Choice experiments for ecosystems and wild birds: A review

Trudy Ann Cameron
January 8, 2023

Abstract

In this detailed review of about two dozen published choice experiments concerning ecosystems and wild birds, I highlight differences across these studies that affect their suitability for benefits-function transfer. Survey-based choice experiments are often used to reveal the types of trade-offs that people are willing to make among the attributes that describe some different alternatives they may face. Studies can be designed merely to illuminate the tradeoffs willingly made by the sampled respondents in their particular context. However, to maximize the usefulness of an expensive survey-based choice experiment, it is important to design the study in a way that maximizes its value for future exercises in "benefits-function transfer." Representativeness of the "study sample" is a crucial issue, as may be methods for systematic sample selection correction. Heterogeneous preferences should be accommodated, and not solely through random-parameters utility-function specifications. Estimated marginal utilities for specific attributes should be allowed to vary systematically with observable characteristics of respondents and/or their neighborhoods, so that benefits-function transfer can accommodate systematic differences in preferences to be expected when the mix of population characteristics differs between the study context and other contexts. Across the set of choice-experiment studies reviewed here, I inventory differences in their characterizations of the ecosystems services related to wild birds, the geographic area of focus, the species of birds studied, the survey design (mode, dates, sample sizes), the choice set structure and elicitation method, the formal choice-set design criteria, sample representativeness, estimation methods, use of continuous versus categorical attributes, interactions between attributes, approach to accommodating heterogeneous preferences, any quantified values for wild birds, caveats (both acknowledged by the authors and not), and recommendations for future research, both offered by the study authors and gleaned from this overall review.

Work-in-progress: In the time since the current draft, I have shared this review with the full set of authors of the papers covered by this review, inviting them to verify that I have not mischaracterized their work. I received a number of replies and will incorporate some minor updates in the next revision of this paper.

1. Contents

Choice experiments for ecosystems and wild birds: A review

2. Introduction

Wild birds are generally a non-market environmental good, with few market-based opportunities to measure rigorously the benefits accruing to society from policies or programs that help to support their abundance and biodiversity. Especially since Rosenberg et al. (2019) highlighted the precipitous population losses for North American avifauna over the last five decades, it has been clear that there is ample reason for concern for a variety of bird species, not just iconic, threatened or endangered species like the Bald Eagle or Spotted Owl. In this review of the literature, we focus on published papers that rely on choice experiments as a non-market valuation method helpful for understanding the tradeoffs that people are willing to make to protect either individual bird species, categories of species, or the habitat upon which these species rely. Our own research interest specifically concerns support for migratory species along the Pacific Flyway on the west coast of North America, but no extant research provides social values for these species themselves or for the types of habitat along this flyway that are essential for sustainable populations.

This review is organized according to aspects of different studies that should be considered in the process of devising any new choice-experiment study, for example, a study concerning land conservation to protect migratory bird species along the Pacific Flyway. We will contrast the various features of the range of published studies, organizing the information into tables where appropriate, and we culminate in the final section with an inventory of the lessons learned from each study in the literature that seem to be the most important to keep in mind in designing future choice experiments in this research area.

3. Characterization of ecosystems services related to birds

Different studies in this literature emphasize different fundamental types of benefits that people derive from wild birds and the habitats that support these birds. Economists would say that the demand for bird sanctuaries or protected areas could be characterized as a derived demand based on what those sanctuaries contribute to the maintenance of wild bird populations, but there can be direct demand for the ecosystems themselves. Furthermore, the demand for wild birds can be conceptualized as just one facet of people's demands for "nature" in general. It can be very difficult to be certain that the values being elicited in a choice experiment are exactly the values that the relevant policy questions require. The preamble to a set of choice experiments can control, to some extent, what people have in mind when they are asked to choose between specific alternatives, but it is not possible to know for sure what is in their minds.[1]

[1] In other valuation contexts, for example, the research might be seeking to identify people's willingness to pay for a policy that prevents damages to a specific environmental good, but it may be difficult to keep respondents from factoring in the associated likely human health impacts. People make choices based on their understanding of what will be affected by their choice, not necessarily based on what we tell them to think about. This tendency has been referred-to as subjective "scenario adjustment." It is sometimes possible to follow up with debriefing questions to confirm whether respondents made their choices as instructed, or whether they admit factoring in other concerns. This additional information makes it possible to control to some extent for scenario adjustment ex post, in the

In many papers in the literature, discussion of the variety of ecosystems services provided to humans by wild birds and their habitat is confined to the introduction of the choice experiment survey. Of course, survey respondents need not necessarily have thought much in advance about why they might/should value birds or bird habitat, specifically.

Past efforts to disaggregate willingness to pay for environmental goods into additively separable components for active use demand and passive use (non-use) demand have been largely misguided. Economists tend to avoid delving too much into "why" people value environmental goods, and to focus instead on "how much" people value environmental goods. Acknowledging this disciplinary tendency, however, it can still be helpful to readers if they understand some of the wide variety of implicit reasons why people might be willing to pay positive amounts for protection of wild birds.

It is therefore useful to pull together some of the speculation, in the existing literature on choice experiments related to wild birds, about the ecosystems services that authors believe that people derive from these particular types of environmental goods.

3.1. Bird watching opportunities in general

Some studies focus on the demand of respondents for bird-watching opportunities, so that the main ecosystem services provided to humans is presumably the enjoyment of opportunities to see wild birds. Implicitly, the recreational activity that is birdwatching may also confer enjoyment if the activity is social, or if people merely enjoy being out in nature, and bird-watching is a good excuse to venture out. Guimaraes et al. (2014) focus primarily on bird-watching, as to do Steven et al. (2017) and Xu and He (2022), without venturing any reasons why people enjoy bird-watching.

3.2. Opportunities to see rare or threatened species

Other authors focus on people's willingness to pay to see a species that is rare or threatened. In their study concerning ecological offsets, Rogers and Burton (2017) consider willingness-to-pay differences between more- and less-abundant species among a set of migratory shorebirds. Valasiuk et al. (2018) emphasize one species of bird that is dependent upon the fenland ecosystem they study, and Liu and Yang (2019) study a nature reserve dedicated to the support of one threatened and still-rare species.

3.3. Opportunities to see an iconic species

Iconic species are also of concern in a couple of studies. Yao et al. (2014) and Yao et al. (2019) explore the public benefits of changes in private management of planted forests that enhance habitat for rare native species (i.e. brown kiwi) in New Zealand, where these species are *important to the national identity*. Also, Zambrano-Monserrate (2020) studies the value of habitat for the endangered Andean condor.

---

estimation phase. These strategies can allow the researcher to simulate what would have been an individual's willingness to pay for the environmental good had they scrupulously followed the instructions in the survey.

3.4. Ecological roles of bird species that people may value

A few of these studies go into some detail enumerating the various ecological roles of wild birds that people might value indirectly. Clucas et al. (2015) mention seed dispersal and reduction of insect pests. Concerning Andean condors, Zambrano-Monserrate (2020) mentions the reduction of carrion, eliminating organic remains and contributing to their recycling, as well as helping the ecological succession of other scavenger species and "decomposers," thereby reducing the potential for infection in ecosystems. Sharma and Kreye (2022) specifically talk about seed dispersal, pollination, and pest control.

3.5. Cultural benefits of avian biodiversity

Boeri et al. (2020) itemize some of the non-material cultural benefits of biodiversity, including "*spiritual enrichment, cognitive development, reflection, recreation, and aesthetic experiences*." These authors cite others who have mentioned links between biodiversity and human wellbeing, measured as *psychological restoration*, improved *physiological health and better social relations*.

3.6. Separating benefits specifically from birds and other ecosystem benefits

Cerda et al. (2018a) emphasize the need to consider *less-charismatic species* (amphibians, insects, reptiles and vegetation) that contribute to biodiversity and are essential for ecosystem functioning. Their study includes both well-known and lesser-known mammals, birds, amphibians, reptiles, pollinating insects, succulents, woody shrubs, pristine landscapes and soil quality.

Cerda et al. (2018b) seek to estimate WTP for the protection of animals, plants and soil, but they also explicitly consider the "provision of *ecosystem services related to water resources*," as well as *tourism infrastructure*. But they decide against including "scenic beauty" of the reserve, since people already pay and entrance fee motivated by opportunities to admire scenery via hiking trails.

Sehra and MacMillan (2021) focus on *wildlife-friendly farming* and its contributions to biodiversity, characterized as the presence of animal species, including frogs, birds, or fish.

3.7. Use versus non-use demands for ecosystem services of birds

Garnett et al. (2018) distinguish between a person's jurisdiction of residence and the location of the bird-related environmental good in question, using this difference to motivate their discussion of nonconsumptive "use" demand, while other demand may be only non-user values. Dobson et al. (2022) seem to focus on *non-use demands* by eliciting willingness-to-donate for overseas conservation areas.

3.8. Demand for agricultural products where growing conditions are one attribute

Sehra and MacMillan (2021) focus on wildlife-friendly farming practices for rice. Likewise, Gatti et al. (2022) consider bird-friendly coffee production.

Two other studies, Bennett et al. (2018) and Czajkowski et al. (2021), concern the willingness to accept, by farmers, of compensation contracts that would oblige them to adopt bird-friendly agricultural practices. Indirectly, these farmers must take into account the likely net benefit in terms of the price they can charge for their product (plus the program's compensation) against the cost of implementing different practices.

3.9. Birds as pets

Finally, the study by Krishna et al. (2019) considers hypothetical purchases of caged song-birds as pets, for which there are likely longer-term consequences for populations of these same species in the wild.

3.10. Option demand

Kim et al. (2021) survey Korean people who have visited at least one national park over the previous five years, asking about a project to build an airport on an island where bird habitat would be harmed by the project. While this audience may not have visited the airport site in the past, they may wish to visit it in the future, so the "demand" in question may be either option demand or existence demand. They do not include island residents in their sample.

4. Geographic focus of study

The literature on choice experiments concerning wild birds can be partitioned according to the type and scope of the geographical area where these birds are located.

4.1. Nature reserves and similar specific locations

Research concerning migratory birds along the Pacific Flyway, for example, would need to accommodate the fact that migration over long distances requires regular stop-over locations for migrating birds. While birds do not use these habitats as permanent homes, they need "rest stops" during their northward and southward migrations, twice per year, and/or protection of nesting sites during their nesting seasons. This implies that permanent dedicated refuges are not necessarily required, although migrating individuals or flocks require seasonally protected habitat to serve their needs at different times and in different locations.

Several groups of researchers focus on people's preferences over the attributes of a specific nature reserve or special ecosystem managed specifically to protect bird species. Some of the species in question are migratory, but others are resident populations. For example, Bennett et al. (2018) considers the Jiangsu-Yancheng Coastal Wetlands Rare Birds National Nature Reserve

in China, while Cerda et al. (2018a) the Llanos de Challe National Park, a protected area of the Atacama Desert in northern Chile near Bolivia, and Cerda et al. (2018b) conduct their study with visitors to the Lircay National Reserve, a Mediterranean biodiversity hotspot in Chile. Liu and Yang (2019) focus on a Black-Faced Spoonbill refuge in the Qigu area of Taiwan. Czajkowski et al. (2021) are concerned with the Biebrza Marshes in Poland, one of the largest wildlife refuges in Europe. And Xu and He (2022) ask respondents to consider the Nansha Wetland in Guangzhou, China.

A shortcoming of studying one specific wildlife/bird refuge or a single biodiversity hotspot is that such studies risk producing results that are relevant only for the unique site in question. If this is the case, the study's findings cannot necessarily be generalized for use in benefit transfer from the "study" site to other "policy" sites. It is always possible that unique characteristics of the study site, other than those specifically considered in choice experiments, affect the tradeoffs people are willing to make with respect to this particular habitat for avian species. Best practices also require that choice experiments include a "status quo" option, which is the complement to an "any substantive alternative" indicator. If all the choice tasks in a study concern the same reserve/refuge site, then the "any substantive alternative" indicator will reflect everything that is unique about that particular site, where other people's baseline utility from the different unique features of other sites (to which one might wish to transfer those estimates) will be unknown.

4.2. A specific non-protected locality

Other choice experiments in the literature are focused not on a specific refuge/reserve site, but are still narrowly focused on a particular geographic area. Guimaraes et al. (2014) consider two wetland areas on Terceira Island in the Azores, located in the North Atlantic Ocean west of mainland Portugal. One of these two areas is a well-known bird-watching site where the main attraction is vagrant birds which divert from their usual migratory routes due to storm. Rogers and Burton (2017) explore tradeoffs that Australians are willing to make in programs that involve the use of environmental offsets to make up for lost habitat for two species of birds due to oil and gas development near a beach on the Kimberley coast of northwestern Australia. Valasiuk et al. (2018) focus on the Zvaniec fen mire in Belarus, an open wet grassland habitat that is almost the exclusive habitat of the globally threatened wading that has suffered as a result of widespread draining of fens for intensive agricultural use. Kim et al. (2021) consider the tradeoffs people are willing to make concerning the potential construction of a new airport on an island in an archipelago off the south coast of South Korea (although they note that the area in question is "part of the East Asian-Australasian Flyway, where it serves as a stopover and wintering site for sea-crossing migratory birds and a habitat for resident birds").

4.3. A specific type of ecosystem

Other choice experiments concerning wild birds focus on specific types of ecosystems in general, rather than individual examples of a specific ecosystem. For example Yao et al. (2014) value planted forest management schemes that enhance populations of rare and protected native species in New Zealand, and Yao et al. (2019) extend that earlier study. Clucas et al. (2015) survey people in Berlin, Germany, and in Seattle, WA (USA) about program to change the abundance of common songbirds and corvid species, specifically in urban areas, for which Berlin

and Seattle are their only two examples. And Gatti et al. (2022) survey coffee drinkers in the U.S. to assess their willingness to pay for ecolabels of several types, including a "Bird Friendly" certification. The specific coffee-growing countries do not appear to be specified in the survey, just the farming practices for coffee-growing areas.

4.4. Broader regions (state/country/global region)

Other types of studies use choice experiments concerning bird species across broader geographical areas or jurisdictions. With a sufficiently representative sample, the preferences elicited by these choice experiments could in principle be employed to value the species under study in any habitat area in that region.

Garnett et al. (2018) consider policies to help the adaptation to climate change of four non-charismatic bird species (and subspecies) on the Australian mainland and in Tasmania, although they do not include a cost attribute, so their tradeoff estimates do not extend to any WTP inferences.

In a particularly novel study, Krishna et al. (2019) explore the preferences of consumers for the attributes of caged birds as pets, motivated by market forces leading to the capture of rare birds in the wild endangering some of those species. Their data are collected in a choice experiment in Jambi Province in Sumatra. The preferences they measure are those of people who current own, or have owned, caged birds. It is not clear whether this group's preferences are a good proxy for the general preferences of the region's population.

Stemmer et al. (2022) use a choice experiment concerning destination choices in hypothetical birding trips, fielded to a convenience sample of visitors to the area of an important island-based birding site in Norway. However, they note that while other researchers have used destination-specific bird attributes, they sought to design a study using attributes that "were potentially applicable to various Northern Hemisphere birding destinations." While their proposed birding destinations may be broadly representative, their sample is not necessarily representative of the preferences of all birders, since their respondents were recruited either on-site at a specific birding destination or at a nearby lodging establishment. While the site attributes may be transferable to other localities, the preference parameters estimated using this sample may not be generalizable to other populations of people.

Zambrano-Monserrate (2020) considers habitat for one specific species, the Andean condor. This particular species ranges from northern Colombia and western Venezuela through the Andes south to Tierra del Fuego in Argentina and Chile. However, this author focuses on condor habitat specifically in Ecuador.

Sehra and MacMillan (2021) explore willingness to pay for rice that has been grown with "wildlife-friendly farming" (WFF) certification in two types of rice-growing landscapes in Japan.

Sharma and Kreye (2022) explore the social value of bird conservation on *private forest lands in Pennsylvania*. In that region, most forests are privately owned and forest habitat is an important landscape for migratory birds.

Finally, respondents are sometimes posed with choice experiments designed to determine their (mostly non-use) values for distant bird-supporting ecosystems. Dobson et al. (2022) estimate UK residents' willingness to donate money to support "flagship conservation areas." These are overseas (South African) conservation areas where one attribute of the area is the presence of a threatened bird species.

4.5. Birding in general

A couple of published studies ask respondents to choose among birding sites described generically, rather than associating the featured attributes with any specific destinations. Asking whether a respondent would visit a site with given bird-related characteristics and cost "on their next trip," Steven et al. (2017) implemented their choice experiments to visitors at several birding sites in Australia and at a birding fair in the UK. Stemmer et al. (2022), already mentioned above, also ask respondents about hypothetical birding trips, but they include scenery and "visitor facilitation" at the site in addition to the site's bird-related attributes and the cost.

5. Species of birds studied

In one study, Guimaraes et al. (2014), birds are not one of the attributes used in the choice experiments concerning infrastructure at two wetlands where birders commonly visit for the purpose of observing birds. However, these authors mention that opportunities to spot uncommon vagrant birds that have been blown off-course by storms make the two wetlands in question very popular for birdwatching.

5.1. Abundance or species richness for birds in general

Kim et al. (2021) are not specific about which bird species would be affected, focusing instead on simply the abundance of birds in general. Likewise, Gatti et al. (2022) do not focus on any specific species of birds, just the bird-friendly biodiversity certification attribute for coffee brands. Xu and He (2022) focus on species richness for rare birds at the Nansha Wetland in Guangzhou, China.

5.2. One specific species of bird

Some studies focus on one particular species of bird, in some cases singled out among generic populations of other birds. Bennett et al. (2018) use choice experiments with farmers near a coastal rare bird nature reserve to assess these households' willingness to accept compensation in programs to reduce pesticides that harm wetland birds, in particular *Red Crowned Cranes*. Similarly, Valasiuk et al. (2018) focus specifically on the *Aquatic Warbler* in a particular fenland area of Belarus.

Liu and Yang (2019) consider abundance of a single species, *Black-Faced Spoonbills*, as well as the number of other unspecified bird species at a specific reserve site in Taiwan. Sehra and MacMillan (2021) select the *Black-Crowned Night Heron* as their conservation target "bird" species protected by wildlife-friendly farming practices in rice paddies in Japan.

5.3. Two or more specific species of birds

Rogers and Burton (2017) specifically focus on two migratory shorebird species: Ruddy Turnstones and Eastern Curlews. Yao et al. (2014) and Yao et al. (2019) both consider New Zealand's brown kiwi and bush falcon, as well as three other species (one fish species, one gecko species, and one shrub).

Garnett et al. (2018) consider policies toward the *Rufous Scrub-Bird* and is subspecies, the *Scrubtit*, as well as the *Brown Thornbill* (*mainland form*) and its subspecies the *Brown Thornbill (Tasmanian form*). They purposely choose to feature non-charismatic species.

Krishna et al. (2019) choose representative caged-bird species as representative for each combination of three binary main species attributes (rarity in the wild, trading frequency, and relative position of the species in terms of general price levels). They use four species groups, each containing eight distinct species, but with two repetitions necessary, their choice experiment involves 30 distinct named species.

Czajkowski et al. (2021), in the training module of their survey concerning farming practices, ask their respondents specifically about the Aquatic warbler, Ruff (considered to be symbols for the area), Black-tailed godwit, Eurasian curlew (protected by existing agri-environmental programs), and Northern lapwing, and Eurasian wigeon (both of which are just popular birds). But effects on these species are only implicit across the choice tasks that respondents are asked to consider. The authors do mention that some bird species in the area are undesirable from a farmer's perspective, since farmers desire compensation for the loss of crops due to protected farmland birds (Greylag geese and Cranes).

5.4. Two or more categories/guilds of birds

Clucas et al. (2015) focus on common native urban songbirds—finches (described as "pleasing") and corvids (described as "displeasing")—in Berlin, Germany, and in Seattle Washington.

Cerda et al. (2018a) consider other species in their Atacama desert study as well, but among birds, they focus on inland raptor species (scavengers and passerines) and shorebirds.

6. Survey design

6.1. Survey mode and in-the-field dates, sample sizes

Many surveys in this inventory were implemented in an online format. Other surveys use face-to-face household surveys with some overarching sample design. Others employ convenience samples at specific locations via or face-to-face intercept surveys. In some cases, the authors report the original number of survey invitations issued as well as the number of completed responses used in estimation, but in other cases the number of invitations is not obvious. We

address this distinction later, in our section on sample representativeness. At the end of this section, we include a summary table to facilitate quick comparisons across studies in terms of their different survey modes, fielding dates and sample sizes.

6.1.1. Online delivery using a market research firm

Some researchers mention their use of a survey research firm in the body of the paper. For example, Rogers and Burton (2017) use an online survey distributed by "a market research company," fielded during October-November 2014, to a sample of 1371 respondents stratified by age, gender and location.

Garnett et al. (2018) use an online survey fielded to a sample recruited from a panel maintained by MyOpinions PermissionCorp, where the active panel includes about 300,000 members and is managed for research only according to governance by marketing research industry bodies. They invited 7,816 people during September-October of 2014 and received 1,421 responses and 1,119 completed questionnaires.

Kim et al. (2021) collect their data via an online survey using a panel recruited by a survey research firm in Seoul. During November and December of 2017, almost 13,000 invitations were sent out and 2200 individuals were willing to participate, but quotas for age, gender and regions, and 158 incomplete responses were taken into account, and the survey closed when 1000 eligible respondents completed the questionnaire.

Dobson et al. (2022) recruited their 852 UK-based respondents via an online research platform (Prolific) in May 2019.

Gatti et al. (2022) use a Qualtrics online survey of 774 US coffee drinkers aged 18 or older, fielded in December of 2020. Participants received a financial incentive. There appears to be no discussion of any quota that may have been used. The sample yields 7740 total choices for analysis.

6.1.2. Online delivery, unspecified

For other online surveys, or others the party actually fielding of the survey was not immediately apparent (but may be mentioned in the paper's supplementary documentation).

Yao et al. (2019) follow-up an earlier survey (that used a combination of modes) with an online survey between January and June of 2015. 1,356 respondents completed the survey.

Zambrano-Monserrate (2020) uses an online survey to respondents over 18 years old and able to make financial decisions, fielded between August 10, 2019 and August 30, 2019. After elimination of surveys with incomplete or inconsistent responses (with no discussion in the paper of the exclusion criteria), 825 valid questionnaires were retained for analysis.

Stemmer et al. (2022) conduct an online survey, pilot-tested in English and Norwegian. The final survey was sent to 559 email addresses during March and April 2018 with versions available in

four languages. Adjusting for undeliverables, 521 birders received an invitation and the analysis based on choices made by 205 respondents with complete data.

Sharma and Kreye (2022) field a statewide "web survey" to Pennsylvanians. The survey date is not immediately apparent, but obviously prior to 2022.

6.1.3. Face-to-face surveys (household)

Clucas et al. (2015) use face-to-face interviews of 460 residents across 10 study sites in Berlin from August 2008 to December 2008, and 209 residents in 8 study sites in Seattle from October 2009 to February 2010..

Valasiuk et al. (2018) administered a face-to-face survey with a sample of the Belarusian population, with 270 completed interviews constructed at respondents' houses in January of 2010. The estimating sample included 206 respondents with "valid questionnaires."

Krishna et al. (2019) use computer assisted personal interviews employing the software Surveybe, after translating into the local language (Bahasa Indonesia). Their survey was in the field between February and May of 2016 in Jambi City. Their sample is limited to current or recent owners of caged birds. They contacted 504 households, and report 5,812 observations.[2]

Czajkowski et al. (2021) used computer assisted personal interview s (CAPI) of 463 farmers, conducted by agricultural advisors who normally work in the area. The survey was in the field during June-August of 2017 and in March 2018. They use a stratified quota sampling method.

6.1.4. Face-to-face surveys (on-site intercepts)

On-site intercept surveys take the research team to some location where they are likely to come in contact with a variety of people who have demonstrated by their behavior that they are interested in the ecosystem services supported at that destination.

Guimaraes et al. (2014) survey both on-site birders (during one birding season in October and November of 2011) and off-site birders who have visited their study area in the previous season. Off-site birders were located through "blogs, Facebook, and specialized websites." They used a mixed-mode survey with identical electronic questionnaire that were self-administered. An interviewer assisted in person for the on-site respondents, and by phone/VoIP for off-site respondents.

Steven et al. (2017) intercepted respondents at Australian birding sites and at a UK birding fair between May 2013 and November 2014. They identified birders in the Australian samples by the equipment they were carrying or through casual greetings. Participants at the UK birding fair were assumed to be birders. The have 283 completed responses.

[2] We note that Stata refers to "alternatives" as observations. Each questionnaire in this study contains only 8 choice tasks, so it seems more likely that their 504 respondents (times 8 choice) face a total of no more than 4032 total choices, where perhaps not all respondents completed all the choices offered to them.

Cerda et al. (2018a) surveyed roughly 500 adult income-earning Chilean visitors to the National Park in question between January and March 2013. Cerda et al. (2018b) used about 400 face-to-face interviews with adult income-earning Chilean citizens visiting the reserve in 2013.

Liu and Yang (2019) conduct an on-site intercept survey of visitors to two of three bird-watching pavilions a Black-Faced Spoonbill reserve in Taiwan. Their survey is in the field during March and April of 2013. They use a formal algorithm to systematically select their first contact. Then they approach every third person after that. Potential respondents were informed that the survey's purpose was "for better managing the coastal wetland." They collect 434 completed questionnaires.

6.1.5. Face-to-face surveys (other types of convenience samples)

Sehra and MacMillan (2021) collected responses from a convenience sample of 231 people at a selection of intercept locations (farmers' markets, public parks, train stations and university campuses) during April-June 2019, with 1375 choice observations in their data.

6.1.6. **SUMMARY TABLE**: Survey mode, dates, initial contacts, estimating sample size, quota-based sampling

Study	Mode (company)	Field dates	# Contacts/ invitations	Estimating sample size	Quota sampling?
Online surveys					
Yao et al. (2014)	Phone/mail and phone/internet (unspecified)	Dec 2009 & Aug 2010	2,996 phone calls	209 respondents	
Rogers and Burton (2017)	Online, unspecified	October-November 2014	unspecified	1,371 respondents	Yes: age, gender, location
Garnett et al. (2018)	Online; MyOpinions PermissionCorp, Australia	September-October of 2014	7,816	1,119 respondents	
Yao et al. (2019)	Online; unspecified	January-June of 2015	unspecified	1,356 respondents	
Kim et al. (2021)	Online; survey research firm in Seoul, South Korea	November-December of 2017	12,839	1,000 respondents	Yes: age, gender, regions
Zambrano-Monserrate (2020)	Online; unspecified	August 2019	unspecified	825 respondents	Data from "several cities"
Dobson et al. (2022)	Online, UK, Prolific	May 2019		852 respondents	
Gatti et al. (2022)	Online, Qualtrics	December of 2020		7,740 choices	Usually, with Qualtrics
Stemmer et al. (2022)	Online; unspecified	March-April 2018	521	205 respondents	
Sharma and Kreye (2022)	Online; unspecified	Not reported		690 respondents	
Face-to-face household surveys					
Clucas et al. (2015)	Face-to-face, Berlin	August-December 2008		460 respondents	
Clucas et al. (2015)	Face-to-face, Seattle	Oct 2009-Feb 2010		209 respondents	

Valasiuk et al. (2018)	Face-to-face, Belarus	Jan 2010		206 respondents	
Krishna et al. (2019)	CAPI; Surveybe; current/recent owners of caged birds	Feb-May 2016	504 households	5,812 "observations"	
Czajkowski et al. (2021)	CAPI; area farmers	Jun-Aug 2017 & Mar 2018		463 respondents	Stratified quota sampling
Face-to-face on-site intercepts					
Guimaraes et al. (2014)	Mixed modes; self-administered electronic questionnaire	Oct-Nov 2011		97 respondents	
Steven et al. (2017)	At Australian birding sites and UK birding fair	May 2013 & Nov 2014		283 respondents	
Cerda et al. (2018a)	Chilean National Park visitors	Jan-Mar 2013		493 respondents	
Cerda et al. (2018b)	Chilean reserve visitors	2013		400 respondents	
Liu and Yang (2019)	Visitors to pavilions at bird reserve	Mar-Apr 2013		434 respondents	
Face-to-face off-site intercepts					
Sehra and MacMillan (2021)	Farmers' markets, public parks, train stations, university campuses	April-June 2019		231 respondents	

6.2. Choice set structure (elicitation method and choices per respondent)

The complexity of a choice set, if too great, can lead to non-response or the respondent may start to use heuristics in selecting their preferred alternative. If a choice set is too complex, involving either too many alternatives or too many attributes (or both), some respondents are likely to ignore some attributes, focusing on those which are most salient to them. Some researcher follow up a set of complex choice sets with specific questions about which attributes the respondent mostly ignored in their decisions.

Particular concerns about choice set structure may arise if the alternatives are "branded" or labeled in some way that conveys additional unspecified information to respondents. If all alternatives are "generic," described only by their objective attributes, and not branded in any way, a respondent cannot recruit other knowledge, based on the brand, to inform their choices.

In the subsections to follow, we sort the studies in our inventory according to the structure of their choice sets. We also note the number of these choice sets presented to each respondent. Again, for ease of comparison, we offer a summary table at the end of this section that includes abbreviated versions of the information described below, concerning the context for the choice experiment, its bird-related attributes, non-bird-related attributes, cost attribute, number of alternatives, and the number of choice tasks presented to each respondent.

6.2.1. Pairwise choices (one alternative and the status quo)

Pairwise choice tasks, if the attributes of the non-market good in question are the same for each choice and for everyone in the sample, constitute what is known as "contingent valuation" tasks. Typically, a choice tasks is not called a choice experiment unless there are multiple attributes besides the cost attribute that vary across choices and/or across respondents, so that it is possible to calculate marginal willingness to pay measures for each attribute. If the attributes of the non-market good are unchanged across choice tasks and respondents, so that only a specific bundle of attributes is being valued, then it is not possible to tease out separate marginal WTP estimates for any of these attributes, only a value for the bundle as a whole.

Among these studies, only Zambrano-Monserrate (2020) uses a referendum contingent valuation method, rather than choice experiments. Respondents were first asked if they were willing to pay, in general, for a single "public (hypothetical) program to save the Andean condor." If they answered in the affirmative, they were asked about whether they would pay a specific "bid" amount. Each respondent was asked about just a single randomly assigned bid chosen from six payment levels.

In contrast, two other papers use binary choices, but vary the program/policy under considerations. Cerda et al. (2018a) use choice sets consisting of pairwise comparisons of the alternative park management situations (64 in total), randomly blocked into eight questionnaire versions with 8 choice sets each. Sharma and Kreye (2022) use choice experiments that present

proposed programs one at a time in a dichotomous choice (referendum) format, where respondents either accepted or rejected that program at a specified price. There were eight scenarios or proposed programs.[3]

6.2.1. Two substantive alternatives and a status quo alternative

In their study of infrastructure options for two wetland areas that support bird populations, Guimaraes et al. (2014) asked respondents to choose their favorite among three alternatives, where one of these alternatives was always the status quo. Visual representations were used to convey attribute levels.

Yao et al. (2014), in their study of five threatened plant/animal species in New Zealand, gave each respondent nine choice tasks, each consisting of three alternatives (two substantive alternatives and current conditions). Yao et al. (2019) appears to have used the same choice structure in their follow-up study.

In their choice experiment concerning two pairs of related species in mainland Australia and Tasmania, Garnett et al. (2018) gave their respondents three options, consisting of two substantive management options for climate change adaptation and a status quo option.

In their study of birds as pets, Krishna et al. (2019) use choice sets where each choice set contains two specific birds of different species, plus an option not to purchase either bird. They focus on a set of attributes of wild bird species as pets, and "brand" each potential pet bird with a specific species name, where this species represents different combination of the three basic bird attributes (rarity, trading frequency, and price). All choices included a no-purchase alternative. Each respondent considered eight choice sets.

Kim et al. (2021) use choice sets with two alternatives policies concerning the construction of an airport that will affect bird habitat and a no-airport option, and each respondent was asked to consider four choice sets.

Gatti et al. (2022) asked each respondent 10 choice questions, where each choice appears to have been between two substantive alternatives for certification of coffee-farming practices in the purchase of coffee, and a no-buy option. However, the possible certification options were made mutually exclusive, it seems.

Stemmer et al. (2022) use choice tasks where each choice had three alternatives. Respondents were asked to choose between two different birding destination options and a no-trip option. Each survey version had a block of four randomly assigned choice sets.

6.2.2. Three substantive alternatives and a status quo alternative

[3] It seems that each respondent in the survey was asked about all eight proposed programs, although this is not entirely clear in the main paper.

Rogers and Burton (2017) study people's preferences for environmental offsets when bird populations are threatened by development. They use choice scenarios with three policy alternatives and an opt-out alternative.

Steven et al. (2017) use choice cards that present respondents with a choice among three potential hypothetical birding destinations and the status quo (no visit). Each person considered six different choice cards, drawn from a full set of 18 cards blocked into three sets of six.

In their study concerning preferences for limits on farming practices and conservation areas in a fenland area of Belarus, Valasiuk et al. (2018) use a survey where each respondent made 16 choices, each one including a status quo alternative and three substantive program alternatives. Respondents were asked to pick their most-preferred and least-preferred alternatives (a "best/worst" approach), rather than just their most-preferred option.

Czajkowski et al. (2021) survey farmers and use choice sets with three types of contracts involving compensation for bird-friendly farming practices and a no-contract option, and they ask respondents to completely rank these alternatives from most preferred (1) to lead preferred (4). They presented each respondent with up to 6 choice situations regarding arable land and livestock reduction, and up to 3 choices regarding peatlands and meadows, provided these programs applied to their farm.

In their study of tourists' willingness to pay for additional species of rare birds at a wetland reserve in Guangdong, China, Xu and He (2022) use choice tasks where each choice set was composed of three hypothetical scenarios for attributes of the wetland and a status quo option. Each survey instrument included just one choice set.

6.2.3. Four substantive alternatives and a status quo alternative

Choice sets that involve too many alternatives and/or too many attributes per alternative can strain the cognitive capacity of respondents, especially if they are impatient to complete the survey. Nevertheless, Liu and Yang (2019) offered their respondents choices between four substantive alternatives and "none of these." Each respondent was asked to consider two choice tasks.

6.2.4. Omitting a status quo alternative (with any number of substantive alternatives)

It is generally important for respondents to be offered the option of just keeping what they have now, rather than being forced to pay money to select some other option than what they have at present. Two of the papers in our inventory ask respondents to make force choices between alternatives that do not include the status quo (or a no-purchase or no-trip option, in these cases). First, Sehra and MacMillan (2021) provide each respondent with six choice sets, each including three different types of rice: two are hypothetical wildlife-friendly rice products and one is a non-wildlife friendly product. They do not appear to have used a no-purchase option. Their numeraire good is a non-wildlife product with no species, non-organic, non-special-origin landscape and a price of 2000 JPY, but involves the purchase of the default type of rice. This is a forced-choice scenario. It is conditional on the consumer buying some 5 kg bag of rice.

However, this may be a plausible forced-choice scenario in this context, since almost everyone in Japan would be buying rice.

A different rationale for having no status quo (opt-out) option is employed by Dobson et al. (2022). They use choice experiments involving a forced choice between visits to two hypothetical conservation areas. They do not include a "neither" option because they want to prevent respondents "from earning their reward without weighing up the alternatives." They mention other studies where researchers have reported the respondents disproportionately choose "neither" when faced with complex choices. It is not entirely clear that the benefits of a forced choice outweigh the risk of scenario rejection by respondents. They make their choices "conditional on having to pick one of these alternatives," but we cannot learn from these choices whether they would actually pick either one of these alternatives. It is impossible to estimate population willingness to pay for any of the alternatives, because it is not possible to reflect whether people would choose any alternative over none.

6.2.5. **SUMMARY TABLE**: Context, attributes, alternatives, and choice tasks

Study	Context	# Attrib.	Bird-related attributes	Non-bird-related attributes	Cost attribute	# Non-SQ Alts	Status quo alt?	# choice tasks	Comments
Choice experiments									
Four attributes									
Steven et al. (2017)	Generic birding sites	4	(1) the number of threatened bird species present at a site (either critically endangered or endangered), (2) the diversity of birds at a site (species richness), and (3) the number of regionally endemic bird species at the site (restricted-range species).	None	(4) a site entrance fee	3	Yes	6	
Sehra and MacMillan (2021)	Rice-growing practices in Japan	4	(1) indicators for the presence or absence of three different animal species (including birds)	(2) an indicator for whether the location of the rice paddy is in a "Satoyama" landscape (3) whether the rice is organically or non-organically grown	(4) price for a 5 lb. bag of rice	3	No	6	Omission of a status quo option (because virtually everyone in Japan will buy rice)
Garnett et al. (2018)	Conservation options for Australian bird species	4	How the program would help each of four species (if at	None	None	2	Yes	4	Omission of cost attribute is based on stated desire to force

			all) to adapt to climate change: (1) Brown thornbill, mainland (2) Brown thornbill, Tasmania (3) Scrubtit (4) Rufous scrub-bird						tradeoffs and preclude opting out of choice (even though this precludes monetization for benefit-cost analysis)
Valasiuk et al. (2018)	Agricultural practices for Belarus fenland area	4	None	(1) means of removing shrubs, (2) size of the managed area, (3) size of the conservation area.	(4) an obligatory annual payment for all adult Belarusian residents	3	Yes	16	Best-worst modeling.
Czajkowski et al. (2021)	Agricultural practices in Poland's Biebrza Valley	4	None	Farmer's contract would involve (1) specific agricultural practices, (2) for some number of years, (3) with what opt-out opportunities.	Willingness-to-accept study, so uses (4) a subsidy payment to the farmer, paid annually per hectare enrolled.	3	Yes	Up to 6 choices about arable land and livestock reduction; up to 3 choices regarding peatlands and meadows, as applicable	
Five attributes									
Guimaraes et al. (2014)	Tourist preferences for wetland management in the Azores	5	None	For one site, whether there it has: (1) access by car, (2) garbage bins, (3) new reservoirs, (4) observatories	(5) No explicit cost variable, but willingness to pay is expressed implicitly as the "extra days of stay" incurred	2	Yes	10 (= 2 sites x 5 choices)	Rationale for non-standard cost variable: goal of the study is to demonstrate the local economic

				For their other site, whether it has: (1) changes to slope of lake shore, (2) anew water channel, (3) changes in water circulation, and (4) and interpretive center	(not explicitly monetized)				impact of visitation to these sites, which is best captured by the number of days that tourists stay in the area
Cerda et al. (2018b) – Central Chile site	Management of Chile's Lircay National Reserve	5	(1) wildlife conservation research, where bird species is one option	(2) protection of soil in public use area, (3) manage-ment of water resource for long-term benefits, and (4) recreation infrastructure.	(5) an increase in the entrance fee	1	Yes	8	
Krishna et al. (2019)	Demand for caged birds in Sumatra	5	For each caged bird species: (1) rare/abundant in the wild, (2) captured from the wild, (3) bred in captivity, and (4) trainability for singing.	None	(5) price of the caged bird	2	Yes	8	
Boeri et al. (2020)	Different quantifications of biodiversity in coastal wetlands	5	Levels of avian biodiversity: (1) the number of different types of birds, (2) the number of individual birds,	None	(5) a one-time voluntary donation			6	

			(3) the likelihood of seeing a rare or unusual type of bird, and (4) the probability of seeing a "wildlife spectacle" such as thousands of birds in one flock.						
Kim et al. (2021)	Potential construction of new airport for a Korean island birding destination	5	(1) damage to bird habitats and reductions in the number of birds (low, medium, high).	(2) degradation of forest and marine environments (low, medium, high, measured as "soccer fields" of damaged area), (3) the accessibility of the island for residents and tourists (low=surface transportation, high=air travel), and (4) additional tourism opportunities (low, medium, high, with examples).	(5) an annual additional household income tax charged for the next ten years	2	Yes	4	
Gatti et al. (2022)	Demand for different types of environmental certification for coffee	5	Whether coffee brand is certified as: (1) bird-friendly	Whether coffee brand is certified as: (2) "shade-grown," (3) "organic," and (4) "pesticide-free."	(5) price for a 12-oz bag of ground coffee	2	Yes	10	
Sharma and Kreye (2022)	Value of bird conservation on private	5	(1) the category of birds that will benefit from the intervention	(3) type of forest habitat (young/shrubby or	(5) a new general tax applied to all	1	Yes	8	

	forestlands in Pennsylvania		(common species or rare species), (2) the benefits of birds to humans (ecological or recreational).	mature/old trees), and (4) the type of landowner assistance (incentives or a management plan).	Pennsylvania residents				
Stemmer et al. (2022)	Demand for birding site attributes (Europe)	5	(1) of birding quality, and (2) avian diversity (intervals of species richness).	(3), scenery, and (4) facilitation of birding.	(5) a conservation and maintenance fee	2	Yes	4	
Xu and He (2022)	Tourists' demands for visits to the Nansha Wetland in Guangdong, China	5	(1) number of species of rare birds	(2) water visibility (i.e., depth visible), (3) mangrove coverage (percent of area), and (4) science education and popularization	(5) an increase in travel cost to the site (tourists' WTP)	3	Yes	1	
Rogers and Burton (2017)	Features of ecological offsets to protect birds	5	(1) whether either or both of two bird species would be protected by the ecological offset, and (2) the number of birds that would be protected	(3) percentage of direct offset activity, (4) location of the offset, and (5) the implementer of the offset	None	3	Yes	6	No personal cost described to respondent because ecological offsets are the responsibility of the developer—did not raise the cost of higher prices for the developer's product
Six attributes									

Yao et al. (2014) and Yao et al. (2019)	Threatened birds, animals and plants in New Zealand	6	Encounters with two species of birds: (1) brown kiwi (2) bush falcon	Encounters with three other species (3) a fish species (4) a gecko species (5) a plant (shrub) species	(6) a payment for the biodiversity program that will be paid annually via income tax for five years	2	Yes	9	
Dobson et al. (2022)	Donations by UK residents to fund conservation areas in Africa	6	(1) presence or absence of threatened bird species.	(2) presence or absence of charismatic mammals, (3) presence or absence of legal protections for the conservation area, (4) high or low existing funding, and (5) ownership of conservation area	(6) a donation amount	2	No	9	No status quo/opt-out option "to prevent respondents from earning their reward without weighting up the alternatives"
Seven attributes									
Cerda et al. (2018a) – Atacama site	Programs to protect biodiversity in a National Park in Chile's Atacama Desert	7	(1) category of bird species protected	(2) protection for a set of three well-known and two lesser-known mammals (3) soil protection in one, another, or both of two zones, (4) three groups of other animals (amphibians, reptiles and pollinating insects), (5) plants (cacti and shrubs),	(7) a voluntary payment per month over 5 years	1	Yes	8	

				(6) maintenance of a pristine landscape (tourism infrastructure or not)					
Eight attributes									
Krishna et al. (2019) – continued from above	Demand for caged birds in Sumatra	8	As above, but extended models that also include implicit bird characteristics embodied by the species name: (5) caged bird species' relative price level relative to other captive species, (6) whether that species is rare in the wild, and (7) whether that species in question is frequently traded.	As above	(8) price of the caged bird	2	Yes	8	
Liu and Yang (2019)	Management of the Black-Faced Spoonbill Reserve in Taiwan	8	(1) the number of Black-Faced Spoonbill individuals seen, and (2) the number of other bird species seen.	(3) the degree of crowdedness (10% being uncrowded and 90% being very crowded), (4) the complement of "necessary facilities" such as washrooms, binoculars, parking, trash receptacles, etc., available at the site,	(8) admission fee to the reserve in question	4	Yes	2	

				(5) the complement of "non-necessary facilities" (gift shops cafes, etc.), (6) the quality of the interpretive program, and (7) the amount of waiting time for shuttle buses					
Contingent valuation									
Clucas et al. (2015)	Urban birds of two type in Berlin and Seattle	2	Increase urban bird populations for either songbirds (housefinch or greenfinch) or corvids (American crow or hooded crow) Songbird vs. corvid varied across split samples)	None: Double-bounded dichotomous choice contingent valuation format.	Household-level tax for conservation	1	Yes	2	
Zambrano-Monserrate (2020)	Andean Condor habitat in Ecuador	2	A program to prevent extinction of the Andean condor	None: Simple dichotomous choice contingent valuation format.	Willingness to pay for a public program.	1	Yes	1	

6.3. Formal choice set design criteria

Virtually every choice experiment reviewed here employs some set of criteria for efficient design. These algorithms maximize the useful information obtained from any given sample size, or allow a target level of estimation efficiency to be achieved from the smallest possible (and therefore cheapest) sample. A limited number of choice-set design software packages are used. Note that a couple of papers do not settle for optimizing just one type of efficiency, but conduct their surveys in waves, rotating through different available efficiency criteria in an effort to take advantage of the benefits of all these different criteria.

There is one main shortcoming of maximizing design efficiency. One of the artifacts of using packaged software to maximize some efficiency measure for the experimental design is that most such software expects the research to specify just a *small* number of *discrete* values for each attribute. If the number of attribute values gets large, the number of possible unique combinations of attributes across all designs rapidly becomes huge, which can stall the algorithm or render it so slow that timely outputs of (even fractional) factorial designs of a given size cannot be achieved.

This is a good place to note that there can be a tradeoff between design efficiency and the researcher's ability to specify models that allow for smooth nonlinear MWTP functions that permit interactions between attributes interpolation, or that permit interpolation or modest extrapolation of MWTP if the quantity of the attribute takes on any arbitrary value not included among those few levels specified in the design. Researchers should ask if they are willing to give up some design efficiency to preserve an opportunity to fit MWTP as a nonlinear parametric function of the level of the attribute in question or the levels of other attributes. Economic analyses should generally preserve an option to test for the presence of diminishing marginal utility or the dependence of utility from one attribute on the level of another.

6.3.1. Papers where design software is not clearly specified

Guimaraes et al. (2014) allow for 64 different choice alternatives, so they use a *factorial* design with 15 alternatives and sort these randomly into three groups of five choice sets each. Each respondent thus answered five choice questions (with four choice sets each).

6.3.2. Simpler studies where a full orthogonal design can be used

Sharma and Kreye (2022) use attribute mixes that are orthogonal, consisting of four attributes with two levels and one factor (cost) with four levels. Their study is basically a contingent valuation study, rather than a full-fledged choice experiment. However, they ask about two different levels of the "policy package," rather than just one (permitting a "scope" test, at least).

6.3.3. Papers using Ngene software, from ChoiceMetrics (choice-metrics.com)

The Ngene package for designing choice sets appears to be the most popular among named packages for the set of papers we inventory here.

Yao et al. (2014) employ a "sequential experimental design" by administering their survey in two waves and using the first wave's orthogonal main effects design and the resulting parameter estimates to improve the efficiency of the choice set designs for the second wave. The second wave employed three different designs, all three being Bayesian efficient designs but each optimized according to a different criterion: d-efficiency, c-efficiency, and s-efficiency. They Ngene ChoiceMetrics (vintage 2011). They also use a fourth design, an optimal orthogonal design, also generated using Ngene. They note that their study is novel in using a combination of experimental designs. Both Yao et al. (2014) and Yao et al. (2019) give each respondent nine choice tasks, each consisting of three alternatives with six attributes each.

Rogers and Burton (2017) use Ngene to generate an s-efficient design with parameters from an earlier study by Burton and others as priors. This strategy results in 24 choice scenarios blocked into 4 groups of 6.

Steven et al. (2017) start from a pilot study that uses an optimal orthogonal-in-the-differences (OOD) experimental design and just 27 respondents. They then use the parameter estimates from this pilot study as priors for a more-efficient experimental design (apparently reported in the supplementary materials for the paper). Their design can detect main effects and first-order interaction effects between attributes. They cite Choice Metrics 2014, which is the user manual and reference guide for Ngene 1.1.2.

Dobson et al. (2022) use NGENE 1.2 to assemble unlabeled (unbranded?) choices with a "Dz efficient" design, assuming an MNL model and zero priors. They have 36 choice pairs split into four blocks, so each respondent has just nine questions.

Stemmer et al. (2022) use Ngene to build a d-efficient design of attribute levels across alternatives. Their final design employs 24 choice sets blocked into 6 survey versions, each with 4 randomly assigned choice tasks.

6.3.4. SAS 9.3 algorithm

Kim et al. (2021) worked with Birdlife International to develop a suitable set of attributes. They use a fractional factorial design which generates 36 total choice sets. These are allocated to nine survey versions.

6.3.5. AlgDesign (R algorithm)

Sehra and MacMillan (2021) reduce their design from 72 possible combinations to just 18, which they generate using R (3.5.1) with the package AlgDesign.

6.3.6. Stata

Krishna et al. (2019) use Stata 13.0 to generate their final experimental design. This software employs a modified Fedorov algorithm to maximize the d-efficiency of the design. They use a final design that consists of 24 choice sets, randomly assigned to three different blocks of eight choice tasks. Each choice set contains two birds of different species, each representing a different combination of the three basic bird attributes (rarity, trading frequency, and price) with each combination of these three represented by four different named species, yielding 96 different choice cards. These they organize randomly into 12 books each containing eight choice cards, where all choices include a no-purchase alternative. Each respondent is shown one of these eight cards.

6.3.7. Citing no particular software package, but referencing other experts

6.3.7.1. Hearne/Salinas, Juutinen, Kelly, Lee

Citing design strategies by these earlier authors, Liu and Yang (2019) use a choice task that has eight groups of items, each with three or four levels, so a full factorial design would have a prohibitive number of combinations. Thus they use an orthogonal fractional factorial design to produce 32 groups of items, randomly divided into eight question sets, with each task consisting of a choice with four alternatives. Each questionnaire has two choice sets.

6.3.7.2. The "Rose/Bliemer" algorithm

Garnett et al. (2018) cite work by Rose and Bliemer and use a Bayesian d-efficient design for their choice tasks, informed by a preliminary sample of 200 respondents with uniform priors that yields estimates of the marginal utilities that can be used to update the design.

6.3.7.3. The "Scarpa/Rose/Ferrini" algorithm

Bennett et al. (2018) extract from the full factorial design for their attributes a subset characterized by orthogonal differences between attribute levels, citing a 2008 paper by Scarpa and Rose. This yields 72 choice tasks, blocked into nine task sets, each including eight choice tasks. This approach allows identification of all the main effects of interest but may not be as efficient as designs that employ prior information about parameters (such as those which minimize D-errors). They pretest the efficiency of their design, and a comparison with a d-efficient design does not suggest that the potential improvement is large enough to change their design criterion. They go into more detail than usual about their choice of design algorithm.

Boeri et al. (2020) used the four most common "efficiency under uninformative priors" criteria for choice set design. They also use more than one criterion and update the design in each of six waves to maximize the statistical power in their estimation process and to permit their models to capture as many tradeoffs as possible. They obtained 66 different choice tasks and divided these into 11 blocks of six, with one of these sets of six going to each respondent.

Czajkowski et al. (2021) design their combinations of attribute levels to be Bayesian efficient, minimizing the determinant of the asymptotic variance-covariance matrix of the parameter

estimates (so-called d-efficiency) given the estimated preference parameters from a pilot study. They cite work by Scarpa/Rose and Ferrini/Scarpa.

6.3.7.4. The "Louviere/Hensher" algorithm

Cerda et al. (2018a) do not appear to discuss in the main paper any algorithms they have used to achieve efficiency in the design of the attribute mixes in their choice sets. They do cite papers by Louviere and by Hensher and mention that their 64 possible pairwise comparisons are obtained "following" the strategies of these other researchers. Their scenarios are randomly blocked into eight questionnaire versions with eight choices each.

Cerda et al. (2018b) also report using an orthogonal main effects design to generate a reduced orthogonal experimental design (citing Louviere and Hensher). They generate 64 management options, combined into choice sets with two options and a status quo. These were blocked into eight different versions with eight choice sets per person.

6.3.7.5. The "Street/Burgess" algorithm

Valasiuk et al. (2018) uses choice sets that are prepared according to the "optimal-orthogonal-in-the-difference design." The OOD design apparently precludes alternatives sharing any attribute levels in common. They cite a 2007 paper by Street and Burgess, noting that an advantage of OOD over D-efficient designs is that they require no prior knowledge of preference parameters. [However, it might also be said that they do not *benefit* from prior knowledge of preference parameters, as from a test sample with the survey.]

6.3.7.6. The "Zwerina" algorithm

Gatti et al. (2022) use a factorial design with no prior assumption about the marginal utilities of each attribute. They use a d-efficient design that minimizes the size of the variance-covariance matrix (and features both orthogonality and "level balance"). They use 20 choice tasks in two blocks, and each respondent is asked to complete 10 choice questions.

6.3.8. Contingent valuation cases with bid distributions only

Clucas et al. (2015) do not use multi-alternative choice experiments, but double-bounded dichotomous choice contingent valuation questions. They used pre-tests to design the different bid amounts.

Zambrano-Monserrate (2020) uses only one generic policy and assigns six different bid values randomly in his straightforward contingent valuation study. This not being a choice experiment in the usual sense, the mix of attributes across choice tasks is less of a focus.

7. Sample representativeness

If a study is designed merely to show that there *can* exist *some* kind of systematic relationship between the attributes of different alternatives and people's preferences among those alternatives, then it is unnecessary to use a representative sample. However, if the preferences implied by a study are intended for use in benefit-cost analysis of public projects, or for any other kinds of real-life decision-making, it is *extremely* important that the average willingness to pay for the sample can be scaled to the general population of the jurisdiction in question.

Very few of these studies undertake much in the way of assessment of the extent to which the preferences identified in their choice analysis are representative of the general population, or even of some specific population, although Yao et al. (2019) go to considerable lengths to carry their research through from their sample to benefit-cost analysis for the general population of New Zealand. Some studies report descriptive statistics (proportions) for the characteristics of their respondents. Even fewer compare these sample characteristics to the corresponding proportions in the population of interest. There is very little discussion of representativeness, let alone much evidence of concern about possible selection on unobservables.

Authors are also inclined to calculate response rates very generously, with virtually no consideration of whether (or why) any invited respondents drop out upon learning the topic of the survey. This could be handled much better in most of these studies. A couple of these papers use some kind of weighting scheme, but only one paper implements a Heckman-type selection model in the process of dropping respondents who are deemed to provide protest responses.

7.1. "Stratified" samples

Bennett et al. (2018) appear to use a stratified sample from eight villages, with two small groups sampled from each village, including about 19-22 households from each small group. Their discussion includes the qualification: "Insofar as this sample is sufficiently representative of rural communities in this region…."

7.2. Quota-based samples that match marginal distributions of population characteristics

Rogers and Burton (2017) describe their sample as stratified on age, gender and location, and as being "nationally representative" of Australia. They relegate their sample characteristics to the supporting information for their published paper. There seems to be no information about whether respondents could choose to continue with the survey after they learned its topic. "Response rate" is not mentioned in the published paper.

Valasiuk et al. (2018) conducted a "random door-to-door round" with socioeconomics controlled to be consistent with those of the Belarusian population. They provide descriptive statistics for their estimating sample. In their discussion of their results, however, they scale their average WTP of about $US 20.25 to the total adult population of Belarusians, which would require an assumption that their sample is representative of the entire population.

Boeri et al. (2020) compare their choice experiment sample of 3000 people with the UK population in 2011, considering about 27 proportions. They note some differences, but do not consider selection on unobservables. The survey company they use screened the sample to be

representative in terms of male/female ratio, age and employment status, but there seems to be no attention to attrition of potential respondents upon them learning the topic of the survey.

Czajkowski et al. (2021) select farmers for their survey by a stratified quota sampling method, with quotas for the size of the farm and main type of agricultural production. The study area is the Biebrza Valley, so it is not clear that their findings are quantitatively transferable to other regions.

Kim et al. (2021) use a quota-based sampling strategy implement by the online survey company they use (Embrain.co.kr), targeting potential participants in proportion to the population of each metropolitan area, in pursuit of a nationally representative sample with respect to jurisdictions of residence. But the study population was limited to people who had visited at least one national park in South Korea in the previous five years, although the authors expect that "almost all of the Korean people" have visited a national park at least once in this time period. They report statistics on gender, age, education brackets, income brackets for the estimating sample, but do not compare these to the population values for South Korea (and do not appear to use these variables as systematic shifters for marginal utilities).

7.3. Quota-based samples with attention to possible sampling bias from opt-in

Sharma and Kreye (2022) use the Qualtrics web survey service to collect panelist responses from 690 households in Pennsylvania stratified by gender, age, education and income. Invited panelists can voluntarily opt into the survey. While they do not appear to address the type of systematic selection that is created by people's voluntary decisions to opting into the survey (after learning the topic), they do describe their efforts to deal with disqualified respondents by using a raking procedure to produce weights to reduce sampling bias. They compare sample proportions in sociodemographic categories or brackets against Pennsylvania's 2010 census proportions.

7.4. Include descriptive statistics for the sample without comparison to population

Steven et al. (2017) describe their sample in terms of respondent age, gender, education, country of residence, retirement status, and income (for 88% of respondents). But their intercept survey of on-site participants makes it hard to know what population these respondents represent. There are no comparisons to the general population other than the a mention that 32% of people who revealed their incomes have incomes in excess of $80,000 per year. There seems to be no discussion of selection bias or representativeness at all.

Cerda et al. (2018a) describe the characteristics of their sample, but do not appear to compare this visitor sample with the general population of Chile. They explicitly acknowledge that only in-country visitors to the park comprise their sample, and that their results do not necessarily transfer to foreign tourists or transfer to other parts of the world. They acknowledge in their concluding discussion that a large share of their visitor sample is young and lower-income than average, but that ecotourists worldwide tend to be older and higher-income. They also note that tourist interest may be different from the broader local community's interest. Cerda et al. (2018b) also describe the characteristics of their sample without relating it to the general population.

Garnett et al. (2018) use a consumer panel to sample adults from three Australian states that coincide with the spatial distribution of the bird species that are the focus of their study. However, only 1,421 people of the 7,816 people invited to take the survey provided answers, and only 1,119 people provided completed surveys. It is possible that attrition happens after people learn the general topic of the survey. The authors do not appear to address the potential for people's decisions not to take the survey, or if they take it, the factors that result in them failing to complete it. The relatively low overall response rate would seem to leave their results vulnerable to selection on unobservables. The provide descriptive statistics overall (marginal means), as well as for subsets of respondents, but no population statistics for comparison.

Dobson et al. (2022) report purging their sample of responses with unreasonably short completion times and exploring for the presence of respondent fatigue, but do not seem to address what their omitted responses might do to sample representativeness. In terms of respondent characteristics, they note that they have 69% female respondents, the median age group is 3-39 years, and 61% have at least an undergraduate degree. They describe their sample as "relatively wealthy." They surmise that their data provide a "reliable representation of respondent preferences" for the sample that they use because there is little evidence of respondent fatigue. However, they do not speculate on how well their sample represents the population for which these results might be used to guide policy decisions.

7.5. Convenience samples unlikely to represent any particular general population

Liu and Yang (2019) intercept visitors at bird-watching pavilions at a Black-Faced Spoonbill reserve in Taiwan. They report their criterion for approaching potential interviewees. They tell each potential interviewee that the survey is for managing a coastal wetland. They do not comment on how many total visitors were approached, but report that they handed out 440 questionnaires and of those, 434 were recovered complete. They report a "98.6 percent response rate," but this is only from the people who accepted a survey, so they have no idea about what preferences for bird species exist among visitors who did not agree to take the survey. They also know nothing about non-visitors, so it is not clear what population their sample actually represents. In their Discussion and Conclusions section, they do mention that it would be helpful to survey people who want to visit the reserve but do not. Actually, it would be more helpful to have a representative sample from the population who is being taxed to provide the services available at the reserve, to see whether "social" benefits for the main species and others at the site are concentrated or widely distributed, and whether overall social benefits for site visitors and non-visitors exceed social costs.

Sehra and MacMillan (2021) acknowledge that their sample is not gender-balanced, with 57% females, 38% males and 5% not reported. Their study is conducted in Tokyo, with respondents intercepted "opportunistically" in farmer's markets, public parks, train stations, and university campuses. Questionnaires were self-completed. The majority of their respondents live in Tokyo and are of working age. The main body of the paper does not address whether their sample proportions match any particular population within Japan, or whether their results have any potential for transfer to other regions.

Stemmer et al. (2022) use a convenience sample of people encountered at or near a major birding site on an island in northeastern Norway, where they were permitted to collect email addresses. They recruit respondents for their online survey with a short in-person survey. The local harbor service also made the recruitment form available to a majority of registered visitors to the island in question, and a lodging property near the site also distributed the survey to its guests. They give detailed descriptive statistics for their sample, but do not compare it to the population of Norway or to the population of Europe (the origin for most visitors). They do not appear to discuss selection bias.

Xu and He (2022) use a "convenience sampling approach" in distributing their on-site intercept surveys. They justify this approach by the claim that it is "not only efficient in implementing the survey, but has also been proven to be adequate for recreational value estimation. [It would be prudent to check out the papers they use to support this claim. While this may be true in some cases, it is unlikely to be true in general.] These researchers administer 400 questionnaires with the help of four trained college students over a one-week period in December of 2019. They net 385 valid responses and characterize this as a response rate of 96.25%, without commenting on how many people declined to accept a questionnaire or whether this decision is correlated with their preferences. They also do not consider whether the preference of people who visit the site can be scaled to the general population. For this, visitors would need to be "drawn at random" from the general population, which is unlikely.

7.6. Localized samples with some explicit effort to render estimates more representative

Clucas et al. (2015) admit that their samples were not designed to be an accurate representation of the two cities' populations, so that aggregated WTP based on uncorrected samples would likely lead to bias. The resort to post-stratification to construct weights based on census information about each jurisdiction. They also acknowledge that their surveys were conducted mostly during business hours, leading to a likely over-representation of those who do not work outside of home. They conclude that "unless selection of the sample is based on unobservable respondent characteristics," their weighted estimation results are not likely to be biased.

7.7. Samples from narrowly defined populations without general population characteristics

Guimaraes et al. (2014) were able to identify at least the number of Swedish visitors to the Azores for birdwatching and how many visits per person were made since 2005. Their goal is to estimate information appropriate to conduct economic impact estimates for the local economy, rather than to quantify WTP by birders, so their study has a slightly different purpose. They have only 97 completed questionnaires, but five choices for each of the bird-watching areas for each person in the sample (i.e., ten choices per person), so they are able to estimate marginal utilities with reasonable precision.

Krishna et al. (2019) acknowledge that their sample households represent buyers of caged birds, but this sample is not expected to be representative of the entire population of Jambi City in Sumatra, Indonesia. 504 households from 26 neighborhoods were selected, where the sample was limited to people who either kept caged birds or had kept them in the three years prior to the survey. The identities of these households were established in consultation with neighborhood

heads, bird shop owners and other bird enthusiasts in each neighborhood (i.e., "key informants"). It is not clear whether the average WTP in this group of individuals can be scaled to the general population. One assumes that birds in the wild are valued by people who do not keep caged birds, so if their findings concerning species attributes are to be used beyond just the caged bird market to the general population, it is difficult to predict how representative this sample will be.

In their study concerning bird-friendly certification for coffee, Gatti et al. (2022) do not appear to discuss representativeness of their sample relative to the entire population of the US (or even relative to the population of U.S. coffee-drinkers).

7.8. Recruitment process unreported (but apparently discussed in supplementary materials)

Yao et al. (2019) do not describe their respondent recruitment process in the published paper, but invite interested parties to request information from the corresponding author.[4] This is unfortunate, because the credibility of any population-level inferences from any choice experiment study depend critically on the respondent recruitment process. The link provided in the paper lead only to Elsevier's permanent digital object identifier for the paper, and Elsevier does not appear to include the supplementary information.

7.9. Explicitly address/correct at least some aspect of selection bias

Zambrano-Monserrate (2020) seeks a nationally representative sample within Ecuador by surveying people from "several cities in the country" where seven cities are mentioned explicitly "among others." While he provides descriptive statistics for his sample for age, sex, six income brackets and four levels of educational attainment, he does not compare these to the corresponding national statistics. He does, however, address the problem of sample selection bias as a consequence of eliminating respondents who give "protest responses." He resorts to a two-step Heckman approach, treating protest responses as non-response. He finds no evidence of selection bias in this analysis.

8. Estimation methods

8.1. Selection modeling

The only instance of some type of selection modeling in this inventory of papers appears to be Zambrano-Monserrate (2020), as mentioned above. In general, it is not possible to conclude that any set of survey results is free from sample selection bias unless the research has some means to test statistically whether characteristics of the respondent (observable or unobservable) are unrelated to their responses in the choice experiment (and thus unrelated to their WTP). We have found that Institutional Review Boards can be persuaded to permit elicitation of basic individual sociodemographic characteristics and zip codes prior to the "consent" page that informs respondents about the subject matter of the survey.

[4] This information has been requested as of 1/20/2023.

Of course, IRBs prefer in general that you not use people's responses to any kind of question until you have received their explicit consent to be a participant. However, when the research uses a standing panel of respondents (as with many survey research firms who can be contracted to field research surveys to their panelists), it is clear that these individuals have already given their permission to be used as respondents for a variety of research projects. Given that their individually identifying information is never at risk, and the fact that a zip code is large enough to obscure their identities to any client of the survey research firm, a handful of sociodemographics and zip code information can help the researcher prepare for assessment of systematic selection into the survey once the topic of the survey is announced to potential respondents.

There remains the question of whether people who volunteer to serve as members of a standing consumer panel have systematically different preferences for the environmental good being studies relative to the preferences of people who do not become panel members. High-quality panels are likely to be more diligent in recruiting panelist that span the entire spectrum of the population, and many panels will employ quota-based sampling to ensure that the final sample has descriptive statistics (typically marginal means) that match the marginal means for the corresponding characteristics in the population of interest. However, this does not necessarily mean that the people who show up in the final sample have exactly the same distribution of unobservable characteristics as occurs in the general population. All of these concerns have the potential to bias the estimates of WTP derived for the estimating sample in terms of their suitability for extrapolation (scaling) to the general population of interest.

A key insight about sample selection modeling in choice experiments is that the research must plan in advance, anticipating the need for demonstrating the representativeness of the estimating sample. Only if sufficient information is available for both respondents *and* nonrespondents to the main survey can rigorous selectivity modeling be undertaken. At the very least, if a consumer panel is being used, the researcher should be sure to request basic sociodemographic information not just for the final set of respondents, but for every panelist that received an invitation to do the survey, whether or not they end up in the final sample. This includes people who do not respond to the invitation as well as people who are deemed ineligible because they would cause the final sample to exceed some quota. Geographic information for each respondent, if not provided by the survey research firm at an adequate level of resolution, needs to be asked of each potential respondent if they engage with the survey at all. This information, of course, cannot be known for invitees who do not even begin the survey.

We note that conventional Heckman two-stage correction algorithms based on the inverse Mills ratio (the hazard function for the response model) are technically not appropriate in a conditional logit model because the joint distribution of the errors in the selection equation and the "outcome" equation are uncorrelated due to the properties of the extreme value distribution that underlies choice models. This has not stopped some researchers from treating a fitted inverse Mills ratio from a selection model like any observable respondent characteristic and interacting it with a status-quo indicator to permit the inverse Mills ratio to shift WTP for any non-status-quo alternative. Just using the fitted response propensity, however, would be equally useful. But neither option produces a corrected second-stage model that accounts formally for selection bias. The usual Heckman method is appropriate *only* when the errors in the selection equation and in

the outcome equation are *bivariate normal* (and potentially correlated). Thus a choice experiment involving just one substantive alternative versus the status quo could have its utility function parameters estimated using a binary probit model, and packaged algorithms could be exploited to estimate a probit model with sample selection (as in Stata). But when the choice tasks involve more than just two alternatives, the estimation problem is more complex.

A dissertation by Mitchell-Nelson (2022) includes a chapter on more-appropriate methods for correcting for sample selection in multi-alternative choice models.[5]

8.2. Conditional (multinomial) logit

All studies in this inventory that employ true choice experiments begin with a conventional conditional logit specification, where the respondent is assumed to have representative preferences and the specification of the indirect utility function involves fixed parameters. Only two studies stop there. For the last five years, almost every study has moved on to consider latent-class models or random-parameters logit.

8.3. Mixed logit (random-parameters logit, RPL)

Mixed logit models can be specified to allow the marginal utilities of each attribute to be random across respondents (but constant for any given respondent across multiple choice tasks for that individual). The method of simulated moments is typically employed to estimate the parameters of these distributions, and most researcher employ packaged algorithms for estimation of these models.

Instead of estimating the marginal utility parameter for each attribute for a representative respondent, the mixed logit approach estimates both a mean and a standard deviation, across respondents, of these random parameters, where each estimate in these pairs (parameter mean and parameter standard deviation) has its own statistical standard error which permits hypothesis testing. The usual question is whether the estimated standard deviation for a given marginal utility parameter, across respondents, is statistically significantly different from zero. If so, there is deemed to be "heterogeneity in preferences" with respect to that attribute.

However, no observed characteristics of the respondent are typically employed to explain this heterogeneity, so it is often called "unobserved heterogeneity." The simpler version of these models constrains to zero the off-diagonals of the *parameter* covariance matrix (not to be confused with the usual variance-covariance matrix for the *estimates*, in a conventional conditional logit model estimated by maximum likelihood). Richer specifications, that often take much longer to converge, can have fully unrestricted parameter variance-covariance matrices, permitting the individual marginal utility for one attribute to be correlated, across the sample, with the marginal utilities for other attributes.

If the goal of a research project is to estimate average preferences in the population and to use those estimates to calculate an overall social benefits measure from mean WTP estimates, then

[5] This chapter is entitled "Ample correction for sample selection in the estimation of choice models using online survey panels," and the correction method adapts code available in the R-based software package called Apollo.

mixed logit models can be perfectly adequate. However, the researcher will often be interested in the distributional consequences of some policy, so it will be important to identify which segments of the population are likely to enjoy larger or smaller benefits from the policy. For these types of tasks, it is helpful to be able to distinguish which types of people have higher and lower WTP for the non-market good in question. Simply knowing that "preferences are heterogeneous" as nonzero standard deviations in marginal utilities based on a mixed logit model would imply, is not sufficient for addressing distributional consequences.

8.4. Latent class logit (LC)

Latent class specifications assume that each respondent's preferences can be expressed as a finite mixture of a small number of different classes of preferences (usually between two and four classes, in practice). Rather than assigning each respondent to a specific preference class based on their observable characteristics, each person has only a *probability* of being a member of each latent class. These models have two components, estimated simultaneously. The first component is a multinomial-type logit component where the probabilities of belonging to each latent class other than the baseline class is modeled as a function of the respondent's observable attributes. Then, conditional on membership in each class, a distinct set of marginal utility parameters is estimated. If there are M latent classes, there will be M-1 sets of parameters for the M-1 class membership equations (given that the parameters for the baseline class are normalized to zero).

These models can sometimes be balky to estimate, especially if one is attempting to estimate a model with more latent classes. However, it is often possible to "sneak up" on the final specification by starting with the simplest possible model (with most attributes having initially fixed coefficients and only one or two having coefficients that vary across classes). Likewise, it is possible to start with a membership equation that involves only "intercept" parameters for each class. The researcher can then slowly generalize the specification to free up the constraints that restrict different coefficients to zero (i.e., models that leave out these variables), relying on the previous model's converged parameter estimates as starting values.

Latent class models can thus also be used to reveal that preferences are not homogeneous, but differ systematically with the mix of respondent attributes for each respondent. However, the researcher is left to decide upon a label for each class of preferences based on the respondent attributes that increase or decrease the probability of membership in that class.

We will break at this point to summarize how the different studies in this inventory have chosen the attribute mix for their choice experiments and how they have specified and estimated their choice models.

8.5. **SUMMARY TABLE**: design/estimation software and methods/specifications

Study	CE design software/ tradition	Estimation software	Non-standard	Mixed logit?	Latent Class # classes	Attrib intx?	Pref heterog?

Guimaraes et al. (2014)		NLOGIT 4.0	Only conditional logits			Yes, for one site	Systematic[6]
Yao et al. (2014)	Ngene					No	
Clucas et al. (2015)	CV (DBDC)	SAS 9.2	Interval model				Systematic
Rogers and Burton (2017)	Ngene			Error components		Yes	Systematic
Steven et al. (2017)	Ngene				3 & 4	No (but tried some)	Via LC (limited)
Bennett et al. (2018)	Scarpa/ Rose/ Ferrini			Error components		Yes	Systematic
Cerda et al. (2018a)	Louviere/ Hensher	Limdep/NLogit 9.0		RPL		No	Via RPL
Cerda et al. (2018b)	Louviere/ Hensher		Nested logit			No	
Garnett et al. (2018)	Rose/ Bliemer	Latent Gold Choice 5.0			3	No	Via LC
Valasiuk et al. (2018)	Street/ Burgess		Best/ worst (& just best)	RPL		No	Via RPL and some systematic
Krishna et al. (2019)	Stata 13.0		Only conditional logit	RPL		Yes	RPL and split samples
Liu and Yang (2019)	Hearne/ Salinas, Juutinen, Kelly, Lee					No	No
Yao et al. (2019)	Ngene			RPL		No	Via RPL (two-stage systematic)

[6] "Systematic" means that respondent attributes are explicitly interacted with the level of at least one attribute in the choice set, perhaps only with a status quo indicator, or with other featured attributes that differ across the substantive alternatives.

Boeri et al. (2020)	Scarpa/ Rose/ Ferrini	unspecified			4	No	Via LC
Zambrano-Monserrate (2020)	CV		Spike models			No	Systematic on status quo effect
Czajkowski et al. (2021)	Scarpa/ Rose/ Ferrini			RPL		Yes (exten-sive)	Systematic
Kim et al. (2021)	SAS 9.3	Nlogit 5.0		RPL		No	Via RPL
Sehra and MacMillan (2021)	R (Alg Design)	NLOGIT4.0, R			3	No	Via LC
Dobson et al. (2022)	Ngene	Latent Gold Choice 5.0			3	No	Via LC
Gatti et al. (2022)	Zwerina	Stata 16		Indep and corr. RPL		No (precluded by design)	Via some split-sample estimates
Sharma and Kreye (2022)		Stata 15.1		RPL		No	Systematic (for status quo effect)
Stemmer et al. (2022)	Ngene	Nlogit and Mplus		RPL and MLR		No	Via RPL
Xu and He (2022)	SPSS 24.0			RPL		No	No

9. Richer specifications

9.1. Continuous attributes versus categorical attributes

9.1.1. Artifacts of design software: small numbers of attribute levels

The use of choice-set design software helps to minimize the sample size needed to achieve statistically significant estimates of key preference parameters (or maximize estimation efficiency for any given sample size), but not without requiring the researcher to make some tradeoffs. Design software for choice experiments typically expects a relatively small number of levels for each attribute, so that the number of combinations of attributes in the full factorial design is not impossibly large. Unfortunately, this leads many researchers to convert fundamentally continuous variables, such as the number of species of birds, into a small set of indicators for intervals of the full possible range (e.g. three or four intervals). This forces the

estimation process to assume that the marginal utility from such an attribute takes on only three different values—one for each interval—and is constant within each interval.

This strategy has the advantage of not forcing the portion of utility determined by that attribute to be linear in the level of the attribute, as would be the case if the variable had been specified on a continuous cardinal scale and entered the indirect utility function in additively separable form. However, there is a distinct disadvantage in that smoother patterns of changing marginal utility cannot be estimated for use in interpolation or modest extrapolation beyond the levels used in the survey. For example, with discrete intervals as attribute levels, it is not straightforward to specify utility as being logarithmic or quadratic or cubic in the level of that attribute, to allow for smoothly diminishing marginal utility from an attribute, or to employ a quadratic form that would permit the researcher to identify levels of the attribute at which the fitted marginal utility changes sign, implying a utility-maximizing (or utility-minimizing) level of the attribute exists.

Furthermore, if the researcher wishes to consider the possibility that the marginal utility from this attribute also varies with the level of some other attribute, then many more terms are required to identify these interactions. With continuously measured cardinal attributes, only one interaction term may be sufficient to identify interdependencies between attributes in determining utility levels (and hence willingness-to-pay).

9.1.2. Design of the cost attribute

Unlike other continuous variables that are often represented by intervals or described only in qualitative terms, such as "current," "increased," or "decreased," (such as species richness), the cost attribute in most choice experiments is captured by a small number of specific amounts. These amounts are ideally selected initially to span the range of likely willingness to pay amounts in the population, given the range of all other attribute combinations being used in the experiment. Test samples are often needed to determine whether the distribution of cost amounts is appropriate for the study. Costs that reveal the most about preferences will be close to the respondent's unobserved individual willingness to pay for the combination of attributes in question. In that situation, the decision about whether to choose that particular alternative versus the status quo is not simply a "no-brainer."

Choice experiment design software takes the researcher's chosen levels of each attribute and generates combinations that will minimize the variance of the estimated parameter variance-covariance matrix while preserving the ability to identify "effects" in models of different levels of complexity. But design software cannot tell the researcher what mixes of other attributes will be *plausible to respondents* or what the range of *real* household-level costs would be for a given attribute mix. Thus many researchers resort to design software that requires a preliminary sample to provide some "prior" information about the parameters of the utility function, on average, in the population of interest.

If the researcher wishes to solve the estimated indirect utility function to yield estimates of the maximum willingness to pay for a particular configuration of the other attributes, the cost variable must be modeled as continuous, even if only a small handful of discrete cost levels were employed in the design. Almost all choice experiments also make the indirect utility function

both linear and additively separable in the implicit "difference in net income relative to the status quo" variable that is constituted by the cost attribute. The marginal utility of net income (given by the negative of the resulting cost coefficient) is used to calculate the marginal rate of substitution between other attributes and money (net income, in this case). This particular marginal rate of substitution is interpreted as the marginal willingness to pay for the attribute in question.

A linear and additively separable term in the net income associated with any given alternative is econometrically expedient, but this assumption may not be very realistic. Benefit-cost analysis in North America, for the most part, assumes that everyone shares the same marginal utility of income. (The "distributional effects" of specific programs are generally considered separately.) This uniform-marginal-utility-of-income assumption is consistent with estimating a single common marginal utility of net income for everyone in the sample/population, but it is not likely to be realistic. It is generally accepted that the marginal utility of an extra dollar is greater for a poor person than for a wealthy person.

It is nevertheless entirely possible to generalize a choice model to allow the marginal utility of income to differ systematically with observable respondent attributes (if you wish to distinguish different WTP for different groups, even if the same bundle of attributes is being considered and people have identical tastes for those attributes. It might seem obvious to allow the cost coefficient to differ systematically with the respondent's household income. However, this conflicts with the conventional North American benefit-cost analysis assumption of equal marginal utilities of income.

We must also keep in mind that household income is a consequence of a lifetime of previous household choices and constraints (for the respondent and all their ancestors). Some of these choices may be correlated with preferences for the environmental good in question, making household income and WTP jointly endogenous. Furthermore, if the marginal utility of net income (the cost attribute) and the marginal utility of other program/policy attributes are all determined by the same set of respondent characteristics, it can sometimes be very difficult to tease apart the alternative interactions that best explain choices. With the same characteristics affecting both the marginal utility of net income and the marginal utilities from other attributes, the WTP calculation will involve terms in the same variables in both the numerator and denominator.

9.1.3. Marginal rates of substitution other than WTP

Very few choice experiments specifically explore other marginal rates of substitution revealed by respondents. This is likely because of the strong tendency not to use cardinal measures for other attribute levels. When two attributes are continuously measured and the indirect utility function is specified as being linear and additively separable in both of these attributes, then a single marginal rate of substitution between those two attributes can be calculated analogously to a WTP measure. Given that there are often many important tradeoffs to consider across different programs or policies, these other marginal rates of substitution may be important to any useful description of preferences. But the fact that choice experiments often use intervals or qualitative

descriptions of attribute levels, to keep the number of different "levels" small, opportunities to assess these other marginal rates of substitution are forgone.

9.1.4. Discrete levels for continuous attributes

At the very least, researchers might choose to specify four distinct point values for a cardinally measured attribute, rather than describing its levels as lying within one of four mutually exclusive and exhaustive intervals. In estimation using interval-coded attribute levels, the interval midpoints could be used as proxies for an exact level of such an attribute, but then the researcher does not know whether the respondent has paid attention to the lower or the upper limit of the interval. It is not typically necessary that a cardinally measured attribute be described in terms of a small set of mutually exclusive and exhaustive intervals. If the researcher is comfortable with a utility function that is linear in the levels of the attribute, even just two different values can suffice. If it is important to allow for *curvature*, then three or four different values can be adequate to capture preferences where marginal utilities are not constant. One might argue that preferences are obscured about equally by the use of vague intervals and by the use of a logarithmic or polynomial form to smooth utility between a small finite number of levels.

The next summary table focuses specifically on each study's bird-related attributes and enumerates the different levels employed in the choice experiment. I also give details on how the attribute enters the utility function and include some notes on what seem to be missed opportunities in each specification.

9.1.5. **SUMMARY TABLE**: Characterizations of bird-related attributes

Study	Context	Bird-related attributes	Levels	Enter utility function as	Comments
Choice experiments					
Steven et al. (2017)	Generic birding sites	(1) the number of threatened bird species present at a site (either critically endangered or endangered),	(1) nil present (2) 1-3 CR/EN species (3) >3 CR/EN species	Two indicator variables: (1) omitted category (2) 1(medium number) (3) 1(high number)	Missed opportunity to use a continuous variable and estimate constant or varying marginal WTP per threatened or endangered species present at the site
		(2) the diversity of birds at a site (species richness)	(1) < 20 species (2) 20-60 species (3)>60 species	Two indicator variables: (1) omitted category (2) 1(medium amount) (3) 1(high amount)	Missed opportunity to use a continuous variable and estimate constant or varying marginal WTP for additional bird species present
		(3) the number of regionally endemic bird species at the site (restricted-range species).	(1) nil present (2) 1-6 species (3) >6 species	Two indicator variables: (1) omitted category (2) 1(medium number) (3) 1(high number)	Missed opportunity to use a continuous variable and estimate constant or varying marginal WTP for additional endemic species present
Sehra and MacMillan (2021)	Rice-growing practices in Japan	(1) indicators for the presence or absence of three different animal species (including one bird species)	(1) bird species absent (2) bird species present (Black-crowned night heron)	One indicator variable (1) omitted category (2) 1(bird present)	Missed opportunity to use continuous number of bird species or number of individual birds present and estimate a marginal WTP per species or per bird
Garnett et al. (2018)	Conservation options for Australian bird species.	(1) Brown thornbill, mainland	(1) status quo (2) maintain in the wild (3) assisted colonization (4) keep in a zoo	Three indicators: (1) omitted category (2) 1(maintain in wild) (3) 1(assisted coloniz.) (4) 1(keep in zoo)	No information about the scale of the program for each species. Decisions about real programs of this type would have to consider

	How the program would help each of four species (if at all) to adapt to climate change:	(2) Brown thornbill, Tasmania	(1) status quo (2) maintain in the wild (3) assisted colonization (4) keep in a zoo	Three indicators: (1) omitted category (2) 1(maintain in wild) (3) 1(assisted coloniz.) (4) 1(keep in zoo)	the scale of the program as well as the relative costs of each program. They "assumed…the costs would be met from government tax revenue, … rather than from direct personal donations." This does NOT mean that people don't care about the cost of such programs. They pay the taxes, and the programs likely have different opportunity costs.
		(3) Scrubtit	(1) status quo (2) maintain in the wild (3) assisted colonization (4) keep in a zoo	Three indicators: (1) omitted category (2) 1(maintain in wild) (3) 1(assisted coloniz.) (4) 1(keep in zoo)	
		(4) Rufous scrub-bird	(1) status quo (2) maintain in the wild (3) assisted colonization (4) keep in a zoo	Three indicators: (1) omitted category (2) 1(maintain in wild) (3) 1(assisted coloniz.) (4) 1(keep in zoo)	
Valasiuk et al. (2018)	Agricultural practices for Belarus fenland area	None	n/a	n/a	Bird impacts only implicit
Czajkowski et al. (2021)	Agricultural practices in Poland's Biebrza Valley	None	n/a	n/a	Bird impacts only implicit
Guimaraes et al. (2014)	Tourist preferences for wetland management in the Azores	None	n/a	n/a	Bird impacts only implicit
Cerda et al. (2018b) – Central Chile site	Management of Chile's Lircay National Reserve	(1) wildlife conservation research, where bird species is one option Each group is an "attribute level"	Mutually exclusive: (1) just general knowledge of species (2) birds (3) reptiles (4) insects (5) *Nothofagus* species (6) *Sclerophyllous* species (7) Herbaceous plants	Six indicators: (1) omitted category (2) 1(birds) (3) 1(reptiles) (4) 1(insects) (5) 1(*Nothofagus* species) (6) 1(*Sclerophyllous* species) (7) 1(Herbaceous plants)	There is no measure of the quantity units of wildlife conservation research, only whether there is "any further effort" versus "no further effort" for each species.

Krishna et al. (2019)	Demand for caged birds in Sumatra For each caged bird species:	(1) rarity in the wild	(1) rare (2) abundant	One indicator variable: (1) 1(rare in wild) (2) omitted category	Some attributes were explicitly listed in each choice set; others were associated with the representative species of bird depicted, and were included in the model, but not in the choice task. When some attributes are not explicitly listed along with the others, there is always a question of how much attention the respondent paid to those attribute (or whether they even perceived them). Objective versus subjective implicit attributes.
		(2) trading frequency	(1) frequently traded (2) not traded frequently	Conveyed implicitly by one of 30 distinct species with each mix of these three attributes	
		(3) relative price	(1) high-end prices (2) low-end prices		
		(4) bird origin	(1) no information (2) wild capture (3) captive breeding	Two indicators: (1) omitted category (2) 1(wild capture) (3) 1(captive bred)	
		(4) trainability for singing	(1) cannot be trained (2) can be trained	One indicator variable: (1) omitted category (2) 1(trainable)	
Boeri et al. (2020)	Different quantifications of biodiversity in coastal wetlands Levels of avian biodiversity:	(1) the number of different types of birds	(1) current (2) increase (3) decline	Two indicator variables: (1) omitted category (2) 1(increase) (3) 1(decline)	Potentially cardinal number converted to status quo or a directional change
		(2) the total number of individual birds	(1) current (2) increase (3) decline	Two indicator variables: (1) omitted category (2) 1(increase) (3) 1(decline)	Potentially cardinal number converted to status quo or a directional change
		(3) the likelihood of seeing a rare or unusual type of bird	(1) current (2) higher (3) lower	Two indicator variables: (1) omitted category (2) 1(higher) (3) 1(lower)	Potentially cardinal probability converted to status quo or a directional change

		(4) the probability of seeing a "wildlife spectacle" such as thousands of birds in one flock.	(1) current (2) increase (3) decrease	Two indicator variables: (1) omitted category (2) 1(increase) (3) 1(decrease)	Potentially cardinal probability converted to status quo or a directional change
Kim et al. (2021)	Potential construction of new airport for a Korean island birding destination	(1) reductions in the number of birds	(1) low: no damage to bird habitats and no change in the bird population (2) medium: 5% decline in bird habitats and 15% decrease in total bird population (3) high: 10% decline in bird habitats and 30% decrease in total bird population	Two indicator variables: (1) omitted category (2) 1(medium) (3) 1(high)	Cardinal percentage loss converted to two indicators. Perfect collinearity between habitat decline and percent loss of bird populations, so cannot separate effect of habitat loss from bird loss
Gatti et al. (2022)	Demand for different types of environmental certification for coffee	Type of environmental certification for growing method	Mutually exclusive: (1) no certification (2) bird-friendly (3) shade-grown (4) organic (5) pesticide-free	Four indicators: (1) omitted category (2) 1(bird-friendly) (3) 1(shade-grown) (4) 1(organic) (5) 1(pesticide-free)	Coffee farms can have more than one type of certification. With mutual exclusivity, missed an opportunity to be able to assess marginal utility of one certification if others are also present.
Sharma and Kreye (2022)	Value of bird conservation on private forestlands in Pennsylvania Attributes described in prose form, not tabular	(1) the category of birds that will benefit from the intervention (common species or rare species)	(1) common species (2) rare species	Effects-coded indicator (1) 1 = common (2) -1 = rare	"Effect codes" approach is not typical in econometric analyses; indicator variables for a non-baseline category is more common. As the indicators in this case differ by two units, rather than one, it is expected that the coefficient size will be half as large.
		(2) the benefits of birds to humans (ecological or recreational).	(1) recreation (2) ecological services	Effects-coded indicator (1) 1 = recreation (2) -1 = ecological	
Stemmer et al. (2022)	Demand for birding site attributes (Europe)	(1) of birding quality	(1) birdwatching of good quality (uncommon target species, habitat and/or birding spectacles)	One indicator: (1) omitted category (2) 1(exceptional)	Assumes birding quality at a "birding site" will never be less than "good"? What about the future?

			(2) birdwatching of exceptional quality (unique target species, habitat and/or birding spectacles)		
		(2) avian diversity (intervals of species richness).	(1) fewer than 15 species (2) 15 to 40 species (3) more than 40 species	Two indicators: (1) omitted category (2) 1(medium) (3) 1(high)	Forgone opportunity to employ a cardinal measure, to be able to estimate constant or systematically varying marginal WTP per species
Xu and He (2022)	Tourists' demands for visits to the Nansha Wetland in Guangdong, China	(1) species of rare birds (provincial or national protected birds)	(1) status quo level (45 species) (2) "a worse level" (37 species) (3)"an improved level" (not clear in paper)	Continuous variable for "species of rare birds"	Three of the four non-cost attributes are continuous variables; could have estimated attribute interaction effects or tested assumption of constant marginal utilities
Rogers and Burton (2017)	Features of ecological offsets to protect birds	(1) species protected by the ecological offset	Mutually exclusive? (not clear) (1) Eastern Curlew (2) Ruddy Turnstone	Two indicators: (1) 1(Eastern Curlew protected) (2) 1(Ruddy Turnstone protected)	Study interacts one of the species protected (Ruddy Turnstone) with number of birds protected, allowing for different marginal values for an additional bird of each species.
		(2) the number of birds that would be protected	(1) 0 (status quo?) (2) 500 (2) 1000 (3) 1500 (4) 2000	Continuous variable	
Yao et al. (2014) and Yao et al. (2019)	Threatened birds, animals and plants in New Zealand Encounters with two species of birds:	(1) brown kiwi	(1) current condition: calls heard in 1 out of 200 planted forests (2) level 1: calls heard in 10 out of 200 planted forests (3) level 2: calls heard in 20 out of 200 planted forests	Two indicators: (1) omitted category (2) 1(level 1) (3) 1(level 2)	Continuous variables are converted to categorical variables, precluding estimation of marginal utility per "call heard in 200 planted forests" or per "sighting in 8 drives"
		(2) bush falcon	(1) current condition: sighted in 1 out of 8 drives (2) level 1: sighted in 3 out of 8 drives (3) level 2: sighted in 5 out of 8 drives	Two indicators: (1) omitted category (2) 1(level 1) (3) 1(level 2)	
Dobson et al. (2022)	Donations by UK residents	(1) presence or absence of	(1) absent (2) present	One indicator: (1) omitted category	No distinction between one or many species, or for how

	to fund conservation areas in Africa	threatened bird species.		(1) 1(present)	many individuals of each species are present
Cerda et al. (2018a) – Atacama site	Programs to protect biodiversity in a National Park in Chile's Atacama Desert	(1) category of bird species protected	Mutually exclusive: (1) none (2) interior birds (scavenger raptors) (3) interior birds (passerine raptors) (4) shorebirds	Three indicators (1) omitted category (2) 1(scavenger raptors) (3) 1(passerine raptors) (4) 1(shorebirds)	If indicators are mutually exclusive, cannot determine whether marginal utility from one scavenger is affected by present of the other, for example.
Liu and Yang (2019)	Management of the Black-Faced Spoonbill Reserve in Taiwan	(1) the number of Black-Faced Spoonbill individuals seen	Four levels (also conveyed by images): (1) no visit = 0 individuals? (2) 50 individuals (3) 100 individuals (4) 200 individuals (5) 400 individuals	Continuous variable	Could interact these two measures to determine whether other species can substitute for additional Black-Faced Spoonbills (e.g. if the marginal utility from an extra Spoonbill is lower when other species can be seen as well, or whether the marginal utilities are independent from each other)
		(2) the number of other bird species seen	Four levels (also conveyed by images) (1) 0 other species (also for no visit) (2) 3 other species (3) 7 other species (4) 12 other species	Continuous variable	
Contingent valuation					
Clucas et al. (2015)	Urban birds of two type in Berlin and Seattle	Increase urban bird populations for either songbirds (housefinch or greenfinch) or corvids (American crow or hooded crow) Songbird vs. corvid varied across split samples)	Each respondent is asked about both a finch species and a crow species For the finch species: If they report liking finches, ask about WTP for (1) an increase in finch abundance, vs. (2) no change in finch abundance If they report disliking finches, ask about WTP for (1) a decrease in finch abundance (2) no change in finch abundance	Not immediately clear how answers to the different questions were combined into a single model. With enough indicators, the responses to the different questions could be safely combined. They seem to have estimated separate models for finches and corvids in the two cities, but do not distinguish between increases and	Experiment did not specify the sizes of the changes in abundance, describing only some unquantified direction of the change. This means a forgone opportunity to infer WTP for an additional finch or crow in each city. With continuous abundances, could also have calculated marginal rates of substitution between species, by pooling the data

			If they report a neutral attitude toward finches, ask both questions of the same respondent Likewise for crow species.	decreases in species numbers. Appears to pool all data and estimate in WTP space using interval-censored dependent variable.	and differentiating the species with an indicator.
Zambrano-Monserrate (2020)	Andean Condor habitat in Ecuador	A program to prevent extinction of the Andean condor	(1) no new public program to save the Andean condor (2) a new public program to save the Andean condor	Indicator variable: (1) omitted category (2) 1(new program)	No information about the scope of the program, just an undifferentiated "program bundle" as is typical with strict CV formats.

9.1. Interactions between attributes

If the design of the choice experiment allows the researcher to identify more than just the main effects for each attribute, it seems important to explore for possible interaction effects. A model with just main effects means an implicit assumption that the marginal utility derived from one attribute is independent of the level of any other attribute. Often, this may not be the case. It is disappointing when a study has had the information necessary to test whether marginal utilities are constant or whether they depend on the levels of other attributes, but the authors choose not to consider this possibility.

9.1.1. No attribute interactions included

Among the studies included in this review, those in the following table appear NOT to report marginal utilities (or corresponding WTP amounts) for one attribute that depend systematically on the level of any other attribute. These studies are ordered by their publication dates.

Yao et al. (2014)
Cerda et al. (2018b)
Garnett et al. (2018)
Valasiuk et al. (2018)
Liu and Yang (2019)
Boeri et al. (2020)
Zambrano-Monserrate (2020)
Kim et al. (2021)
Sehra and MacMillan (2021)
Dobson et al. (2022)
Sharma and Kreye (2022)
Stemmer et al. (2022)

9.1.2. Some attribute interactions tested (but rejected)

Steven et al. (2017) considered attribute interactions but report that they find no significant interaction effects among site attributes.

9.1.3. Potentially interesting attribute interactions precluded by design

Gatti et al. (2022) appear to have "switched on" just one sustainability certification at a time for the coffee brands in their choice experiment. They mention in their "limitations" section that "coffees can have multiple eco-labels and each eco-label can have multiple sustainable attributes. They acknowledge that future studies could consider "whether these attributes are substitutes or complements." could easily have interacted their attributes, to permit answers to questions such as whether the marginal utility from one kind of certification depends on whether other types of certification are also present. But they include no interactions in the results they report in the paper.

9.1.4. Some interactions among attributes included in specification

Guimaraes et al. (2014) employ no interactions between attributes for one of their sites, but for the other, they find that the marginal utility for cars being admitted varies systematically with whether there are also new water bodies within the wetland and whether there are observation towers at the site. For some reason, however, the presence of observation towers, by itself, is not included as a stand-alone attribute. Perhaps the marginal utility of the estimated coefficient on the stand-alone attribute was not statistically significant, so the coefficient was constrained to be zero. However, it is generally best practice in modeling to retain lower-order terms in variables when a higher-order term is included. The coefficient on the lower-order term may be statistically insignificant due to a lack of statistical power, not because it is truly zero.

Rogers and Burton (2017) include interactions between the species protected and the number of birds to determine whether the marginal utility of abundance depends on the species being protected.

In their study of eco-compensation (WTA) to participate in conservation-friendly agricultural practices that may affect the Jiangsu-Yancheng Coastal Wetlands Rare Birds National Nature Reserve, Bennett et al. (2018) employ a rich array of interaction terms between program attributes in their estimating specifications. Many of the coefficients on their interaction terms are statistically significant, highlighting the importance of these attribute interactions in understanding how the features of an eco-compensation program can affect the size of the subsidy required to induce participation.

Krishna et al. (2019) consider interactions among the attributes of the caged bird species in their study of demand for species rarity. They interact relative market position with rarity and with trade frequency for the species. They are also forced by a positive price coefficient to consider a quadratic form in the price of the species, which reveals that the marginal utility of price eventually becomes negative. They attribute the initially positive price coefficient to a Veblen effect at lower prices that gives way to a downward-sloping demand curve when prices get high enough.

9.2. Heterogeneity in preferences

It is the norm now for choice experiments for authors to move quickly away from the homogeneous preferences implicit in a straightforward conditional logit model to explain choices. The mixed-logit/random-parameters-logit generalization, as well as the latent class generalization, were introduced above in the section on estimation methods. Three basic strategies are used to introduce preference heterogeneity:

(a.) Random parameters (RPL) models, also called mixed logit models, allow for a distribution of marginal utilities for each attribute across the sample (and thus the population). These models may assume independent marginal utilities across attributes, or correlated marginal utilities across attributes. The marginal (dis)utility of the cost

attribute is often constrained to be a fixed parameter, mostly just to make calculations of WTP easier.

(b.) Finite mixture models for preferences (latent class models) are a less flexible alternative, since they assume everyone's preferences are some probabilistic mixture of a small set of underlying preference classes. But these models can be interesting because explicit respondent characteristics can be employed in the "class membership" sub-model that explains probabilities of being in any given latent class (relative to a numeraire class of preferences). If there are M different preference classes (typically just two or three), then there are M-1 sets of coefficients on respondent characteristics that can explain class membership. These models can be balky to estimate, but a strategy of generalizing the model from an MNL specification little by little can be more successful. Start with the simplest possible specification and work up.

(c.) A third way to accommodate heterogeneity in preferences is to allow specific marginal utility parameters to depend systematically on particular respondent characteristics. This can be accomplished by interacting respondent characteristics with one or more attributes, to allow the marginal utility of an attribute to vary with the level of that characteristic. Caution must be used in interpreting these models, however. Many respondent characteristics are self-selected and can therefore be correlated with the respondent's latent willingness-to-pay amount. This is *especially* true for attitudinal or opinion variables, or knowledge variables, or any respondent current or past behavior relative to the environmental good in question. Models using characteristics that are jointly endogenous with willingness to pay can be helpful to demonstrate the "construct validity" of the demand information in the model. However, endogenous regressors, as always, can introduce simultaneity bias. It is best to stick with forms of respondent characteristics that are exogenous or at the very least, pre-determined relative to the choice task (like gender and age and ethnicity), rather than recreational habits, or "user" status, for example, or people's attitudes toward the non-market good in question.

9.2.1. Homogeneous preferences

Liu and Yang (2019), in their intercept survey at the Taiwan Black-Faced Spoonbill Reserve, collect respondent characteristics data on gender, age, levels of education, annual income, and the number of visits they have made to the reserve. However, they use this information only to describe their sample, not to introduce observed heterogeneity or latent class membership into their specifications.

Kim et al. (2021) report descriptive statistics for a large number of "explanatory variables," including gender, age, education brackets, income brackets, along with indicators for prior visits to the island in question and trust in government environmental policies. But they do not appear to use the sociodemographic variables in the estimating specifications.

9.2.2. Random parameters models

Cerda et al. (2018a) accommodate heterogeneous preferences through a random parameters logit specification. Some other researchers go beyond just switching from all fixed parameters to mostly random parameters, estimating both a mean and a standard deviation for the random parameters.

9.2.2.1. RPL with differences across split samples

Krishna et al. (2019) identify unobserved heterogeneity in preferences via their random parameters logit model. The **explore split samples** along the dimensions of household wealth, respondent education, age, marital status, but relegate these results to online appendices. They note that across the different wealth groups, most coefficients on the standard deviation parameters become statistically insignificant, which they interpret as evidence that household wealth is one of the major sources of preference heterogeneity. [This could simply be a loss of power for smaller subsets of data. Why do they not interact sociodemographics with the bird attributes and look for systematically varying preferences, preserving more degrees of freedom?]

Gatti et al. (2022) have consumer characteristics including age brackets, gender, educational attainment, household size and income brackets. The report **split-sample estimates** for groups of consumers that are and are not familiar with the Bird Friendly certification/logo, and for groups high-income versus low-income groups. For some reason, they omit to report the parameter standard deviations or correlations when they report their mixed logit results.

9.2.2.2. RPL with systematic varying marginal utilities

Yao et al. (2014) use their random parameters logit model, which includes unobserved heterogeneity in the cost coefficient (marginal utility of income), to calculate mean marginal WTP for each attribute (level) in their study. They then investigate heterogeneity in a *second-stage* panel-data model which uses as a dependent variable the predicted marginal mean WTP for each respondent, calculated from their choice model. They regress the marginal mean WTP on respondent characteristics (including distances from the resource to be enhanced by the program).

Yao et al. (2019) take the analysis of their augmented New Zealand sample much further, with much more geographical indexing of values and aggregated mean WTP by small census jurisdictions, which they can now map. This emphasis with the much larger sample permits them to explore benefit-cost comparison.

Cerda et al. (2018a) say that they interact eight socioeconomic variables (age, sex, income, number of children, years of education, rural/urban residence, region of residents and subjective probability of real payment) with their any-program (non-status-quo) indicator to assess systematic preferences for or against any program, regardless of its attribute levels. However, these interactions do not appear to be reported in the main paper. They also explored respondents' interest in visiting the park and their subjective perceptions of the role of the park as shifters of the any-program ASC effect (although this is typically for construct validity only, since these variables are jointly endogenous with WTP). These interactions do appear in their main table of results.

Valasiuk et al. (2018) consider mixed logit specifications for their main attributes and are careful to report point estimates of the correlations between their normally distributed non-cost marginal utilities (mentioning that all standard deviations are statistically significant. They discuss the interpretation of the positive and negative correlations between different marginal utility estimates. However, they also interact some of the attributes of their alternatives with a few characteristics of their respondents. These authors allow their cost coefficient to be random as well (a modeling choice that can make it more complicated to calculate MWTP amounts, since the denominator of the WTP formula, as well as the numerator, are random, and in their case, the cost parameter is lognormal.

Stemmer et al. (2022) use a random-parameters logit model to accommodate heterogeneity. But they also put considerable effort into quantifying "birding specialization" based on three dimensions: behavior (two measures), skills and knowledge (four measures), and commitment (centrality of birding to their life, four measures). They talk about having difficulty "integrating" a first-order three-dimensional model of birding specialization. [I suspect this is some form of weighted sum of responses on constituent variables, used to build "factors" with specific labels. Not sure about this technique.] They use interaction terms between their latent "specialization" variable and each attribute-level dummy variable (including the fee). It appears that efforts to estimate the hybrid model simultaneously were unsuccessful, so the estimate their specialization model first and use the predicted value (not clear what that is) in a separate conditional logit model. They say they cannot compare the fit of their attributes-only, RPL, and hybrid models because "the additional model variables inflate LL values in HC models." But I think they could sum the log-likelihood models for the independently estimated specialization model and the MNL model, and compare this to their estimates for the hybrid model.

9.2.3. Latent class models (finite mixtures of preferences)

9.2.3.1. Three classes of preferences

Garnett et al. (2018) identify three different latent classes of preferences: "wild preferred," "no extinction," and "status quo." It seems that they somehow use a wide range of respondent characteristics to explain class membership, but their reporting of their results is non-standard for the economic literature. They somehow manage to report descriptive statistics for each of their three preference classes. It is unclear, however, how they assign each individual to a specific class when class membership is only probabilistic. Perhaps they assign each individual to the preference class for which they have the highest estimated probability?

Sehra and MacMillan (2021) introduce preference heterogeneity using latent class models with three classes. The presence of birds has a 5% statistically significant marginal utility only for one group, whereas in the homogeneous preferences model, the presence of birds has a positive marginal utility that is significant at the 0.001 level. Perhaps this explains why they calculate marginal WTP amounts only for their MNL model. It is not clear, however, why they report only one set of membership coefficients, rather than two, for their three-latent-class specification.

Dobson et al. (2022) allow their 3-class latent class specification to have class membership determined by education, environmental organization membership, and the logarithm of income.

9.2.3.2. Four classes of preferences

Steven et al. (2017) find that sociodemographic characteristics (at least the ones they had available) were not good predictors of group membership in their 4-class latent class models. The only thing that made a difference to class membership appears to have been non-attendance to the cost attribute.

Boeri et al. (2020) use latent class analysis to identify four classes of preferences, with class membership influenced by income, age, environmental activity, visits to environmental settings, and gender. However, they appear not to discuss the likely endogeneity of environmental activity and visits to environmental settings.

9.2.4. Marginal utility parameters vary systematically with respondent characteristics

Guimaraes et al. (2014) use their survey to collect a considerable number of respondent attributes, although many of them are likely to be strongly endogenous with willingness to stay. For their Cabo da Praia quarry site on Terceira Island, the marginal (dis)utility from extra days of stay varies systematically with an indicator for the respondent's first visit to the area, the year of their last visit and whether they are visiting from Europe. For their Paul da Praia wetland site, this marginal disutility varies with whether the respondent donated to conservation or wildlife causes in the previous year, whether they participate in birding festivals, whether this is their first visit to the area, the number of islands visited, the year of the last visit, whether they are from Northern Europe, whether they have a university degree, and their income.

Clucas et al. (2015) consider that demographic, cultural and socioeconomic factors (e.g. age, population density, employment status, years of education, foreign-born) influence people's choices in the SP study, but they also explore the relationship between willingness to pay and attitudes towards birds as well as general attitudes about conservation. Their reported estimates for Seattle suggest that a "positive conservation attitude) has a significant effect on WTP for finches, whereas WTP for crows is lower for women and for homeowners, but higher with higher population density, higher with conservation organization membership. For Berlin, WTP for finches also decreases with population density, and WTP for corvids (magpies) has several more significant determinants.

In their study with farmers near a coastal wetland reserve, Bennett et al. (2018) find that willingness to accept compensation for modified pesticide use depends on program attributes, and that these attributes interact with farmer characteristics to affect WTA.

Rogers and Burton (2017) include individual attitudinal heterogeneity constructed from batteries of questions designed to measure "social license to operate" (SLO) for the oil and gas industry of two types: whether the industry is perceived to provide economic benefits and whether the industry will improve community wellbeing and act in the community's interests. But they also include interactions between the location for the ecological offset activity and the jurisdiction of

residence for the respondent interacted with the location of the offset. The SLO variables are interacted with the entity implementing the offsets and with the status-quo indicator. The jobs provided by the oil and gas development (a factor that did not vary across alternatives, and is thus used like a respondent characteristic) did not shift the status-quo effect.

Cerda et al. (2018b) interact education, age, and the respondent's subjective probability of being able to pay the higher fee with the indicator for their non-status-quo alternatives.

Zambrano-Monserrate (2020) uses a binary contingent valuation elicitation format, so a whole range of respondent characteristics and behaviors/attitudes can potentially be interacted with an implicit "any program" indicator (when they are simply added to this list of regressors in an already-differenced binary choice model for contingent valuation analysis.

Czajkowski et al. (2021) identify substantial unobserved heterogeneity in preferences indicated by the statistically significant standard deviations in their marginal WTA parameters. They then undertake to explain some of this heterogeneity using observable farmer and farm characteristics. They include observed heterogeneity by augmenting the random marginal WTP parameters with systematic variation via interaction terms in farm/farmer characteristics and program attributes. They employ a full set of interactions with 14 different farm/farmer characteristics and identify numerous statistically significant interactions. However, substantial amounts of unobserved heterogeneity still manifest in the statistical significance of the standard deviations of the baseline random marginal utilities.

Sharma and Kreye (2022) collect a lot of attitudinal data in their survey. They use a set of statements with answers on a five-point Likert scale, with several statements about the respondent's subjective knowledge and opinions, and other factual knowledge questions about bird populations. Other statements related to opinions about present and future conditions for bird populations. They also elicit answers for 27 statements that pertain to nine attitude dimensions towards birds. These attitudes are described as "naturalistic, humanistic, moralistic, aesthetic, ecologistic, scientistic, symbolic, negativistic and dominionistic." They also collect attitudes towards timber harvesting, motivations for harvesting, and government involvement in private timber management decisions (landowner assistance or topdown/regulatory approaches). Sharma and Kreye (2022) implicitly interact respondent characteristics with an "any program" indicator (relative to the status quo) in their binary choice question format.

9.2.5. **SUMMARY TABLE**: Types of systematic heterogeneity included

Study with heterogeneity by respondent characteristics	**Exogenous/ predetermined heterogeneity**	**Behavioral heterogeneity**	**Attitudinal/belief heterogeneity**
Guimaraes et al. (2014)	- whether respondent has a university degree	- year of last visit - number of islands visited	

	- whether respondent is from Northern Europe - income	- whether this is first visit to the area - whether money was donated to conservation or wildlife causes in the previous year - whether respondent participates in ornithological events	
Clucas et al. (2015)	- age - area population density - years of education - income category	- notice birds at least weekly - conservation organization membership	- positive conservation attitude
Bennett et al. (2018)		- whether household received training before in the use of pesticides - whether household is in the top third of total pesticide expenditures - whether household is in the bottom third of total pesticide expenditures	- whether household believes that pesticides have an impact on household health - whether household believes that pesticide boost crop yields
Rogers and Burton (2017)	- residence in Western Australia		- constructed social license to operate (economic legitimacy) - constructed social license to operate (social legitimacy)
Cerda et al. (2018a)	- age - education		- probability of real payment - thoughts concerning the role of the park
Cerda et al. (2018b)	- age - education		- probability of real payment
Zambrano-Monserrate (2020)	- knowledge - sex - age	- charitable contributions - attendance at talks	- importance of animals

	- income bracket - education		
Czajkowski et al. (2021)	- arable land - farm with streams/rivers - farm with areas subject to flooding - share of farm in Natura 2000 and National Park - bird knowledge - number of birds recognized - information treatment	- crop production - livestock production - number of crops - livestock/ha - work force - has participated in agri-env schemes	- subjective AE practices for birds
Sharma and Kreye (2022)	- age - gender - race = White - income categories - education categories - factual knowledge - current conditions poor		- subjective knowledge - future condition poor - attitudes toward birds - attitudes toward harvesting - attitudes toward government

10. Quantified values for birds

Every study in this inventory that directly calculates marginal willingness-to-pay estimates for birds seeks to value a different bird-related attribute. Some do not calculate marginal willingness to pay for anything because they omit a cost attribute from their choice sets. Other studies do not have any directly bird-related attributes in their choice tasks. While they may seek to associate a WTA or WTP measure with some sort of ecosystem attribute, they do not quantify bird values.

Consequently, this is not a context where we can make straight-across comparisons between the wild bird values estimated in different studies. Given the wide range of things being valued, and the variety of issues that arise in these efforts, we do not construct a summary table for this information. There would be too many columns and too many empty cells. Instead, we review the monetized values of different bird-related attributes used in these different studies in four categories, with studies sorted by publication date in each category. We begin each paragraph with the citation to the paper, to make it easier to locate the different types of value estimates for each paper in our inventory.

10.1. Studies that do not directly monetize bird-related attributes

Rogers and Burton (2017) value program attributes in terms of numbers of their numeraire species, Ruddy Turnstones, rather than (more usually) in terms of money.

Garnett et al. (2018) avoid using a cost variable, arguing that "costs would be met from government tax revenue, which is usually the case for conservation in Australia, rather than from direct personal donations. Of course, the government budget is provided by taxes on households, so households will bear at least an opportunity cost if the government pays for these programs. These authors state that they "had no interest in the willingness-to-pay for each attribute but were interested in the trade-off between attributes regardless of cost." This is problematic because it leaves the different costs of the alternative programs implicit, and respondents will not necessarily assume that each alternative is identically costly. Forcing the marginal utility of cost to be zero for all respondents, if respondents do not assume that all cost differences are zero, is potentially problematic.

Stemmer et al. (2022) find that birding quality matters in all their models, but bird biodiversity matters most to more highly specialized birders (more likely to be "listers"). Curiously, they seem not to bother calculating any willingness to pay measures from their fitted coefficients, but this is probably because they could not identify a statistically significant marginal (dis)utility for their cost (fee) attribute, even though they talk about WTP in their introductory material. The fee does matter, however, when interacted with specialization. But its coefficient is positive, not negative. This suggests that they have some problems with the range of fees they proposed. Given the income levels among their respondents, perhaps none of the proposed fees were high enough to influence people's choices?

10.2. Studies that provide monetized marginal values for some bird-related attributes

Yao et al. (2014) find that their "level 2" increases in encounters for bush falcon populations in New Zealand are valued at $24/year (with $14 to $91 as a 90% interval), and for brown kiwi at $21/year (with $13 to $76 as a 90% interval).

Steven et al. (2017) identify one of their four latent classes of preferences as "Price-is-no-object birders." For this group, they fix the marginal (dis)utility of the cost attribute at zero and do not estimate WTP or marginal WTP amounts for this group. However, for the other groups, they can estimate WTP differences by level for their various measures of bird diversity. "Quantity-driven birders" are willing to pay $105 and $135 for sites with medium and high biodiversity, and $33 and $66 for sites with medium and high levels of endemic species. "Special-bird seekers" are willing to pay less ($18 and $36) for sites with medium and high levels of bird biodiversity, but they are willing to pay the most for threatened species ($31 and $45) for medium and high levels, and about the same for endemic species. Their fourth preference class, "Confused," had relatively low WTP for endemic species.

Cerda et al. (2018a) find that, among different types of birds, respondents were more interested in supporting research about shorebirds (mean WTP US $6.3), although they were also willing to pay to support research on birds occurring in the interior areas of the Atacama Park in Chile.

(WTP for birds did not differ much from WTP for some lesser-known mammals, amphibians, reptiles or pollinating insects, unlike the differences for birds in some of their other work.)

Cerda et al. (2018b) find that people's willingness to pay for conservations of birds through in-depth research is greater than for other species. They discuss the possibility that birds are "tourism flagship species." They find that the marginal mean WTP/visitor/visit for "protection of birds through investigation" is about US$7.

Krishna et al. (2019) appear to miss an opportunity to calculate *total* willingness to pay for caged birds *of every named species* in their study, as a function of the birds explicit or implicit attributes and their estimated preference parameters. Instead, they focus mostly on the signs of the *marginal* willingnesses to pay for rarity. It would have been interesting to see whether their estimated fitted WTP for each species bears any relation to the current market prices for that type of caged bird. They might have been able to display point estimates and confidence intervals for the predicted price of each species plotted against the actual mean or median price and the 95% range. This could inform the market about whether specific species appear to be under-priced or over-priced (how much consumer surplus is enjoyed by buyers of caged birds). Perhaps they have other work that explores this dimension.

Liu and Yang (2019) estimate the marginal WTP for an additional Black-Faced Spoonbill seen at the reserve to be 0.3 NTD, where 1 US dollar = 30.68 NTD. Respondents' marginal willingness to pay for one more other species seen is 6.4 NTD. These authors observe that the marginal WTP per individual Spoonbill is $1/20^{th}$ of the marginal WTP for one more species of other birds. The authors point out that given that spoonbills are migratory, their number is not a decision variable for the site managers. However, site managers could act to increase the number of species of local resident birds. Unfortunately, this ratio does not reflect a bird-for-bird tradeoff, but a bird-for-species tradeoff.

Zambrano-Monserrate (2020) finds, across alternative functional forms for the model, that predicted mean WTP to protect the Andean condor ranges rather widely, but he elects to focus on the results from the best-fitting spike log-normal specification. He finds individual WTP for condor protection is about US $18 to $35, with a median of $24.83.

Kim et al. (2021) report that a "medium" reduction in the number of birds is equivalent to -1380 KRW of income, and a "high" reduction in the number of birds is equivalent to -2520 KRW of income. "Medium" is a 5% decline in habitat and a 15% decrease in the total bird population. "High" is a 10% decline in bird habitat and a 30% decrease in the total bird population. Without baseline numbers of birds, the MWTP for medium and high reductions cannot be converted into WTP per bird per year.

Sehra and MacMillan (2021) find, in their homogeneous and additively separable specification, that the marginal willingness to pay for rice that is grown in wildlife-friendly paddies that support bird species is about 1060 JPY higher than WTP for rice that is not grown using wildlife-friendly farming. In contrast, WTP to support fish species is only about 250 JPY higher. This is on the same order as WTP to support frog species, but that marginal utility is not statistically significantly different from zero in their homogeneous preferences model.

Dobson et al. (2022) estimate a marginal utility (or latent class marginal utilities) for conservation areas that contain threatened bird species. They cannot estimate a marginal willingness to pay when a conservation area has a threatened bird species for their Latent Class 1 because its cost coefficient is positive (although not statistically significantly different from zero). For class 2, the WTP for a threatened bird species is small and positive, but for class 3, the willingness to pay appears to be more on the order of $75.

Gatti et al. (2022) estimate a WTP premium of about $2.23 for a 12-oz bag of ground coffee if the coffee is grown with "Bird-Friendly" certification.

Sharma and Kreye (2022) report mean household WTP for bird conservation across all of the randomized program scenarios used in their study OF $11.83 with a 95% confidence interval of $10.65 through $13.01. But they also estimate marginal WTP amounts (part-worths) for many attributes.

Xu and He (2022) estimate non-use values related to bird populations in the Nansha Wetland in China. They calculate a per-capita marginal willingness to pay for an additional species of rare bird of about US $1.19, although they do not provide an interval estimate for this WTP. However, they use a convenience sample of on-site visitors. Transferring the estimates from their sample to the entire population of the region is thus questionable.

10.3. Studies that make some effort to extend their bird values to broader populations

Clucas et al. (2015) allowed for positive and negative WTP amounts by valuing both finches (which most people like in both Berlin and Seattle) and corvids (which were associated with negative WTP for Seattle). Based on people's willingness to pay for either increases or decreases in these two types of bird species, they estimate that the economic value derived from people's enjoyment of native urban songbirds is about $US 120 million/year in Seattle and $US 70 million in Berlin. This assumes that the respondents they recruited in each city are representative of the entire population of each city.

Yao et al. (2019) make a heroic effort to scale their WTP estimates to the population of New Zealand as a whole, to permit benefit-cost analysis of a biodiversity initiative. Using individual-specific means of marginal WTP amounts and modeling these as a function of respondent characteristics and geographical factors allows them to predict WTP per year for brown kiwi conservation at fine-resolution census levels and display their findings spatially in map form. For their five-year conservation program, they ultimately arrive at a net present value estimate of NZD 507 million, using a social discount rate of 3%.

Boeri et al. (2020) find that considering the 22 million households in the UK, the average WTP for an increase in the number of species of birds in coastal areas (species richness, specifically) is about £5 per household, implying that about £110 million overall could be donated to environmental organizations concerned with bird conservation in coastal areas of the UK. The values for increases in their other measures of biodiversity were lower and not statistically

different from each other, valued at about £3.5. However, it is not clear how to interpret these WTP amounts because the units are not cardinal, just qualitative (higher, current, decreased).

10.4. Values of some ecosystem attributes when bird-related attributes are only implicit

Guimaraes et al. (2014) do not directly value birds, focusing instead on measuring the likely economic impact (increased days of stay) for visiting birders. If these days-of-stay can be converted using some average cost-per-day for local stays, then the estimates could be used to calculate respondents' total and marginal willingnesses to pay for management alternatives and their attributes. None of the attributes are bird populations, however. Just infrastructure for avitourism.

Bennett et al. (2018) consider marginal WTA estimates for farmers to participate in conservation-related practices. While these practices will have an effect on bird populations, they do not address bird populations directly.

Valasiuk et al. (2018) do not value the Aquatic Warbler species directly, but they do make some back-of-the-envelope calculations based on their estimates of average WTP in their choice scenarios for different land-use management policies. They scale their results to the adult population of Belarus and and then calculate that this yields an annual WTP of more than 8240 $US per hectare of the Zvaniec mire, which they note is a high estimate compared to other conservation management estimates in the international literature.

Czajkowski et al. (2021) estimate WTA values, in 100 EUR per hectare per year, for farmers' participation in agri-environmental program. The large number of interaction effects mean that the marginal WTA for each program attribute depends on other variables. The interact farming practices with respondent's subjective bird knowledge, as captured by three separately measured variables, and consider how farmers' marginal WTA to adopt specific practices depends to some extent on the farmer's knowledge concerning birds. The bird-knowledge-related heterogeneity in preferences does affect WTA amounts, but these are not measures of bird values specifically.

11. Caveats to the study

It is generally good practice to acknowledge the limitations of a study, of only to pre-empt criticisms from referees. Not all papers include a separate section where they acknowledge some shortcomings of their approach. In the following summary table, I collect the limitations that authors acknowledge in their write-ups, but also note some others that have not been identified explicitly in each paper. These additional points are incomplete, but may be helpful to future researchers.

11.1. **SUMMARY TABLE**: Acknowledged and other caveats

	Caveats
Yao et al. (2014)	*Acknowledged*: The study has a rather low response rate. Given the small sample size, the quantitative results should not be aggregated over the total population of New Zealand.
Clucas et al. (2015)	*Acknowledged*: Analyses focus on expected WTP, but individual-level preference heterogeneity was sizeable Choice experiments concern only finches and corvids, not all wild bird species Several variables expected to be significant were not One source of potential bias is the presence of respondents at home at the time of the survey. They attempt to correct using age and employment status, and state that "unless the selection of the sample is based on unobservable respondent characteristics, we do not believe our results were biased as a result." Concerns about stated preference methods being subject to known biases.
Rogers and Burton (2017)	*Acknowledged*: Theirs is a new area of study. They recommend caution in attempting to extrapolate their results to other biodiversity contexts or to policy settings outside Australia. *Unacknowledged:* They treat their attitudinal variables as though they were exogenous to the individual's preferences about biodiversity offsets. It is likely that attitudes and preferences are jointly endogenous, and both are influenced by other individual characteristics and prior experiences.
Steven et al. (2017)	*Acknowledged:* They were unable to identify sociodemographic characteristics (i.e., age, gender, country of residence) as important determinants of WTP and birding destination preferences.
Bennett et al. (2018)	n/a
Garnett et al. (2018)	n/a
Valasiuk et al. (2018)	*Unacknowledged:* These authors are not very careful to distinguish the distributions of their variables from the assumed distributions of their random utility parameters.
Krishna et al. (2019)	n/a
Liu and Yang (2019)	n/a
Yao et al. (2019)	
Czajkowski et al. (2021)	*Acknowledged*: Their study is concerned mostly with farmers' preferences, but this is only a first step towards a full cost-benefit analysis of a more complex policy design.
Sehra and MacMillan (2021)	*Acknowledged*: There may be significant problems associated with delivering biodiversity benefits at a landscape scale. What constitutes a "Satoyama" landscape is context-specific, and there is variation in biodiversity across Satoyama systems. Each

	flagship species scheme might need to specify a different flagship species appropriate to specific areas.
Dobson et al. (2022)	*Acknowledged*: A flagship campaign involves many considerations other than selecting a site with the most attractive attributes. South African conservation areas provided a good example, but preferences may have been different had a different country been chosen. Their results are not "a rigid description of the most effective flagship areas for a UK audience."
Gatti et al. (2022)	*Acknowledged*: This is an exploratory survey conducted online, so respondents are forced to consider the alternative types of coffee certification without having a "visual or tasting experience."
Sharma and Kreye (2022)	*Acknowledged*: It is unclear whether there is a real difference in utility from bird conservation across races or if this finding is an artifact of sampling bias. Non-white residents represent a small share of the Pennsylvania population, and non-white respondents were also under-sampled relative to this share. *Unacknowledged*: They report that attitudes towards birds also had the strongest influence on WTP in their model. This is unsurprising since WTP is itself a type of attitude towards birds. Again, attitudes and preferences are jointly determined, and while consistency between attitudes and preferences is helpful to demonstrate "construct validity," attitudes are not generally observable in the general population (limiting model transfer), and simultaneity bias can certainly create biases in the estimates of key parameters.
Stemmer et al. (2022)	*Acknowledged:* The study presents findings only for a relatively under-examined region, Northern Europe. The study relies on a relatively small sample, with 205 completed responses, and with some groups over-represented and others under-represented. These authors find a lack of price sensitivity, and regret not having revised the scale of fees to include higher amounts, since the lack of price sensitivity was evident in their pilot study. They speculate that some people may have found their proposed conservation and maintenance fee a "good thing" because it would contribute to the protection and management of these ecosystem services. Their measures of birding quality and scenery attributes were vague, "with the potential for diverse perceptions of meaning across respondents." Their results may be "limited by the specific context (e.g., spatial scale, birder population, and season). *Unacknowledged:* These authors build three dimensions of "birding specialization" to describe each respondent's engagement with wild birds. The variables that are ingredients for these measures are not generally observable for individuals, so the measures are helpful for assessing "construct validity," but render the model not transferable to other populations. These birding specialization variables are jointly

	endogenous with people's values for bird species diversity. Thus the model uses how much people care about birds to explain another measure (WTP) of how much people care about birds.
Xu and He (2022)	*n/a*

12. Recommendations for future research

12.1. Recommendations by study authors

In the following summary table, I have extracted from each paper the main suggestions by the study's authors about issues that future research might be designed to address.

12.1.1. **SUMMARY TABLE**: Author recommendations for future research

Study	Future research should:
Guimaraes et al. (2014)	- Contemplate the estimation of both direct and induced effects of avitourism on the local economy, beyond just the direct benefits from longer average stays in terms of higher expenditures at local hotels and restaurants.
Yao et al. (2014)	- Investigate the use, option, and non-use values of biodiversity enhancement through distance effects, and calculate whether the estimated WTP amount for the types of planted-forest species protections would justify the costs of implementing programs with these effects.
Clucas et al. (2015)	- Recognize that areas of greater socio-economic status have higher bird diversity, creating the potential for an ecological inequality, thus public investments in restoration and maintenance of green spaces that increase bird habitat and diversity in urban areas may benefit humans in deprived areas.
Rogers and Burton (2017)	- Explore the increasing the flexibility of the characteristics of offset programs for development that affects migratory bird species, since people are willing to make more tradeoffs than the legislation currently allows.
Steven et al. (2017)	- Examine under-researched avitourist source markets (e.g., China and India) - Continue to use multi-disciplinary approaches, as in this study, to "inform the development of tangible and practical conservation strategies on both public and private lands."
Bennett et al. (2018)	- Use the insights from the current study to compare direct-subsidy approaches in China that focus instead on agricultural non-point-source pollution from small farms.

	- Develop a regional framework that improves use of pesticides, given that pest populations are becoming pesticide-resistant.
Cerda et al. (2018a)	- Assess the preferences of local communities, not just National Park visitors. - Explore heterogeneity in the demand for broader biodiversity by using split samples of visitors, and should focus "either on well-known species or on a broader spectrum of species."
Cerda et al. (2018b)	- Conduct a deeper exploration of attitudes of visitors towards different species and the informedness of visitors about current threats to these species and how these factors affect WTP "in the context of conservation and nature-based tourism. - Investigate whether visitors to protected areas perceive conflicts over the uses of water and the conservation of biodiversity. - Explore the preferences of visitors for changes in the infrastructure of tourism. - Consider the role of attitudes and opinions in driving WTP
Garnett et al. (2018)	n/a
Valasiuk et al. (2018)	
Krishna et al. (2019)	- Explore the nature and determinants of market demand for wildlife forms, and to better understand wildlife supply chains—what motivates rural households to engage in the collection and trade of wildlife products. - Pursue interdisciplinary inquiries at the species level to explore the conservation effects of relevant attribute interactions for caged birds.
Liu and Yang (2019)	- Conduct a survey of people who want to visit the Reserve but do not, to help managers understand the impact of transportation convenience (i.e. public transportation services) on the tourism benefits of the Reserve.
Yao et al. (2019)	- Apply the spatial economic tool called "Forest Investment Framework" to estimate the public net benefits of enhancing other iconic species in New Zealand, where the framework developed in this paper can be expanded "to account for the ecological benefits of the programme through bird population modelling based on the proportion of native and exotic trees per conservation site."
Boeri et al. (2020)	- Explore further the precision of metrics which could be used to evaluate the functional form of preferences related to the number of species [verify]
Czajkowski et al. (2021)	- Seek to provide a better overview of farmers' preferences for various contract characteristics in different settings, so that there is more information to support land use policy with respect agricultural and ecological uses. - Explore the cognitive process of knowledge formation when information is provided to farmers, since object knowledge

	seems to have a positive effect on farmers' choices, but information provision was not as effective in this case.
Kim et al. (2021)	- Include residents, as well as visitors, since their priorities (e.g. for getting off the island, rather than visiting it) might be different. - Express benefits measures, including harm to bird habitat and population, in quantitative terms, rather than just qualitative terms [verify]
Sehra and MacMillan (2021)	- Investigate the challenges of landscape-scale policies, for example, to standardize a definition of "Satoyama" landscapes (a socio-ecological concept) since this characterization seems to be desirable for ecological certification of rice in Japan.
Dobson et al. (2022)	- Involve both researchers and practitioners to co-develop research on how best to use flagship conservation areas to address issues that affect broader conservation area networks.
Gatti et al. (2022)	- Allow for multiple types of sustainability certifications for each product, so it would be possible to learn whether the different certifications are substitutes or complements. - Try to approximate real-life choices using laboratory experiments. - Explore whether WTP differs between environmentally concerned and non-concerned consumers, with a focus on which characteristics identify these two groups. - Explore to what extent consumers perceive the different types of certification as sustainable. - Whether the strong preference for "organic" over "bird-friendly" certification is different between general consumers and bird watchers.
Sharma and Kreye (2022)	- Assess whether asymmetric information about the status of wild bird populations may permit market failures to persist. Is it possible to help people learn more about the status of birds without them becoming anxious about land confiscation if their property is scarce habitat for an important species?
Stemmer et al. (2022)	- Consider a holistic perspective concerning birding-related travel destinations, where destination attributes include "experiential elements" that are not strictly birding-related. - Replicate the study in different regions. - Develop more quantitative, or at least more specific, measures of the bird-related destination attributes. - Increase the ranges of levels of bird diversity and the proposed fee (specific to the context). - Further develop attributes and levels relating to guiding preferences and other aspect of nature-based attractions.
Xu and He (2022)	- Generalize their study, because their study of the Nansha Wetland may not be generalizable to other coastal wetland parks (although it provides an example of how the attributes of

	recreational areas and their levels should be analyzed in future research). The relevant attributes and their levels at other recreational areas may be different.

12.2. Notable lessons from each study

In this section, I review, in general, some additional limitations of each study from the perspective of how one might choose to conduct a choice experiment intended to value land conservation programs to protect and enhance migratory bird species along a specific flyway.

I include commentary on how the design of future studies might be adapted to make it relatively easier to value avian biodiversity and abundance in ways that are more likely to be generalizable to other similar contexts, even if the affected human population has a different distribution of basic characteristics.

This final summary table concludes this review of the literature using choice experiments to value some aspect of wild birds.

12.2.1. **SUMMARY TABLE**: Lessons for future research based on this review

Study/*Journal*/topic	Lessons
Guimaraes et al. (2014) *Ecological Economics* (longer local stays by avitourists)	This study illustrates that while policy-makers may prefer to consider the overall welfare effects of ecosystem policies that affect birds, other constituencies (especially local stakeholders in the vicinity of land acquisition projects) might be much more interested in the positive "economic impacts" of avitourism. It would be desirable to design choice experiments that could support estimates of both welfare effects and economic impacts, to demonstrate to local authorities the value of the resource in attracting tourism and tourist spending. Would need to figure out how to stipulate WTP in terms of local costs for visitors. Would a study with this feature permit the researcher to generalize their model to predict local avitourism spending for visitors to other birding hotspots?
Yao et al. (2014) *Ecological Economics* (brown kiwis in planted forests)	Their survey displays quantitative measures of abundance of brown kiwi and other threatened species, but unfortunately converts this quantitative information into qualitative direction-of-change information in the analysis. Try to preserve more different values of bird-related attributes so a continuous function is possible for MWTP. This may strain the usual choice set design software, but preserves the ability to fit smooth curvature for the MWTP function and to test whether it is constant or diminishing.

Clucas et al. (2015) *Urban Ecosystems* (urban wild birds, Seattle & Berlin)	Use a better sampling procedure, so that findings are potentially generalizable to some known population. Online survey will allow sample to be representative in terms of a half-dozen observable respondent attributes. On the west coast of North America, for example, a high proportion of the population will be urban, so researchers may want to stratify their samples by county or county group to get adequate representation of rural population, using exogenous weights to balance influence in estimation. Use quantitative attribute levels, not just "increase" or "decrease" in abundance of finches and corvids.
Rogers and Burton (2017) *Conservation Biology* (offsets for development that affects migratory birds)	Try to describe more than just one or two species of birds, and use more points of support for the range of "number of birds protected," rather than just three or four levels. Collect basic attitudinal data to allow for demonstration of construct validity, but make sure to collect as many exogenous and predetermined respondent characteristics as is practicable. Make sure to have precise enough locational information on each respondent so that you can locate them in their zip code (or census tract). This is the way to specify demand in terms of neighborhood characteristics, so that the model can be transferred to areas with different neighborhood characteristics. You cannot transfer to a different region any WTP model that depends upon individual information available only from survey respondents.
Steven et al. (2017) *Conservation Biology* (birding sites in Australia)	Avoid using quantitative intervals (e.g. >3 endangered species, 20-60 species), perhaps in favor of "5 endangered species" and "about 40 species." It is impossible to know what "point" value is appropriate for people who choose based on an offered range. You want to minimize respondents' ability to pick their own effective levels of attributes because you have not specified it sufficiently well. It is possible to describe numbers in terms of a most-likely amount and a possible range, and to test whether people make choices based solely on the most-likely amount or whether their WTP also depends on the range. Some respondents could hear about a narrower range of "uncertainty" and others could hear about a wider range of "uncertainty." But there should be sufficient numbers of different "most-likely" amounts to permit the fitting of a smooth function for MWTP (and testing of the smooth function against a set of indicators that yield a non-parametric MWTP (step) function. Step functions can be less useful because it is harder to interpolate (or to extrapolate modestly at the edges of the sample domain). Between point values, it is possible to fill in a piecewise linear function, too.
Bennett et al. (2018) *Ecological Economics*	This study uses attributes of the alternatives that do not explicitly include biodiversity or abundance measures for birds. They describe birds as a reason why farmers should reduce their

(eco-compensation for farmers for coastal wetlands in China)	pesticide use, but birds are not valued directly. This study concerns WTA a subsidy to uses less pesticide. They collect qualitative information about respondents' *subjective* perceptions of how harmful pesticides are for large birds and small birds (i.e. "not harmful," "slightly harmful," "moderately harmful," or "very harmful") but this information is *not* sufficient to map reductions in pesticide use into quantifiable effects on birds. Presumably, a lower WTA compensation for changing pesticide usage means that the respondent derives greater utility from protecting birds, ceteris paribus. But the preamble to the choices also talks about "ground- and surface-water pollution [that] can adversely impact rural household health…can damage and unbalance the regional ecology by harming important bird and animal species, by killing off natural predators of pests…" This obscures WTP to protect birds specifically.
Cerda et al. (2018a) *Environmental Conservation* (desert biodiversity hotspot in Atacama desert, northern Chile)	Their choice experiment mentions only "scavenger raptors," "passerine raptors," and "shorebirds," along with a large number of other types of plants and animals. The program they describe to respondents involves "additional research efforts" concerning each species they describe. It is not clear how to map "research at the species level" into any given measure of biodiversity or abundance among bird species in particular. At best, this study elicits from respondents information about where people think research is needed, producing estimates of the value of research about birds, as distinct from the value of the birds themselves.
Cerda et al. (2018b) *Biodiversity and Conservation* (protected area, central Chile)	This study also seeks to value "protection of wildlife through investigation." The quantity of protection, or its effectiveness, is not well-defined. Soil quality, water benefits, and infrastructure for recreation are also attributes of the different programs. On average, visitors are willing to pay US$6.78 per visit for "protection through investigation" of birds, and these numbers may help prioritize investigation across birds, reptiles, insects and rodents, but it does not measure WTP for any particular measure of biodiversity or abundance.
Garnett et al. (2018) *Oryx* (Australian birds, adaptation to climate change)	This study focuses on four bird species with relatively low "public profiles" where the species differ in their taxonomic distinctiveness. The policy options include in situ conservation, assisted colonization, and establishment of captive populations (e.g., in zoos). The problem with this study is that the cost of each program is not mentioned as an attribute, rationalized by the claim that "the costs would be met from government tax revenue…rather than from direct personal donations." This leaves respondents to impute whatever they wish about the real opportunity costs of each program and whether these costs are likely to differ across programs. Tax revenues are collected from

	households and have other uses. With no monetized attribute, it is not possible to assess WTP for any of the programs or the MWTP to protect any of the species being considered. This was a lost opportunity.
Valasiuk et al. (2018) *Wetlands Ecology and Management* (wet grassland restoration in Belarus)	These authors motivate concern about birds by focusing on just one bird species, the Aquatic Warbler. The programs in the choice experiment, however, are described in terms of the method used to remove shrubs, the managed area, enlargement of the conservation area and the annual cost per person. There is insufficient information about the effects of these different practices on bird populations to permit mapping WTP for each of these attributes into the corresponding MWTP per Aquatic Warbler, or for other bird species that might also benefit from changes in practices or managed areas. If respondents had been informed, specifically, about the expected effects of each program on bird populations, there may have been some scope for valuing bird biodiversity or abundance using their results, but they only go as far as WTP for "manual scything," "mechanical mowing," "controlled burning," "chemical treatment," "1000 ha of managed area," and "1000 ha additional conservation area." Their main specifications include limited socio-demographic heterogeneity, but not of a form that would permit model transfer to similar ecosystems in other regions.
Krishna et al. (2019) *Environment and Development Economics* (caged-bird demand in Sumatra)	Caged-bird interest in Sumatra appears to be much higher than elsewhere in the world, but the prevalence of interest in caged birds creates a unique opportunity to measure WTP for particular species attributes. However, it cannot automatically be assumed that WTP for these attributes amongst pet birds maps directly into WTP for the same attributes among wild birds, which is a more relevant concern for any study of habitat preservation or enhancement along continental flyways. This study provides useful information, but it would be helpful to figure out how to test whether the results could transfer to wild populations, which are enjoyed in different ways from caged birds.
Liu and Yang (2019) *Cogent Social Sciences* (Black-faced Spoonbill reserve in Taiwan)	This study focuses on management options for a popular Reserve area dedicated to a particular iconic species, but at which other bird species can also be observed. The choice scenarios, unfortunately, describe the abundance of the main species, but only the species richness among other types of birds. The sample is limited to visitors to the bird-focuse reserve in question, rather than being a representative population sample, and one would expect demand for bird species richness to be higher among these people who have already self-selected to travel to the Reserve. However, WTP per trip for other kinds of birds is only about $0.20 per species, which seems rather low.

	So perhaps visitors to this reserve are far more interested in its featured species than in other types of birds.
Yao et al. (2019) *Ecosystem Services* (brown kiwi in planted forests in New Zealand)	This paper extends earlier work with a larger sample and a more comprehensive analysis and application of the results. This work carries the preference parameter estimates through an aggregation of public benefits from a biodiversity conservation program at each of 12 "ecologically and economically feasible ecosystem-service sites" and calculates their costs, continuing all the way through to a benefit-cost analysis at the national level. *This persistence is laudable and should be emulated where possible*. To accomplish this, however, it is necessary to use a representative population sample and to describe the ecosystem(s) in question in a way that makes the results plausibly transferable across ecosystems and populations throughout the region in question. These authors "estimated conditional means of WTP distributions for each biodiversity attribute and for each respondent in the sample." Rather than allowing the key marginal utilities that determine MWTP to vary systematically in estimating the choice model, they use their calculated individual MWTP estimates from their mixed logit in WTP-space as a dependent variable in subsequent OLS and spatial regression models that use individual respondent characteristics as well as the sociodemographic and locational characteristics of Census units to explain WTP. This step allows them to transfer their fitted model across all Census units in the country (assuming a representative individual for each Census unit). It is not clear in the paper, though, how they assign the "representative individual" for each Census unit. Presumably, this is someone whose individual characteristics coincide with the mean in the population of each of these characteristics. Since Census characteristics are typically counts within different intervals, there may have been considerable additional work required to identified the "representative individual" in terms of each of the individual respondent attributes employed in the follow-up WTP models estimated as a function of individual and Census unit characteristics.
Boeri et al. (2020) *Estuarine, Coastal and Shelf Science* (different measures of biodiversity, coastal birds)	This work recognizes that there are different ways to describe "biodiversity" among bird species, and asks whether any of these measures are given more or less attention as respondents choose between alternative programs. Unfortunately, the levels of these attributes are described only in terms of their increase or decrease relative to the status quo, with no quantitative information about the sizes of these changes, which could be anything from miniscule to huge in the minds of respondents. It is unfortunate that these authors did not employ two or more numerical measures of (for example) "Number of different types

	of birds you can see when you visit." These numbers could even have been quantified numerical changes relative to some unstated status quo (provided that the researcher can reconstruct an approximate absolute status quo level in each case). As it stands, this research can only indicate whether people care about, or pay attention to, one or more of these directional changes. Knowing that somebody is willing to pay some predicted amount for "an increase of unspecified size" is not very helpful compared to knowing that somebody is willing to pay some predicted amount "per species," or "per bird," or "per rare or unusual bird," or "for a 10% increase in the probability of seeing a wildlife spectacle."
Zambrano-Monserrate (2020) *Journal for Nature Conservation* (WTP to protect Andean Condor)	This is a contingent valuation study concerning willingness to pay to "prevent the extinction of the Andean condor in Ecuador." It is not clear whether the continuing existence of the Andean condor in other countries is considered a substitute by respondents. It is generally important to be explicit with respondents about the consequences of any given program for potential substitute environmental goods.
Czajkowski et al. (2021) *Land Use Policy* (willingness to adopt bird-friendly farming, Biebrza Marshes, Poland)	This study concerns farmers' willingness to accept compensation (a subsidy) to engage in more environmentally friendly agricultural practices. Their preamble to the choice experiments included information about the goals and environmental benefits of selected programs (including mention of wetland birds and endangered birds, generically, for different types of farmland). Their follow-up questions included quiz on their knowledge of local bird species, but the impacts of the various programs directly on birds were left implicit in the program descriptions in the choice experiments. This makes it hard to translate MWTA compensation for certain practices into WTP to protect birds (where, again, WTP to protect birds would presumably be embedded in a lower MWTA compensation for undertaking measures that protect birds (among other species).
Kim et al. (2021) *Land* (new airport, Korean island, national park along East Asian-Australasian Flyway)	Attribute levels in this study are characterized as high-medium-low without specific quantities, either as absolute amounts or relative to an unspecified (but knowable) status quo level. Thus the MWTP estimates for numbers of birds (negative, since the scenarios involve decreases from the status quo) are just one amount for "medium" and a larger amount for "high," where there seems to be nothing to reveal what each respondent assumes is meant by the medium and high designations for reductions in the number of birds. Each respondent is left to make their own assumption and the researcher does not know what this is. These results could be so much more generalizable if they had used a general population sample, rather than a sample of people who had revealed a greater interest in national

	parks by having visited at least one in the previous five years. With only 1000 completes after 12,839 contacts, it should have been a top priority to model people's decisions about whether or not to take the survey.
Sehra and MacMillan (2021) *Journal of Environmental Management* (wildlife-friendly rice farming in Japan)	The choice tasks posed by these authors include birds as an attribute, but the bird attribute consists only of a presence/absence indicator for any bird species being present in the rice paddy where the package of rice in question was grown. It is thus possible to infer a respondent's MWTP for a type of rice where there is at least one bird of any type in residence where the rice was grown. It is not clear whether this is at least one bird per paddy, on average, or an average of one resident bird per farm, or one bird visiting the farm per year. This is an any-versus-none distinction that does not lend itself to valuing the biodiversity or abundance of birds of different species. It is uncertain whether people's WTP extra for wildlife-friendly practices for other types of agricultural products would be similar, or whether this applies only to rice.
Dobson et al. (2022) *Oryx* (UK donors and flagship conservation area in Africa)	This study likewise includes the presence of at least one threatened bird species at the conservation area in question, or the absence of any threatened birds. There is a lack of specificity concerning whether the threatened bird is an occasional visitor or a permanent resident, how many individuals of that species might typically be present, and whether there is more than one threatened species in the area. Thus the choice experiment can differentiate between the amounts people would be willing to donate for an area with no threatened birds versus the amount they would donate with at least one threatened bird (or any species). The marginal value (measured as willingness-to-donate) of an additional individual bird cannot be addressed with these data.
Gatti et al. (2022) *Food Quality and Preference* (Bird-friendly certification for coffee)	Like the earlier studies of farmers' willingness to accept subsidies in exchange for bird-friendly land-use practices, this study concerns consumers' willingness to pay a premium for coffee for the knowledge that the coffee farmer in question qualifies for "Bird-Friendly" certification. This is a willingness to pay for an increased probability that bird biodiversity and abundance may be higher as a result of these measures. Respondents get no information about avian biodiversity measures and how they might differ from coffee plantations with this certification, or what these measures might do for the abundance of specific species of birds. The coffee in question could be grown anywhere, and the certification would make more or less of a difference in different regions. Thus it is hard to use the results from this study to infer a MWTP for specific improvements in avian biodiversity or abundance.

Sharma and Kreye (2022) *Ecological Economics* (bird conservation on private lands in Pennsylvania)	This research distinguishes only between common species of birds and rare species of birds. They focus on which type of bird would benefit from the policy. No information is provided concerning the *extent* to which either type of bird would benefit from the policy. But it is interesting that they stipulate whether a policy would have bird-associated human benefits via ecological services provided by these birds or simply due to human interactions directly with birds. But this structure of a choice experiment does not permit estimation of MWTP per bird or per bird species.
Stemmer et al. (2022) *Journal of Outdoor Recreation and Tourism* (birding destination preferences among visitors to an island destination in Norway)	The sampling strategy for this study consists of convenience samples among visitors to a prominent Norwegian birding site, so any quantitative estimates cannot necessarily be transferred to the whole country or to Europe more broadly (or other continents). Their site attributes include indicators for two qualitative levels of birding quality, and three intervals of species richness. The use of intervals makes it impossible to fit (or test whether there is) any smooth parametric curve to the relationship between utility and either of the bird-related attributes. Disaggregation of the "quality" categories into quantitative measures of numbers of birding spectacles, or indicators for the numbers of species that are "uncommon" or "unique" might have made the data suitable for estimating MWTP for these specific bird-related qualities of each destination.
Xu and He (2022) Journal of Environmental Management (tourists' MWTP for rare birds in Nansha Wetland, China)	These authors seek to estimate the recreational value of a specific coastal wetland park, so it is not clear whether their estimates are general enough to permit transfer of an estimated benefits function to other contexts involving coastal wetland parks. They describe the sociodemographic profile for their sample, but these individual characteristics are not used to estimate systematically varying preference parameters that might allow the model to be tailored to a different population of consumers. If there happened to be sufficient location information for where each respondent lived, it would be possible (as in Yao et al. (2019)) to see whether individual or neighborhood characteristics have systematic effects on preferences, and thus create an opportunity to simulate values of other coastal wetland parks of similar types to different populations of consumers.

REFERENCES

Bennett, M. T., Y. Z. Gong and R. Scarpa. 2018. "Hungry Birds and Angry Farmers: Using Choice Experiments to Assess "Eco-compensation" for Coastal Wetlands Protection in China." *Ecological Economics* 154: 71-87.

Boeri, M., T. A. Stojanovic, L. J. Wright, N. H. K. Burton, N. Hockley and R. B. Bradbury. 2020. "Public preferences for multiple dimensions of bird biodiversity at the coast: insights for the cultural ecosystem services framework." *Estuarine Coastal and Shelf Science* 235.

Cerda, C., J. P. Fuentes, C. L. De la Maza, C. Louit and A. Araos. 2018a. "Assessing visitors' preferences for ecosystem features in a desert biodiversity hotspot." *Environmental Conservation* 45(1): 75-82.

Cerda, C., J. P. Fuentes and G. Mancilla. 2018b. "Can conservation in protected areas and visitor preferences converge? An empirical study in Central Chile." *Biodiversity and Conservation* 27(6): 1431-1451.

Clucas, B., S. Rabotyagov and J. M. Marzluff. 2015. "How much is that birdie in my backyard? A cross-continental economic valuation of native urban songbirds." *Urban Ecosystems* 18(1): 251-266.

Czajkowski, M., K. Zagorska, N. Letki, P. Tryjanowski and A. Was. 2021. "Drivers of farmers' willingness to adopt extensive farming practices in a globally important bird area." *Land Use Policy* 107.

Dobson, F., I. Fraser and R. J. Smith. 2022. "Identifying the characteristics of conservation areas that appeal to potential flagship campaign donors." *Oryx* 56(4): 555-563.

Garnett, S. T., K. K. Zander, S. Hagerman, T. A. Satterfield and J. Meyerhoff. 2018. "Social preferences for adaptation measures to conserve Australian birds threatened by climate change." *Oryx* 52(2): 325-335.

Gatti, N., M. I. Gomez, R. E. Bennett, T. S. Sillett and J. Bowe. 2022. "Eco-labels matter: Coffee consumers value agrochemical-free attributes over biodiversity conservation." *Food Quality and Preference* 98.

Guimaraes, M. H., L. Madureira, L. C. Nunes, J. L. Santos, C. Sousa, T. Boski and T. Dentinho. 2014. "Using Choice Modeling to estimate the effects of environmental improvements on local development: When the purpose modifies the tool." *Ecological Economics* 108: 79-90.

Kim, M., S. Lee and C. O. Oh. 2021. "Assessing Tradeoffs between Development and Conservation: A Case of Land Use Change in a National Park of Korea." *Land* 10(2).

Krishna, V. V., K. Darras, I. Grass, Y. A. Mulyani, D. M. Prawiradilaga, T. Tscharntke and M. Qaim. 2019. "Wildlife trade and consumer preference for species rarity: an examination of caged-bird markets in Sumatra." *Environment and Development Economics* 24(4): 339-360.

Liu, T. M. and W. C. Yang. 2019. "Using choice experiments to inform management of black-faced spoonbill reserve in Taiwan." *Cogent Social Sciences* 5(1).

Mitchell-Nelson, J. (2022). Essays on Development and Health Economics. Ph.D., University of Oregon.

Rogers, A. A. and M. P. Burton. 2017. "Social preferences for the design of biodiversity offsets for shorebirds in Australia." *Conservation Biology* 31(4): 828-836.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr and P. P. Marra. 2019. "Decline of the North American avifauna." *Science* 366(6461): 120-+.

Sehra, K. K. and D. C. MacMillan. 2021. "Wildlife-friendly food requires a multi-stakeholder approach to deliver landscape-scale biodiversity conservation in the Satoyama landscape of Japan." *Journal of Environmental Management* 297.

Sharma, S. and M. M. Kreye. 2022. "Social value of bird conservation on private forest lands in Pennsylvania, USA." *Ecological Economics* 196.

Stemmer, K., O. Aas, K. Veisten and K. Lindberg. 2022. "Assessing recreation specialization to guide nature-based tourism development: A hybrid choice model of birder destination preferences." *Journal of Outdoor Recreation and Tourism-Research Planning and Management* 39.

Steven, R., J. C. R. Smart, C. Morrison and J. G. Castley. 2017. "Using a choice experiment and birder preferences to guide bird-conservation funding." *Conservation Biology* 31(4): 818-827.

Valasiuk, S., M. Giergiczny, T. Zylicz, A. Klimkowska and P. Angelstam. 2018. "Conservation of disappearing cultural landscape's biodiversity: are people in Belarus willing to pay for wet grassland restoration?" *Wetlands Ecology and Management* 26(5): 943-960.

Xu, S. J. and X. L. He. 2022. "Estimating the recreational value of a coastal wetland park: Application of the choice experiment method and travel cost interval analysis." *Journal of Environmental Management* 304.

Yao, R. T., R. Scarpa, D. R. Harrison and R. J. Burns. 2019. "Does the economic benefit of biodiversity enhancement exceed the cost of conservation in planted forests?" *Ecosystem Services* 38.

Yao, R. T., R. Scarpa, J. A. Turner, T. D. Barnard, J. M. Rose, J. H. N. Palma and D. R. Harrison. 2014. "Valuing biodiversity enhancement in New Zealand's planted forests: Socioeconomic and spatial determinants of willingness-to-pay." *Ecological Economics* 98: 90-101.

Zambrano-Monserrate, M. A. 2020. "The economic value of the Andean Condor: The national symbol of South America." *Journal for Nature Conservation* 54.