This study investigated discrimination of familiar and unfamiliar voices in non-domesticated cat species. I enjoyed reading this manuscript, feel that it is generally well written, and that it provides an interesting insight into the effects of domestication, versus individual experience on animal social cognition of human cues. That being said, I have a number of points I would like addressed before I would consider this manuscript publication ready.

ABSTRACT

MINOR:

Lines 32-36: It initially says: "We tested whether 25 members of 10 non-domestic Felidae species recognized familiar human voices," before later saying: "we presented 24 cats of 10 species with unfamiliar and then familiar voice playbacks using a dishabituation paradigm." These statements appear to be contradictory and also a bit repetitive. I would recommend cutting one of these lines, while also providing the correct sample size. Also, later in the text you refer to the paradigm as a habituation-dishabituation paradigm, so I would advise amending the text here for consistency.

Line 40: Can you provide some examples, perhaps in brackets to give readers an idea of what a more intense response constitutes?

INTRODUCTION

I generally feel the introduction has been well-written and does a good job of introducing the background to the study. I do however, think it could be a little bit more concise and have a few minor points I would like addressed in the final manuscript.

MINOR:

Lines 55-56: Can you explain what you mean by "highly similar neighbours' vocalisations"?

Lines 61-62: Should this be amended to: "With humans becoming a more 'commonly encountered' heterospecific for many animal species" or similar?

Lines 64 + 110 & Elsewhere: I would amend "non-domestic" cats to "non-domesticated" cats.

Line 86: By preserves, do you mean reserves, like nature reserves?

Lines 87-90: Please provide the Latin name (as well as the common name) for all the species you studied, unless they are mentioned above. Provide the short-hand Latin name for the species of clouded leopard you studied (*N. nebulosa* or *N. diardi*), unless you studied them both, in which case state as such.

Line 112-120: To give an example of how you can streamline your writing, currently you give many different examples over many lines of how hand-rearing impacts the breeding success and behaviour in general of big cats. This information is less relevant to the current study which is more interested in the effect of hand-rearing on responses to human cues. Therefore, less relevant information like this should be summarised for conciseness (or even excluded), for example, "hand-rearing has been shown to affect species-normative behaviours and the reproductive success of big cats." Similarly, later on in the paragraph you state: "For instance, in cases of maternal neglect, wild orphans, infanticide, or large litters that are physically taxing on the mother, animal care staff will rear the cats

to increase their chances of survival" (**Lines 126-128**). This level of detail although interesting, is unnecessary, and I would perhaps suggest cutting this line and at the end of the previous sentence write something along the lines of:**,** "there are some circumstances that require human intervention (e.g., to increase chances of neonate survival)" (**Line 126**).

Line 114: Give each species common name unless this has not previously been introduced in the text above, in which case, please provide the Latin name as well.

Line 132: Should "Saito and colleagues (2019)" be Saito et al. (2019)?

Lines 176-178: This is also a common theme in your results section, but I think it would enhance your writings readability if you make the cat the subject of the sentence rather than the response measured. For example, instead of "We hypothesized that familiarity would lead to greater attention as measured by faster latencies to respond and greater intensity and duration of responses following the familiar voice relative to the unfamiliar voices" something more along the lines of: "We hypothesised that cats would respond more quickly, for longer and with greater intensity following a familiar, compared to an unfamiliar human voice."

Lines 178-180: This brings me on to my second point. "Measures of attentiveness and responsiveness included behaviours such as head, ear, and body orientation, movement towards or away from the sound, and vocalizations." These are measures described later and mentioned in the above sentence as response "intensity" so to avoid confusion, I would suggest amending this sentence to something like: "To measure the intensity of a cats response…"

METHODS

MAJOR

Lines 272-277: Although this is briefly mentioned in the introduction section (**Lines 168-170;** and then never again) I think it is important to explain the specific predictions of the habituation-dishabituation playback paradigm used for your main experiment in the Methods section, i.e., when an animal is repeatedly presented with stimuli sharing similar properties they are expected to habituate, but when presented with a stimulus differing to that previously presented, given they can perceive a difference, they are expected to dishabituate and renew their responsiveness accordingly. I would suggest going back to Saito & Shinozuka (2013) referenced in the text, or reading the methods section of Baciadonna et al. (2019[1]; and references therein) who explains the predictions of this paradigm in detail. Indeed, the habituation-dishabituation paradigm used and your findings in relation to its predicted outcomes is not really discussed throughout your manuscript. Looking at your results, there does not seem to be much evidence of one of its key predictions, habituation (a decline in response over the habituation phase) and you do not seem to employ analyses to measure this (looking at how responses changed over the habituation phase). I'm not saying the authors should go into detail, but it worth discussing the findings in relation to the specific paradigm used in the discussion section and potentially in your results and data analysis sections. Indeed, in the Results section especially you use simple contrasts, pooling cat responses to unfamiliar voices, and comparing these against their responses to familiar ones. To justify this approach, I would assume it would be first necessary to compare cat responses over the habituation phase. Without first showing

[1] Baciadonna, L., Briefer, E. F., Favaro, L., & McElligott, A. G. (2019). Goats distinguish between positive and negative emotion-linked vocalisations. *Frontiers in Zoology*, *16*, 1-11.

cat responses to unfamiliar playbacks were similar (and more specifically that they did not habituate – which I do not think they did), I would have more reservations about this approach.

Line 293-306: Measures of intensity would be a qualitative measure, and like many articles which use qualitative behaviour analysis, I would expect a bit more detail as to how coders scored this behaviour, preferably in a table. I know that a reference is provided (Stanton, Sullivan, & Fazio, 2015), but would say that readers shouldn't have to read a different reference to know precisely how behaviour was scored.

MINOR

I just wanted to ask, were all animals tested individually or were some tested together in a group setting? Lions are generally kept as groups, and tigers sometimes are at least. I may have missed this, but think this should be made clearer as it could affect the level of independence between measurements taken from different individuals, although I appreciate sample size concerns and time constraints may have made this necessary. Ultimately, although these would be entirely understandable constraints, it is important to at least acknowledge these things if they apply.

For the cats which were tested in both the pilot and main study can you explain what was the approximate time between these two studies? It would be a good idea to just mention this as there could be potential carry-over effects between them.

Line 190-191: I would suggest including this as a separate section called "Ethics Statement" or similar to better highlight that the ethical implications of your research were considered, and that you gained the necessary approval to undertake this research rather that have it as a free-floating statement.

Line 193-194: Please give the species common names so readers can know which species you are referring to without having to remember the Latin name.

Line 208 + Line 303: Please provide a relevant reference for the BORIS software at least at its first mention.

Lines 211-227: You mention that "The UF speakers were four different sex-matched people that the cat had never encountered or heard before", but for clarity can you also state whether the less familiar and most familiar person in each playback were also gender-matched?

Line 235: I think the authors meant e.g., rather than i.e.,

Line 244-245: Which version of Audacity?

Lines 262-277: I would recommend producing a schematic illustrating the playback procedure used in both the pilot study and the main study, because for the pilot study in particular I find it quite difficult to understand what was done. Alternatively, please revise the text to make the explanation a bit clearer.

Line 286: By the end of the recording, do you mean the end of the playback series?

RESULTS

MAJOR

I may have missed this, but one important thing that needs to be mentioned when explaining the data analysis, is the software package used (i.e., R, SPSS or SAS) along with the version used and a citation where relevant.

Throughout the results section you provide the F-value, the degrees of freedom, a *p*-value and a pi-squared value (π^2). Instead of π^2 ([3.1415…]^2), are you trying to provide an eta squared value (a measure of effect size), or η^2? Please change the next version of your manuscript to provide the correct notation of the measure of effect size used (including in tables).

Lines 316-318: I appreciate the *p*-value threshold of less than 0.05 is very arbitrary and that a *p*-value of 0.052 is so frustratingly close to significance. However, it is not, and the wording of these lines need to be amended to reflect this. This was previously mentioned by Reviewer 1, so I just think the authors just missed this line.

MINOR

It is more conventional to explain how the data was analysed in the methods section under a subheading such as data or statistical analysis rather than in the results section. I would prefer for the authors to do likewise but am happy to concede this point if the author's have a specific reason to use their current layout. Moving details of the statistical analysis to the Methods section may also help make the cats the subject of each sentence, rather than the statistical analysis which will enhance readability (like you have done in e.g., **Lines 331-332: "**Cats responded significantly more quickly to familiar versus unfamiliar voices").

Throughout the results section you refer to mean as *M* (I think), but this is not a common abbreviation. I would prefer you either use the word mean, or a more commonly used symbol (e.g., $\bar{x}$) or explain this abbreviation on its first use. It would also be better to provide units, for example mean latency or mean duration of response in seconds (s).

This is just a preference, and may help with conciseness, but throughout the results sections you mention "statistically significant effects" which could just be referred to as significant effects.

Line 308: It would be better to provide the link to your data set in a data availability statement below the full text body, i.e., along with the Acknowledgements etc. section.

Lines 316-318: You mention this below for your main study, but did you need to perform any corrections for sphericity in the pilot study (e.g., Greenhouse–Geisser, Huynh–Feldt)?

Line 388-389: "the file was split by subfamily and the analysis was re-run without including subfamily as a factor." Do you mean splitting the data set?

DISCUSSION

The discussion section is well-written and relevant, but I feel it can be a bit more concise, especially in the study Limitations and Future Directions section which is a bit too long.

MAJOR

Line 417-419: "Although it is possible that they responded to specific phrasing rather than recognition of the speaker per se, the results nonetheless suggest that the cats recognize when a voice is familiar." This sentence begins to touch on one important constraint for the current research and one that was addressed, by the previous reviewers, however, I feel it would be useful to go into this in a bit more detail (perhaps in the Limitations and future directions section). The current research used a phrase that should have been familiar to cats which means we currently cannot establish whether the cats have a general recognition of human voices, or whether this recognition was limited to specific words or phrases. For further information regarding this constraint, please consult Kriengwatana et al. (2015)[2].

Lines 435-436: "Similar research with domesticated species has suggested that this ability may be a consequence of long-term domestication,…" You cannot say what similar research has suggested without providing a few example references supporting this statement.

MINOR

Throughout the discussion you discuss things as being significant, which would be more expected in the results section. I suggest removing at least some of the significantlys from the discussion section for conciseness.

Line 392-393: I feel rather than starting the discussion section by discussing the sample size of your pilot study, it would be better to give a very brief summary of the methods you used or your experimental aims to remind readers of what was done.

Line 408-409: "Future studies will need to explore whether Felinae are generally more reactive than Pantherinae, which are typically larger." You discuss the need for future studies to follow up this finding later on in the Limitations and Future Directions section, so it is repetitive to discuss this here.

Line 413-415: ". Most importantly, name did not interact with familiarity such that the cats demonstrated differential responding to the familiar voice regardless of whether their name was spoken" – this sentence should be amended to make it easier for readers to understand.

Lines 423-424: It seems out of place to mention it's the first study to investigate the effect of early socialisation and sex on voice responsiveness in exotic cats and then go on immediately to state that you found no effect of these variables. I would suggest deleting everything in this sentence following: "This is the first time such an ability has been demonstrated in nine additional exotic cat species (Table 1)."

Lines 459-474: You discuss the effect of name usage earlier on (**Lines 410-419**). For flow, I would suggest perhaps putting these paragraphs together just before the Limitations and Future Directions Section.

Line 507: Maybe put in brackets what cross-modal matching is – as readers may be unfamiliar.

[2] Kriengwatana, B., Escudero, P., & Ten Cate, C. (2015). Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Frontiers in Psychology*, *5*, 1543.

OTHER COMMENTS

MAJOR

In your graphs (Figures 2-5) you do not provide the information needed to let readers know what they are showing, i.e., means and standard errors etc. In some cases, you also do not give the units these responses were measured in (Figure 2). Please change your figures and legends accordingly. Remember also, you should be able to understand figures somewhat independently of the text and other figures, and although you state in the legends of Figures 2 and 3 that they are showing the results of the pilot study (in the text Study One) and the main study respectively, this is not the case for Figures 4 and 5.

MINOR

You should try to be a bit more consistent in style between figures, for example, in some figures the y-label is in capitals whereas sometimes it is in lower case. You can also consider positioning Figures 3-5 side-by-side and combining these as three parts of a single figures (e.g., Figure a) b) and c)) to reduce the number of separate figures, but this latter point would not be a mandatory correction, more of a preference.

Table 1. Format this table in line with Table 2, both for consistency and because it makes it clearer which are the table headings when they are separated from the body of the table with a divider. Also, please make sure words are not carried over two lines (i.e., table heading, Sex).

Table 2. The first rows look a bit squashed together (because of the single line spacing) and in the later rows, variables carry over multiple rows. I would suggest using 1.15 line spacing or greater and to either ensure variable headings are all on a single line or put small additional spaces between them to make the boundaries between the different variable headings clearer.

Figure 1: Please can you use the metric system in your experimental diagram? It is more scientific than imperial and can be understood by an international audience.

Figure 2: Following reviewer suggestions you refer to your first study as a pilot study. Amend the text legend for Figure 2 to reflect this.