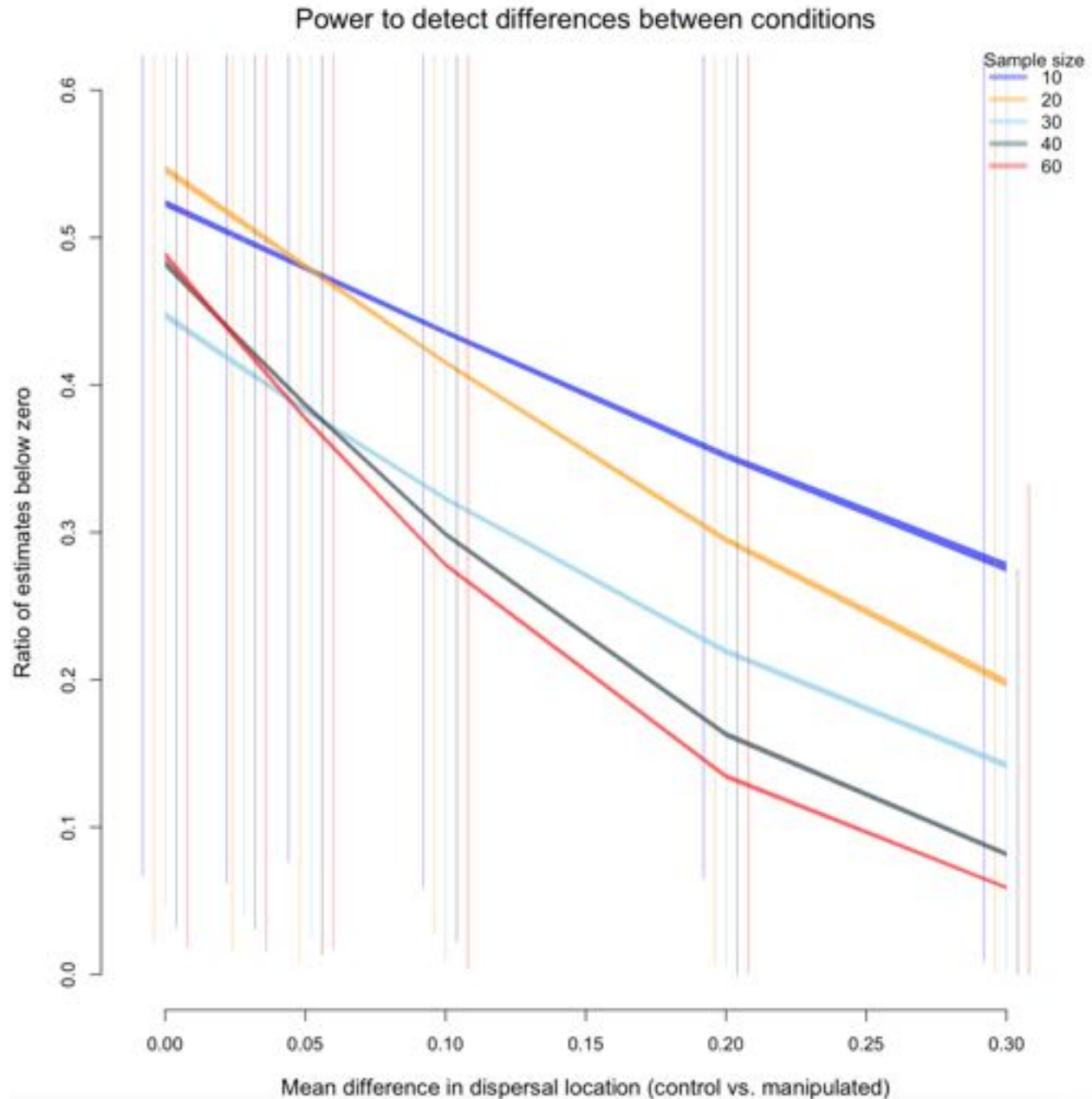


Figure 15. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

Run this model on the actual data (1.3)

Run the code below to determine whether there were differences between the control and manipulated conditions in their survival status (alive or dead when individuals are or would be 112 days old).

Code

T.Q2 Compared to control individuals, are flexibility manipulated individuals more likely to survive their first 16 weeks post-fledging, particularly if they disperse into the urban reserves outside the sanctuary fence?

The model

Bayesian models with a binomial distribution:

Examining the difference between conditions:

```
status ~ dbinom(1,p)
```

```
logit(p) <- g[condition]
```

The response variable is survival status (0=dead, 1=alive) on day 112 post-fledging. $g[\text{condition}]$ is the average log odds for each condition (control or flexibility manipulated). A binomial distribution was used because the response variable is binomial (see Figure 10.6 in McElreath, 2020b). The Bayesian model was developed using McElreath (2020b) as a guide.

Examining the difference between conditions in each of the two areas:

```
status ~ dbinom(1,p)
```

```
logit(p) <- g[habitat] +  $\beta$ [habitat,condition]
```

The response variable is survival status (0=dead, 1=alive) on day 112 post-fledging. $g[\text{habitat}]$ is the average log odds for each area (inside Zealandia, which is a natural area, or outside Zealandia, which can include natural/suburban/and urban areas), and $\beta[\text{habitat,condition}]$ is an effect for each area (inside Zealandia or outside Zealandia) in each condition (control or flexibility manipulated).

Power analysis

We estimated our power to detect differences between conditions at different sample sizes and with different mean changes in the survival status in the control vs. manipulated conditions (Figure 16). We analyzed simulated data with the model we will use to analyze the actual data, estimating the change in survival status between the conditions. From the posterior estimates of the model, we extracted both the mean change as well as the ratio of the posterior estimates that were below zero.

If the mean ratio of estimates below zero is close to 0.5, the model assumes that the change in the number of food types taken before the flexibility manipulation is similar to after. If the ratio is close to zero, the model assumes individuals have changed their behavior. For differences in the survival status at or larger than 0.3, models are likely to assume that no changes occurred even with large sample sizes. If the difference in survival status between control and manipulated conditions is 0.3, on average 88% of the posterior of the model based on a sample size of 30 individuals will be larger than zero. This means that the model is quite certain there is a difference that is larger than zero.

In general, with sample sizes at or above 30 and mean differences in the survival status of 0.3 or larger, it is likely that the model will indicate that the flexibility experiment influenced their behavior in a way that affected their survival. Mean changes below 0.3 can still be detected, however there is a higher risk that there will be a false negative. For example, 17% of the models for a sample size of 30 at the mean change of 0.2 have a ratio larger than 0.3, meaning there is a large risk of having a false negative.

With small mean changes in the response variable, some individuals might not increase or even decrease their response after the manipulation because there is variation around the mean change in individual responses. With small sample sizes, there is a risk that only individuals who did not clearly increase their response will be studied, whereas larger sample sizes are more likely to include a wider spectrum of individuals.

To estimate the risk of detecting false positives, we set the mean change in survival to zero so there was no change between the conditions. As expected, the average ratio of estimates below zero is close to 0.5 and independent of sample size. With a sample size of 30, 17% have a ratio smaller than 0.3, meaning that the risk of having a false positive is high.

We also estimated the power to detect differences between conditions in different habitats (inside or outside Zealandia) at different sample sizes and with different mean changes in the survival status (Figure 17). We found that mean changes in survival status of 0.3 are likely to be detected with a sample size of at least 40 when considering the individuals inside Zealandia, but we will not be able to detect differences outside of Zealandia because the subset sample size will likely be too small.

Code
Code

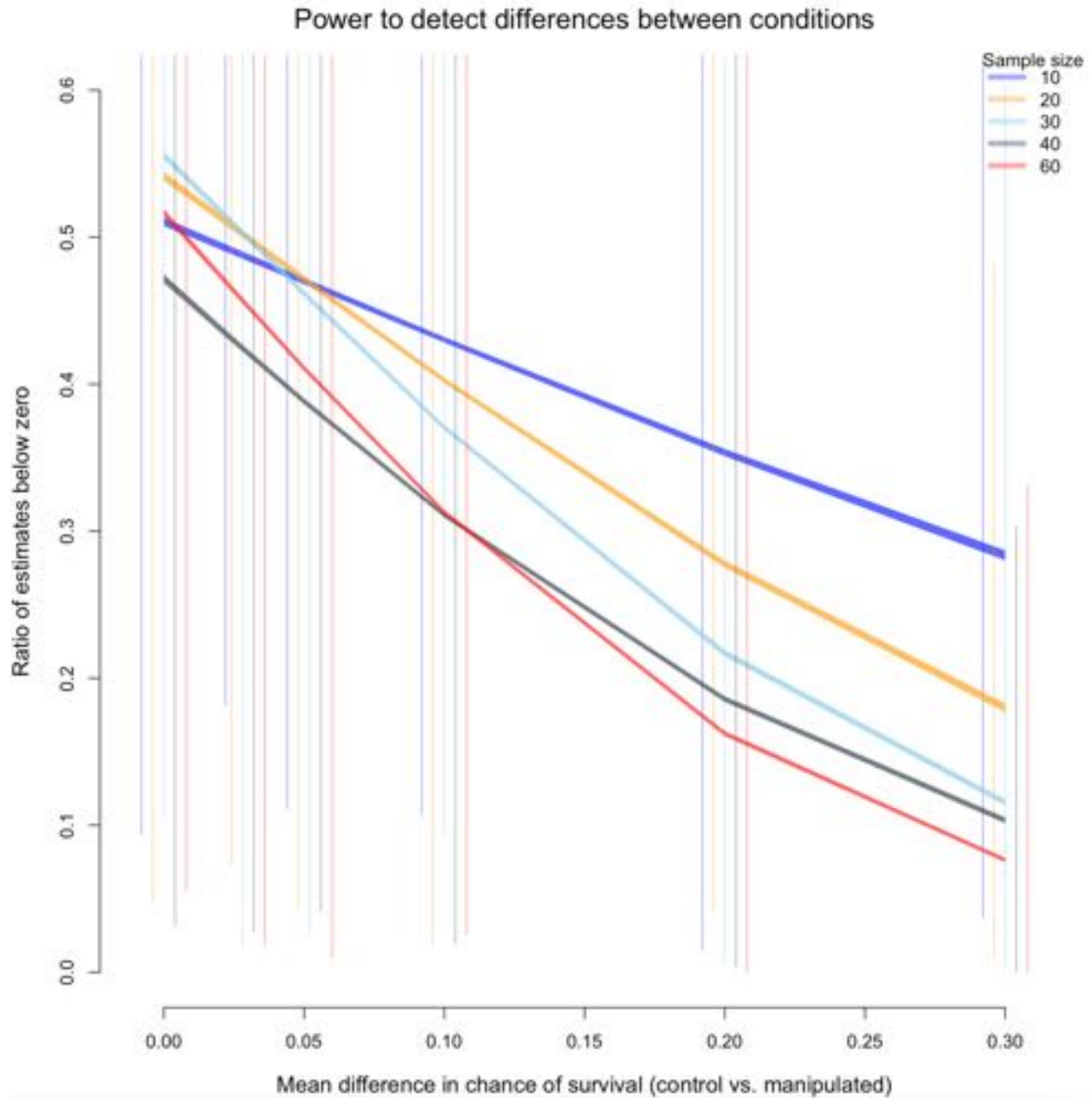


Figure 16. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions at different sample sizes. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each other.

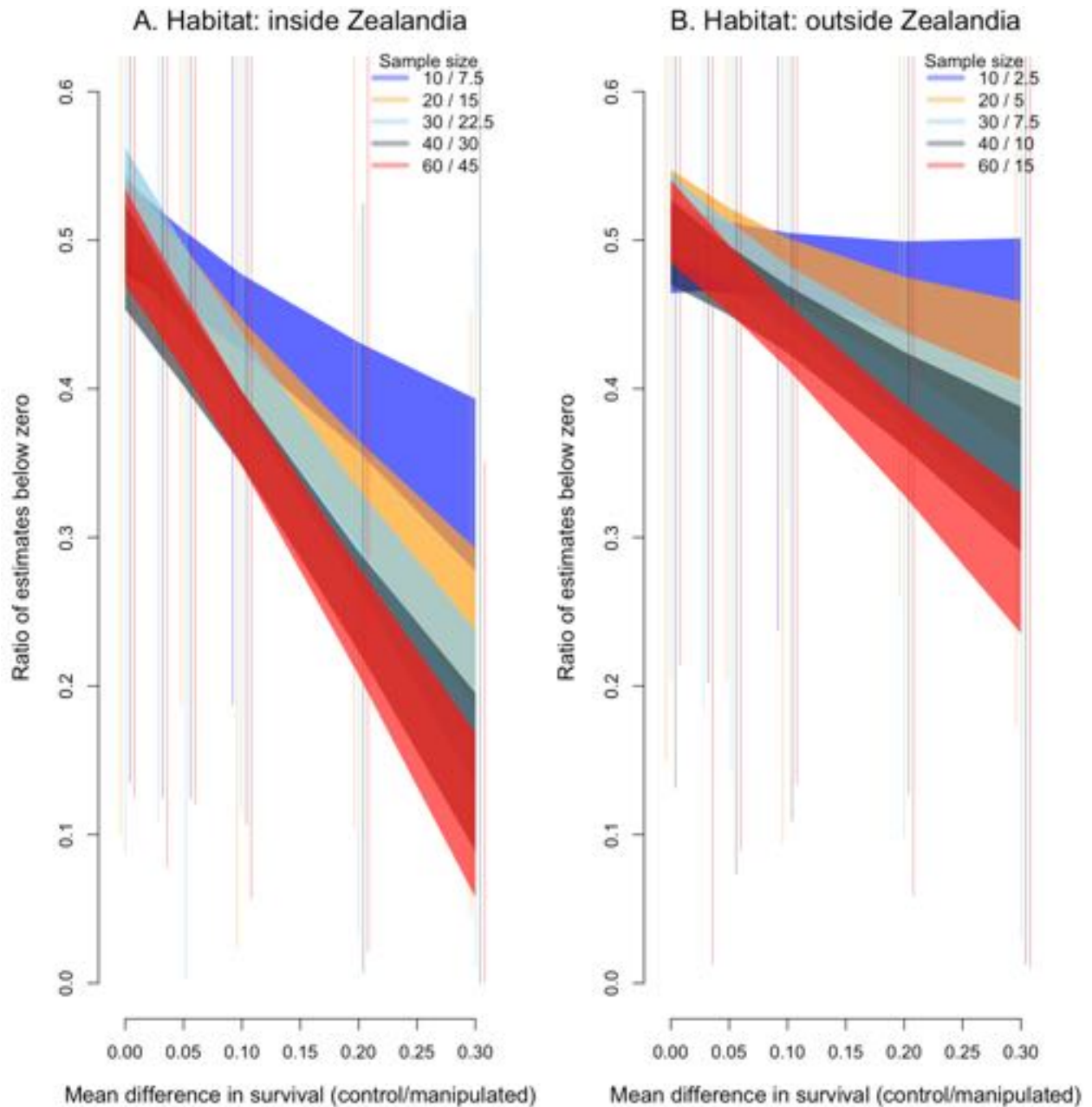


Figure 17. Risk of false positives and false negatives depending on sample and effect sizes. Curves of the mean ratio of estimates below zero (vertical lines show the minimum and maximum ratios), which illustrates the power we have to detect differences between conditions (control vs. manipulated) in different habitats (inside or outside Zealania) at different sample sizes. The sample size was set at 75% inside Zealania and 25% outside Zealania and the associated numbers with these percentages are shown for each overall sample size in the legend. The curves show the model estimates as the effect increases (larger changes in the mean proportion time spent after versus before on the x-axis) for different potential sample sizes (10-60, illustrated by different colors). When there is no change (x value of zero), estimates suggest that, as expected, half of the estimates are below zero because the before and after conditions are not different from each other. As the change increases, the estimates decrease because models are able to reliably tell that the before and after conditions differ from each

other.

Run this model on the actual data

Run the code below to determine whether there were differences between the control and manipulated conditions in their survival status (alive or dead when individuals are or would be 112 days old).

Code

ETHICS

This research is carried out in accordance with permits from the:

1. US Fish and Wildlife Service (scientific collecting permit number MBPER0039225)
2. US Geological Survey Bird Banding Laboratory (federal bird banding permit number 23872 for grackles and 24273 for jays)
3. Institutional Animal Care and Use Committee at the University of California Santa Barbara (protocol number 958)
4. California Department of Fish and Wildlife (scientific collecting permit [specific use] number S-192100001-19210-001 for grackles and S-192100002-20329-001 for jays)
5. Oregon Department of Fish and Wildlife (scientific collecting permit 125-22)
6. The New Zealand Department of Conservation (permit number 97554-FAU)

AUTHOR CONTRIBUTIONS

Logan: Hypothesis development, data collection, data analysis and interpretation, write up, revising/editing, materials/funding.

Shaw: Hypothesis development, data collection, data analysis and interpretation, write up, revising/editing, materials/funding.

Lukas: Analysis design and code, sample size modeling, validated reversal passing criteria, write up, revising/editing

McCune: Hypothesis development, data collection, data analysis and interpretation, write up, revising/editing.

FUNDING

This research is funded by the Department of Human Behavior, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology (to Logan) and the Rutherford Discovery Fellowship from the Royal Society Te Apārangi (to Shaw).

CONFLICT OF INTEREST DISCLOSURE

We, the authors, declare that we have no financial conflicts of interest with the content of this article. CJ Logan is a co-founder of and on the Managing Board at PCI Registered Reports.

ACKNOWLEDGEMENTS

We are grateful to our funders: Richard McElreath at the Max Planck Institute for Evolutionary Anthropology (whole project) and the Rutherford Discovery Fellowship from the Royal Society Te Apārangi (Shaw).

REFERENCES

- Alberti, M. (2015). Eco-evolutionary dynamics in an urbanizing planet. *Trends in Ecology & Evolution*, 30(2), 114–126.
- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49(3-4), 227–266.
- Blair, R. B. (1996). Land use and avian species diversity along an urban gradient. *Ecological Applications*, 6(2), 506–519.
- Blaisdell, A., Seitz, B., Rowney, C., Folsom, M., MacPherson, M., Deffner, D., & Logan, C. J. (2021). *Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles (version 5 of this preprint has been peer reviewed and recommended by peer community in ecology [https://doi.org/10.24072/pci.ecology.100076]).* <https://doi.org/10.31234/osf.io/z4p6s>
- 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. <https://doi.org/10.1037/0735-7036.121.4.372>
- Chejanovski, Z. A., Avilés-Rodríguez, K. J., Lapiedra, O., Preisser, E. L., & Kolbe, J. J. (2017). An experimental evaluation of foraging decisions in urban and natural forest populations of anolis lizards. *Urban Ecosystems*, 20(5), 1011–1018.
- Ciani, A. C. (1986). Intertroop agonistic behavior of a feral rhesus macaque troop ranging in town and forest areas in India. *Aggressive Behavior*, 12(6), 433–439.
- Curry, R., Townsend Peterson, A., & Langen, T. (2017). *California scrub-jay (Aphelocoma californica)*, *Birds of North America*.
- Daniel, M. M., & Schluessel, V. (2020). Serial reversal learning in freshwater stingrays (potamotrygon motoro). *Animal Cognition*, 23(1), 109–119.
- Dingemanse, N. J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within populations: Causes and consequences. *Animal Behaviour*, 85(5), 1031–1039.

- Duckworth, R. A. (2009). The role of behavior in evolution: A search for mechanism. *Evolutionary Ecology*, 23(4), 513–531.
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, 306(5703), 1903–1907.
- Federspiel, I. G., Garland, A., Guez, D., Bugnyar, T., Healy, S. D., Güntürkün, O., & Griffin, A. S. (2017). Adjusting foraging strategies: A comparison of rural and urban common mynas (*acridotheres tristis*). *Animal Cognition*, 20(1), 65–74.
- Gabry, J., & Češnovar, R. and. (2022). *Cmdstanr: R interface to 'CmdStan'*. <https://mc-stan.org/cmdstanr/>
- Galbraith, J. A., Beggs, J. R., Jones, D. N., & Stanley, M. C. (2015). Supplementary feeding restructures urban bird communities. *Proceedings of the National Academy of Sciences*, 112(20), E2648–E2657.
- Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package "irr." *Various Coefficients of Interrater Reliability and Agreement*, 22.
- Goldewijk, K. K. (2001). Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles*, 15(2), 417–433.
- Howard, W. E. (1960). Innate and environmental dispersal of individual vertebrates. *American Midland Naturalist*, 152–161.
- Hutcheon, J. A., Chiolerio, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *Bmj*, 340, c2289.
- Johnson, K., DuVal, E., Kiehl, M., & Hughes, C. (2000). Male mating strategies and the mating system of great-tailed grackles. *Behavioral Ecology*, 11(2), 132–141.
- Johnson, K., & Peer, B. D. (2001). *Great-tailed grackle: Quiscalus mexicanus*. Birds of North America, Incorporated.
- Jolly, C. J., Kelly, E., Gillespie, G. R., Phillips, B., & Webb, J. K. (2018). Out of the frying pan: Reintroduction of toad-smart northern quolls to southern kakadu national park. *Austral Ecology*, 43(2), 139–149.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
- Lapiedra, O., Chejanovski, Z., & Kolbe, J. J. (2017). Urbanization and biological invasion shape animal personalities. *Global Change Biology*, 23(2), 592–603.
- Lee, V. E., & Thornton, A. (2021). Animal cognition in an urbanised world. *Frontiers in Ecology and Evolution*, 9, 120.
- Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A. D., Chen, A. others. (2020). High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 1–7.
- Liu, Y., Day, L. B., Summers, K., & Burmeister, S. S. (2016). Learning to learn: Advanced behavioural flexibility in a poison frog. *Animal Behaviour*, 111, 167–172.
- Logan, C. J. (2016). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, 4, e1975.
- Logan, C. J., Lukas, D., Bergeron, L., Folsom, M., & McCune, K. B. (2019). Is behavioral flexibility related to foraging and social behavior in a rapidly expanding species? *In Principle Acceptance by PCI Ecology of the Version on 6 Aug 2019*. http://corinalogan.com/Preregistrations/g_flexforaging.html

- Logan, C. J., MacPherson, M., Rowney, C., Bergeron, L., Seitz, B., Blaisdell, A., Folsom, M., Johnson-Ulrich, Z., & McCune, K. B. (2019). Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem solving in a new context? *In Principle Acceptance by PCI Ecology of the Version on 26 Mar 2019*.
http://corinalogan.com/Preregistrations/g_flexmanip.html
- Logan, C., Blaisdell, A., Johnson-Ulrich, Z., Lukas, D., MacPherson, M., Seitz, B., Sevchik, A., & McCune, K. (2022). Behavioral flexibility is manipulatable and it improves flexibility and problem solving in a new context. *EcoEvoRxiv*.
<https://doi.org/https://doi.org/10.32942/osf.io/5z8xs>
- Logan, CJ, McCune, KB, Breen, A, Chen, N, & Lukas, D. (2020). Implementing a rapid geographic range expansion - the role of behavior and habitat changes. *In Principle Acceptance by PCI Ecology of the Version on 6 Oct 2020*.
<http://corinalogan.com/Preregistrations/gxpopbehaviorhabitat.html>
- Lucon-Xiccato, T., & Bisazza, A. (2014). Discrimination reversal learning reveals greater female behavioural flexibility in guppies. *Biology Letters*, 10(6), 20140206.
- Mackintosh, N., Mcgonigle, B., & Holgate, V. (1968). Factors underlying improvement in serial reversal learning. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 22(2), 85.
- McCune, K. B. (2018). *Cognition gone wild: A test of the social intelligence hypothesis in wild birds* [PhD thesis].
- McCune, KB, Folsom, M, Ross, C, Bergeron, L, & Logan, C. (2020). Does great-tailed grackle space use behavior reflect individual differences in exploration? *In Principle Acceptance by PCI Ecology of the Version on 23 Sep 2020*.
<http://corinalogan.com/Preregistrations/gspaceuse.html>
- McCune, KB, MacPherson, M, Rowney, C, Bergeron, L, Folsom, M, & Logan, C. (2019). Is behavioral flexibility linked with exploration, but not boldness, persistence, or motor diversity? *In Principle Acceptance by PCI Ecology of the Version on 27 Mar 2019*.
http://corinalogan.com/Preregistrations/g_exploration.html
- McElreath, R. (2020a). *Rethinking: Statistical rethinking book package*.
- McElreath, R. (2020b). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.
- Midford, P. E., Hailman, J. P., & Woolfenden, G. E. (2000). Social learning of a novel foraging patch in families of free-living florida scrub-jays. *Animal Behaviour*, 59(6), 1199–1207.
- Mikhalevich, I., Powell, R., & Logan, C. (2017). Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*, 7(3), 20160121. <https://doi.org/10.1098/rsfs.2016.0121>
- Miskelly, C., Empson, R., & Wright, K. (2005). Forest birds recolonising Wellington. *Notornis*, 52(1), 21–26.
- Moseby, K. E., Blumstein, D. T., & Letnic, M. (2016). Harnessing natural selection to tackle the problem of prey naïveté. *Evolutionary Applications*, 9(2), 334–343.
- Moseby, K. E., Cameron, A., & Crisp, H. A. (2012). Can predator avoidance training improve reintroduction outcomes for the greater bilby in arid australia? *Animal Behaviour*, 83(4), 1011–1021.

- Pasinelli, G., Schiegg, K., & Walters, J. R. (2004). Genetic and environmental influences on natal dispersal distance in a resident bird species. *The American Naturalist*, 164(5), 660–669.
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Rice, N. H., Martinez-Meyer, E., & Peterson, A. T. (2003). Ecological niche differentiation in the aphelocoma jays: A phylogenetic perspective. *Biological Journal of the Linnean Society*, 80(3), 369–383.
- Richard, Y. (2007). *Demography and distribution of the north island robin (petroica longipes) in a fragmented agricultural landscape of new zealand: A thesis presented in partial fulfilment of the requirements for the degree of doctorate of philosophy in ecology at massey university, palmerston north, new zealand* [PhD thesis]. Massey University.
- Richard, Y., & Armstrong, D. P. (2010). Cost distance modelling of landscape connectivity and gap-crossing ability using radio-tracking data. *Journal of Applied Ecology*, 47, 603–610.
- Robertson, H. A., Baird, K., Dowding, J. E., Elliott, G. P., Hitchmough, R. A., Miskelly, C. M., McArthur, N., O'Donnell, C. F. J., Sagar, P. M., Scofield, R. P., & Taylor, G. A. (2016). Conservation status of New Zealand birds 2016. *New Zealand Threat Classification Series 19*, 26 p.
<http://www.doc.govt.nz/upload/documents/science-and-technical/nztcs4entire.pdf>
- Rodrigo, A., Ávila-Chauvet, L., & Buriticá, J. (2021). Daily patterns of foraging and aggressive behaviors in great-tailed grackle (*quiscallus mexicanus*) at an urban patch with availability or absence of resources. *bioRxiv*.
<https://doi.org/10.1101/2021.06.14.448443>
- Ross, A. K., Letnic, M., Blumstein, D. T., & Moseby, K. E. (2019). Reversing the effects of evolutionary prey naiveté through controlled predator exposure. *Journal of Applied Ecology*, 56(7), 1761–1769.
- Saunders, A., & Norton, D. A. (2001). Ecological restoration at Mainland Islands in New Zealand. *Biological Conservation*, 99(1), 109–119. [https://doi.org/10.1016/S0006-3207\(00\)00192-0](https://doi.org/10.1016/S0006-3207(00)00192-0)
- Sevchik, A, Logan, CJ, Bergeron, L, Blackwell, A, Rowney, C, & Lukas, D. (2019). Investigating sex differences in genetic relatedness in great-tailed grackles in tempe, arizona to infer potential sex biases in dispersal. *In Principle Acceptance by PCI Ecology of the Version on 29 Nov 2019*. <http://corinalogan.com/Preregistrations/gdispersal.html>
- 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.
- Shaw, R. C., & Harvey, A. (2018). *Toutouwai population monitoring and community engagement in WCC Reserves adjoining Zealandia , 2017-2018. A report submitted to the Wellington City Council* (pp. 1–15). Wellington City Council.
- Shaw, R. C., & Harvey, A. (2020). Long term memory for a learned behaviour in a wild bird. *Biology Letters*, 16, 20190912. <https://doi.org/https://doi.org/10.1098/rsbl.2019.0912>

- Shawkey, M. D., Bowman, R., & Woolfenden, G. E. (2004). Why is brood reduction in florida scrub-jays higher in suburban than in wildland habitats? *Canadian Journal of Zoology*, 82(9), 1427–1435.
- Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, 63(3), 495–502.
- Stan Development Team. (2020). *RStan: The R interface to Stan*. <http://mc-stan.org/>
- Strang, C. G., & Sherry, D. F. (2014). Serial reversal learning in bumblebees (*bombus impatiens*). *Animal Cognition*, 17(3), 723–734.
- Tetzlaff, S. J., Sperry, J. H., & DeGregorio, B. A. (2019). Effects of antipredator training, environmental enrichment, and soft release on wildlife translocations: A review and meta-analysis. *Biological Conservation*, 236, 324–331.
- Tuomainen, U., & Candolin, U. (2011). Behavioural responses to human-induced environmental change. *Biological Reviews*, 86(3), 640–657.
- Valente, J. J., LeGrande-Rolls, C. L., Rivers, J. W., Tucker, A. M., Fischer, R. A., & Betts, M. G. (2021). Conspecific attraction for conservation and management of terrestrial breeding birds: Current knowledge and future research directions. *The Condor*, 123(2), duab007.
- Wehtje, W. (2003). The range expansion of the great-tailed grackle (*quiscalus mexicanus gmelin*) in north america since 1880. *Journal of Biogeography*, 30(10), 1593–1607.
- West, R., Letnic, M., Blumstein, D. T., & Moseby, K. E. (2018). Predator exposure improves anti-predator responses in a threatened mammal. *Journal of Applied Ecology*, 55(1), 147–156.
- Woolfenden, G. E., & Fitzpatrick, J. W. (1996). *Florida scrub-jay (aphelocoma coerulescens)*, *birds of north america*.
- Wright, T. F., Eberhard, J. R., Hobson, E. A., Avery, M. L., & Russello, M. A. (2010). Behavioral flexibility and species invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, 22(4), 393–404.
- Wu, J., Jenerette, G. D., Buyantuyev, A., & Redman, C. L. (2011). Quantifying spatiotemporal patterns of urbanization: The case of the two fastest growing metropolitan regions in the united states. *Ecological Complexity*, 8(1), 1–8.
- Xie, Y. (2018). Knitr: A comprehensive tool for reproducible research in r. In *Implementing reproducible research* (pp. 3–31). Chapman; Hall/CRC.