Public Opinion of Captive Cetacean Attractions: A Critique of Wassermann et al. (2018)

Heather M. Manitzas Hill[1] & Kelly Jaakkola[2]

[1]St. Mary's University, Psychology Department, San Antonio, Texas

[2]Dolphin Research Center, Grassy Keys, Florida

Author note. Please address all correspondence to Dr. Heather Hill. Email: hhill1@stmarytx.edu

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27852v1 | CC BY 4.0 Open Access | rec: 11 Jul 2019, publ: 11 Jul 2019

Abstract

Wassermann et al. (2018) argued that previous public opinion research about marine mammal attractions should be considered unreliable due to possible biases in study design, which may have influenced participants' responses. As in all scientific endeavors, reducing bias in order to gather more objective, evidence-based information is a worthy and commendable goal. Unfortunately, Wassermann et al. fell short in their efforts to produce an unbiased investigation into the beliefs of the general public about captive marine mammal attractions, due to a number of methodological flaws and biases in their own study. Specific concerns include a non-representative sample, methodological issues with data collection and coding procedures, a lack of reliability between data published and data provided, a failure to demonstrate inter-coder reliability, a failure to control for sequence effects in quantitative data, misrepresentation of data between text and tables, and biased over-interpretation of qualitative responses. These errors undermine the authors' conclusions and indeed render their findings uninterpretable. To achieve the goal of an unbiased understanding of public opinion about marine mammal attractions, further research on this topic is warranted using rigorous and sound scientific methodology.

In the article "Reassessing public opinion of captive cetacean attractions with a photo elicitation survey," Wassermann et al. (2018) argued that previous public opinion research about marine mammal attractions should be considered unreliable due to possible biases in study design, which may have influenced participants' responses. They suggested that such biases can be avoided by utilizing a photo elicitation technique in which respondents are asked for their open-ended opinion rather than being asked to choose among a selection of researcher-designed choices to questions that might have biased wording. To demonstrate this approach, the researchers conducted a study in which they showed tourists on the Turks and Caicos Islands images of six possible future attractions, including a marine mammal park (MMP) killer whale show and a swim-with-the-dolphins (SWTD) attraction. The research protocol entailed first asking each tourist the open-ended question, "What are your opinions on any of these six attractions being introduced in the Turks and Caicos Islands?" followed by a second series of questions in which responding tourists indicated their likelihood of visiting each attraction on a 4-pt scale ranging from "very unlikely" to "very likely."

As in all scientific endeavors, reducing bias in order to gather more objective, evidence-based information is a worthy and commendable goal. We also commend the authors for explicitly stating their own "anti-captivity" bias (p. 6). To be fair, we should note that each of us conducts research at marine mammal facilities. As such, it is a legitimate question whether our critique might simply reflect the opposite bias from Wassermann et al. By itself, having an opinion on a research topic (in either direction) should not invalidate any researcher's scientific contribution. In fact, it seems noncontroversial to suggest that most scientists who publish research likely have opinions about their topic. However, opinions are not the same as scientific claims. To be credible, any scientific claim must be derived from rigorous and sound scientific

methodology, including valid and reliable data collection, analysis, reporting, and argumentation. Unfortunately, in the case of the Wassermann et al. (2018) study, although some of their techniques (e.g., hiding the true purpose of the study by asking about both cetacean and non-cetacean attractions) undoubtedly supported their stated goal of obtaining a non-biased measure of public opinion on this topic, there are also quite a few errors in methodology and reporting that both bias their results and conclusions in other ways, and in some cases even make it impossible to ascertain the actual results of their survey. Below, we highlight several major methodological issues and potential misrepresentation of the results from the study by Wassermann et al. (2018).

Methodological Problems

Biases in Data Collection and Coding

Wassermann et al. (2018) argued that asking respondents an open-ended question would "minimise the chance introduction of various researcher biases during survey design" (p. 7). At face value, this claim is certainly reasonable, as considerable research has shown that people's responses are primed or affected by associations triggered by question wording, question order, format of response options, alternatives given, order of the alternatives, etc (Schwarz & Hippler, 1991; Strack 1992; Schwarz, 1996; Sudman, Bradburn, & Schwarz, 1996). However, open-ended questions by themselves are not a panacea for bias. Instead of possible bias introduced in the wording and response alternatives of closed-ended questions, the locus for bias in open-ended questions is transferred to response interpretation and coding (Burns, 1989; Campbell, Quincy, Osserman, & Pederson, 2013; Passer, 2017). In the case of Wassermann et al. (2018), such bias may have been introduced on two levels.

First, rather than using a recording device to produce an accurate transcript of respondents' comments, surveyors simply "took notes" on these comments. The problem is that such note-taking is subject to recording biases and interpretation modifications, which can lead to inaccurate recording of an individual's actual response that then affects the coding of the theme inherent in the response (Burns, 1989; Hammersley, 2010; Campbell et al., 2013). As the authors of the study have an admittedly clear anti-captivity bias, it is not unlikely that the surveyors may have had an unconscious tendency to interpret and summarize these responses in ways that aligned with their own beliefs. Unfortunately, the surveyors did not audiotape record the participant responses for later verbatim transcriptions (Burns, 1989).

Second, these summarized notes on respondents' comments were then coded as expressing either positive or negative views of each type of marine mammal attraction, and also for any expressed reasons for these opinions, such as *appropriate for children* (positive) or *animal welfare concerns* (negative). Wassermann et al. (2018) neither provided details about their criteria for classifying these opinions (i.e., operational definitions) nor any measure of inter-coder reliability. The lack of these details is extremely problematic, as studies have shown that the high degree of inference needed to categorize open-ended responses can lead to a high probability of initial error and bias in interpretation, often resulting in low levels of agreement during initial coding (e.g., Burns, 1989; Carey et al., 1996; Hagelin, 1999; Hruschka et al., 2004; Passer, 2017). Multiple studies and methodological textbooks emphasize the necessity of establishing strong inter-coder agreement, which often includes an iterative revision process for the coding criteria, to ensure that the final coding results are valid and credible, and cannot be explained as the idiosyncratic result of the coder's beliefs, biases, or imagination (e.g., Burns, 1989; Gorden, 1992; Carey et al., 1996; Hruschka et al., 2004; Passer, 2017).

Clear examples of bias in Wassermann et al.'s coding can be seen by consulting their raw data, provided in Supplementary Materials (https://doi.org/10.7717/peerj.5953/supp-2). For example, the authors noted in the paper that "Only five respondents who were 'likely' or 'very likely' to visit an MMP killer whale show gave qualitative feedback, all stating 'entertainment' as their reason for wanting to visit" (p. 11). It is not obvious, however, how this *entertainment* code matches the written notes on these five responses, reproduced here:

(1) visited in FL keys--fun

(2) would be mean

(3) interested, similar response to dolphin experience (excited)

(4) thought a lot of people would visit

(5) go to seaworld if she wanted

Similarly, their code for *animal welfare concerns* was applied not only to cases in which respondents referred to animals' treatment or welfare (e.g., "bad for animals") but also if they simply expressed that they were against the idea of captivity without giving any further rationale (e.g., "not into captivity") or that they would prefer to see animals in the wild (e.g., "see in ocean over captivity").

All of these examples represent clear overinterpretations of the respondents' qualitative statements, in which the underlying intention of a statement is inferred beyond any information given. For example, in the five responses that were coded as entertainment, only the first statement satisfies that theme clearly, and the third statement arguably so. The three remaining statements have nothing to do with the entertainment theme. Clear operational definitions and inter-rater reliability would have addressed these issues (Burns, 1989; Campbell et al., 2013).

Unreliable Data Recording and Analysis

Data recording and tabulation. The raw data files provided in Wassermann et al.'s supplementary materials include two spreadsheets. The first sheet lists the tourists' responses to demographic questions, their ratings for how likely they would be to visit each hypothetical attraction, and a binary code for whether they provided a free response about the MMP and SWTD attractions. The second sheet lists the surveyors' notes for each free response about the MMP and SWTD attractions, the codes assigned to each response, and the rating for how likely the person said they were to visit that attraction. Both sheets also include the respondent's survey number, which should allow for the responses from the two sheets to be matched. Unfortunately, comparing the responses from Sheet 2 with the binary codes for whether that person gave a free response from Sheet 1 shows that these two sources of information did not match 11% of the time. Specifically, both sheets agreed in 262 of 292 cases for the MMP, and in 259 of 292 cases for SWTD, regarding whether or not that person gave a free response. To complicate matters further, neither set of data matches the numbers that Wassermann et al. presented in their text (pp. 11-12 and Table 4) regarding how people who gave free responses rated their likelihood of visiting the attractions (Table 1). Moreover, in addition to nonmatching numbers of responses, there are also large discrepancies between the codes for these responses in the raw data versus their tabulations as presented in Table 4 of Wassermann et al.'s article (Table 2).

[Insert Table 1 about here]

[Insert Table 2 about here]

Inaccurate data presentation. Furthermore, there are also cases in which the data presented in different parts of Wassermann et al.'s paper do not match each other. For example, the section of Table 5 dealing with Accommodation Type is supposed to portray the proportion

of tourists who reported being likely or unlikely to visit each cetacean attraction for: (a) those staying at all inclusive resorts, and (b) everyone else. An examination of the data portrayed, however, shows that for both attractions, Wassermann et al. apparently switched the proportions for those likely to visit with the proportions for those unlikely to visit. For SWTD, this is obvious, as both demographic groups are listed as more unlikely to visit, which is impossible if 60.3% of respondents overall responded that they were likely (or very likely) to visit as reported in Table 3, Figure 2, and the text. For MMP this reversal of proportions is less obvious at a glance, however multiplying the number of people at inclusive versus other types of accommodation as reported in Table 2 by the proportions presented in Table 5 shows that the totals are also opposite what they should be. This reversal of proportions means that Wassermann et al.'s purported finding that "Tourists staying in all-inclusive resorts were significantly more 'unlikely' to visit a potential SWTD attraction than respondents in other accommodations ($p < 0.001$)" should instead read that they were significantly more *likely*. Similarly, the claim that "Those staying in all-inclusive resorts were also significantly less interested in visiting MMP killer whale shows ($p < 0.001$)" should instead read that they were significantly *more* interested. It is unfortunate that this mistake was not noticed before publication as the inaccurate representation is being perpetuated with each subsequent citation.

Potential sequence effects. Another potential concern relates to the presentation protocol of the image grid used to depict the various attractions. Wassermann et al. (2018) report that the images were arranged in the same order for all participants. Given that they do not report that the order of questioning was randomized in any way, one can only assume that a typical presentation order (i.e., left to right top row, followed by left to right bottom row of the grid) was utilized. This interpretation is bolstered by the fact that the order of image labeling (A – F) on the

grid in their supplementary materials, and the attraction numbers (1 – 6) in their supplementary data files correspond to this presentation order. The problem is that with a fixed presentation order such as this, earlier placed images such as the aquarium (#1) and botanical garden (#2) may have received more conscious responses than later placed images such as the killer whale show (#5) and maritime museum (#6), as respondents tend to give more automatic responses during repetitive questioning (Tourangeau & Rasinski, 1988; Schwarz & Hippler, 1991; Strack 1992; Schwarz, 1996; Sudman, Bradburn, & Schwarz, 1996). In addition, because the list of forced choice options always began with the most negative anchor response (i.e., "very unlikely"), respondents may have experienced a priming effect toward negative options. If this were the case, it would result in a growing bias for more negative responses towards the end of questioning as responding became more automatic. Research on survey development and cognitive biases encourages researchers to control for these issues by randomizing questions and reversing answer options (Tourangeau & Rasinski, 1988; Schwarz & Hippler, 1991; Strack 1992; Schwarz, 1996; Sudman, Bradburn, & Schwarz, 1996).

A visual inspection of Wassermann et al.'s Figure 2 suggests that such a sequence effect occurred. (As one moves across the stacked bars, the blue bars depicting negative responses increase in percentage over the six images). To examine this possibility statistically, we ran correlation analyses on the percentage of negative responses reported in Wassermann et al.'s Figure 2/Table 3 with the presentation order of the images. The results showed significant correlations between negative responses and image order, whether examined for only the most negative response category (i.e., very unlikely to visit), $r(4) = .92$, $p = .01$, or for both negative responses combined (i.e., unlikely or very unlikely to visit), $r(4) = .90$, $p = .015$, Thus, more respondents were likely to endorse images at the end more negatively. This sequence effect was

also demonstrated by a significant correlation between the scores each respondent gave for each attraction (from Wassermann et al.'s supplementary materials) and image order, $r(1708) = -.21$, $p < .001$. Respondents rated images presented later as less likely to visit than images that were presented earlier in the grid.

Lack of Sample Representativeness

Wassermann et al. argue that "allowing open-ended responses… seems to have given a more representative understanding of what is at the forefront of the public's mind than closed questioning." (p. 1). The truth of this claim relies on two underlying assumptions: (1) Whether open-ended questions are better at accessing the "true worldview" (p. 7) of respondents than are closed-ended questions; and (2) whether respondents who chose to provide open-ended responses to particular questions have worldviews that are representative of the rest of the public. For purposes of this commentary, we do not take any particular stand on the first assumption. It certainly seems reasonable that – *assuming* valid and reliable data collection and coding procedures – open-ended questions likely provide a legitimate methodology for discovering a person's beliefs.

The second assumption – that the people who chose to provide open-ended responses in this study are representative of the general public -- is much more problematic. To explain this second assumption further, we break it down into two components. First, there is the issue of whether Wassermann et al.'s sample can reasonably be described as representative of "the public." We would argue that it cannot, given that the respondents in this study were comprised specifically of North American beach-going adults (97% of sample) with the financial means to take a vacation to an exotic destination. The second problem is that the people who provided

open-ended responses in this study were not representative of the rest of the people in the study. Rather, for both the MMP and SWTD attractions, those respondents who were unlikely (or very unlikely) to visit the attraction answered the open-ended question significantly more often than those who were likely (or very likely) to visit (χ^2 = 19.55 and 14.45 for MMP and SWTD respectively, p's <0.001[1]). Therefore, the open-ended responses in Wassermann et al. cannot even be described as representative of their own sample, much less of the general public.

Conclusion

Wasserman et al. set out to investigate public opinion on cetacean captivity "using a methodological approach known for reducing the introduction of some forms of bias. . ." (p. 1). In doing so the authors "aimed to contribute to an accurate and up-to-date baseline of public opinion on cetacean captivity." (p. 1). We commend the authors for identifying and targeting potential biases from previous research and agree that more work needs to be done. However, despite their stated objective, the Wassermann et al. (2018) study contained a number of flaws in data collection, coding, analysis, and interpretation of results. Specifically, (1) coding of open-ended responses may have been unreliable and biased as participant responses were summarized by hand and not audio recorded for accurate transcription, and were then coded without identified operational definitions or inter-coder reliability; (2) the data presentation contained numerous inaccuracies in which the text did not match the raw data provided, or the data presented in different parts of the text were contradictory; (3) supplementary analyses with the

[1] These statistics were calculated using the respondents from Wassermann et al.'s raw data Sheet 2 with their associated likelihood ratings from Sheet 1. If we instead use the binary response codes from Sheet 1 as the measure of who gave an open-ended response, the same pattern of results is found (χ^2 = 19.17 and 21.70 for MMP and SWTD respectively, p's <0.001).

raw data indicated the presence of sequence effects that were not controlled for or acknowledged by Wassermann et al.; and (4) the participants who chose to provide open-ended responses (on which many of the conclusions presented in the paper relied heavily) were not representative of even the skewed demographic in this study, much less the general public. The presence of such pervasive methodological flaws casts serious doubt on the study's conclusions (including cases in which data transpositions meant that the conclusions should have been the exact opposite from what was presented in the paper). Further research on this topic, with a particular eye toward rigorous scientific methodology that remedies the methodological threats to internal and external validity we have identified, will be needed to achieve the goal of an unbiased understanding of public opinion.

References

Burns, N. (1989). Standards for qualitative research. *Nursing Science Quarterly*, *2*(1), 44-52.

Campbell, J., Quincy, C., Osserman, J. & Pedersen, O. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42, 294-320.

Carey, J. W., Morgan, M., & Oxtoby, M. J. (1996). Intercoder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *CAM Journal*, *8*(3), 1-5.

Gorden, R. (1992). *Basic Interviewing Skills*. Itasca, IL: F. E. Peacock.

Hagelin, E. M. (1999). Coding data from child health records: The relationship between interrater agreement and interpretive burden. *Journal of Pediatric Nursing*, *14*(5), 313-321.

Hammersley, M. (2010). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, *15*(1), 1-7. Sociology Press.

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, *16*(3), 307-331.

Passer, M. W. (2017). *Research methods: Concepts and connections*. New York, NY: Worth Publishers, Macmillan Learning.

Schwarz, N. (1996). Survey research: collecting data by asking questions. In G. R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 65-90). London, UK: Sage.

268 Schwarz, N., & Hippler, H.J. (1991). Response alternatives: the impact of their choice and
269 presentation order. In P. B. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S.
270 Sudman (Eds.), *Measurement errors in surveys* (pp. 41-56). New York, NY: Wiley.
271 Strack, F. (1992). 'Order effects' in survey research: activation and information functions of
272 preceding questions", In N. Schwarz & S. Sudman (Eds.), *Context effects in social and*
273 *psychological research* (pp. 23-34). New York, NY: Springer-Verlag.
274 Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application*
275 *of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.
276 Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in
277 attitude measurement. *Psychological Bulletin, 103*(3), 299-314.
278 Wassermann, S. N., Hind-Ozan, E. J., & Seaman, J. (2018). Reassessing public opinion of
279 captive cetacean attractions with a photo elicitation survey. *PeerJ*, *6*, e5953.
280
281
282
283
284
285

Table 1:

Number of respondents who gave qualitative responses to MMP and SWTD attractions, and their likelihood of visiting each, as reported from three different sources.

Rating	Article[a]	Sheet 1[b]	Sheet 2[c]
		MMP	
Likely/ Very Likely	5	10	5
Unlikely/ Very Unlikely	48	54	47
		SWTD	
Likely/ Very Likely	26	43	32
Unlikely/ Very Unlikely	51	59	46

Notes:

(a) As presented in Wassermann et al. (2018), pp. 11-12 and Table 4

(b) Calculated from Wassermann et al. (2018) supplement 2 (Raw data) Sheet 1

(c) Calculated from Wassermann et al. (2018) supplement 2 (Raw data) Sheet 2

298 Table 2:

299 Proportion of respondents in each category who provided each type of reason for their opinion

300 of MMP and SWTD attractions, as reported from two different sources.

301

Opinion codes	Likely/ Very Likely		Unlikely/ Very Unlikely	
	Article[a]	Sheet 2[b]	Article[a]	Sheet 2[b]
		MMP		
Animal welfare	20.0	20.0	72.9	74.5
Not entertaining	.	.	10.4	8.5
Human welfare concerns	.	.	4.2	2.1
Overly commercial	.	.	14.6	14.9
Conservation concerns	.	.	4.2	4.3
Too costly	.	.	.	.
Unclear reasoning	.	.	4.2	4.3
Entertaining	100.0	100.0	.	.
Appropriate for children	20.0	.	.	4.3
Cited media influence	.	.	16.7	17.0
		SWTD		
Animal welfare	15.4	28.1	56.9	52.2
Not entertaining	.	.	21.6	21.7
Human welfare concerns	3.8	3.1	9.8	10.9
Overly commercial	.	3.1	3.9	2.2
Conservation concerns	3.8	.	2	2.2

Too costly	.	3.1	2	2.2
Unclear reasoning	.	.	5.9	10.9
Entertaining	96.2	78.1	2	.
Appropriate for children	34.6	28.1	11.8	13
Cited media influence	3.8	.	2	.

Notes:

(a) As presented in Wassermann et al. (2018), Table 4

(b) Calculated from Wassermann et al. (2018) supplement 2 (Raw data) Sheet 2