Towards improved behavioural research in aquatic toxicology: Acclimation and observational timeframes are important considerations when designing behavioural tests with fish

Steven D. Melvin[1], Marie A. Petit[1], Marion C. Duvignacq[1], John P. Sumpter[2]

[1]Australian Rivers Institute, Griffith University, Southport, Queensland, 4222, Australia
[2]Institute for Environment, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom

*Address for correspondence:
Steven D. Melvin
Australian Rivers Institute
Building G51, Griffith University, QLD 4222 Australia
Tel: (+61) 755 528 272
Email: s.melvin@griffith.edu.au

ABSTRACT

The quality and reproducibility of science has recently come under scrutiny, with criticisms spanning disciplines. In the field of aquatic toxicology, behavioural tests are currently an area of controversy since inconsistent findings have been highlighted and attributed to poor quality science. The problem ultimately boils down to our lack of understanding about basic behavioural patterns, which limits our ability to effectively design statistically robust experiments that yield ecologically relevant data. The present study takes a first step towards understanding baseline behaviours in fish, including how basic choices in experimental design might influence behavioural outcomes and interpretations in aquatic toxicology. Specifically, we explore how fish acclimate to experimental arenas used for behavioural analysis and how different lengths of data acquisition influence estimates of basic swimming parameters (i.e., average, maximum and angular velocity). We evaluate these factors qualitatively in terms of fundamental behavioural characteristics and quantitatively in terms of the theoretical statistical power achievable with different design choices. We also performed a semi-quantitative literature review to place our findings in the context of the published literature describing behavioural tests with fish. Our experimental findings reveal that fish fundamentally change their swimming behaviour over time, and that our choices surrounding acclimation and observational timeframes may therefore have implications for influencing both the ecological relevance and the statistical robustness of behavioural toxicity tests. Our review identified 165 studies describing behavioural responses in fish exposed to various stressors. Importantly, these data reveal that the majority of publications describing fish behavioural responses report the use of extremely brief acclimation times and observational durations, which helps explain the inconsistencies identified across studies. We recommend researchers applying behavioural tests with fish, and other aquatic species, apply a similar framework to better understand baseline behaviours and the implications of design choices for influencing study outcomes.

KEY WORDS

Behavioural analysis, experimental design, acclimation, Ethovision, toxicity testing

INTRODUCTION

Behavioural analysis is being increasingly applied towards contemporary aquatic toxicology research. The growing popularity of behavioural testing largely stems from recent technological advancements (Parker, 2015), which have made commercial and open-source analysis software widely accessible. Additionally, the general consensus is that behavioural tests are rapid and sensitive to a wide range of pollutants (Melvin and Wilson, 2013), and offer a novel approach that may help link sub-lethal physiological effects with population-level outcomes (Pyle and Ford, 2017). However, while there are several perceived benefits to studying behavioural changes in wildlife exposed to environmental pollutants, our understanding of the factors governing animal behaviour is still very limited and caution is therefore necessary when applying such tests (Melvin, 2017; Sumpter et al., 2014). Considering the marked increase in behavioural techniques in aquatic toxicology testing, it now seems prudent to evaluate how such tests are being applied, including factors that might influence the validity and repeatability of behavioural outcomes amongst studies (McCallum et al., 2017).

One of the most notable subjective aspects of modern-day behavioural toxicity research relates to the wide range of study designs and test methodologies being applied. One the one hand, the flexibility of behavioural testing can be viewed as a positive attribute since this allows diverse ecological processes to be studied (Parker, 2015). On the other hand, a lack of standardisation makes it very difficult to ensure the validity of different experimental designs and may lead to inconsistency in documented response patterns amongst studies (Huerta et al., 2016; Sumpter et al., 2014). The latter holds consequence for the progression of science because variable study outcomes can lead to continued exploration of potentially unimportant stressors, thereby resulting in unnecessary animal usage and resource expenditure. As a starting point, there is a pressing need to establish basic knowledge about baseline behavioural characteristics for species being used in behavioural toxicology (Melvin et al., 2016).

Since the increased prevalence of behavioural tests in aquatic toxicology seems to largely correspond with the wide accessibility of computational software tools (Bae and Park, 2014), basic approaches for using these technologies must be carefully evaluated. The most straightforward application of specialised behavioural analysis software involves measurement of basic swimming characteristics, such as velocity and other aspects of animal movement. As such, the most obvious areas where subjectivity in study design might be introduced are in the timeframes for acclimation, exposure, and data collection. Indeed, Kane et al. (2005) identified 1) the timeframe for acclimation to experimental arenas and 2) the duration of observation as key factors that require consideration when designing behavioural toxicity tests with fish. This was reinforced by a recent study demonstrating how different observational timeframes can influence overall conclusions of behavioural analysis (Melvin, 2017). However, despite the identified importance of these factors, there have been no studies explicitly focused on understanding how choices in experimental methodology, and specifically acclimation time and the duration of observation, might influence fish behaviour and subsequent study outcomes in aquatic toxicology.

The present study explores swimming performance and temporal behavioural variability of adult mosquitofish (*Gambusia holbrooki*) using commercially available behavioural analysis software, to investigate the importance of adequately acclimating fish to observational arenas for testing. The theoretical statistical power achievable with different acclimation times, and observational durations, was determined to explore how these factors might influence the quality of behavioural toxicity tests. Finally, a literature review was performed to document acclimation timeframes and observational durations reported in published behavioural toxicity tests using fish.

MATERIALS AND METHODS

Experimental fish

Adult mosquitofish (*Gambusia holbrooki*) were used for the experiment, due to their wide geographical distribution (Pyke, 2008; 2005) and recent interest into the use of this species for behavioural testing (Jakka et al., 2008; Magellan et al., 2014; Melvin, 2017; Melvin et al., 2016; Saaristo et al., 2014; Sismeiro-Vivas et al., 2007). Fish were collected from a local woodland pond near Griffith University's Gold Coast campus, and transported to the laboratory in water from the collection site where they were separated by sex and size and acclimation to experimental conditions for several months prior to experimentation. Moderately hard testing water was used for holding and experimentation (USEPA, 1994) and temperature and photoperiod were maintained at 22.2 ± 0.8 °C and 12: 12-h light: dark, respectively. Experiments were approved by the Griffith University Animal Ethics Committee (Protocol No. ENV/03/16/AEC), and performed in accordance with the guidelines of the Australian Code for the Care and Use of Animals for Scientific Purposes.

Video recording fish swimming behaviour

Our experimental setup consisted of 20 square glass dishes (21 × 21 × 6 cm; Pyrex®) arranged in a 4 × 5 array. Dishes were placed on a large LED panel providing dim backlighting to increase contrast and achieve optimal tracking of the fish. The fish were fed staple flaked food *ad libitum* in their holding aquaria first thing in the morning on the day of testing, while setting up the behavioural arenas and software. After feeding, twenty sexually mature females weighting 730.65 ± 105.82 mg and measuring 32.43 ± 1.38 mm (standard length), were transferred to behavioural arenas filled with 800mL control water using a fine mesh dip-net. Mosquitofish are well known to prefer shallow, calm waters where risk from predation is low (Casterlin and Reynolds, 1977; Pyke, 2008). This volume was therefore chosen because it offered a semblance of ecological relevance and provided ample depth for free movement (3cm), while also preventing vertical movement and thereby limiting the study to two-dimensional behaviour to simplify the analysis. Video recording commenced immediately after the fish were placed into their respective test arenas at 9:00am and continued for a period of eight hours. Recordings were made using Ethovision XT 9.0 (Noldus Technologies, Inc) connected to an acA1300-30gc GigE camera (Basler AG, Germany) mounted above the test arenas. The experiment was performed in an empty laboratory behind closed doors, and no one entered the room during filming.

Following video recording the data was analysed over both 5min and 2hr intervals and the results were exported as excel files. Standard behavioural endpoints generated by the software were chosen for our assessment, including average and maximum swimming velocities (mm/s), and angular velocity (°/s). Since they are automatically produced these are commonly reported endpoints in behavioural studies. However, they also provide useful information including assessment of basic swimming performance, and when combined indicate behavioural complexity and occurrences of erratic movements (Benhaïm et al., 2012).

Acclimation characteristics and statistical power analysis

Behavioural data were plotted over time to reveal temporal patterns in how fish acclimate to experimental arenas, and to facilitate comparison of short (5min) and longer (2hr) observational timeframes. Basic statistical comparisons of differences in each behavioural endpoint over time were assessed via non-parametric Kruskall-Wallis test, using the data collected over 2hr observational timeframes. This was merely done to demonstrate statistical differences in behaviour during and after acclimation. Non-parametric analysis was used due to common violation of the assumption of equal variance for fish behaviour, which was observed in our study.

G*Power (v 3.1.9.2) software was used to calculate the achievable statistical power (1 – β) for each of the three behavioural endpoints over time (Faul et al., 2007). For the power analysis, a two-tailed t-test comparing independent means was used, with an alpha error probability of 0.05 and sample size n = 20 for each group. Effect size (d) was calculated by choosing a hypothetical Group 1 average reflecting the actual range of behaviours observed for each of the endpoints and a Group 2 average that resulted in maximum power (i.e., 1 – β = 0.99) when the lowest observed population standard deviation (σ) was used. Accordingly, means used for Group 1 and Group 2 were: 24 and 20 mm/s for average velocity, 150 and 100 mm/s for maximum velocity, and 2500 and 2000 °/s for angular velocity. These values remained constant, and the achievable statistical power for each data point was calculated by substituting the population standard deviation at each observation time.

Literature review – Current status of behavioural testing globally

A literature review was performed to identify laboratory-based experimental studies documenting how environmental stressors influence behavioural responses in fish. We searched the Web of Science™ and Google Scholar™ databases using various combinations of the following search terms relating to fish behavioural toxicity testing: ‘fish’, ‘aquatic’, ‘exposure’, ‘toxicity’, ‘contaminant’, ‘behaviour’, ‘behavioural’, ‘swimming’, ‘activity’, ‘movement’, ‘mating’, ‘feeding’, ‘foraging’, ‘aggression’, ‘avoidance’, ‘attraction’, ‘predation’. We included any studies that described behavioural responses in fish exposed to environmental stressors, provided they reported either the timeframe that animals were acclimated to exposure arenas (i.e., the aquaria or chambers used for behavioural observations) or the duration of time behaviours were monitored, or both. Our search predominantly returned studies assessing chemical pollutants, but also a handful of cases exploring animal cues (e.g., predator cues) and physical stimuli (e.g., noise pollution). Observational time was clearly reported in all cases, but there were several instances where acclimation time was somewhat ambiguous. Wherever possible, we made a conservative estimate of acclimation time based on the data provided. For example, if a study reported 'daily measurements of behaviour throughout the entire exposure', we assumed at least 24hr acclimation prior to the first measurement as a conservative estimate of acclimation.

The information obtained through our literature review was used to qualitatively assess and evaluate the current status of behavioural test methodologies. We classified studies based on the primary goal or behavioural endpoint, which resulted in 6 categories of behavioural test: avoidance/attraction, learning/memory, feeding/predation, mating/courtship/aggression, basic swimming, and anxiety. Basic descriptive data was calculated (i.e., N, mean and median) for acclimation time and observational timeframe for each category of behavioural test.

RESULTS

Acclimation characteristics and statistical power analysis

Fish displayed clear differences in exploratory behaviour immediately after introduction to behavioural arenas compared to that exhibited after several hours of acclimation. This was characterised by increased average velocity (Figure 1a) and decreased angular velocity (Figure 2a) during the first 3-4hrs in the behavioural arenas. Kruskal-Wallis analysis of data from 2hr observational timeframes revealed these differences to be statistically significant for both average (Figure 1b; $p < 0.0001$) and angular velocity (Figure 2b; $p < 0.0001$). Considering data based on 5min observation timeframes, population variance was observed to decrease for average velocity (Figure 1a) and increased for angular velocity (Figure 2a) following the acclimation period. Maximum velocity was consistent overall throughout the 8hr test when data was measured over 5min timeframes, but the population exhibited periods of consistency interspersed with periods of increased variance for this endpoint (Figure 3a). When data was collected over a longer observation time (2hr) maximum velocity differed significantly over time (Figure 3b; p = 0.0001).

Power analysis was used to explore the implications of acclimation time and observational duration on the reliability of behavioural tests. For average velocity, the theoretically achievable statistical power was generally lower during acclimation (i.e., first 3-4hrs), compared to that achievable in acclimated fish (Figure 1a). Increasing observation time from 5min to 2hr yielded an overall increase in achievable power, most notably during acclimation (Figure 2b). An opposite trend was identified for angular velocity, where initial consistency amongst fish resulted in very high power that quickly tapered off after the first hour in the behavioural arenas (Figure 2a). Power subsequently increased as the fish apparently became acclimated to the arenas, and longer observational time also yielded greater achievable statistical power for this endpoint (Figure 2b). The intermittent variability in maximum velocity resulted in predominately high power, but this was interspersed with periodic timeframes where very low power was achievable (Figure 3a). Contrary to average and angular velocity, this patchiness resulted in longer observational times (2hrs) exhibiting relatively low statistical power for this endpoint (Figure 3b).

Literature review – Current status of behavioural testing globally

Our review identified 165 studies describing behavioural endpoints in fish exposed to environmental pollutants or other imposed stimuli. Of these, the greatest number of studies focussed on basic swimming parameters (77 studies), followed by mating/courtship/aggression (42 studies), anxiety (27 studies), feeding/predation (10 studies), avoidance/attraction (5 studies), and learning/memory (4 studies).

Results identified that greater than 70% of all studies assessing basic swimming parameters (73.6%), learning/memory (80%), and anxiety (88.5%) acclimated fish for less than 60min prior to behavioural recording (Figure 4a). The same is true for more than half of all studies assessing mating/courtship/aggression (58.5%), feeding/predation (55.6%), and learning/memory (50%). Overall, 89.6% of studies assessing basic swimming parameters and 100% of the studies in all other categories observed fish for less than 60min (Figure 4b). The median acclimation time for studies assessing basic swimming parameters was 0.33hr (mean 18.1hrs) and the median observational time was 5min (average 59.4min). Studies assessing mating/courtship/aggression had a median acclimation time of 1hr (mean 19.0hr) and median observation time of 10min (mean 15.6min). Anxiety studies had median acclimation time of 0.01hr (mean 1.95) and median observation time of 6min (mean 10.4min). Feeding/predation studies report median acclimation time of 1hr (mean 16.6hr) and median observational time of 12.5min (mean 19.6min). Avoidance/attraction and learning/memory studies reported median acclimation times of 0.33hr (mean 0.6hr) and 1hr (mean 1hr), and median observation time of 5min (mean 28.5) and 5.5min (mean 6.5min), respectively (Figure 4a,b).

DISCUSSION

Criticisms about irreproducible science are currently widespread and span disciplines (Baker, 2016). In most cases this problem is believed to relate to lax publication of false-positive data, as opposed to researcher misconduct or data falsification (Forstmeier et al., 2016; Loken and Gelman, 2017). Regardless the cause, this is a major problem that can seriously undermine scientific progress. Aquatic toxicologists are generally well informed about the need for caution and scientific rigour when applying behavioural tests with fish (Marentette et al., 2012; Maximino et al., 2010; Melvin, 2017; Parker, 2015; Sumpter et al., 2014; Thompson et al., 2016). However, despite this comprehension, very few studies have explicitly investigated (or discussed) how basic choices in study design might influence behavioural patterns, and subsequently affect outcomes and interpretations. Considering the increased prevalence of behavioural tests in the field of aquatic ecotoxicology, concerns regarding the validity, reliability and repeatability of these approaches must be addressed (Melvin, 2017; Parker, 2015; Sumpter et al., 2014). The present study represents a first step towards addressing the issue of irreproducibility in behavioural toxicology research, by considering how a fundamental design element in any exposure study – time – influences baseline behavioural patterns in fish. Results demonstrate that acclimation time (i.e., to behavioural arenas) and observational duration (i.e., the length of time for data acquisition) are two basic factors that may influence the quality and validity of behavioural toxicity tests with fish.

Acclimation characteristics and statistical power analysis

Due to limitations of available analysis tools and housing requirements for long-term experiments, it is common for animals to be exposed to contaminants in one location and subsequently transferred to ‘behavioural arenas’ for behavioural analysis. In the present study, visualisation of swimming patterns over an 8hr observational period yielded several basic insights regarding acclimation of fish following introduction into novel behavioural arenas. Consideration of average and angular velocity together are particularly useful for exploring the complexity of swimming patterns (Benhaïm et al., 2012) or the occurrence of erratic movements

(Kim and Wardle, 2005). Evaluation of these endpoints indicated that, upon transference to a new environment, fish actively explored for several hours and generally exhibited what can be interpreted as anxious behaviour until about 4hrs after being introduced to behavioural arenas. Swimming patterns during the first 4hrs were characterised by greater average velocity (Figure 1a) and reduced angular velocity (Figure 2a) compared to the last 4hrs, when fish were apparently more acclimated. Functionally, this reveals that fish swam rapidly in more or less straight lines and performed relatively few slow turns immediately after being introduced into the behavioural arenas. Contrarily, following approximately 4hrs acclimation, the fish consistently swam slower overall and perform a greater number of more rapid directional changes. This is significant because it reveals that swimming patterns fundamentally change as fish acclimate to new environments for testing, and raises a key question – when is it appropriate to assess and compare fish behaviour?

At this stage, we unfortunately do not have sufficient information to formulate an answer, but we emphasis the importance of exploring and documenting baseline behaviours prior to using a species for testing. There are several possibilities that could explain increased activity at certain times compared to others, and each warrants further investigation to ensure robust behavioural testing in aquatic toxicology research. As one example, previous studies have established that hungry fish tend to swim faster than those that are satiated (Hansen et al., 2015; Pang et al., 2010). This would be a particularly important consideration for cases where it is necessary to measure behaviours of individual animals separately, for example throughout the course of a day (or several days). Classical conditioning (Pavlovian activity) is another concept related to feeding regime that could result in increased activity at certain times of day. Several fish species have been shown to anticipate feeding, displaying increased activity as established feeding time approaches (Chen and Tabata, 2002; Gee et al., 1994). This could conceivable influence behavioural parameters even with synchronised measurement of all animals from a test. Animals were fed just prior to commencing behavioural recordings in the present study, suggesting that the observed increase in average velocity early compared to later in the test is likely not associated with hunger or anticipatory behaviour. Nevertheless, hunger and satiation play an important role in fish foraging behaviour (Priyadarshana et al., 2006) and should be carefully considered when designing behavioural tests.

The importance of suitable acclimation time has been outlined in standard guidelines for toxicity testing (ASTM, 2014; OECD, 1992), and the specific relevance for behavioural research with fish has been further emphasised (Kane and Salierno, 2005). Despite this, ours is the first study to our knowledge explicitly assessing how fish acclimate to behavioural testing arenas, or more importantly speculating what this might mean for outcomes and interpretations in the context of toxicity testing. On a basic level, irrespective of any implications related to data analysis or statistical power, failure to adequately acclimate fish may preclude a study from assessing changes to ‘true’ baseline behaviours. It is unclear to what extent this might influence the ecological validity of behavioural toxicology research, and this will intuitively depend on the goal of a study. For example, short or even zero acclimation time are intuitively appropriate for studies assessing anxiety or exploratory behaviour, such as the widely used novel tank diving test (Blaser and Gerlai, 2006; Gerlai, 2003). However, as evidenced in this and other studies, a large proportion of behavioural toxicity research with fish has focused on basic swimming parameters (Melvin and Wilson, 2013). In such cases acclimation time may have important implications for our ability to extrapolate behavioural responses to natural populations. Defining what constitutes ‘normal’ behaviour in fish is very difficult and we hesitate to attempt this herein. Rather, we hope that the present study sparks much needed thought and discussion regarding appropriate design and use of behavioural tests in toxicology.

Contrary to average and angular velocity, mean maximum velocity was identified as being quite stable throughout the 8hr observational duration (Figure 3a). However, the variability (standard deviation) associated with maximum velocity differed markedly between observation time points. If we consider the achievable statistical power for maximum velocity, it is generally quite high overall (Figure 3a), but interspersed with short periods of high variance that in turn yield low achievable statistical power. This may help to shed some light as to why discrepancies are sometimes described amongst behavioural studies, and supports criticisms that the inconsistency in published behavioural responses likely reflects the occurrence of non-reproducible science (Sumpter et al., 2014). With the observed variability in maximum velocity, power analysis effectively revealed how data collected over short temporal scales (e.g., 5min observations) could reach completely different outcomes from one moment to the next. This should serve as an example to caution researchers from publishing findings from behavioural toxicity tests simply due to the fact that a significant difference was observed (Forstmeier et al., 2016), and particularly if the observational timeframe was very short. Perhaps more importantly, in the absence of standardised approaches for behavioural research, this highlights the need to repeat studies to verify short-term behavioural responses. From a design perspective, these results support the recent suggestion that Repeated Measures designs may be more appropriate for behavioural toxicity tests than one-off comparisons of group means (e.g., ANOVA), since the inclusion of multiple time points will help to account for the natural temporal variability that exists in fish behaviour (Melvin, 2017).

Behavioural data was acquired and analysed over both long (2hrs) and short durations (5min) to further explore how the timeframe for data acquisition might influence behavioural estimates and statistical robustness. These observation times were arbitrarily chosen and merely serve to explore how difference in the length of data acquisition might influence behavioural estimates. In the case of average and angular velocity, longer observational time helped account for the population variance in these endpoints, and thus increased the achievable statistical power. This seems intuitive, but based on the findings of our review (discussed later) the published literature is wrought with studies recording fish behaviour over comparatively short timeframes. Interestingly, in the case of maximum velocity, the intermittent occurrences of high variance observed from short-term (5min) acquisition were emphasised when data was collected over longer observation time (2hrs). Maximum velocity subsequently yielded lower power when measured over longer durations. Upon first consideration, this seems to suggest that short observation times would be optimal (for maximum velocity), since this at least offers a chance for robust analysis. However, we stress that longer observational durations are critical to improve the repeatability of behavioural studies, since the alternative is inconsistency over short observation times. The latter is more damaging to science because it introduces the likelihood of producing inconsistent or anomalous findings (Begley, 2013).

Literature review – Current status of behavioural testing globally

We performed a semi-quantitative review of studies describing behavioural tests with fish, to place our experimental findings in the context of the published literature. Advocates of behavioural toxicity testing with fish argued the need for standardisation over 30 years ago (Atchison et al., 1987), but our review clearly reveals that little progress has been made towards achieving consistent and comparable methodologies (Figure 4). Somewhat contrary to the views of Atchison et al. (1987), we support embracing the flexibility that behavioural toxicology testing can offer, including the ability to evaluate a range of effects related to mating, aggression, predator avoidance, anxiety and more. However, our experimental results demonstrate the importance of exploring and understanding basic baseline behavioural characteristics of focal test species. The present study provides a simple framework for evaluating how basic choices in experimental design might influence fish behaviour and demonstrates that this has relevance for toxicity testing. Importantly, if there is a ‘correct’ approach to acclimate and observe behaviours in fish, our review indicates that a large proportion of studies may not be achieving this.

Our review identified a wide range of test methodologies being applied in behavioural toxicity research, with many studies acclimating fish and acquiring data over what likely constitutes insufficient timeframes. A ‘one size fits all’ standardised test runs the risk of limiting the applicability of behavioural analysis towards aquatic toxicity testing, but there is a clear need to understand baseline behaviours and sources of variability to validate design choices and ensure robust science. Species differences will certainly exist, and our study with mosquitofish therefore only serves as an example and starting point for improving the validity of behavioural toxicity tests. Similar evaluations should be performed for other common test species. We reiterate that short acclimation timeframes may be well suited to anxiety tests where exploratory behaviour in a novel environment is the focus. Nevertheless, even in such cases it may be more meaningful and informative to assess behaviour until (control) fish reach a baseline behavioural state, and the timeframe required to reach the identified behavioural state could serve as an additional endpoint.

CONCLUSIONS

The quality of scientific research is under attack and the occurrence of irreproducible findings is a major point of criticism. As such, researchers across disciplines must be critical and comprehensive in their efforts to ensure the highest quality science. Behavioural toxicology testing is a rapidly growing field and, as such, is currently under scrutiny due to many seemingly haphazard studies. We used a basic approach to evaluate how basic choices in the design of behavioural tests might influence outcomes, and placed our findings in the context of what is currently the status quo in the published literature. Results demonstrate the importance of appropriate acclimation timeframes and observational durations when designing behavioural tests with fish. Specifically, acclimation for several hours may be necessary if the goal is to evaluate ‘normal’ baseline behaviours (arguable the most ecologically relevant starting point), and demonstrates how appropriate acclimation may effectively increase the robustness and validity of behavioural studies by increasing statistical power. Longer observational timeframes may be necessary to account for the high temporal variance that can exist for certain behavioural endpoints, and we further hypothesise that inappropriately short observation times may be a factor contributing to discrepancies commonly identified in the literature. It is our hope that these findings provokes critical thought, and stimulates discussion, regarding the appropriate application of behavioural tests in aquatic toxicology.

ACKNOWLEDGEMENTS

Funding was provided through a 2016 Griffith University Postdoctoral Fellowship (No. 219059) awarded to SDM. We extend our sincere thanks to Stewart Owen and Luigi Margiotta-Casaluci for providing thoughtful and critical feedback that improved the quality of the manuscript.

FIGURE CAPTIONS

Figure 1. Average velocity (mm/s) of mosquitofish (*Gambusia holbrooki*) recorded for 8hrs immediately after transference into novel behavioural arenas. Data was acquired at both a) 5min and b) 2hr intervals. Error bars represent standard deviation and were used to calculate achievable statistical power for each observational time interval. Letters represent groups that differ significantly, with α = 0.05.

Figure 2. Angular velocity (°/s) of mosquitofish (*Gambusia holbrooki*) recorded for 8hrs immediately after transference into novel behavioural arenas. Data was acquired at both a) 5min and b) 2hr intervals. Error bars represent standard deviation and were used to calculate achievable statistical power for each observational time interval. Letters represent groups that differ significantly, with α = 0.05.

Figure 3. Maximum velocity (mm/s) of mosquitofish (*Gambusia holbrooki*) recorded for 8hrs immediately after transference into novel behavioural arenas. Data was acquired at both a) 5min and b) 2hr intervals. Error bars represent standard deviation and were used to calculate achievable statistical power for each observational time interval. Letters represent groups that differ significantly, with α = 0.05.

Figure 4. Summary of a) acclimation time to behavioural arenas and b) observational duration for acquiring data, from 165 published behavioural tests with fish. Lines represent mean ± 1SD and individual data points are shown. The number of cases falling outside the graphed region is indicated in brackets next to the appropriate study classification.

FIGURES

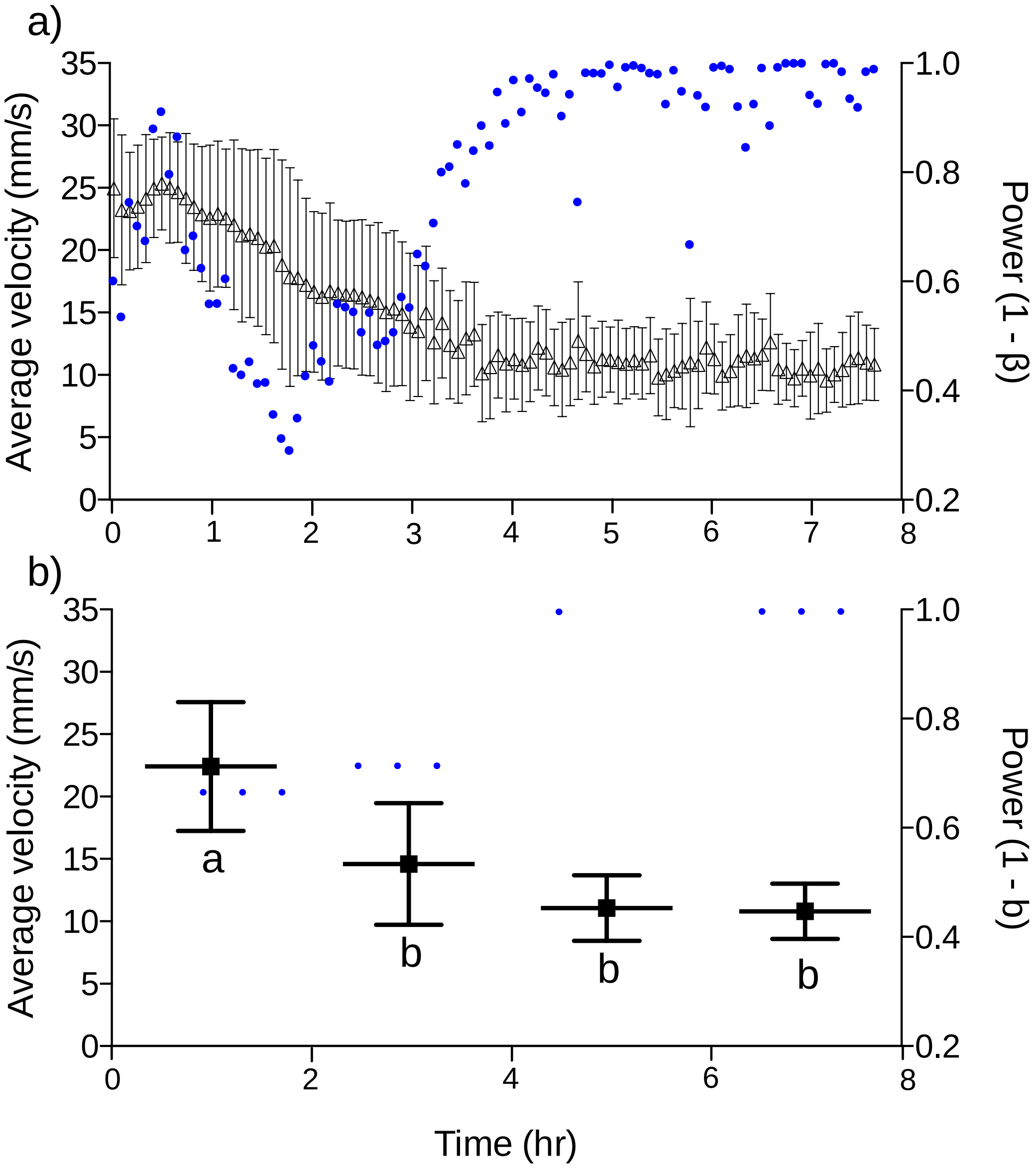
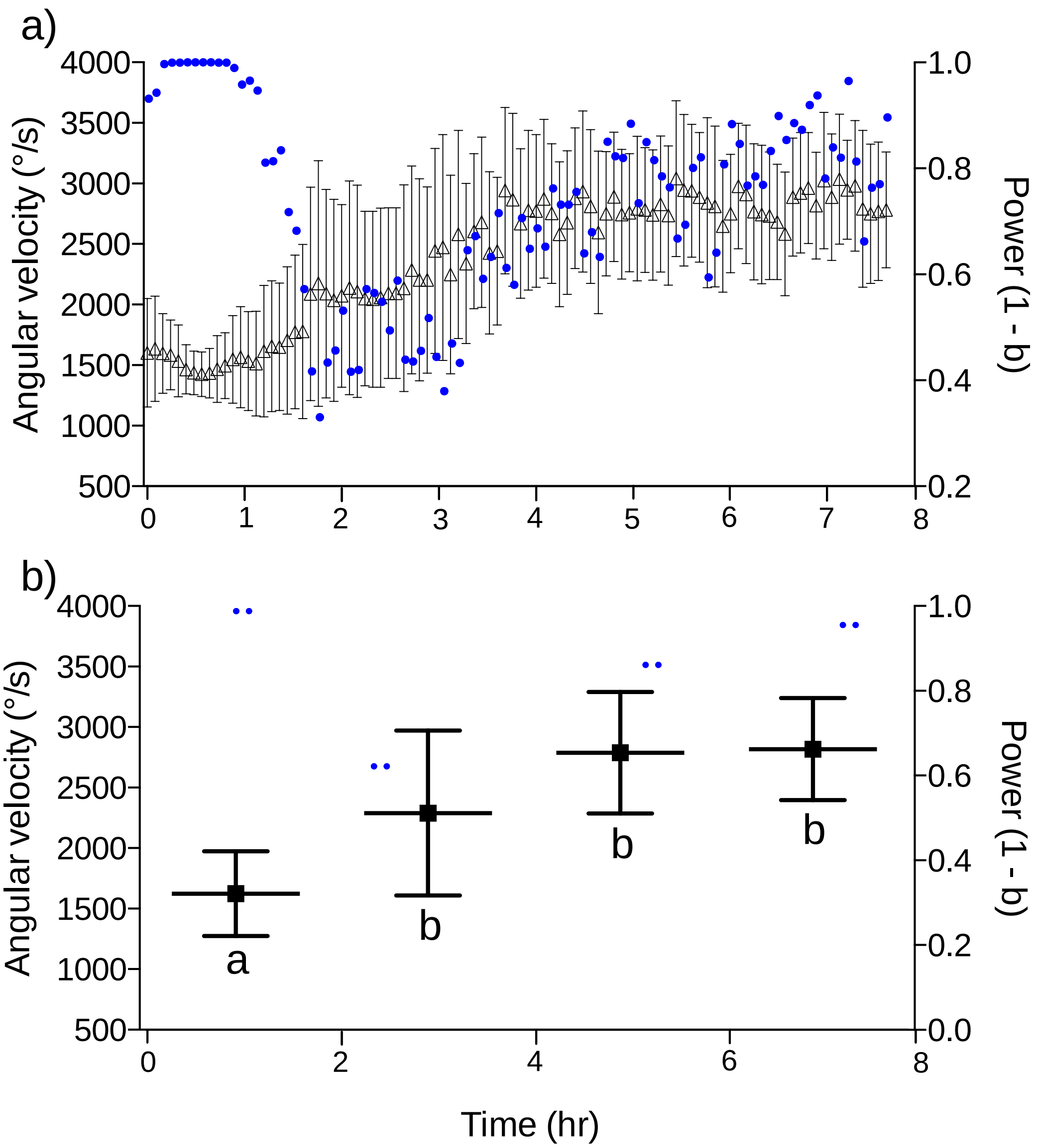


Figure 1.



473

474 Figure 2.

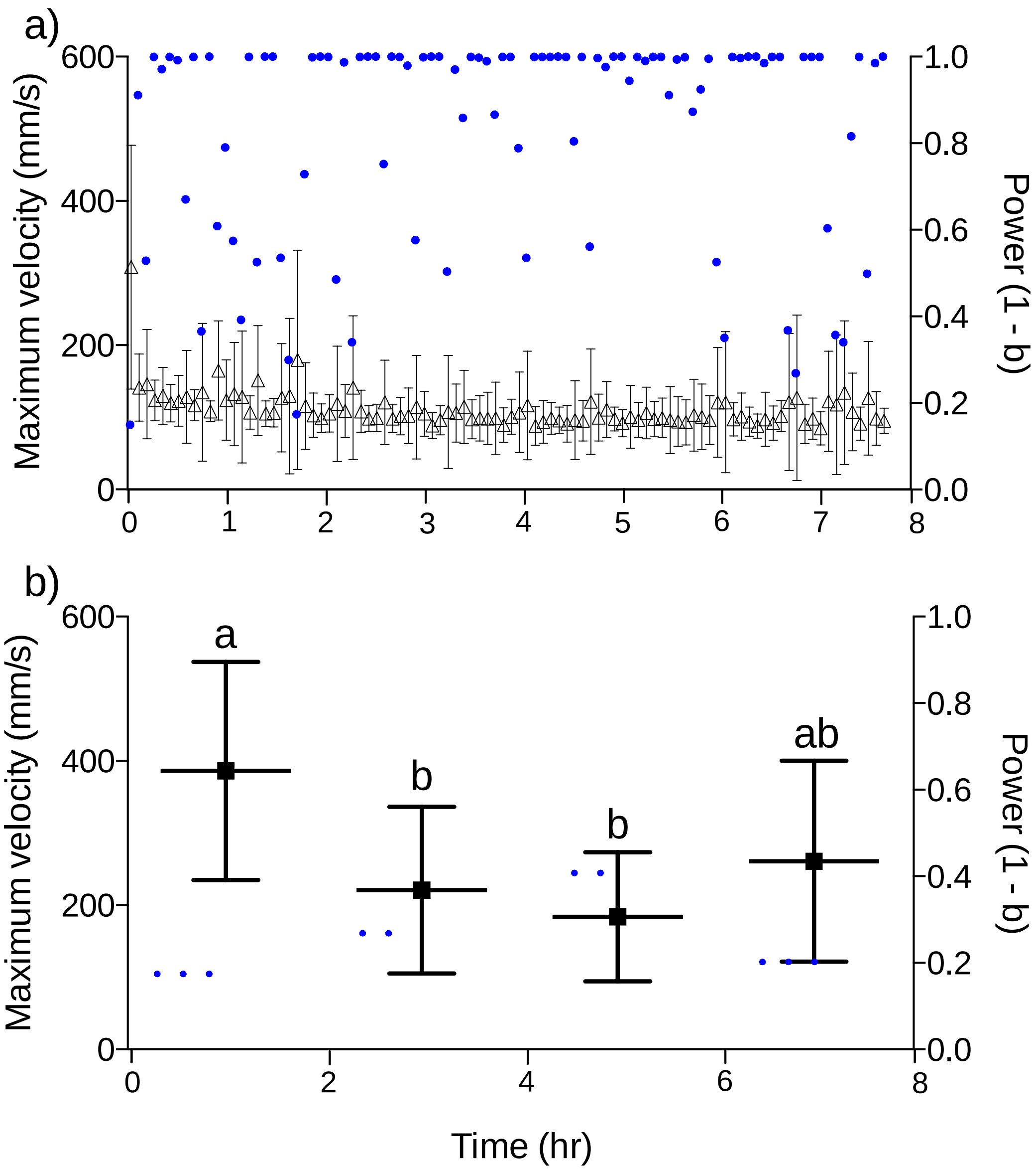

a)
b)
Maximum velocity (mm/s)
Power (1 - b)
0
200
400
600
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
8
a
b
b
ab
Time (hr)

Figure 3.

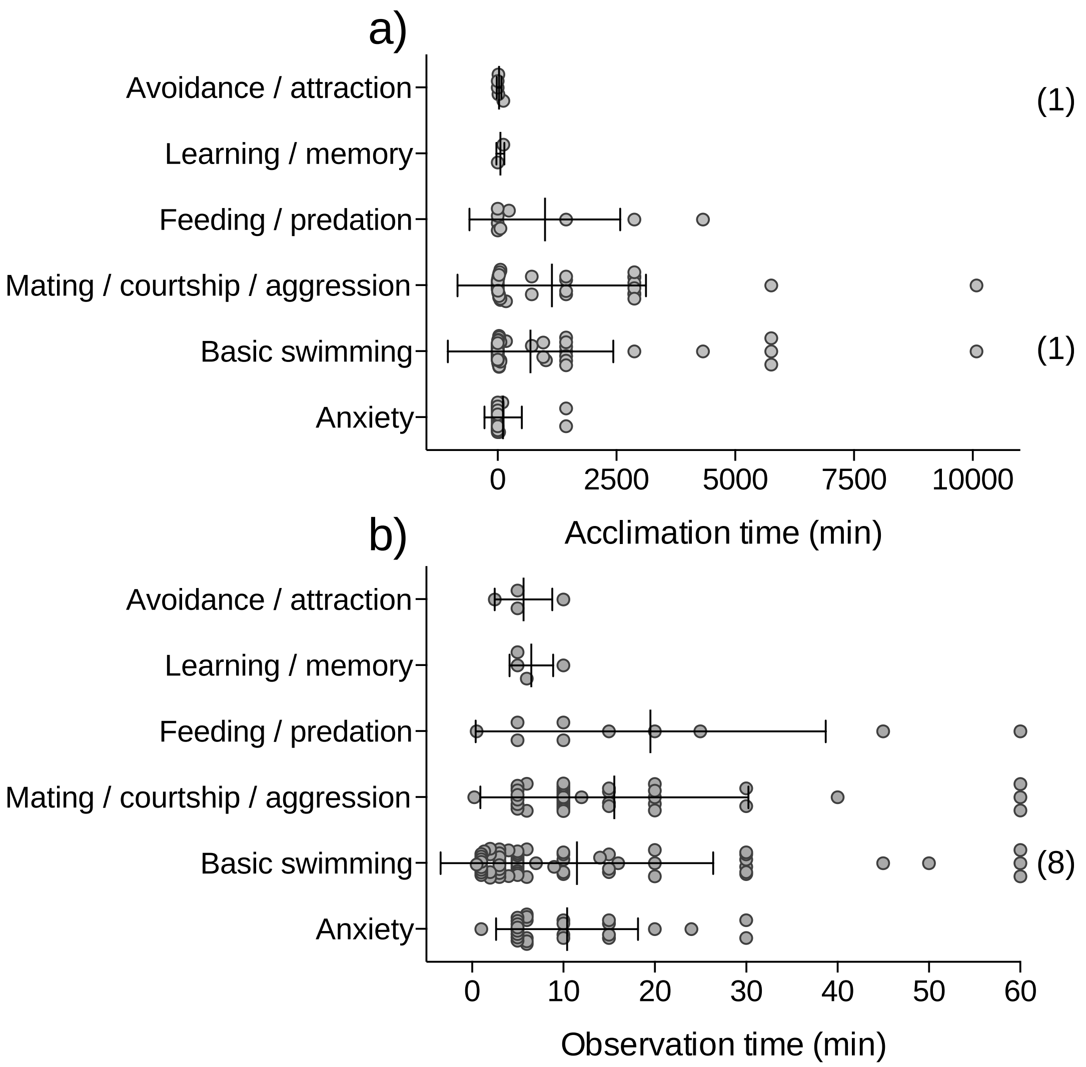

a)
Avoidance / attraction
Learning / memory
Feeding / predation
Mating / courtship / aggression
Basic swimming
Anxiety
(1)
(1)
0
2500
5000
7500
10000
Acclimation time (min)
b)
Avoidance / attraction
Learning / memory
Feeding / predation
Mating / courtship / aggression
Basic swimming
Anxiety
(8)
0
10
20
30
40
50
60
Observation time (min)

Figure 4.

References:

ASTM, 2014. Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians. ASTM International, West Conshohocken, PA. doi:10.1520/E0729-96R14

Atchison, G.J., Henry, M.G., Sandheinrich, M.B., 1987. Effects of metals on fish behavior: a review. Environmental Biology of Fishes. 18, 11-25.

Bae, M.-J., Park, Y.-S., 2014. Biological early warning system based on the responses of aquatic organisms to disturbances: A review. The Science of the Total Environment 466-467, 635–649. doi:10.1016/j.scitotenv.2013.07.075

Baker, M., 2016. Is there a reproducibility crisis? Nature 533, 452–454.

Begley, C.G., 2013. Six red flags for suspect work. Nature 497, 433–434. doi:10.1038/497433a

Benhaïm, D., Péan, S., Lucas, G., Blanc, N., Chatain, B., Bégout, M.-L., 2012. Early life behavioural differences in wild caught and domesticated sea bass (*Dicentrarchus labrax*). Applied Animal Behaviour Science 141, 79–90. doi:10.1016/j.applanim.2012.07.002

Blaser, R., Gerlai, R., 2006. Behavioral phenotyping in zebrafish: Comparison of three behavioral quantification methods. Behavior Research Methods, Instruments & Computers 38, 456–469.

Casterlin, M.E., Reynolds, W.W., 1977. Aspects of habitat selection in the mosquitofish *Gambusia affinis*. Hydrobiologia. doi:10.2307/2423931?ref=search-gateway:213ec55a7fa15d0f328961bf42b9be8c

Chen, W.M., Tabata, M., 2002. Individual rainbow trout can learn and anticipate multiple daily feeding times. Journal of Fish Biology 61, 1410–1422. doi:10.1006/jfbi.2002.2152

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, Instruments & Computers 39, 175–191.

Forstmeier, W., Wagenmakers, E.-J., Parker, T.H., 2016. Detecting and avoiding likely false-positive findings - a practical guide. Biological Reviews 1–28. doi:10.1111/brv.12315

Gee, P., Stephenson, D., Wright, D.E., 1994. Temporal discrimination-learning of operant feeding in goldfish (*Carassius auratus*). Journal of the Experimental Analysis of Behavior 62, 1–13.

Gerlai, R., 2003. Zebra fish: An uncharted behavior genetic model. Behavior Genetics 33, 461–468. doi:10.2307/20486953?ref=search-gateway:ec3cb6e920cdfaf2a9dbca04867189fb

Hansen, M.J., Schaerf, T.M., Ward, A.J.W., 2015. The effect of hunger on the exploratory behaviour of shoals of mosquitofish *Gambusia holbrooki*. Behaviour 152, 1659–1677. doi:10.1163/1568539X-00003298

Huerta, B., Margiotta-Casaluci, L., Rodríguez-Mozaz, S., Scholze, M., Winter, M.J., Barceló, D., Sumpter, J.P., 2016. Anti-anxiety drugs and fish behavior: Establishing the link between internal concentrations of oxazepam and behavioral effects. Environmental Toxicology and Chemistry 35, 2782–2790. doi:10.1002/etc.3448

Jakka, N.M., Gnaneshwar Rao, T., Venkateswara Rao, J., 2008. Locomotor behavioral response of mosquitofish (*Gambusia affinis*) to subacute mercury stress monitored by video tracking system. Drug and Chemical Toxicology 30, 383–397. doi:10.1080/01480540701522551

Kane, A.S., Salierno, J.D., 2005. Fish models in behavioral toxicology: Automated techniques, updates and perspectives. In: Ostrander, G.K., (Ed). Methods in Aquatic Toxicology, Volume 2 (Chapter 32). Lewis Publishers, Boca Raton, FL.

Kim, Y.-H., Wardle, C.S., 2005. Basic modelling of fish behaviour in a towed trawl based on chaos in decision-making. Fisheries Research 73, 217–229. doi:10.1016/j.fishres.2004.12.003

Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355, 584-585. doi:10.1126/science.aam5409

Magellan, K., Barral-Fraga, L., Rovira, M., Srean, P., Urrea, G., García-Berthou, E., Guasch, H., 2014. Behavioural and physical effects of arsenic exposure in fish are aggravated by aquatic algae. Aquatic Toxicology 156, 116–124. doi:10.1016/j.aquatox.2014.08.006

Marentette, J.R., Tong, S., Wang, G., Sopinka, N.M., Taves, M.D., Koops, M.A., Balshine, S., 2012. Behavior as biomarker? Laboratory versus field movement in round goby (*Neogobius melanostomus*) from highly contaminated habitats. Ecotoxicology 21, 1003–1012. doi:10.1007/s10646-012-0854-y

Maximino, C., de Brito, T.M., da Silva Batista, A.W., Herculano, A.M., Morato, S., Gouveia, A., Jr, 2010. Measuring anxiety in zebrafish: A critical review. Behavioural Brain Research 214, 157–171. doi:10.1016/j.bbr.2010.05.031

McCallum, E.S., Bose, A.P.H., Warriner, T.R., Balshine, S., 2017. An evaluation of behavioural endpoints: The pharmaceutical pollutant fluoxetine decreases aggression across multiple contexts in round goby (*Neogobius melanostomus*). Chemosphere 1–35. doi:10.1016/j.chemosphere.2017.02.059

Melvin, S.D., 2017. Effect of antidepressants on circadian rhythms in fish: Insights and implications regarding the design of behavioural toxicity tests. Aquatic Toxicology 182, 20–30. doi:10.1016/j.aquatox.2016.11.007

Melvin, S.D., Buck, D.R., Fabbro, L.D., 2016. Diurnal activity patterns as a sensitive behavioural outcome in fish: Effect of short-term exposure to treated sewage and a sub-lethal PPCP mixture. Journal of Applied Toxicology 36, 1173-1182.

Melvin, S.D., Wilson, S.P., 2013. The utility of behavioral studies for aquatic toxicology testing: A meta-analysis. Chemosphere 93, 2217–2223. doi:10.1016/j.chemosphere.2013.07.036

OECD, 1992. OECD Guidelines for the Testing of Chemicals, Section 2: Effects on Biotic Systems. OECD Publishing. doi:10.1787/9789264069961-en

Pang, X., Cao, Z.-D., Peng, J.-L., Fu, S.-J., 2010. The effects of feeding on the swimming performance and metabolic response of juvenile southern catfish, Silurus meridionalis, acclimated at different temperatures. Comparative Biochemistry and Physiology, Part A 155, 253–258. doi:10.1016/j.cbpa.2009.11.005

Parker, M.O., 2015. Adult vertebrate behavioural aquatic toxicology: Reliability and validity. Aquatic Toxicology 170, 323-329. doi:10.1016/j.aquatox.2015.09.001

Priyadarshana, T., Asaeda, T., Manatunge, J., 2006. Hunger-induced foraging behavior of two cyprinid fish: *Pseudorasbora parva* and *Rasbora daniconius*. Hydrobiologia 568, 341–352. doi:10.1007/s10750-006-0201-5

Pyke, G.H., 2008. Plague Minnow or mosquito Fish? A review of the biology and impacts of introduced *Gambusia* species. Annual Review of Ecology, Evolution, and Systematics 39, 171–191. doi:10.1146/annurev.ecolsys.39.110707.173451

Pyke, G.H., 2005. A review of the biology of *Gambusia affinis* and *G. holbrooki*. Reviews in Fish Biology and Fisheries 15, 339–365. doi:10.1007/s11160-006-6394-x

Pyle, G., Ford, A.T., 2017. Behaviour revised: Contaminant effects on aquatic animal behaviour. Aquatic Toxicology 182, 226–228. doi:10.1016/j.aquatox.2016.11.008

Saaristo, M., Myers, J., Jacques-Hamilton, R., Allinson, M., Yamamoto, A., Allinson, G., Pettigrove, V., Wong, B.B.M., 2014. Altered reproductive behaviours in male mosquitofish living downstream from a sewage treatment plant. Aquatic Toxicology 149, 58–64. doi:10.1016/j.aquatox.2014.02.001

Sismeiro-Vivas, J., Abrantes, N., Pereira, J.L., Castro, B.B., Gonçalves, F., 2007. Short-term effects of Quirlan® (chlorfenvinphos) on the behavior and acetylcholinesterase activity of *Gambusia holbrooki*. Environmental Toxicology 22, 194–202. doi:10.1002/tox.20256

Sumpter, J.P., Donnachie, R.L., Johnson, A.C., 2014. The apparently very variable potency of the anti-depressant fluoxetine. Aquatic Toxicology 151, 57–60. doi:10.1016/j.aquatox.2013.12.010

Thompson, R.R.J., Paul, E.S., Radford, A.N., Purser, J., Mendl, M., 2016. Routine handling methods affect behaviour of three-spined sticklebacks in a novel test of anxiety. Behavioural Brain Research 306, 26–35. doi:10.1016/j.bbr.2016.03.015

USEPA, 1994. Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms (No. EPA-600-4-91-002). United States Environmental Protection Agency.