Pharmacology, Biochemistry and Behavior 178 (2019) 30–38

Contents lists available at ScienceDirect

Pharmacology, Biochemistry and Behavior

journal homepage: www.elsevier.com/locate/pharmbiochembeh

Review

Reproducibility and replicability in zebrafish behavioral neuroscience research

Robert Gerlai*

Department of Psychology, University of Toronto Mississauga, Canada

ARTICLE INFO

Keywords:
Reproducibility
Replicability
Zebrafish
Behavior
Brain function
Standardization
Husbandry

ABSTRACT

Reproducibility and replicability are fundamentally important aspects of the scientific method. From time to time the discussion about whether scientific findings are replicable enough flares up. In fact, some recent publications claim we are witnessing a replication crisis. This is a particularly important problem in laboratory organisms that are relatively new, i.e., for which only limited amount of information is available, and for which only a limited number of methods have been developed. The zebrafish is a relative newcomer in behavioral neuroscience. This review considers four distinct reasons as possibly underlying reproducibility issues in behavioral neuroscience studies using the zebrafish. One, publication bias for positive results. Two, statistical issues that surround the question of how to address type 1 and type 2 errors, and how to make statistical inference. Three, inappropriate control of factors that are known to potentially influence results. And four, methodological issues stemming from insufficient understanding of factors that may influence experimental results. The review will mainly focus on experimental issues and solutions, i.e. the latter two reasons listed above. It is not intended to be comprehensive, and its examples are drawn mainly from the author's own studies and experience with zebrafish. Nevertheless, most issues discussed are not unique to his laboratory, to the zebrafish, or even to behavioral neuroscience.

1. Introduction

A fundamental aspect of the scientific method that concerns experimental or empirical research is that it relies upon testable hypotheses. Testing hypotheses, or simply generating data upon which future hypotheses will be based, is expected to be achieved in an objective manner, i.e., in a way that allows replication of results. In other words, the research method, procedure, technique, or experiment is expected to yield fundamentally the same answer irrespective of when and where the study was performed and who conducted it (Fisher, 1935). Replicability is a core requirement of the scientific method. In recent years, replicability is distinguished from reproducibility (see e.g. Leek and Peng, 2015). Reproducibility is defined as the ability of obtaining the same results repeatedly when using identical methods. Replicability is similarly defined, except the requirement of sameness of methods and results is made somewhat less stringent. Briefly, replicability refers to being able to reach fundamentally similar conclusions even when slight variations in methodology exist. In the current paper, I use the term reproducibility interchangeably with replicability. That is, I do not sharply distinguish these two terms, mainly because in real-life empirical research their definitions actually overlap.

2. The replicability "crisis"

In recent years, several scientists rang the alarm bell and claimed that science was experiencing replicability and reproducibility crises (Baker, 2016a, 2016b; Peng, 2015; Savalei and Dunn, 2015). According to a recent survey conducted on behalf of the magazine Nature, for example, 50% of scientists have experienced issues with reproducibility of their own results, and 70% of them had trouble reproducing findings obtained by others (Baker, 2016a). And the problem was not limited to a particular field of science, but was found in chemistry, physics and engineering, earth and environmental sciences, biology, medicine, as well as in other scientific fields. From cancer prognosis (Baggerly and Coombes, 2009), through brain imaging (Vul et al., 2009), to global economic policies (Herndon et al., 2014) results have been found unreplicable. The problems are deemed so acute that entire journal issues have been published on this topic. A recent example is the ILAR Journal (Vol 58, issue 1, July 2017, pages 115–128), whose special issue was titled "Bridging the gap between reproducibility and translation, data resources and approaches". Another example is the very issue of the journal Pharmacology, Biochemistry and Behavior in which this paper is published. In fact, an entire conference was organized recently to

* Rm CCT4004, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada.
E-mail address: robert.gerlai@utoronto.ca.

https://doi.org/10.1016/j.pbb.2018.02.005
Received 15 October 2017; Received in revised form 6 February 2018; Accepted 22 February 2018
Available online 23 February 2018

discuss these issues (https://en-exact-sciences.tau.ac.il/sites/exactsci_en.tau.ac.il/files/media_server/Exact_Science/math/events/Replicability_Reproducibility_of_Discoveries.pdf), and a comprehensive review based upon this conference expressing the opinion of the behavioral neuroscience and behavioral genetics community's views has just been published (Kafkafi et al., 2018). Furthermore, major funding agencies, including the NIH, have also taken notice, and have been implementing guidelines for scientists on how to enhance rigor and reproducibility in grant applications (https://grants.nih.gov/reproducibility/index.htm).

But is the problem of diminished reproducibility and replicability new? Are we, scientists, doing our job less well than before? I suspect the answer to these questions is no. I suspect the main reason for the "crisis" is that we communicate more efficiently, generate more data and more complex data, publish more, and have access to a wider spectrum of findings and studies. Thus, we have better means of discovering and of discussing these problems. That is, the replicability crisis is not new. This, of course, does not diminish the importance of the issue. Reproducibility and replicability are fundamentally important, and should be considered. In order to eliminate or reduce the problem with replication of results, one must first understand its root causes.

3. The root causes of reproducibility and replicability issues

I consider four somewhat overlapping, but still distinct causes that may lead to reduced reproducibility and replicability. One, publication bias for positive results. Two, statistical issues that surround the question of how to address type 1 and type 2 errors and the asymmetry of statistical decision making, or, in general, the way we make statistical inferences. Three, inappropriate control of factors that are known to potentially influence results. This latter point partially overlaps with the second one, as it includes questions about inappropriate randomization or sampling bias as well as issues associated with individual differences among study subjects, an increasingly fashionable topic not discussed in this review. And four, methodological issues stemming from insufficient understanding of factors that may influence experimental results. In the analysis of brain function and behavior, the focus of this review, the latter two points represent a particularly vexing problem. The brain is perhaps the most complex organ in the organism, and thus it is potentially responsive to a large number of environmental and genetic factors, some of which may not be well controlled, and others may not even be known. In this review, I will focus mainly on experimental issues affecting reproducibility and replicability in zebrafish behavioral neuroscience, i.e. points three and four above, but first I will briefly discuss the other causes, i.e. those resulting from publication bias and statistical decision making.

4. Publication bias: is there a solution?

A large proportion of published studies report positive results, i.e. results that show significant differences among groups, significant effects of treatment or of the manipulation used in the study. Negative results, i.e. results that show lack of such significance, are rarely reported. Where does this bias come from? Perhaps it is basic human nature that we are interested in finding new things. Novelty preference is almost universal in the animal kingdom to which our own species also belongs (Thompson et al., 1991; Bardo et al., 1989; Gerlai, 1998; Gómez-Laplaza and Gerlai, 2010). Similarly, most of us like being correct. Thus, when a scientist formulates a working hypothesis, he or she is usually more pleased when it is proven correct. But even when the scientist is completely unbiased, which most of us claim we are, there are forces that may bias us towards publishing positive, i.e. significant, results. We all know the slogan "publish or perish". Funding agencies, tenure promotion committees, performance evaluation boards in Academia and Industry all look at number and quality of studies completed and/or published. But a large majority of peer reviewed scientific journals and their editors prefer studies with significant results. A case in point was a recent publication from my laboratory (Seguin et al., 2016), in which we found a lack of significant effect of embryonic alcohol exposure on fear and anxiety-related behavioral responses in zebrafish. Although this study did get published at the end, the initial response from referees and editors was what we expected: the study only presents negative findings and thus it is not very interesting. Only after we further emphasized the point that the negative results presented in the study are crucial for the interpretation of prior findings showing significant (positive) effects of embryonic alcohol treatment on another phenotype, social behavioral responses, did the journal finally accept the paper. What has this got to do with replication? Studies aimed at showing reproducibility and replicability are inherently about "old" stuff, findings that have been obtained before. They are also about showing lack of difference between the past and current results, an essentially negative type finding. Old and negative findings are not the favorite of referees and editors of peer reviewed scientific journals.

Could this publication bias be reversed? And if yes, how? My own impression is that it will be difficult to go against the current tradition. Nevertheless, science literature is becoming more and more open, i.e., the number of peer reviewed journals that allow, or in some cases demand, open access to data is increasing, so is the number of publicly searchable electronic data bases that often allow deposition and second party analysis of raw data. These changes facilitate openness, re-evaluation of results and more objective discussion of findings, but also importantly, publication of negative results. Briefly, the importance of publishing negative results is increasingly better appreciated (Baxter and Burwell, 2017).

5. Statistics: what does significance or the lack of it mean?

The second main reason for lack of ability to replicate findings lies deeply in the way we make statistical inferences. Although all university programs designed for students of behavioral neuroscience and related fields, I know of, include courses on statistics, and although all researchers of behavioral neuroscience routinely utilize a variety of methods of statistics, it appears that occasionally basic concepts of this field are forgotten. Let me briefly examine one of the most basic of all, the concept of significance itself.

Most scientists accept the classical threshold of significance being $p < 0.05$, which means that the probability of the experimental manipulation not having an effect, or of groups of animals not differing from each other, is less than five percent. If the probability of the null hypothesis (lack of effect or lack of difference) is this small or smaller, we accept the result as being significant. But what happens when the result is non-significant? How do we interpret that? For example, if the p value we obtain using a statistical test equals, say, 0.09, i.e. proving lack of significance. Lack of significance demonstrated by this p value would be interpreted quite often as proof that the treatment had no effect, or that the groups of animals did not differ from each other. But, is this interpretation correct? $p = 0.09$ means that the probability of the treatment having an effect, or the probability of the groups of animals differing, is 91%. In other words, statistical inference is asymmetrical. It is designed to detect presence of effect or difference with high certainty, but it is not designed to prove absence of difference or absence of effect. A "significant" absence of effect would require an equally stringent p threshold, i.e. $p > 0.95$. And there lies one of the issues with our ability to replicate negative findings.

But why cannot we replicate positive findings, i.e. those results that did turn out to reach the conventional level of significance? There may be several reasons for this inability. Some of these relate to experimental design and sampling methods. Others are more fundamental to our statistical methods. I deal, briefly, with the latter first.

Experiment-wise false-discovery rate is one issue that has been discussed in the literature (Ioannidis, 2005). This issue is known as

type-1 error in statistics, i.e., accepting significance of results when in fact they are not significant. False positive results may be obtained in a study if the study utilizes many statistical comparisons, each with the traditional $p = 0.05$ threshold for accepting significance. For example, if a study has a large number of groups to compare in a pair-wise manner, or if it uses a large number of variables along which small number of groups are compared, the cumulative false discovery rate will be n x p, where n is the number of comparisons and p is the significance threshold value. There are two simple statistical methods by which one can minimize this error. One, in case of multiple group comparisons, one needs to use post-hoc multiple-comparison procedures, e.g. the Tukey Honestly Significant Difference Test, or the Holm-Bonferroni correction (Holm, 1979), methods that minimize type-1 error without inflating type-2 error (the latter concerns claiming lack of significance when in fact the result was significant). Two, in case of multiple variables used, one can utilize a multivariate method, e.g. MANOVA, or correlation coefficients-based reduction method, including Factor analysis. However, all the above methods are expected to be appropriate when the number of groups or the number of variables are not very large, which is most often the case in behavioral neuroscience studies, but not when one deals with hundreds or thousands of comparisons, such as the case, for example, in DNA microarray or other large scale systematic analyses (Sham and Purcell, 2014). An alternative to significance level correction or multivariate statistical methods, others have suggested, is to abandon the use of $p = 0.05$ threshold requirement, and replace it with the use of confidence interval (Savalei and Dunn, 2015). Last, some even advocate complete abandonment of the black or white statistical hypothesis testing inherent in classical statistical methods (Szücs and Ioannidis, 2017). These authors suggest that the dichotomous decision making associated with null hypothesis significance testing is partially responsible for the replication crisis seen in several fields of neuroscience including behavioral neuroscience. They recommend that instead of relying on the all or nothing hypothesis tests, other types of inferential methods, e.g. Bayesian statistical approaches (also see Wellek, 2017), may be employed. They recommend that the use of classical null hypothesis-based statistical inference must be properly justified by pre-study power calculations as well as estimation of effect sizes.

Although these statistical considerations are important, one must not forget that irrespective of what statistical approach one uses, the validity (and thus reproducibility and replicability) of one's results will depend upon the quality of data collected, i.e. the manner in which data were obtained. For example, proper experimental design, including appropriate randomization and sampling methods and proper sample sizes must be employed. While these basic principles of experimental design and procedure are well known in theory, in practice they may not be implemented for two main reasons. One, the factors that potentially influence the outcome of the experiment are not controlled, or two, perhaps not even known. Randomization and proper sampling can only be achieved if we understand along what factors we must properly randomize, and which subjects may be differentially affected by such factors. These are basic empirical questions that are still being addressed in the field of neuroscience, neurobehavioral genetics and psychopharmacology of the mouse, despite that this research organism is perhaps the most well studied among vertebrate laboratory species. The problems are even more severe in case of the zebrafish, which is a relative newcomer in these fields.

6. Factors we know may influence brain function and behavior: an example of anxiety in rodents

One reason why findings within a laboratory or across laboratories may not be reproducible or replicable is that not all factors that potentially influence the outcome of the experiment are properly controlled. To experimentally test the question how well one can control such factors, Crabbe et al. (1999) conducted a coordinated set of experiments with mice that involved three geographically separated laboratories. In this study, the authors investigated reproducibility in the context of genotype x environment interaction. They set up their behavioral experiments exactly the same manner, using the same set of apparati, test procedures and the same inbred and genetically modified mouse strains, and conducted their studies at the same time, but across the three different laboratories. They found significant laboratory-specific differences in the absolute values of the behaviors they measured. But, even more importantly, in some behavioral tests, the relative rank order of their mouse strains was found laboratory dependent. That is, they found genotype x environment interaction. These results were widely discussed by the scientific community, and some even made the general conclusion that conducting behavioral analysis appropriately, i.e. in a replicable and reproducible manner, is just not possible. However, closer examination of the data and procedures of the Crabbe et al. (1999) study yielded some interesting insights that did not support this harsh view. Some behavioral tests produced a remarkably similar strain rank order across all three laboratories. These were the tests that required minimal handling of the animals or in which the assumed chance of induction of fear or anxiety was lower. Tests in which the latter factors were potential driving force behind the behavioral responses provided more inconsistent findings. The authors of this study made the excellent observation, for example, that specific experimenters performing the testing of their mice were unique to each laboratory, and could have had a laboratory-specific effect on the behavior of the experimental animals. They stated, for example, that the experimenter in one of the laboratories was highly allergic to mice, and performed all tests while wearing a respirator. An experimenter isolated from airborne allergens is also isolated from the mice in terms of olfactory communication. Experimenters in the other laboratories were not. Mice, as several other rodent species too, are highly sensitive to olfactory cues, and tend to respond with anxiety-like or fear behaviors to the smell of large mammals, potential predators. Thus, the Crabbe et al. (1999) study exemplified a rather difficult experimental issue, controlling or standardizing the effect of the human experimenter.

Another interesting, and perhaps underappreciated, aspect of the Crabbe et al. (1999) study concerns the behavioral analysis of the null mutant mouse line they studied. Crabbe et al. detected laboratory specific differences in these mice, a finding that many interpreted again as resulting from laboratory specific unknown environmental factors. However, there was another possibility: The genetic background of the studied null mutant mice was not identical across the three laboratories. Without delving into too much detail, the genetic background of these mice suffered from the same issue most other classical knock out mice had: hybrid genetic origin and hitch-hiking alleles around the target locus (Gerlai, 1996). And because the mouse lines used in the three different laboratories were not obtained from the same source, the possibility existed that their genetic background was not the same across the three laboratories. Briefly, the observed differences could potentially be due to not laboratory environment or procedure specific effects, but rather to genetic factors. Zebrafish studies also suffer from these two above exemplified issues: experimental error arising from the difficulty to standardize or eliminate human handling, and potentially uncontrolled genetic effects. I discuss the former first.

7. Human handling: the issue of fear and anxiety in zebrafish research

I list this problem under the category of factors that are known to influence our results. But this does not mean we have good solutions for it. In most studies, the experimental zebrafish have to be removed from their home tank, and have to be placed in an experimental apparatus. How the test fish is caught and transported to the test tank, can make a huge difference in how it will behave during the test. Zebrafish, unlike rodents, have not gone through domestication, intentional or unintentional artificial selection that would make them better adapted to

human handling. Also, when zebrafish are removed from their home tank, they experience a short period of hypoxia as well as restraint stress: they cannot properly breathe and cannot properly move in the net. Last, zebrafish are small but very fast little creatures, and thus catching them is not always an easy task. Why does this matter from the perspectives of reproducibility and replicability? Because depending on how fast and how delicate the experimenter is with zebrafish, and what method he/she uses for catching and transporting the experimental fish, he/she can introduce large differences in the level of fear and anxiety in the test fish.

There are two possible ways one may be able to address this source of error variation. One is that the behavioral test may be preceded by habituation sessions (Sison and Gerlai, 2011). Multiple exposures to human handling without any deleterious consequences may allow the zebrafish to habituate to the handling procedure. However, in our own experience, such habituation trials may be counterproductive. If not conducted with utmost care, they can achieve the opposite, i.e. they can lead to sensitization, enhanced fear and anxiety responses to human handling in zebrafish. This is because handling of zebrafish may represent too strong an aversive stimulation for this species. The second way one could avoid handling induced anxiety and fear in zebrafish is to not handle the fish at all. To avoid handling, the fish may be tested in its home-tank, an idea that has been implemented for the mouse (de Visser et al., 2006), but not yet for the zebrafish.

The above discussed factors, i.e., the genetic background, and environmental factors that affect fear and anxiety, are known to influence a variety of behavioral responses in zebrafish (Gerlai, 2010; Luca and Gerlai, 2012; Ahmed et al., 2012; Mahabir et al., 2013; Gerlai et al., 2008). In addition to these factors, there may be many others that have not yet been explored, or whose potential effects on brain function and behavior we do not even suspect yet. Unfortunately, given the relative newcomer status of the zebrafish in behavioral neuroscience, there may be many such factors, and this requires future investigation. But before I return to these yet to be characterized environmental factors, I would like to focus first on another important potential confound, often ignored or misunderstood in zebrafish research, the genetic background.

8. The genome as confound in zebrafish research

Unlike in case of the laboratory mouse, zebrafish researchers do not have a large selection of genetically well-defined inbred strains. Standard strains of zebrafish do exist, but the proportion of homozygous loci in these strains rarely exceeds 80% (Johnson and Zon, 1999). Briefly, inbred zebrafish strains are lacking. AB is one of the most frequently used zebrafish strains, one which has been bred in several laboratories all around the Globe. But because the strain is not fully inbred, due to random genetic drift, subpopulations of AB strains are expected to differ genetically, a problem ignored in zebrafish research, but one which can easily lead to lack of replicability. Others utilize so called “wild-type” populations, obtained from pet-stores. This is, of course, not a problem in itself. For example, we established a zebrafish population in our own zebrafish facility at the University of Toronto Mississauga starting the breeding from zebrafish obtained from a local pet store (Big Al’s Aquarium Warehouse, Mississauga). Our rationale was that such a population is likely to possess large genetic variability across individuals and large percentage of heterozygosity within each individual (e.g. Miller and Gerlai, 2007; Bass and Gerlai, 2008; Speedie and Gerlai, 2008; Al-Imari and Gerlai, 2008). As such, this population, unlike inbred strains, may not be genetically unique, and, we argued, should better represent species-specific features of the “prototypical” zebrafish. The problem, however, is that when one studies pet-store fish, the actual genetic make-up of these individuals can vary from one location to the other and from study to study, further contributing to the issue of lack of replicability. Furthermore, some researchers use the very fish they purchased from pet stores (as opposed to breeding them and establishing a laboratory population for their study). These pet store fish thus are potentially influenced by a large number of unknown and uncontrolled environmental factors that affected them at the commercial breeding facility and pet store before their transport to the laboratory.

Another issue, still concerning genetics, is the occasional mixing and misunderstanding of the meaning of genetic marker and genetic background. A study conducted by Dlugos and Rabin (2003) represents a case in point. These authors compared three different groups of zebrafish for their behavioral responses to acute and chronic alcohol (ethanol) treatment, and reported what they called significant “strain differences”. Even the title of their study claims “model for genetic investigations”. However, the authors purchased the three groups of zebrafish from pet stores, and used these purchased fish for their studies. Thus, as explained above, they could not distinguish potential fish group specific differential environmental effects from genetic effects. But even more importantly, the claim that they were comparing different strains of zebrafish had no foundation. The authors called their “strains” WT (wild-type, zebrafish that have short wild type fin and wild type color and stripe pattern), LFS (long-fin striped, zebrafish whose fins are elongated but their body coloration and pattern is wild type), and BLF (blue long fin, zebrafish whose fins are elongated and whose body is uniformly bluish, lacking the wild type color and striped pattern). What is the issue with all this? The problem is that the phenotypical traits that the authors used to distinguish their “strains” are just genetic markers. These markers represent a very limited number of loci, perhaps as few as only two, and do not define the genome, the genetic background of the studied populations of fish. In fact, the three “strains” could have been different in alleles of only these two loci (one determining color-pattern and the other fin length), or the LFS could have differed from the BLF at thousands of loci but not from WT, or WT could have differed from BLF at a few hundred loci, etc. No one can tell. Perhaps even more importantly, an LFS fish obtained from one pet store could have an entirely different genetic background compared to an LFS fish obtained from another pet store. In fact, two LFS fish obtained from the same pet store at the same time could have entirely different genetic backgrounds. And there lies the issue of replicability. A superficial phenotypical feature, a genetic marker, does not define the genetic background of the population or individual in these fish. Briefly, it is crucial that one uses genetically well-defined strains, if one wants to make inferences about genetic effects.

But even if one uses genetically well-defined strains of zebrafish, one may have to face the fact that there may be numerous experimental procedure related or zebrafish maintenance-dependent environmental factors whose effects we do not understand, or even worse, do not even know of yet. The subsequent section focuses on the former, experimental factors we do not fully understand.

9. Experimental procedures in zebrafish research: do we know what we are doing?

Without trying to be too negative, below I list a few examples that show how many things can go wrong when we try to manipulate zebrafish. Some of these examples illustrate simple and correctable experimental errors, others discuss more vexing issues to which we do not yet have answers. But all of them exemplify how replicability may be compromised in a laboratory organism that is quite novel for behavioral neuroscience, the zebrafish. The first example concerns lack of replication of psychopharmacological effects of a well-known drug of abuse, nicotine.

Nicotine has been employed in rodent research and has been found to possess anxiolytic as well as memory altering properties. Given the evolutionary conservation of neurotransmitter receptors, including nicotinic acetylcholine receptors, nicotine (in the form of nicotine ditartrate) has been also employed using zebrafish with success. For example, Levin et al. have shown significant anxiolytic effects (Levin et al., 2007) as well as learning (Levin et al., 2006) and memory (Levin

and Chen, 2004) altering effects of this drug in zebrafish. We (Miller et al., 2012) have also found nicotine to be efficacious, this time in altering shoaling (group forming) behavior of zebrafish. Our findings made sense, we argued, because the main function of shoaling is believed to be predator-avoidance, and a drug that affects anxiety, like nicotine, should thus influence shoaling (Miller et al., 2012). These results appear to show good replicability, at least for the anxiolytic properties of nicotine in zebrafish. However, we encountered a major problem. The doses routinely employed in the above cited studies (50–800 mg/l nicotine ditartrate bath concentration administered via immersion of the fish) conducted by Levin et al. led to significant deleterious effects in my own laboratory. In fact, in our hand, bath concentration of as low as 12 mg/l of nicotine ditartrate already induced twitching, lethargy and even leading to death in some cases in our zebrafish. That is, despite that we employed the same administration method (immersion), same nicotine salt (nicotine ditartrate) and same length of immersion (3 min), the concentrations of nicotine that did not induce grossly deleterious effects in zebrafish were one to two orders of magnitude lower that those employed in the Levin et al. studies cited above. How could such discrepancies arise? We currently have no answer to this question. But notably, the age and strain origin of the experimental zebrafish were not disclosed in the Levin et al. studies. Genotype as well as the age of the fish may influence drug effects, but it is unlikely these factors could explain the huge dose effect differences. The source of nicotine ditartrate is also not disclosed in the Levin et al. studies, but again, one would hope that large dose effect differences should not arise because the drug is obtained from different vendors. Perhaps, the most reasonable explanation for the discrepancies is the possibility that the obtained drugs were stored differently and perhaps for different periods of time. In my laboratory, we used the drug soon after we obtained it from Fisher Scientific, but in other laboratories the drug may have been kept for prolonged period of time on the shelf. Although nicotine ditartrate is considered fairly stable, exposing it to higher temperatures, or oxidative conditions, may have reduced its potency.

Despite all precautions my students and I attempt to make, however, mistakes are also made in my laboratory. The following example concerns such a mistake, a simple one, but nevertheless one that led to irreproducible results in my laboratory when we wanted to study circadian rhythm in zebrafish. Contrary to common belief, fish sleep too. They show circadian activity pattern typical of diurnal species, high activity during subjective daytime and low activity or absence of activity during subjective night-time. We were interested in the potential circadian/sleep altering effects of embryonic alcohol exposure in zebrafish, and ran a pilot study in which we wanted to quantify activity patterns of adult zebrafish followed continuously for 48 h. Given that sleep, and in general typical diurnal circadian activity changes have been demonstrated in the zebrafish (for a review see e.g. Zhdanova, 2011), we expected the classic activity pattern. The pilot study included setting up a newly acquired infrared light source illuminating the fish with light invisible to them, and digital cameras equipped with new infrared filters and a video-tracking system that allowed us to quantify numerous parameters of the swim path of the experimental fish. The set up seemed to be well designed by a member of my laboratory, so she set out to record circadian activity of our fish. The surprize came when we analyzed the data: our fish remained active throughout the 48 h period, no sleep, only minimal activity pattern change between day and night. How was this possible? We went through numerous procedural details and at the end realized we had made a mistake. The student who set up the cameras needed to adjust the angle of the camera so that it properly viewed the fish, and she used the built-in view screen of the camera. The only way she could view this screen was to open it and turn it towards the outside. However, the viewscreen emitted light frequencies that were visible not only to her but also to the zebrafish. In effect, we were monitoring circadian activity under constant light.

The next example I illustrate here for lack of replicability due to experimental procedures exemplifies a problem that is somewhat more complex, and less easy to fix, than making a trivial mistake like the above one. It concerns the use of light-dark choice tasks for the analysis of anxiety. Light-dark test paradigms have been successfully utilized in rodent research. The nocturnal rodent has been found to avoid illuminated areas and prefer dark places, presumably because the latter offers protection from visually hunting predators. Irrespective of the ecological or adaptive aspect of this behavior, it has been well utilized for screening for anxiolytic compounds in psychopharmacology research in academia and industry alike. A similar reasoning prompted us to study how the diurnal zebrafish would behave. We conducted a light-dark study using adult zebrafish in which our subjects were offered a choice in a shuttle box with one side being well illuminated and having transparent glass walls and the other completely covered and dark. Also, we exposed some of our zebrafish acutely to low concentrations of ethanol, an anxiolytic drug at these doses, and compared their responses in the light-dark task to control fish. Our results demonstrated a significant preference for the well illuminated side of the apparatus in control fish (zebrafish is diurnal and we argued would avoid dark places where it cannot see predators), and this dark avoidance was reduced by acute alcohol treatment (Gerlai et al., 2000). These results made sense both in the broader context of what we knew about diurnal versus nocturnal species and also from the perspective of the natural habitat and sympatric predators of the zebrafish (Gerlai et al., 2000). However, our results were not always replicated. In fact, Maximino et al. (2010) found the opposite in zebrafish, dark preference, and named the behavior scototaxis (movement towards the dark). Subsequently, a number of studies have been published in which one or the opposite finding was supported. From the controversies the only main conclusion one could draw was that adopting a task from the rodent literature for the zebrafish was not as simple as it originally seemed to be. By now, however, some studies have started to disentangle the controversies and are starting to provide answers (e.g. Blaser and Peñalosa, 2011; Facciol et al., 2017). Briefly, it appears that the zebrafish does prefer well illuminated areas, but also prefer dark backgrounds to light backgrounds, and they also avoid cave like environments. The light-dark paradigm confused three independent factors: level of illumination, background/substrate shade, and openness of the environment.

Another example from my own laboratory on lack of reproducibility also shows how simple and trivial mistakes may lead to dramatic problems in the study of animal behavior. The example again concerns zebrafish shoaling. Zebrafish are highly social, and thus when placed singly into a test tank and shown a group of conspecifics, attempt to join the group and thus swim close to the stimulus. In one of our studies, however, we obtained a highly unexpected finding. Upon presentation of the computer animated images of zebrafish, the experimental zebrafish moved away from the stimulus presentation screen, a highly significant change, but one which was the exact opposite of what we saw before (Gerlai unpublished results). We examined and scrutinized every possible parameter of this study, rearing conditions, experimental procedures, everything we could think of. After a long and arduous examination of our methods, we discovered the root cause of the problem: the experimenter, who set up the image presentation, forgot to double check what the images looked like on the actual presentation screen. He used a new computer monitor whose aspect ratio differed from that of the computer that ran the image presentation software we developed in-house (Saverino and Gerlai, 2008). This software allows us to precisely set several parameters of image presentation, including the size, pattern, color, shape of the images, how many conspecific images are shown, how fast and where exactly they would be moving, how long they are shown, and when exactly they are turned on and off. This level of control allows us to present a stimulus that is consistent across all our studies, but it also allows us to test what aspects of the stimulus zebrafish responds to, and how modification of certain aspects may alter the behavior of our subjects. As it turned out, the images set up on the computer screen correctly, showed differently on the actual

presentation screen the experimental fish could see. These images appeared abnormally elongated and slim compared to normal looking zebrafish. The discovery of this error explained well why we could not reproduce our previous results. Again, we used this error to our advantage, and ran a study in which we systematically altered different aspects of the conspecific stimulus images (Saverino and Gerlai, 2008). For example, we presented images that were colored red or yellow, that wore vertical, as opposed to the natural horizontal, stripes, or had no stripes at all, and also presented images of zebrafish that appeared "fat" or slim while keeping the overall surface area of the image constant. We found that elongated images were highly aversive, and speculated that this may be because they resembled the needle fish, a natural predator of zebrafish (Saverino and Gerlai, 2008). The above examples show how serendipitous findings leading to irreproducibility may be used to advance our understanding of zebrafish behavior and also that understanding of the responses of a relatively unknown laboratory species to environmental stimuli requires thorough and systematic analyses.

The next experimental example is also drawn from my own laboratory. It is based upon a line of research that led to excellent reproducibility and replicability, but also to an unexpected surprize. We have been investigating the effect of alcohol exposure during embryonic development in zebrafish (e.g. Fernandes and Gerlai, 2009; Seguin and Gerlai, 2017). The hope has been to establish a zebrafish model of the less severe and most prevalent forms of Fetal Alcohol Spectrum Disorders. We have discovered that if we expose zebrafish embryos at their 24th hour post-fertilization stage to alcohol for only 2 h at a concentration not exceeding 1% (vol/vol) in the external bath (into which intact, i.e. not dechorionated, eggs were placed), zebrafish develop without any obvious signs of defects. But when these previously alcohol exposed fish are tested at their adult stage (i.e. several months after the alcohol exposure) they exhibit significant impairment in their shoaling (social or group forming) behavior (Fernandes and Gerlai, 2009). We measured how a single adult zebrafish placed in a test tank alone responds to social stimuli, moving images of zebrafish. Control zebrafish move towards and stay close to these images, a sign of social cohesion, or shoaling. Our results indicated that zebrafish exposed to alcohol exhibited a reduced response to moving images of conspecifics, i.e. the adult fish that were exposed to a small amount of alcohol for a very short period of time during their embryonic development did not approach and did not stay close to the conspecific images. And this change was found alcohol concentration dependent (Fernandes and Gerlai, 2009). In fact, the dose response curve was almost perfectly linear: the higher the concentration of embryonic alcohol exposure, the more robust the impairment. This dose-dependency itself represents excellent replicability, but we also found that when the embryonic alcohol exposed zebrafish were tested at their adult stage in another manner, i.e. by allowing these fish to swim freely in natural shoals, they too showed reduced shoaling, i.e. exhibited increased inter-individual distances compared to control fish (Buske and Gerlai, 2011). Thus, we have had good replicability for the social behavior disrupting effect of embryonic alcohol exposure. Other laboratories, using the same dosing regimen and concentrations as ours, but a different strain (population) of zebrafish and a slightly different method of inducing and measuring shoaling, also found similar results (Baggio et al., 2017), again confirming replicability. We too found practically identical results in our own laboratory when we repeated our experiments testing fish that were older than the ones we studied before, suggesting good reproducibility for a lasting embryonic alcohol effect on shoaling in zebrafish (Fernandes et al., 2015a, 2015b; Fernandes and Gerlai, 2009). Why are the above findings relevant for a discussion about issues with reproducibility and replicability? They are, because we did not always find good reproducibility in our laboratory. In one of our studies, we observed lack of changes in shoaling behavior in zebrafish exposed to embryonic alcohol. This was a surprize to us, as all procedures, we thought, including the timing and concentration of alcohol administered, the manner in which the adult fish were tested, etc., were performed exactly the same way as before (Fernandes & Gerlai unpublished results). Upon closer examination of our methods, however, we discovered that the rearing conditions were slightly different (more enriched environment) in this study compared to those of previous ones. We still do not know, and thus have not published about, what specific factors during rearing rescued the embryonic alcohol induced impairment. We view our inability to reproduce our findings, however, not as a problem, but rather as an opportunity. It will potentially allow us to discover environmental manipulation methods that may counteract the deleterious effects of embryonic alcohol exposure in zebrafish.

Understanding how environmental factors may influence zebrafish is also crucial from the perspective of maintenance and breeding of this species. The subsequent section discusses some of these factors focussing mainly on what we do not yet know.

10. Environmental factors whose effects on zebrafish behavior have yet to be explored

There are numerous environmental factors that may affect brain function and behavior of zebrafish, but we do not yet know them, or have not yet explored how their effects may manifest. Even what would constitute optimal parameters of zebrafish maintenance and breeding is not well understood (Tsang et al., 2017). This is an important shortcoming in zebrafish research, because if one does not know about such factors, one may not be able to properly randomize the experiment, or sample experimental subjects. For example, it is possible that differing illumination levels at different heights of the frequently employed multi-shelved high-density zebrafish rack has an effect on the development of the visual system of the zebrafish or in the functioning of many brain areas of the adult (Villamizar et al., 2014). But most do not report on, or even consider, randomizing experimental zebrafish according to the location of their holding tank on the rack system. Briefly, exploring the potential role of numerous environmental factors on zebrafish brain and behavior, and increasing our understanding of their effects, will allow us to better design our experiments, and to reduce issues with reproducibility and replicability.

Although the zebrafish has been in the forefront of developmental biology and genetics (Grunwald and Eisen, 2002), for the past four decades it has essentially been looked at as an egg producing machine. Scientists were concerned about how frequently they could make zebrafish spawn, how many eggs they could squeeze out each time from their subjects, and how they could save on costs and space while achieving maximal egg yield. Unfortunately, the design of zebrafish facilities, high density holding systems, and the environmental parameters in these systems, were all established with this view in mind. Nowadays, however, when zebrafish are being used in a variety of subdisciplines of biology, including neuroscience, behavioral neuroscience, behavior genetics and psychopharmacology (Kalueff et al., 2014), the question of what environmental parameters may be optimal for the zebrafish has become urgently pertinent. The level of urgency is demonstrated, for example, by a recent workshop spearheaded and organized by the Office of Research and Infrastructure Programs of the NIH, titled "Zebrafish and Other Fish Models: Extrinsic Environmental Factors for Rigorous Experiments and Reproducible Results" (https://orip.nih.gov/zebrafish-and-other-fish-models-extrinsic-environmental-factors-rigorous-experiments-and). Briefly, very little is known about how many and which environmental factors, parameters of zebrafish husbandry, and in what way may affect brain function and behavior of this species. What are considered optimal holding, maintenance or breeding conditions is often a matter of tradition and not the result of systematic and controlled analyses (Tsang et al., 2017).

For example, zebrafish are usually housed in small tanks in large numbers, an efficient and cost effective high-density housing method typical of most zebrafish facilities, including my own. However, in a recent study, we stumbled upon a surprising finding that suggested this

housing method may be far from optimal. We found socially isolated fish to exhibit reduced signs of anxiety and stress compared to those housed in a standard manner in groups (Shams et al., 2015), the exact opposite of what we originally expected. Others have also obtained similarly paradoxical results showing that socially isolated zebrafish exhibit lower level of cortisol compared to socially housed zebrafish (Parker et al., 2012). A similarly surprising result was recently communicated at a workshop by Stephen Ekker who reported that one of his colleagues (Karl Clark of Mayo Clinic, Rochester) found larval zebrafish in the 96 well plate, the standard high-throughput screening set up, to exhibit signs of elevated stress, as compared to larval fish tested in the plates that had smaller number of, and thus larger volume, wells (Ekker and Clark, personal communication, Workshop on Zebrafish and Other Fish Models: Extrinsic Environmental Factors for Rigorous Experiments and Reproducible Results, NIH/ORIP, Bethesday MD, 2017). Larval or adult zebrafish used as control that are housed and tested in the traditional manner thus cannot be considered normal or representing baseline. These fish may suffer from elevated levels of stress and/or anxiety, a condition that may interfere or interact with whatever experimental manipulation the investigator wishes to employ. Given that housing density varies greatly across laboratories, within a laboratory, and perhaps even within an experiment, reproducibility and replicability of results may suffer.

Numerous other environmental factors may also need to be considered, examined, and re-examined. Most zebrafish researchers report on the light-cycle employed in their facilities, but almost never on light intensity, and especially not on the variation in light intensity across the tanks in which the experimental zebrafish are housed. The type of light, i.e. whether it is provided by incandescent, fluorescent or LED lamps, is also usually not considered. These different lamps produce a source-specific idiosyncratic spectral distribution pattern. Although we do not exactly know how the source or the type of light affects behavior and brain function of the zebrafish, we do know that these different light sources provide a light frequency spectrum signature that grossly differs according to the light source. We also have some evidence that brain function can significantly be affected by what wavelengths of light are employed (Villamizar et al., 2014; Sovrano et al., 2016; Di Rosa et al., 2015). Furthermore, the effects of rearing zebrafish under specific restricted spectral conditions have already been demonstrated on retinal physiology in zebrafish (Dixon et al., 2004), a finding whose implications are wide-ranging for the behavior of this primarily visual species.

Another environmental parameter that must be systematically analyzed, controlled, and also importantly, precisely reported in publications is salinity (conductivity) as well as the salt composition of the holding (system) water. Most zebrafish facilities employ reverse osmosis (R/O) filtered water. This allows the experimenter to remove any and all potential impurities, trace elements, chlorine, etc., from the local water source. These impurities or chemicals are not harmful for humans, and thus are present in tap-water in fluctuating amounts, but could alter the biology of zebrafish, and could also introduce error variation in zebrafish research. However, R/O water is close to pure, distilled, water, and as such, it is devoid of crucial dissolved elements required for the health, normal development and functioning of zebrafish. Thus, most researchers reconstitute salinity of the "system water" used for maintaining zebrafish by adding sea salt back to the R/O purified water. Surprisingly little is known, however, about what the optimal level of salinity is, and what type of salts should be added. It is also troubling that zebrafish facilities utilize widely different salt concentrations in zebrafish husbandry, ranging among facilities by an order of magnitude, i.e. from 200 to 2000 μS conductivity. Some studies report, while others do not even disclose, the salinity of water used. It is also noteworthy that most facilities, including our own, use sea salt (e.g. Fernandes and Gerlai, 2009). But sea salt is made up mostly of sodium chloride (about 85% of the salts), a salt that is not the main constituent of dissolved molecules in natural waters of the zebrafish. Zebrafish live in a variety of waters with sandy or muddy bottoms (Spence et al., 2006), where the main salt constituent is likely calcium carbonate which zebrafish facilities do not add to R/O water, perhaps because it can form deposits and stains that can clog up valves and water lines. Last, the exact salt composition of the sea salt employed by zebrafish facilities is not precisely controlled. Thus, depending on the brand of salt, the batch, the time of year when the purchase was made, or the local store where the salt was obtained from, salt composition may vary without the knowledge of the experimenter, which may all contribute to problems with replication of results. Similar issues may arise because of fish food (source, type, amount, frequency of feeding), parameters that may vary across laboratories, food sources, and even across time periods for the same food obtained from the same company.

But how do we know if this all matters or not? The problem is that we do not. Analyses controlling and systematically modifying these factors are needed before we can decide what may be important, and what aspects of the environment we may be able to ignore. I recommend the guiding principle in this quest should be ecology, i.e. understanding of the conditions in the natural habitat of the zebrafish. Not all agree. Some argue that as long as we keep parameters controlled and stable, it does not matter how close these parameters match those characterizing the natural habitat of zebrafish. I disagree. If we consider the ecology and ethology of the species we study in our laboratory, and set up the maintenance and test conditions according to the specific features and requirements of the species under study, we will likely experience less fluctuations in the biological, including behavioral, responses we record, a sentiment that has been voiced in the context of behavioral neuroscience research before (Gerlai and Clayton, 1999a, 1999b). The main argument is that by not exposing our subjects to abnormally artificial conditions, we can expect better buffering from our animals against inevitable fluctuations in the environment. In other words, if we keep our subjects within the range of environmental parameters they are evolutionarily adapted to, we can expect better reproducibility and replicability.

Therefore, we need to better understand what may be the optimal range of parameters we must employ, and how deviating from the optimum may influence our research with zebrafish. Although the above appear to be important problems that should be addressed to increase reproducibility and replicability in zebrafish research, and in fact in research with any laboratory organisms in general, not all agree. Some argue that standardization is a fallacy (Würbel, 2000).

11. Should we standardize?

At first sight the answer to this question seems trivial: Of course, we should standardize. After all, reducing unwanted variation, and making procedures, conditions, methods, equipment and genotypes of animals comparable across laboratories should enhance reproducibility and replicability, and thus should enhance the validity of our findings and advance our knowledge. However, others argue this may not be so (Würbel, 2000; Richter et al., 2009). Consider the example of genotype x environment interaction. We may assume that inbred strains should exhibit reduced phenotypical variance. This is because phenotypical variance is made up of two main components, environmental variance and genetic variance. If the latter is zero (as it is expected in a fully inbred strain), the total phenotypical variance is reduced. However, this is often not what we see in empirical research. Outbred (genetically heterogeneous) populations often exhibit smaller phenotypical variance than inbred strains. Why? Because phenotypical variance has a third component arising from genotype x environment interaction. For example, an individual whose genome has a high heterozygosity ratio is expected to be able to cope with fluctuating environmental conditions better than a fully inbred individual whose loci are in a homozygous form. This is simply because many of these loci (in the inbred individual) may have fixated an allele that is sub-optimal from a homeostatic control standpoint. Thus, this inbred individual may exhibit higher intra-individual variance, or put it in other way, multiple

inbred subjects within a strain may show higher inter-individual variance as a result of fluctuating environmental conditions. Whereas, individuals with high number of their loci in a heterozygous form, or individuals within a genetically heterogeneous population, may be buffered against such fluctuations, and thus will remain relatively uniform (exhibiting reduced environmentally induced variance). Should we prefer standardized inbred strains of zebrafish to outbred, genetically heterogeneous stock, then?

Similarly, one may argue that although rigorous control of environmental conditions may reduce environmentally induced error variation in our experiments, and thus may increase reproducibility and replicability, such rigor may come with a price. If the environment is fixed to a particular set of parameters, the investigator's results will be valid only in this fixed context. The problem with narrowly defining and controlling environmental conditions is that we will not know how robust our findings may be. That is, how well such results would generalize to conditions that even slightly differ from the one we employed. And we know that it is very difficult to set all relevant environmental parameters and maintain them at the same pre-set value across multiple laboratories, because we often do not even know what parameters we should worry about. One could then ask: should not we allow the environment to vary, and should not we let ourselves suffer from irreproducibility on the short run at least? Allowing fluctuation of, or variability in, environmental conditions (or genomes) would allow us to weed out results that are highly dependent upon idiosyncratic environmental (or genetic) contexts/conditions, and would allow us to keep those results that are more robust against fluctuations in such factors. The latter argument essentially invokes the idea of natural selection in scientific discoveries. Over the long run, given the myriad of variations around methods, procedures and conditions employed by scientists, only those results will stand the test of time (that is will be regarded as valid, and thus will be kept in mainstream science) that are indeed robust against such fluctuations. A similar debate has occurred in the context of mouse behavioral phenotyping (Richter et al., 2009; Würbel, 2000; van der Staay and Steckler, 2002), a field that is much more mature than zebrafish behavioral neuroscience. Despite the maturity of this former field, however, answers to the above questions still remain, and the above points are still debated.

My own impression is that asking whether we should or should not standardize, or how much we should worry about reproducibility or replicability is perhaps too general and too esoteric. The answer to these questions will likely be dependent upon specific studies and particular research contexts. But one thing is certain, understanding which factors may contribute to increasing our ability to reproduce and replicate our results, which factors may need to be controlled tightly, and which ones may be ignored, will require detailed and systematic analyses, research that will keep zebrafish behavioral neuroscientists busy for decades to come.

12. Conclusions

Issues concerning reproducibility and replicability are neither unique to zebrafish behavioral neuroscience research, nor new. The issues have multiple sources, which include bias for publishing only positive (significant) findings, the way we make statistical inferences, and numerous empirical or experimental problems associated with not controlling or not being able to control potentially important factors that influence results. Some advocate statistical solutions to address reproducibility and replicability related problems, but I believe the most stressing issue is to improve our methods of data collection. Even the most appropriate statistical procedure is meaningless unless the data it is used to analyze are inappropriately obtained. Appropriate experimental data collection means that one needs to investigate, understand, and, ultimately, control as many factors that may influence the outcome of research as possible. For zebrafish, a relative newcomer in behavioral neuroscience, this is an especially urgent and crucial goal.

References

Ahmed, T.S., Fernandes, Y., Gerlai, R., 2012. Effects of animated images of sympatric predators and abstract shapes on fear responses in zebrafish. Behaviour 149, 1125–1153.

Al-Imari, L., Gerlai, R., 2008. Sight of conspecifics as reward in associative learning tasks for zebrafish (*Danio rerio*). Behav. Brain Res. 189, 216–219.

Baggerly, K.A., Coombes, K.R., 2009. Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. Ann. Appl. Stat. 3, 1309–1334.

Baggio, S., Mussulini, M., Gerlai, R., de Oliveira, D.L., 2017. Anxiety-like behavior is related to social avoidance phenotype of adult zebrafish FASD, and is modulated by buspirone. Behav. Brain Res. http://dx.doi.org/10.1016/j.bbr.2017.08.039. (in press), pii: S0166-4328(17)30613-7.

Baker, M., 2016a. Reproducibility, Seek out stronger science. Nature 537, 703–704.

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

Bardo, M.T., Neisewander, J.L., Pierce, R.C., 1989. Novelty-induced place preference behavior in rats - effects of opiate and dopaminergic drugs. Pharmacol. Biochem. Behav. 32, 683–689.

Bass, S.L.S., Gerlai, R., 2008. Zebrafish (*Danio rerio*) responds differentially to stimulus fish: the effects of sympatric and allopatric predators and harmless fish. Behav. Brain Res. 186, 107–117.

Baxter, Mark G., Burwell, Rebecca D., 2017. Promoting transparency and reproducibility in behavioral neuroscience: publishing replications, registered reports, and null results. Behav. Neurosci. 131, 275–276.

Blaser, R.E., Peñalosa, Y.M., 2011. Stimuli affecting zebrafish (*Danio rerio*) behavior in the light/dark preference test. Physiol. Behav. 104, 831–837.

Buske, C., Gerlai, R., 2011. Early embryonic ethanol exposure impairs shoaling and the dopaminergic and serotoninergic systems in adult zebrafish. Neurotoxicol. Teratol. 33, 698–707.

Crabbe, J.C., Wahlsten, D., Dudek, B.C., 1999. Genetics of mouse behavior: interactions with laboratory environment. Science 284, 1670–1672.

de Visser, L., van den Bos, R., Kuurman, W.W., Kas, M.J., Spruijt, B.M., 2006. Novel approach to the behavioural characterization of inbred mice: automated home cage observations. Genes Brain Behav. 5, 458–466.

Di Rosa, V., Frigato, E., López-Olmeda, J.F., Sánchez-Vázquez, F.J., Bertolucci, C., 2015 Jul 6. The light wavelength affects the ontogeny of clock gene expression and activity rhythms in zebrafish larvae. PLoS One 10 (7), e0132235. http://dx.doi.org/10.1371/journal.pone.0132235.

Dixon, L.J., McDowell, A.L., Houchins, J.D., Bilotta, J., 2004. Effects of restricted spectral rearing on the development of zebrafish retinal physiology. Doc. Ophthalmol. 109, 17–33.

Dlugos, C.A., Rabin, R.A., 2003. Ethanol effects on three strains of zebrafish: model system for genetic investigations. Pharmacol. Biochem. Behav. 74, 471–480.

Facciol, A., Tran, S., Gerlai, R., 2017. Re-examining the factors affecting choice in the light-dark preference test in zebrafish. Behav. Brain Res. 327, 21–28.

Fernandes, Y., Gerlai, R., 2009. Long-term behavioral changes in response to early developmental exposure to ethanol in zebrafish. Alcohol. Clin. Exp. Res. 33, 601–609.

Fernandes, Y., Rampersad, M., Gerlai, R., 2015a. Impairment of social behaviour persists two years after embryonic alcohol exposure in zebrafish: a model of fetal alcohol spectrum disorders. Behav. Brain Res. 292, 102–108.

Fernandes, Y., Rampersad, M., Gerlai, R., 2015b. Embryonic alcohol exposure impairs the dopaminergic system and social behavioural responses in adult zebrafish. Int. J. Neuropsychopharmacol. 18 (6), 1–8. http://dx.doi.org/10.1093/ijnp/pyu089. (pii: pyu089).

Fisher, R.A., 1935. The Design of Experiments. Oliver & Boyd, Oxford, England, pp. 14.

Gerlai, R., 1996. Gene targeting studies of mammalian behavior: is it the mutation or the background genotype? Trends Neurosci. 19, 177–181.

Gerlai, R., 1998. A new continuous alternation task in T-maze detects hippocampal dysfunction in mice: a strain comparison and lesion study. Behavioral. Brain Res. 95, 91–101.

Gerlai, R., 2010. Zebrafish antipredatory responses: a future for translational research? Behav. Brain Res. 207, 223–231.

Gerlai, R., Clayton, N.S., 1999a. Analysing hippocampal function in transgenic mice: an ethological perspective. Trends Neurosci. 22, 47–51.

Gerlai, R., Clayton, N.S., 1999b. Tapping artificially into natural talents – reply. Trends Neurosci. 22, 301–302.

Gerlai, R., Lahav, M., Guo, S., Rosenthal, A., 2000. Drinks like a fish: zebra fish (*Danio rerio*) as a behavior genetic model to study alcohol effects. Pharmacol. Biochem. Behav. 67, 773–782.

Gerlai, R., Prajapati, S., Ahmad, F., 2008. Differences in acute alcohol induced behavioral responses among zebrafish populations. Alcohol. Clin. Exp. Res. 32, 1763–1773.

Gómez-Laplaza, L.M., Gerlai, R., 2010. Latent learning in zebrafish (*Danio rerio*). Behav. Brain Res. 208, 509–515.

Grunwald, D.J., Eisen, J.S., 2002. Headwaters of the zebrafish—emergence of a new model vertebrate. Nat. Rev. Genet. 3, 717–724.

Herndon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Camb. J. Econ. 38, 257–279.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70.

Ioannidis, J., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124.

Johnson, S.L., Zon, L.I., 1999. Genetic backgrounds and some standard stocks and strains used in zebrafish developmental biology and genetics. Methods Cell Biol. 60, 357–359.

Kafkafi, N., Agassi, J., Chesler, E.J., Crabbe, J.C., Crusio, W.E., Eilam, D., Gerlai, R., Golani, I., Gomez-Marin, A., Heller, R., Iraqi, F., Jaljuli, I., Karp, N.A., Morgan, H., Nicholson, G., Pfaff, D.W., Richter, S.H., Stark, P.B., Stiedl, O., Stodden, V., Tarantino, L.M., Tucci, V., Valdar, W., Williams, R.W., Würbel, H., Benjamini, Y., 2018. Reproducibility and Replicability of mouse phenotyping in pre-clinical studies. Biobehav. Rev. http://dx.doi.org/10.1016/j.neubiorev.2018.01.003.Neuroscience. (in press). (S0149-7634(16)30657-1).

Kalueff, A.V., Stewart, A.M., Gerlai, R., 2014. Zebrafish as an emerging model for studying complex brain disorders. Trends Pharmacol. Sci. 35, 63–75.

Leek, J.T., Peng, R.D., 2015. Opinion: Reproducible research can still be wrong: adopting a prevention approach. Proc. Natl. Acad. Sci. U. S. A. 112, 1645–1646.

Levin, E.D., Chen, E., 2004. Nicotinic involvement in memory function in zebrafish. Neurotoxicol. Teratol. 26, 731–735.

Levin, E.D., Limpuangthip, J., Rachakonda, T., Peterson, M., 2006. Timing of nicotine effects on learning in zebrafish. Psychopharmacology 184, 547–552.

Levin, E.D., Bencan, Z., Cerutti, D.T., 2007. Anxiolytic effects of nicotine in zebrafish. Physiol. Behav. 90, 54–58.

Luca, R., Gerlai, R., 2012. In search of optimal fear inducing stimuli: differential behavioral responses to computer animated images in zebrafish. Behav. Brain Res. 226, 66–76.

Mahabir, S., Chatterjee, D., Buske, C., Gerlai, R., 2013. Maturation of shoaling in two zebrafish strains: a behavioral and neurochemical analysis. Behav. Brain Res. 247, 1–8.

Maximino, C., de Brito, T.M., Colmanetti, R., Pontes, A.A., de Castro, H.M., de Lacerda, R.I., Morato, S., Gouveia Jr., A., 2010. Parametric analyses of anxiety in zebrafish scototaxis. Behav. Brain Res. 210, 1–7.

Miller, N., Gerlai, R., 2007. Quantification of shoaling behaviour in zebrafish (*Danio rerio*). Behav. Brain Res. 184, 157–166.

Miller, N., Greene, K., Dydynski, A., Gerlai, R., 2012. Effects of nicotine and alcohol on zebrafish (*Danio rerio*) shoaling. Behav. Brain Res. 240, 192–196.

Parker, M.O., Millington, M.E., Combe, F.J., Brennan, C.H., 2012. Housing conditions differentially affect physiological and behavioural stress responses of zebrafish, as well as the response to anxiolytics. PLoS One 7 (4), e34992. http://dx.doi.org/10.1371/journal.pone.0034992.

Peng, R., 2015. The reproducibility crisis in science: a statistical counterattack. Significance 12, 30–32.

Richter, S.H., Garner, J.P., Wuerbel, H., 2009. Environmental standardization: cure or cause of poor reproducibility in animal experiments? Nat. Methods 6, 257–261.

Savalei, V., Dunn, E., 2015. Is the call to abandon p-values the red herring of the replicability crisis? Front. Psychol. 6 (245). http://dx.doi.org/10.3389/fpsyg.2015.00245.

Saverino, C., Gerlai, R., 2008. The social zebrafish: behavioral responses to conspecific, heterospecific, and computer animated fish. Behav. Brain Res. 191, 77–87.

Seguin, D., Gerlai, R., 2017. Fetal alcohol spectrum disorders: zebrafish in the analysis of the milder and more prevalent form of the disease. Behav. Brain Res. http://dx.doi.org/10.1016/j.bbr.2017.10.005. (in press), pii: S0166-4328(17)31429-8.

Seguin, D., Shams, S., Gerlai, R., 2016. Behavioural responses to novelty or to a predator stimulus are not altered in adult zebrafish by early embryonic alcohol exposure. Alcohol. Clin. Exp. Res. 2667-26 (96), 40.

Sham, P.C., Purcell, S.M., 2014. Statistical power and significance testing in large-scale genetic studies. Nat. Rev. Genet. 15, 335–346.

Shams, S., Chatterjee, D., Gerlai, R., 2015. Chronic social isolation affects thigmotaxis and whole-brain serotonin levels in adult zebrafish. Behav. Brain Res. 292, 283–287.

Sison, M., Gerlai, R., 2011. Associative learning performance is impaired in zebrafish (*Danio rerio*) by the NMDA-R antagonist MK-801. Neurobiol. Learn. Mem. 96 (75), 230–237 (121).

Sovrano, V.A., Bertolucci, C., Frigato, E., Foà, A., Rogers, L.J., 2016. Influence of exposure in ovo to different light wavelengths on the lateralization of social response in zebrafish larvae. Physiol. Behav. 157, 258–264.

Speedie, N., Gerlai, R., 2008. Alarm substance induced behavioral responses in zebrafish (*Danio rerio*). Behav. Brain Res. 188, 168–177.

Spence, R., Fatema, M.K., Reichard, M., Huq, K.A., Wahab, M.A., Ahmed, Z.F., Smith, C., 2006. Distribution and habitat preferences of the zebrafish in Bangladesh. J. Fish Biol. 69, 1435–1448.

Szücs, D., Ioannidis, J.P.A., 2017. When null hypothesis significance testing is unsuitable for research: a reassessment. Front. Hum. Neurosci. 11, 390. http://dx.doi.org/10.1016/j.beproc.2017.08.004.

Thompson, L.A., Fagan, J.F., Fulker, D.W., 1991. Longitudinal prediction of specific cognitive-abilities from infant novelty preference. Child Dev. 62, 530–538.

Tsang, B., Zahid, H., Ansari, R., Lee, R., Partap, A., Gerlai, R., 2017. Breeding zebrafish: a review of different methods and a discussion on standardization. Zebrafish. http://dx.doi.org/10.1089/zeb.2017.1477. (in press).

van der Staay, F.J., Steckler, T., 2002. The fallacy of behavioral phenotyping without standardisation. Genes Brain Behav. 1, 9–13.

Villamizar, N., Vera, L.M., Foulkes, N.S., Sánchez-Vázquez, F.J., 2014. Effect of lighting conditions on zebrafish growth and development. Zebrafish 11, 173–181.

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect. Psychol. Sci. 4, 274–290.

Wellek, S., 2017. A critical evaluation of the current “p-value controversy”. Biom. J. 59, 854–872.

Würbel, H., 2000. Behaviour and the standardization fallacy. Nat. Genet. 26, 263.

Zhdanova, I.V., 2011. Sleep and its regulation in zebrafish. Rev. Neurosci. 22, 27–36.