**We would like to thank both reviewers for the comments and suggestions. The reviews greatly improved the manuscript. Please note that we also changed the minor comments suggested by Reviewer C within the manuscript.**

Manuscript #5051

**Reviewer B:**

Switzer and Combes present a study on the effects of marking bees with
numbered bee tags with super glue gel or paint on foraging behavior.  Their
study provides much needed insight on how methods in which we study bumble
bee behavior (e.g., marking with numbered tags) may facilitate behaviors
that could affect a researcher’s end result. That is, how we chose to
label bees for unique identification may prove to be a confounding variable
in our analyses, and most importantly, affect the interpretation of the
results. In their study they found no significant difference in wing beat
frequency, sonication frequency, or sonication length in bees that were
either paint marked or had numbered tags glued to their mesosoma. However,
they did find that the probability of collecting pollen via sonication was
much lower in bees when paint marked versus having a numbered tag glued to
their mesosma.

I do find the research idea and goal of this study to be (1) of  high value
to those studying bumble bee foraging behavior and (2) to fit the scope of
JoM. However, upon review of the methods and results, I do not recommend
publication of this study in its current state. However, if the authors take
the time to consider my suggestions and concerns, I recommend the study to
be resubmitted to JoM. I have several major concerns in regards to how
methods are reported and chosen, and the interpretation and reporting of the
results. My biggest concerns include the following:

(1) The authors never reported how many bees were actually used in their
study or how many bees were used in each treatment (paint vs. numbered
tags). This is critical.

**Response: We have updated the manuscript to reflect the sample sizes for each of the analyses.**

(2) There may be operator error in administering glue/paint to the mesosoma
of smaller bees. Specifically, the authors found that smaller “average”
sized bees (bees with smaller IT spans) were less likely to sonicate than
larger bees (pg. 6). From their study “A one-mm increase in IT span in
bees that were marked with bee tags is associated with a hazard ratio of
4.79, which means tagged bees whose IT span was one mm higher than the
average size were about 5 times more likely to engage in further sonication
behavior than bees of average IT span (β(IT Span) = 1.57; z = 4.19, p <<
0.05).” The authors need to confirm that glueing bee wings together did
not occur in their study. It seems that smaller bees may be more vulnerable
to having their wings damaged on accident by the operator than larger bees,
which could have affected their results. Furthermore, the authors lack any
sort of control (unmarked bees)  in their study, which I think is
critically important considering their goals.

**Response: We agree with this statement, and we address it in the manuscript – it is possible that gluing bee tags to smaller bees is more problematic that gluing bee tags to large bees, because it is difficult to administer glue without getting any onto the wings. We have added some clarification to this point.**

**In addition, we wrote more clarifying details in the methods. We wrote that we made sure that bees could flap their wings as we released them back into the cage – we let them fly out of the queen-marking cage), so we know that at the time of release, their wings were able to function properly.**

**Though paint and glue could both potentially run into the wing bases after during marking, the problems caused by glue in the wing bases may be much worse than paint getting in the wing bases.**

**Regarding having a control group – we don’t know of any way to identify individual bees throughout a multi-day experiment without either marking them or removing them from the colony. Instead, we treat the painted bees as the reference group. In the multivariate multiple regression we make all comparisons, by treating the bee tag group as the treatment and painted bees as reference. We changed the reference level for the Cox regression to be the painted group, so all comparisons are relative to the painted group.**

 (3) The use of the phrase “nearly significant” for a P value of 0.08 and
“not significant” for a P value of 0.056 is odd and misleading (Pg. 6).
The authors should reconsider how they interpret significant results in
their study.

**Response: We have clarified the language regarding significant p-values.**

(4) There is no description or report of how old the bees are in their
study. On Pg. 7, they report the number of bees that were alive, dead, and
missing from each of the two treatments. They perform a chi-square test to
determine whether marking/painting bees had an effect on mortality. I think
this approach is flawed as some bees may have died of natural causes (old
age). However, because the authors do not report how old the bees are in
their study, there is no way of knowing whether painting/marking actually
had an effect on their mortality. In a next experiment the authors can
document the emergence of the bee and see if marked bees have shorter lives
than bees that are not marked. However, without knowing old each bee is,
there is no way, I think, that one could objectively measure the effect of
painting/marking bees on mortality.

**Response: We clarify that we are not trying to compare marked vs unmarked bees, but we instead report tagged vs. painted bees. By effectively randomly assigning the treatment (i.e. alternating marking foragers with either paint or bee tags), we expect that both groups have equivalent distributions of ages.**

**The major source of bias (which we acknowledge) is that we quantified mortality for only bees that did not sonicate after being marked. The sample sizes for painted vs tagged bees in this mortality analysis were highly skewed (22 painted vs. 72 tagged). We removed the ChiSq test for equality of proportion from the results, since it may be inappropriate in this case (due to a biased sample). We clarify that our results do not link mortality with marking treatments, and that this was not one of the main objectives of the study.**

**We changed the section titled, “Are tagged bees more likely to die?” to “Bee mortality”, because we felt that the former title was misleading.**

(5) General inconsistency in reporting significant digits and P-values, see
comments throughout. Also, there are statements in the manuscript like “a
reasonable range for sonication frequency”, which is awkward and not
followed by a citation. Why is this reasonable? There is another phrase like
“noticeably different”, which is awkward and is not fully discussed or
defended.

**Response: We changed all p-values to be accurate to the 0.001 place. For very low p-values we used “p-values < 0.001” instead of using scientific notation. We added citations for reasonable frequency ranges. We elaborated on how we checked the sounds for accuracy and removed the wording, “noticeably different”**

Again, while I do think this a fantastic idea, and should be pursued, I
think the authors need to re-evaluate and clarify their methods and
interpretation of the results. I do recommend that the authors continue
their study, and I look forward to seeing it in the future.

**Reviewer C:**

#5051 Review

*Bombus impatiens* (Hymenoptera: Apidae) reduce pollen foraging behavior when marked with bee tags

**Summary & Recommendation**:

This paper compares changes in sonication mechanics and behavior in paint-marked and tagged bees, and provides a novel and important evaluation of these commonly used techniques. The study is sound, and I recommend that it be accepted for publication. However, there are a few minor adjustments to the manuscript that should be applied to improve clarity and address what I see as a problematic inference from the results of the wing-beat frequency comparison. These comments are outlined below.

**Minor Comments:**

* Line 124: Add in the total numbers of bees marked with tags and paint here. The basic information on sample sizes that is included in the results section would be easier to digest if it was presented sooner.
	+ **Response: We put sample sizes in the materials and methods and reported sample sizes used for each of the analyses (since each analysis had a different sample size – due to i.e. inability to find bees at the end of the experiment)**
* Line 140: Add in the total number of marked-individual observations here.
	+ **Response: We added the sample sizes into various parts of the manuscript**
* Line 145: Add in the number of bees retained for analysis here.
	+ **Response: We added the number of bees excluded from analysis here and put the number retained for analysis later in the manuscript.**
* Line 163: Add a reference for the peak sonication frequency range considered to be reasonable, and note that the range is restricted to bumble bees (*Xylocopa* can sonicate at lower frequencies). Table 1 in the following reference would be an appropriate choice: De Luca, P.A., & M. Vallejo-Marín. 2013. What's the ‘buzz’about? The ecology and evolutionary significance of buzz-pollination. Current Opinion in Plant Biology 16(4): 429-435.
	+ **Response: We added this reference and some additional citations.**
* Lines 209-210: strike “with cancer”. A specific reason that “patients…die” in other studies using this analytical technique is a bit distracting and out of place here.
	+ **Response: We removed the part of the sentence about patients with cancer dying.**
* Line 233: “figures” There is only a single figure to which this statement would apply.
	+ **Response: We rewrote the sentence to say that we used the R packages to make figure 2.**
* Line 248: Reword “nearly significant”. It seems a stretch to emphasize a p-value of 0.08 as close to significant when you’ve already told us that you didn’t correct for multiple comparisons and your threshold value is likely much lower than 0.05 for this test.
	+ **Response: We reworded the sentence and a few sentences before to make clear that these results provide hypotheses that may warrant future investigation.**
* Lines 262-263: Give us some warning that IT stands for intertegular, (*i.e.,* “intertegular (IT) span”).
	+ **Response: We wrote “intertegular (IT)” to clarify what “IT” means.**
* Line 294 “relatively low p-value of 0.08 for the change in wing beat frequency” Emphasize the values of the data rather than the p-value. After all, it is the trend that you think you observe that you wish to emphasize, not the failure of the data to pass your statistical test. Looking at the standard deviations in Table 1, I’m not sure that I buy that there is a difference, but I’d be willing to accept that you could test for one in follow-up studies if the observed deviation is due to experimental procedure, rather than actual biological variability. In the results, you mention that there is a small change in the average wing beat frequency with tagged bees, compared to painted bees. Mention this and mention that this difference was not statistically significant, but that you think there is enough of a suggestion in the data that follow-up studies may be needed to determine if there is a real effect.
	+ **Response: We changed the wording to say that our results suggest a hypothesis for future studies.**
* Lines 309-311 move the year of citation to just after the author: Hagen *et al.* (2011).
	+ **Response: We changed citation to this: Hagen *et al.* (2011)**
* Line 317-336: One potential way to test for the effects of glue alone would be to add the drop of glue, but not the tag.
	+ **Response: We added a sentence saying: “we may be able to test the effect of superglue alone by marking bees with superglue (perhaps colored) without a bee tag to determine if the glue or the tag is more problematic.”**
* Line 340: The data support using either tags or paint to mark bees, based on the lack of changes in mechanical differences.
	+ **Response: We added “or paint” to clarify that either could be used.**
* Lines 337-348: You are leaving out an important difference that you observed: bees with tags sonicate less often than those with paint! It is the most interesting finding that you have, it does have implications for experiments using marked bees and it deserves to be mentioned here. Also, there isn’t really anything in this study to suggest that tagged bees ever recover the ability to sonicate after being marked. Thus, that last sentence seems unsupported by the included data.
	+ **Response: We did see evidence that some tagged bees recovered – we show that on the survival curves (fig. 2). Most of them were observed sonicating a few days after being marked. Our last sentence is basically meant to warn people to not tag bees and perform experiments immediately after tagging. We suggest waiting to make sure that the tagged individuals will still perform the behavior after being tagged.**