Research Article

Assessing Reliability for Quantifying Social Interactions among Crayfish

Lee Ann Smith,[1] Jeremy Nadolski,[2] Grace Jacobs,[3] Jodi M. Ogle,[3] Madhusudan P. Srinivasan,[3] Hannah N. Tanner,[3] Elizabeth R. Steele,[3] Nicole T. Marguerite,[4] Sonya Bierbower,[5] Slane Steen,[6] Isaac Easterling,[7] Abigail Greenhalgh,[8] Cecilia Pankau,[9] Shelby McCubbin,[10] Brad Behymer,[11] and Robin L. Cooper [ID] [3]

[1]*Department of Biological Sciences, Benedictine University, Lisle, IL 60532, USA*

[2]*Department of Mathematical and Computational Sciences, Benedictine University, Lisle, IL 60532, USA*

[3]*Department of Biology, University of Kentucky, Lexington, KY 40506, USA*

[4]*Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37235, USA*

[5]*Department of Chemistry and Life Science, United States Military Academy, West Point, NY 10996, USA*

[6]*Department of Psychology, University of Kentucky, Lexington, KY 40506, USA*

[7]*Department of Biology, Eastern Kentucky University, Richmond, Kentucky 40475, USA*

[8]*Oregon State University, College of Agricultural Sciences, Corvallis, Oregon 97331, USA*

[9]*College of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA*

[10]*College of Medicine, University of Kentucky, Lexington, KY, USA*

[11]*Department of Biological Sciences, Eastern Kentucky University, Richmond, KY 40475, USA*

Correspondence should be addressed to Robin L. Cooper; rlcoop1@uky.edu

Animal behavior is a useful way to evaluate the environment and can be a predictive tool to assess not only the effects of treatments in a laboratory setting, but also the status of ecological habitats. As invasive species of crayfish encroach on territories of native species, the social behaviors and interactions can be informative for ecological studies. For a wider and more impactful effect, training community scientists using a scoring system to record the social interactions of crayfish that includes both the level of aggression and intensity would provide useable data to monitor the environment. Amateur scientists with little training were fairly reliable in their average scoring of the crayfish and the maximum behavior score with an expert as well as among themselves. However, the number of interactions was not as a reliable metric to compare with the expert or just among the amateurs.

1. Introduction

Behavioral ecology and animal behavior studies use scoring indices to quantify simple to intricate behavior across animal taxa. However, scoring indices must be specific to behavioral context as well as taxa. Numerical quantification of behavior lends itself to comparisons among treatment groups that can be supported by statistical inference.

Behavior in crayfish has been studied in solitary and social settings [1–3]. Aggressiveness based on differences in size, response to missing cheliped, and differential exposure to pharmacological or hormonal agents has been studied in pairs of crayfish (dyads) and in groups [4, 5]. The extremely aggressive behaviors seen in an intense battle, such as removing the opponent's limbs or flipping them over to cut the articulation membrane in the abdomen, were ranked as the

most aggressive score as this would inflict a life-threatening wound [6]. More recently, stability of the social hierarchy over time has also been assessed using behavioral indices [3, 7]. Furthermore, Gherardi and Pieraccini [8] developed detailed indices for 20 different behavioral patterns for quantifying agonistic encounters in crayfish. Ranking the levels of each behavior allows for comparing individuals in multivariate space (i.e., across multiple behavioral axes).

Scoring aggressiveness and the intensity of the interaction may expedite quantifying social behaviors in both lab and field settings. The reliability and reproducibility of the scoring system can be assessed when the same digital video recordings are scored by multiple observers, and this allows for employing crowd sourcing and community science efforts. For community science data to be publishable, it must be reliable and accurate. The scoring system to rank interactions between crayfish can be tested for reliability by having multiple observers score the same video recording. "JWatcher" (https://www.jwatcher.ucla.edu/) is a free software developed by Daniel T. Blumstein, Janice C. Daniel, and Christopher S. Evans for scoring behavioral interactions in real time. Here, we describe a standardized step-by-step protocol to use JWatcher for scoring crayfish interactions to ensure reliability. The objective of this study is to analyze the scoring of crayfish social interactions by multiple participants for both level of aggressiveness and the intensity of the interaction to determine the reliability of the scoring paradigm for potential community science projects. This preliminary study used three different dyads of crayfish: *Faxonius virilis* versus *Faxonius virilis* of similar size; large *Cherax quadricarinatus* versus large *Cherax quadricarinatus*; and small *Cherax quadricarinatus* versus small *Cherax quadricarinatus*. The *Faxonius virilis* is commonly known as the Northern crayfish and the *Cherax quadricarinatus* as the Australian crayfish. The scoring was used to address the potential of survival for the invasive species (i.e., Australian crayfish and displacing a resident species). The purpose was not to have an extensive scoring system of interaction time, duration, and subtle types of interactions, such has how the chelipeds pinch an opponent on a walking leg or body part, but to use a very generalized scoring system to provide an overview of which species and size was most aggressive in a general sense.

2. Experimental Design

2.1. Animals. The large Australian crayfish were obtained from a supplier in Kentucky (Crystal Bridge Fish Farm; 4111 South Highway 53, Crestwood, KY 40014). The small Australian crayfish were obtained from Live Aquaponics (3800 CR 13, S. Elkton, Florida, 32033, USA). Northern crayfish were obtained from Carolina Biological Supply Company (PO Box 6010, Burlington, NC, USA, 27216-6010). They were housed in individual standardized plastic containers with weekly exchanged dry fish food and aerated water (20-21°C).

2.2. Experimental Arena and Video Collection. Three sets of crayfish dyads (Northern crayfish versus Northern crayfish, large Australian crayfish versus large Australian crayfish, and small Australian crayfish versus small Australian crayfish) were recorded with a Panasonic video camera (model HC-VX870 4K) from the top down over an aquarium (52 cm length × 24 cm width × 30 cm high) filled with half dechlorinated and half aerated water. Large Australian crayfish measured postorbital carapace length (33.94 ± 1.28 mm) and weight (28.66 ± 3.33 grams), and small Australian crayfish measured postorbital carapace length (16.18 ± 0.29 mm) and weight (2.82 ± 0.15 grams).

The city water from Lexington, Kentucky, USA, was first passed through a carbon-filled tank from Callaghan (150 cm tall) to remove chloramines and then placed into a 50-gallon plastic container to be aerated for three days. The water in the aquarium was changed out between each pair of crayfish. The two crayfish were placed in the tank at essentially the same time to avoid residency and hence territorial behavior. However, the first crayfish placed in the tank was always marked as "left" in the software for consistency in following each crayfish interaction separately. The bottom of the aquarium was layered with sand to provide a substratum to grip while walking. There was no shelter for the crayfish to hide, but a flat rock was placed on one side to provide some additional enrichment to the environment. The interactions were recorded for 20 minutes.

2.3. Scoring. Each crayfish in a pair was individually followed throughout the 20 minutes and ranked individually for the various types of interaction. The video was replayed to rank the behaviors of the second crayfish. Three pairs of crayfish interactions were monitored, and each crayfish was only paired once. We used the same ranking method as in Huber et al. [9] and Jimenez and Faulkes [10] with the exception of starting the first scoring level at 1 instead of 0. The following is the ranking scheme used herein (Table 1).

To examine the consistency and reliability of ranking the interactions when participants are given a scale and directions to follow for analysis, we enlisted 15 volunteers, also called amateurs, to watch the three videos and rank the interactions of all six crayfish using the JWatcher software (1.0 version). Volunteers were provided with instructions to use JWatcher for ranking crayfish behaviors (see Appendix A for directions). The rankings provided by the 15 volunteer observers were then compared to those recorded by an expert. All rankings of crayfish were blind to the rankings of other participants to avoid bias. The expert was a person who has spent a considerable amount of time in scoring 100s of Australian crayfish for a more in-depth investigation into the behaviors of this species in using shelters and pairings with various species of crayfish. Also, this expert is a person who modified the JWatcher program for the scoring regime used in this study. This person also made movies and posted them on YouTube to educate the amateurs in what the

TABLE 1: Ranking and behavior descriptions used.

Ranking	Behavior description
1	No fighting
2	Threat displays with no claw contact
3	Claw lock, where both animals contested the encounter and at least one animal used its claws to grab its opponent
4	Strike and rip, where both animals contested the encounter and at least one animal made unrestrained use of the claws

scoring system is related to in sample interactions. Since the expert had a vested interest in developing this study and to use the scoring system for the in-depth studies to follow, this person was very precise in scoring.

The explanation of how to install and set up the software was provided in a text file, as well as a movie going through the procedure (see Appendix A; Video https://youtu.be/VfbpRIGKCmc).

Recordings of three pairs of crayfish were provided via YouTube, and each participant was provided with a link (https://www.youtube.com/watch?v=-j79LR4-qII&list=PL-plR67-pM5S3r-XZDXMLzSovTnlh1ZY9) to use these three pairings for determining reliability in the analysis.

2.4. Statistical Analysis

2.4.1. Distribution of Interaction Scores between Amateurs and the Expert.
The distribution of the interaction scores was examined to determine whether there was a difference between the percent distribution of time all amateurs scored at given level of interaction and the distribution of percentage the expert scored in each category. A chi-squared test of homogeneity was used to test whether the distribution of all the amateurs was the same as the expert. When a significant result was found, a post hoc analysis was conducted with a Bonferroni adjustment to determine which levels were significantly different. The chisq.posthoc.test function in R was used to conduct the post hoc analysis [11, 12].

2.4.2. Comparing the Frequency of Crayfish Interactions.
For comparing the number of interactions recorded between the amateurs and the expert, a Shapiro–Wilk test was conducted to test the assumption of normality for each crayfish data set of the number of interactions recorded by the amateurs. A Wilcoxon signed-rank test, tied observations were treated with the average rank, was used when the data were not normally distributed; otherwise, a one-sample t-test was used. For both tests, the null value was set to the number of interactions recorded by the expert.

2.4.3. Comparing Different Metrics on the Inter-Reliability between the Expert and All the Amateurs and Intra-Reliability of the Amateurs.
Since the number of interactions recorded by the amateurs and the expert varied (Table 1), we used different metrics to compare the reliability between the expert and the amateurs as well as among the amateurs. Instead of looking at all the scores for each reviewer and comparing them to the expert or another amateur, we focused on other metrics due to the wide range of total recordings of the reviewers. These metrics include the average ranking score of each of the six crayfish subjects was recorded with the 1 expert and 15 amateurs, the highest behavior score recorded for each of subjects and each scorer (1 expert and 15 amateurs), and finally, the total number of interactions recorded for the 6 crayfish for the 16 reviewers. Analysis of the consistency of the metrics used intraclass correlation coefficient (ICC) with a 95% confidence interval using the R package irr and the R function icc. When looking at the reliability of the expert versus the amateurs, all 6 crayfish and 16 data points were used in computing the ICC. Finally, when reporting on the intra-reliability of the amateurs, the same data set was used with the expert data excluded.

All data analysis was conducted using R version 4.1.0 and R Studio 2022.12.0.

3. Results

The distribution of rankings over the entire observed interaction duration for each crayfish (i.e., "left" and "right") produced no significant differences for the Northern crayfish pairing (Figures 1(a) and 1(b)) nor the large Australian crayfish pairing (Figures 2(a) and 2(b)). Scoring between the amateurs and experts for either first or second Northern crayfish in the aquarium had no statistical difference. However, with the smaller Australian crayfish grouping, there was a significant difference in the distribution of rankings for both left and right crayfish (Chi-square, $p < 0.001$; Figures 3(a) and 3(b)). When conducting the post hoc analysis, there was a significant difference for the "left" small Australian crayfish between the expert and the amateurs with regard to the percentage of interactions at all ranking levels (Chi-square, all p values <0.003; Figures 3(a) and 3(b)). In general, amateurs scored both crayfish at a higher level of aggressiveness (Figure 3) and with more interactions (Table 2) than the expert (Chi-square, $p < 0.001$). Furthermore, in all levels, except 2 and 4, for the "right" crayfish, a post hoc Chi-square test found a significant difference between the experts and the amateurs. Results suggested that the expert had a constant or equal percentage in each of the ranking levels, while the amateurs found that level 2 was the least likely interaction pattern and level 3 was the highest. Figure 3(b) shows that the expert and the amateurs found statistically the same rate of ranking for
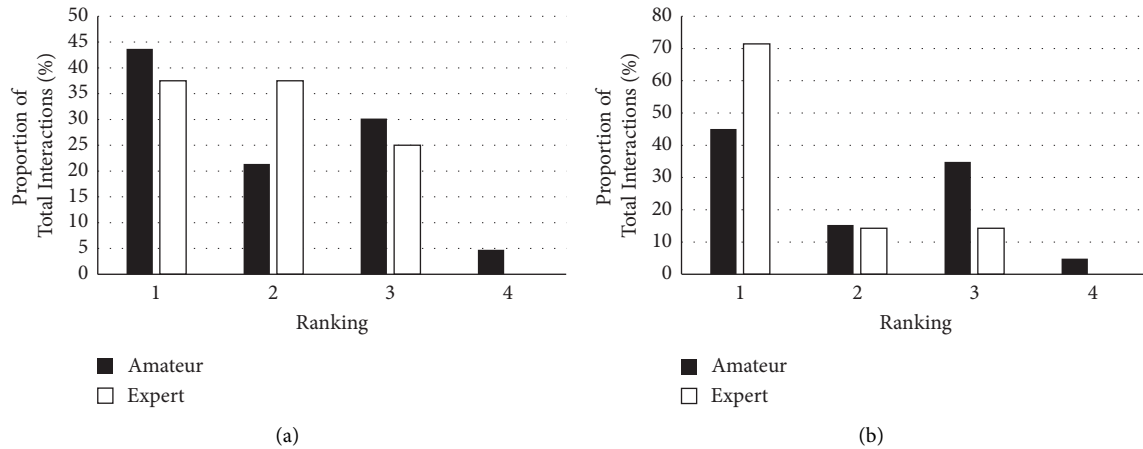
(a)

(b)

FIGURE 1: The distribution of rankings for the pairing Northern crayfish vs Northern crayfish. (a) the "left" or first crayfish entering the aquarium and (b) the "right" or second crayfish entering the aquarium. Each rank represents a different type of behavior that was scored. For Northern crayfish, there was no statistical difference in the distribution of rankings between the amateurs and expert for either the first or second crayfish into the aquarium.
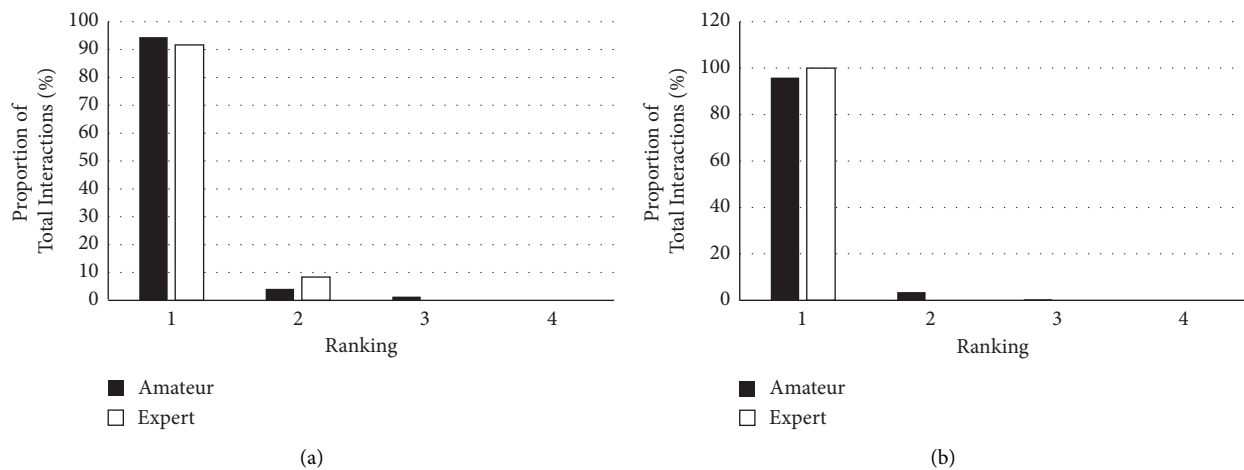

(a)

(b)

FIGURE 2: The distribution of rankings for large Australian crayfish vs large Australian crayfish. (a) the "left" or first crayfish to be monitored or entered the aquarium and (b) the "right" or second crayfish entered into the aquarium. Each rank represents a different type of behavior that was scored.

levels 2 and 4 but were different at levels 1 and 3. The expert tended to rank the interactions lower in this instance than the amateurs (*significant finding at the $p < 0.05$).

The left crayfish is presented in the A panels, and the right crayfish is presented in the B panels in the following figures.

When comparing the average number of interactions between the amateurs and the expert, only the "B" or second large Australian crayfish had a significant difference (t-test, $p = 0.016$). However, the *F. virilis* crayfish pair and the small Australian crayfish pair had high variability in the number of interactions recorded by the amateurs as demonstrated by the range of counts (Table 1). This variability may have caused the lack of significance especially in the small Australian crayfish groups. A total of 15 amateurs scored in each group.

For the reliability of the averages between the expert and the amateurs for the six crayfish subjects, we find a 0.704 (0.457–0.937, 95% CI) intraclass correlation coefficient (ICC), which is a good amount of reliability. Based on the maximum score for each subject interaction, the expert and the amateurs have an ICC of 0.684 (0.432–0.931, 95% CI), which is a moderate reliability in detecting a similar max score of aggressiveness. Finally, for the number of interactions each amateur and the expert recorded for their video of the crayfish subjects, the ICC was 0.285 (0.103–0.728, 95% CI), which is indicative of poor reliability.

Looking at the consistency of the data for just the amateurs, we find similar results at all three metrics. The ICC for the average aggression score among the amateurs is 0.711 (0.464–0.939, 95% CI), which is good reliability. The maximum behavior recorded for each amateur for the six
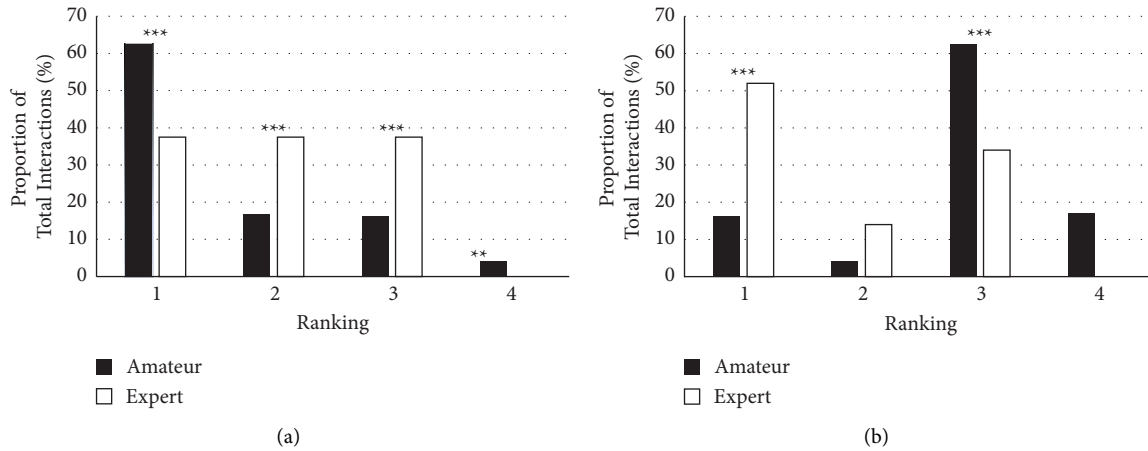
FIGURE 3: The distribution of rankings for small Australian crayfish vs small Australian crayfish. (a) the "left" or first crayfish into the aquarium and (b) the "right" or second crayfish into the aquarium. Each level represents a different type of behavior that was ranked (all $p < 0.003$; ** $= p < 0.01$; *** $= p < 0.001$).

TABLE 2: Frequency of crayfish interactions.

Group	Expert count	Amateurs average count (min-max)
"Left" Northern crayfish	24	29.6 (7–104)
"Right" Northern crayfish	21	26.2 (9–86)
"Left" large Australian crayfish	12	14.47 (9–28)
"Right" large Australian crayfish*	10	13 (6–21)
"Left" small Australian crayfish	63	88.87 (12–321)
"Right" small Australian crayfish	50	79.4 (14–297)

subjects had an ICC of 0.679 (0.424–0.93, 95% CI), which is moderately reliable, while the number of interactions ICC computation was 0.28 (0.098–0.725, 95% CI), which is poor reliability.

4. Discussion

This study examined the reliability of observers, including amateurs, in ranking aggressive interactions in crayfish as a means to develop protocols for community science projects aimed to engage a wider range of participants in scientific endeavors. The potential theme of this wider community science project is to assess the potential of the Australian crayfish as an invasive species in their ability of being more aggressive than local domestic species. It was shown that various participants ranking the same sets of video data of crayfish interactions for the level of aggressiveness did fairly well in reliability in the distributions of scores when the interactions were relatively few and when there were not as many highly aggressive interactions. The group with the highest level of aggression was the pairing of the small-to-small Australian crayfish. The term "small" is a reflection of the animal's size and developmental stage. In this pairing, there was a low reliability in the scoring distributions. More explicit instructions and sample videos highlighting the scoring with sample sets will be needed prior to releasing this project as an official community science project for higher accuracy in the scoring methods.

Furthermore, analysis demonstrated the most reliable metric to use when comparing the expert with the amateurs is the mean or average scoring value with the maximum being the next reliable. The number of interactions other than looking at the distributions was a very unreliable measure as some amateurs would overly score a movement versus others.

The expert had extensive experience ranking crayfish interaction videos from various projects. The expert was more likely to notice nuanced behaviors and be adept distinguishing cheliped hold vs a pinch. This might explain why amateurs recorded greater instances of aggressive behaviors (levels 3 and 4) compared to the expert. Though we provided sample videos to illustrate the ranks for the major behaviors, we did not think it was vital to include training resources for some of the subtle behaviors. Furthermore, including the term "aggression" while introducing the project to the participants may have resulted in an implicit bias that may have caused some participants to classify instances of nonaggressive physical contact erroneously. For example, an interaction in which one crayfish pushed the other out of the way when walking around the tank may have been scored as a "threat" (level 2) when in fact it was actually a level 1 (no fighting). Usually, crayfish tend to explore the boundaries of the container when they are first introduced into it [13, 14]. This increases the chance of interactions, some of which may not necessarily be antagonistic. For instance, one crayfish might pass over or under the other with chelipeds in front

but not necessarily in an aggressive state (i.e., close chelipeds and low to the surface). A passing individual might touch their chelipeds to the other's side pushing them slightly. Such an interaction could be misread as threat displays (level 2) by an observer. These interactions would need to be clarified in an introductory video and listed as level 1.

It is becoming more common practice to provide raw data with publications for various scientific studies. Even with an explanation in video format, as done in this study, explicit examples may be needed to cover a range in conditions. This will be of particular importance if relying on contributions from a community science project or a crowdsourcing project to analyze data sets [15]. A large number of repetitive measures will likely be required for accuracy [16]. Comments by individual participants in conducting this project are shown in Appendix B.

5. Conclusion

Indexing animal behavior for quantitative analysis can be complex depending on the level of detail desired. In the field of animal science, it is common to index standard behavior such as feeding initiation and consumption. This is a relatively easy measure, but this does not necessarily capture feeding duration. One might want to know if the animal is just standing there or standing and chewing food or with the head in the feeding trough moving around the food or ingesting the food. One may also want to index how much the animal eats by measuring a change in weight of the trough. It might even be of interest to know how an adjacent animal might alter the behavior of the focal individual. In determining an index for social interactions can be very complex as one may use visual observations, but it fails to incorporate other vital communication cues that are auditory or chemical in nature. Physiological measures (i.e., heart rate or hormonal levels) would provide further detail in the responses to correlate with behavioral interaction; however, obtaining such physiological measurements may interfere with normal behavior. Thus, depending on the goal of a project, indices can be adjusted to optimize effort in relation to reliability and accuracy.

Appendix

A. Modifications and Clarifications in the User Guidelines of JWatcher Software

How to install JWatcher based on text provided on the website https://www.jwatcher.ucla.edu/

Video https://youtu.be/VfbpRIGKCmc

For installing on a windows-based computer:

Version 1.0 runs under Mac OS-X (at least through 10.12 (Sierra), but it is not compatible with 10.3.9), Windows OS from XP through (at least) Windows 10, and Linux OS (thanks to Julien Martin for making the new builds).

Are you using Vista or Windows 7 or later. Some users have reported that the program "hangs" on installation. Here is a user-written trick to install on these platforms: use JWatcher Version 1.0 and download the version with the JRE. Navigate to the installer that saved from the website. Right click on it and select Properties (should be the bottom option). In the window that opens, select the second tab (Compatibility). There is a Compatibility Mode listed on that tab near the top of the page. Check the "Run this program in compatibility mode for:" checkbox. Select "Windows XP (Service Pack 2)" or "Windows XP (Service Pack 3)". Click the OK button in the bottom of the window. Run the installer.

For installing on a MacIntosh-based computer:

If the Mac rejects the JWatcher download, try these steps:

(i) Click the "JWatcher 1.0 for Mac" link in the gray box on the website https://www.jwatcher.ucla.edu/download-jwatcher/

(ii) When the window pops up saying that it can't be run click "OK."

(iii) Open "System preferences." Then, select "Security and Privacy."

(iv) Click "open anyway" toward the bottom of that window.

(v) When it asks you if you are sure, click "open."

From there, it should download like normal (and ask you to set up your language preferences). You can continue clicking "next" on the dialogue boxes until it looks like the JWatcher window.

B. Comments and Suggestion by Participants in Conducting the Scoring

The views of those conducting this study are very helpful in developing an improved analysis. Such comments are summarized next. (1) The use of gravel or sand is preferable to dirt as the dirt can cause the water to be cloudy with the individual when then tend to dig. (2) Tend to keep the overhead lighting to a minimum as bright lighting will likely imped the interactions of the crayfish due to hesitation to be exposed. (3) Avoid movements by the person setting up the arena and recording as the crayfish are very visual and will respond to shadows and changing in lighting or noises in the background. (4) Some of the interactions are very quick and the tail flips make it difficult to know that one is following the same individual. Thus, the video recording and JWatcher have to be stopped at the same time and then note the time on the video in order to scroll back and advance slowly to see which crayfish is which. Then one can proceed with starting JWatcher and the video in the correct time window to complete the analysis. A way to avoid this would be to mark the crayfish with a tag such a drop of fingernail polish, of a discernible color, on the telson. Just a dot for example would be very helpful. (5) When viewing a recording from the side of a small aquarium, it is difficult to spatial tell how close or if a slight touch occurs between two individuals. The viewing of the interactions is preferred from the top down over the interactions.

Data Availability

The movies of the interactions used to support the findings of this study are included within the article in the YouTube links herein. Also, all data for the graphs are available from the corresponding author upon request. Recordings of three pairs of crayfish were provided via YouTube, and each participant was provided with a link (https://www.youtube.com/watch?v=-j79LR4-qII&list=PL-plR67-pM5S3r-XZDXMLzSovTnlh1ZY9). How to manage the software https://youtu.be/VfbpRIGKCmc.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

G.J. and R.C. conceptualized the study. L.A.S., J.N., G.J., J.M.O., M.P.S., H.N.T., E.S., N.T.M., S.B., S.S., I.E., A.G., C.P., S.M., B.B., and R.L.C. performed the methodology. G.J. provided software. L.A.S., J.N., G.J, J.M.O., M.P.S., H.N.T., E.S., N.T.M, S.B., S.S., I.E., A.G., C.P., S.M., B.B., and R.C. contributed to validation. J.N. performed formal analysis. L.A.S., J.N., G.J., J.M.O., M.P.S., H.N.T., E.S., N.T.M., S.B., S.S., I.E., A.G., C.P., S.M., B.B., and R.L.C. performed investigation. J.N., G.J., and R.L.C. provided resources. R.L.C. contributed to data curation. L.A.S., J.N., G.J., M.P.S., N.T.M, and R.L.C. wrote the original draft. L.A.S., J.N., G.J., M.P.S., N.T.M., and R.L.C. reviewed and edited the manuscript. R.L.C. contributed to visualization. R.L.C. performed supervision. J.N., G.J., and R.L.C. performed project administration. R.L.C. provided funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] R. V. Bovbjerg, "Dominance order in the crayfish *Orconectes virilis* (Hagen)," *Physiological Zoology*, vol. 26, no. 2, pp. 173–178, 1953.

[2] R. V. Bovbjerg, "Some factors affecting aggressive behavior in crayfish," *Physiological Zoology*, vol. 29, no. 2, pp. 127–136, 1956.

[3] F. A. Issa, D. J. Adamson, and D. H. Edwards, "Dominance hierarchy formation in juvenile crayfish *Procambarus clarkii*," *Journal of Experimental Biology*, vol. 202, no. 24, pp. 3497–3506, 1999.

[4] R. Huber, M. Orzeszyna, N. Pokorny, and E. A. Kravitz, "Biogenic amines and aggression: experimental approaches in crustaceans," *Brain, Behavior and Evolution*, vol. 50, no. 1, pp. 60–68, 1997.

[5] M. R. Smith and D. W. Dunham, "Chela posture and vision: compensation for sensory deficit in the crayfish Orconectes propinquus (girard) (Decapoda, cambaridae)," *Crustaceana*, vol. 59, no. 3, pp. 309–313, 1990.

[6] R. Huber and E. A. Kravitz, "A quantitative analysis of agonistic behavior in juvenile American lobsters <i>(*Homarus americanus* L)," *Brain, Behavior and Evolution*, vol. 46, no. 2, pp. 72–83, 1995.

[7] C. Goessmann, C. Hemelrijk, and R. Huber, "The formation and maintenance of crayfish hierarchies: behavioral and self-structuring properties," *Behavioral Ecology and Sociobiology*, vol. 48, no. 6, pp. 418–428, 2000.

[8] F. Gherardi and R. Pieraccini, "Using information theory to assess dynamics, structure, and organization of crayfish agonistic repertoire," *Behavioural Processes*, vol. 65, no. 2, pp. 163–178, 2004.

[9] R. Huber, K. Smith, A. Delago, K. Isaksson, and E. A. Kravitz, "Serotonin and aggressive motivation in crustaceans: altering the decision to retreat," *Proceedings of the National Academy of Sciences*, vol. 94, no. 11, pp. 5939–5942, 1997.

[10] S. A. Jimenez and Z. Faulkes, "Can the parthenogenetic marbled crayfish Marmorkrebs compete with other crayfish species in fights?" *Journal of Ethology*, vol. 29, no. 1, pp. 115–120, 2011.

[11] R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.

[12] D. Ebbert, "chisq.posthoc.test: a post hoc analysis for pearson's chi-squared test for count data," *R package version 0.1.2*, 2019, https://CRAN.R-project.org/package=chisq.posthoc.test.

[13] H. Li and R. L. Cooper, "Spatial familiarity in the blind cave crayfish, Orconectes australis packardi," *Crustaceana*, vol. 74, no. 5, pp. 417–433, 2001.

[14] S. J. G. Kellie, J. Greer, and R. L. Cooper, "Alterations in habituation of the tail flip response in epigean and troglobitic crayfish," *Journal of Experimental Zoology*, vol. 290, no. 2, pp. 163–176, 2001.

[15] J. D. J. Clare, P. A. Townsend, C. Anhalt-Depies et al., "Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved?" *Ecological Applications: A Publication of the Ecological Society of America*, vol. 29, no. 2, Article ID e01849, 2019.

[16] F. Leal, B. Malheiro, H. González-Vélez, and J. C. Burguillo, "Trust-based modelling of multi-criteria crowdsourced data," *Data science*, vol. 2, no. 3, pp. 199–209, 2017.