The University of Southern Mississippi
The Aquila Digital Community

Master's Theses

Summer 8-2011

Individual Differences in Zebrafish (Danio rerio) Behavior: Implications for Personality

Christina Nicole Toms
University of Southern Mississippi

Follow this and additional works at: https://aquila.usm.edu/masters_theses

Part of the Experimental Analysis of Behavior Commons

Recommended Citation

Toms, Christina Nicole, "Individual Differences in Zebrafish (Danio rerio) Behavior: Implications for Personality" (2011). Master's Theses. 249.
https://aquila.usm.edu/masters_theses/249

This Masters Thesis is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Master's Theses by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
INDIVIDUAL DIFFERENCES IN ZEBRAFISH (Danio rerio) BEHAVIOR:
IMPLICATIONS FOR PERSONALITY

by

Christina Nicole Toms

A Thesis
Submitted to the Graduate School
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts

Approved:

Dr. David Echaverria_____________
Director

Dr. Stan Kuczaj_____________

Dr. Virgil Zeigler-Hill_____________

Susan A. Siltanen_____________
Dean of the Graduate School

August 2011
ABSTRACT

INDIVIDUAL DIFFERENCES IN ZEBRAFISH (DANIO RERIO) BEHAVIOR: IMPLICATIONS FOR PERSONALITY

by Christina Nicole Toms

August 2011

Individual differences (IDs) in zebrafish (Danio rerio) behavior were examined across time and contexts in order to investigate personality traits. Zebrafish (N = 30) were exposed to six behavioral tasks that were collectively expected to capture (a) aggression, (b) boldness/shyness, and (c) fear. The tasks included a small open field, mirror exposure, emergence, large open field, novel object, and predator exposure tasks. IDs in behavior were found to be consistent across time for a majority of behaviors and consistent across contexts for all but two behaviors. Convergent and discriminate evidence was examined for the three constructs. There was little evidence for an aggressive trait, which may largely be attributed to the behaviors chosen for measurement in this study. There were mixed results for identifying a separate bold vs. fearless construct; however, results largely indicate that the emergence, novel object, and predator exposure tasks may all have been capturing a tradeoff between boldness and fear. Results suggest that responses to novel objects depend on the object used (the basis for which is still unknown), suggesting that more research needs to be done to determine how best to utilize a novel object task for this species. Although the results from this project did not allow for specific traits to be labeled with confidence, the predator task demonstrated high internal consistency and may be particular useful for capturing fear/anxiety-like traits. Overall, this study provides a comprehensive examination of zebrafish behavior, support for
consistent IDs and personality traits, and a much-needed foundation for further personality research in a key model organism, the zebrafish.
ACKNOWLEDGMENTS

The writer would like to thank the thesis director, Dr. David Echevarria, and the other committee members, Dr. Stan Kuczaj and Dr. Virgil Zeigler-Hill, for their advice and support throughout the duration of this project. I would especially like to thank Dr. David Echevarria for keeping me sane, supporting me through the rough patches, being there whenever I needed, and believing in me enough to offer me such a wonderful opportunity. Special thanks go to my undergrad assistants for all of their help and for the endless days of mundane behavioral coding. I could not have done it without you: Daniel Miles, Abby Regan, Laura Sumrall, Amanda Welch, Shanley Wilson, and Andrew Wooton. Appreciation must also be expressed to Dr. Virgil Zeigler-Hill, Dr. John Harsh, and Dr. Richard Mohn, who were more than willing to chat and offer advice regarding statistics and analytical approaches for this project. And finally to David Rosenfield for taking the time to create a Matlab program used to double check all of my databases.
# TABLE OF CONTENTS

ABSTRACT ............................................................................................................................................. ii

ACKNOWLEDGMENTS ......................................................................................................................... iv

LIST OF TABLES ................................................................................................................................... vi

LIST OF ILLUSTRATIONS .................................................................................................................... vii

CHAPTER

I. INTRODUCTION ...............................................................................................................................1

Species Choice
Current Project and Hypotheses

II. METHODS .........................................................................................................................................8

Subjects and Housing
General Tests and Procedures
Behavioral Tasks
Behavioral Measures

III. RESULTS ......................................................................................................................................32

Reliability
Exploratory Data Analysis
Examining Relationships Across Time
Examining Relationships Across Contexts
Relationship Between Behaviors and Underlying Constructs

IV. DISCUSSION ...............................................................................................................................67

Behaviors
Consistency
Examining Constructs
Considerations and Caveats
Conclusions and Broader Implications

APPENDIX ........................................................................................................................................86

REFERENCES ....................................................................................................................................87
LIST OF TABLES

Table

1. Summary Experimental Timeline for Completion of Each Task .........................13

2. Summary of Behaviors Recorded, Operational Definition and Measurement Criteria Used ..................................................................................................................23

3. Raw Frequency and Proportion Data for Freezing and Drift Behaviors ...............36

4. Consistency across Time (Trials 1 and 2) for Each Behavior on Each Task; Data Prior to Correction for Multiple Tests ...........................................................................41

5. Consistency across Time (Trials 1 and 2) for Each Behavior on Each Task; Final Results after Correction for Multiple Tests .........................................................44

6. Results from Examining Differences in Magnitude across Time (Trials 1 and 2) for Each Behavior on Each Task .................................................................46

7. Results for Examining Main Effects of Differences in Magnitude across Contexts and for Examining Consistency across Contexts................................................49

8. Posthoc Results from Tests Indicating Significant Differences in Behaviors across Treatments (for Trial 1 Only) .................................................................50

9. Results of Pairwise Posthoc Comparisons after Significant Friedman’s ANOVA .............................................................................................................................51
LIST OF ILLUSTRATIONS

Figure

1. A Male Zebrafish (*Danio rerio*) ................................................................................8
2. Convict Cichlid (*Amatitlania nigrofasciata*) .................................................................10
3. The Testing Apparatus for Four Behavioral Tasks........................................................11
4. Explanation of INSPT and APP Behavior. (A) Mirror Task in 5-gallon Tank; (B) Novel Object Task with Pebble in 10-gallon Tank; (C) Predator Exposure Task in 10-gallon tank ................................................................................31
5. (A) Average Percent Time Fish Spent Freezing in Trials 1 and 2 for Each Treatment Condition; (B) Average Time Fish Spent Drifting in Trials 1 and 2 Each Treatment Condition ..................................................................................37
6. Average Percent Time Fish Spent Drifting and Freezing in Trial 1 for Each Treatment Condition ..................................................................................................................38
7. Comparing the Proportion of Time Inspecting and Approaching Stimuli across Tasks (INSPT_p, APP_p; Trial 1) .................................................................................................56
8. Comparison of the Frequency to Inspect for Each Task (INSPT_i; Trial 1) ...............57
9. Comparison of Latency Measures (LatINSPT, LatAPP and LatEMRG) across Tasks (Trial 1) ..........................................................................................................................57
10. Average Percent of Environment Explored across Tasks (EXP; Trial 1) ...............61
11. Average Ambulation Rate across Tasks (Average Square Crossing Per Minute; AMB_r; Trial 1) .........................................................................................................................62
CHAPTER I

INTRODUCTION

Individual differences (IDs) among non-human animals in a population are often ignored. Instead, in classic ecological and behavioral research and models, scientists tend to focus on groups as a unit, seeking answers based on average behavior and treating conspecifics as ecologically equivalent. Consequently, this has left research in specific differences in individuals, and their contribution to collective behavior, underrepresented and largely unexplored (Bolnick et al., 2003; Dussutour, Nicolis, Despland, & Simpson, 2008; Romey, 1996; Sih, Bell, Johnson, & Ziemba, 2004). If individual differences in behavior reflect more than just noise around an average, then traditional approaches to behavioral ecology, behavioral biology and how we approach questions about evolution may be challenged. Dussutour et al. (2008) summarized this importance well in their statement, “even slight differences in the tendency of individual animals to show a given behaviour may be at the heart of decision-making processes and may have consequences for the ecology and evolution of populations” (p. 5). Consequently, there is a growing interest in IDs as researchers realize the importance of understanding behavior at this level (Mather, 1998; Slater, 1981). As a result, many researchers have begun to turn to personality-related questions to explain individual differences (IDs) in animal behavior.

Although the concept of personality in nonhuman animals is only starting to make headway in the animal research community, and pales in comparison to human research, there is growing evidence supporting the existence and importance of personality in animals. Despite criticism and the fear of anthropomorphism, accumulating evidence suggests “personality” and “temperament” can be defined objectively and precisely in
animals (Budaev, 1997; Gosling, 2001). In research with humans, a broad working
definition of personality that satisfies most, can be defined as “those characteristics of the
person[s]...generally that account for consistent patterns of behavior” (Pervin, 1984). In
human research, communication and self-report are key tools for understanding and
testing the theories surrounding personality. Animal personality research however, is
obviously limited by our inability to communicate with animals about their experiences.
Therefore the majority of personality phenomena studied in animals is focused on traits,
which for both humans and animals refers to the consistency of individual response to a
variety of situations (Pervin, 1984; Vazire & Gosling, 2004). In this manner, individual
variation is distinct from temporal, situational and environmental changes and is more
indicative of the underlying characteristics of individuals. The trait approach is
particularly appropriate for nonhumans because one can assess personality traits by
measuring observable behaviors (Rouff, Sussmani, & Strube, 2005). For those that do
support personality research in animals, personality is generally defined as an
individual’s distinguishing pattern of behavior which remains consistent over time and
across situations/contexts (Budaev, 1997; Dall, Houston & McNamara, 2004; Gosling,
2001; Highfill & Kuczaj, 2007; Pervin, 1984; Vazire & Gosling, 2004).

An important consideration is the use of temperament, described in human
literature as the biological foundation for personality, commonly appearing early in
development and interacting with environmental influences (Gosling & John, 1999;
Vazire & Gosling, 2004). Unfortunately, animal researchers frequently use personality
and temperament interchangeably without distinction. For our purposes, we continue to
use personality and refrain from making suggestions about temperament. One other
important consideration is that there is a notable and fundamental difference between behavioral stability (behavior that does not change) and consistency (a behavioral measure that is predictable across time and/or contexts even if the degree or level of the behavior changes) (Budaev & Zworykin, 2002). By definition, consistency in behavior across time and context is the targeted concept in personality studies (also see Toms, Echevarria, & Jouandot, 2010 for a more thorough review). Thus as Budaev and Zworykin (2002) concluded, “behavior may be situation specific whereas individual differences, consistent” (p. S190).

Burgeoning interest from psychologists, behavioral ecologists and related fields have created a boom in animal personality-related research in a wide range of taxa (for a review, see Gosling, 2001). Unfortunately unified approaches and methodology are lacking. Psychologists tend to base research on hierarchical models and employ multivariate statistical techniques to extract behavioral factors or components (Gosling, 2001; Itoh, 2002). This approach focuses primarily on personality structure and has been criticized for not exploring ecological implications. On the other hand, ecologists tend to examine the ecological relevance of personality without first providing a solid foundation for personality structure; i.e., the reliability and validity of the measurements employed is often ignored. Furthermore, behavioral neuroscientists conduct a vast amount of animal behavior research and have recognized the importance of accounting for behavioral paradigms (Wright, Nakamichi, Krause & Butlin, 2006) and phenotypes (Blaser & Gerlai, 2006) in modeling genetic and neurological mechanisms. However, most still investigate behavior based on averages, ignoring individual differences. Generally,
individual-level analysis has not been widely accepted and the concept of personality has not been incorporated into behavioral models in this field.

This current project examines consistent individual differences in behavior across time and across contexts in an up-and-coming model organism, the zebrafish (*Danio rerio*), and examine supporting evidence for the validity of the tasks used and constructs investigated.

**Species Choice**

Several species of fish have become popular models for research in behavior, behavioral ecology, and personality (e.g., guppies, sticklebacks and cichlids). Fish models are economical (low maintenance, cost, and space required), easily manipulated for high experimental control and allow for large samples sizes (thus high-powered analysis). Many species are independent immediately following birth, which when combined with a quick developmental period, minimizes the social influences on behavior and temperament from the mother or other individuals during rearing (Warren & Callaghan, 1975). For ecologists, research in the natural environment is not impossible, and experimental re-creation of natural environments is conceivable. Therefore it is not surprising that research in individual differences and personality has been vastly expanded upon using fish models in recent years.

The zebrafish (*Danio rerio*) is a particularly excellent candidate as a model organism. They exhibit a broad range of behaviors, are a shoaling (schooling) species (offering a simple model for other social species as opposed to primates with complex social systems), have a large number of offspring, and exhibit external development, allowing observations of behavior at stages unavailable in mammals (Spence, Gerlach,
Lawrence, & Smith, 2008). They have become a favored model organism and recently, one of the most important species used in genetics, developmental biology, neurophysiology and biomedicine (Lawrence, 2007; Spence et al., 2008). They have a well-characterized, simple nervous system and a genetic sequence that is similar to most other vertebrates (Gerlai, Lee, & Blaser, 2006). As Spence et al. (2008) reports, the zebrafish was the first to be used for a large-scale random mutagenesis screen in a vertebrate, which has since led to the identification of over 400 genes controlling vertebrate development, and numerous technological advances, increasing the application of zebrafish as a model for human diseases. With a solid understanding of the genetic and biological structure of this species, one can see the potential for the level of understanding that could be further obtained from behavioral research and the comparative power across fields that could be utilized.

Despite these advantages however, their application has not yet been capitalized on by behavioral or personality researchers. Of the many studies conducted on a variety of fresh and salt water fish species, only a handful have employed zebrafish. Behavioral research in zebrafish has been vastly expanded upon by behavioral neuroscientists in recent years. However, a large portion of the research available examines responses of group behavior using “classic methodology” in which different individuals are utilized in different tasks. Within a given task, all behaviors of interest are averaged for interpretation. It is clear that there is a need for research that quantifies zebrafish behavior at the individual level that considers individual differences. This is important even for modeling human behavior, since after all, it is humans that we strongly attribute personality and individual differences to.
Current Project and Hypotheses

Six commonly used behavioral tasks were utilized for this project: a small open field, mirror exposure task, an emergence task (i.e., latency to enter), a large open field, a novel object task and a predator response task. These tasks provide a variety of test conditions which have been used to elicit behavioral responses that can be used to measure aggression, exploration, activity, and measures of boldness and fear. The specifics of these tasks were chosen based on the supporting literature and the variety of expected responses that could be used to help provide a comprehensive picture of zebrafish behavior.

In much of the personality work reviewed for this project, researchers often decide upon one or a couple of measures per task, often specific to that task, which are expected to represent the underlying trait of interest. For example, if using a novel object, open-field and mirror task, one might measure latency to approach, activity and bites to the mirror as measures expected to capture boldness, activity and aggression, respectively. This fails to consider other behaviors within each task and how behaviors might be related. If a foundation has not already been established to understand how behaviors are related and expressed in given tasks, there’s no certainty that these selected behaviors are capturing boldness, activity and aggression, respectively. Why not measure activity on all three tasks? This way one can directly compare individual behaviors across different types of tasks.

The aim for this study was for a more comprehensive picture of zebrafish behavior in order to best quantify not only the behavioral responses, but to compile evidence to support conclusions regarding personality. For example, the large and small
open tasks are expected to be different variations of the same task such that behaviors should be consistent across them. Similarly, the novel object and emergence tasks have both been utilized to investigate boldness and shyness, and if behaviors are positively correlated across these, there would be support that they are capturing similar underlying motivations. The array of tasks utilized allowed for further examination of convergent and discriminant validity of the constructs of interest through measuring multiple behaviors across contexts, an important component of this research.

In summary, the current project is largely exploratory and descriptive with the aim of exploring potential personality traits in this zebrafish. If personality-like traits exist, it should be possible to show that individual differences in behavior are (1) consistent across time and (2) consistent across contexts. Furthermore, one should be able to support conclusions with evidence of convergent and discriminant validity for the constructs of interest. The primary constructs targeted in this project are (1) aggression, (2) bold/shyness, and (3) fear. Exploration is also explored, with the understanding that it may be closely linked with one or more of these other constructs.
CHAPTER II

METHODS

Subjects and Housing

The zebrafish (Figure 1; *Danio rerio*) is a small freshwater teleost belonging to the family Cyprinidae. They are indigenous to South Asia but can be found across India, Bangladesh, Nepal, Myanmar, and Pakistan (Lawrence, 2007). They occur in a variety of habitats, including irrigation ditches, man-made ponds, rice fields, upper ends of rivers and fast flowing streams (Lawrence, 2007; Spence et al., 2008). The temperature of their natural habitat ranges widely from 6° C in winter to over 38° C in the summer (Spence et al., 2008). They average approximately 3-4 cm standard length (SL: the measurement from tip of snout to the origin of the caudal fin) (Spence et al., 2008).

![Zebrafish](image)

*Figure 1. A male zebrafish (Dani rerio).*

A total of 30 adult zebrafish were used for this project and were obtained from a local aquarium (Pet Superstore, Hattiesburg, MS) whose exact origin is not known. The advantage of purchasing from a local vendor and not utilizing a particular strain from a carefully monitored lab population is to obtain a sample that is likely more genetically heterogeneous and therefore more phenotypically variable. Increased phenotypic variance means that there should be increased behavioral variability between individuals, crucial
to this project. Furthermore, this population is expected to be more similar to natural populations compared to using lab-bred sample.

Fish were originally housed in community tank systems with temperatures maintained at approximately 27° C. Individuals were randomly selected from the original stock system (comprised of two, 20-gallon tanks; 24.25L x 12.5W x 16H cm), sexed, measured and placed into individual housing units (8L x 4.5W x 5.5H cm), in a separate housing system, a minimum of 48 hours prior to experiments. The compartment chosen for each individual’s housing was randomized to reduce sampling bias due to the capture order (Budaev, 1997). Individual housing units were side by side and clear, providing visual contact to reduce potential stress due to separation.

Both tank systems contain aeration and filtration units with de-chlorinated water. Lighting was set to 14-hour light and 10-hour dark cycles. Zebrafish were fed (Omega One, Freshwater Flakes for Tropical Fish, Sitka, AK) once daily, in the evenings to reduce behavioral bias from hunger (or from complete satiation if they had been fed just before trials). Only male zebrafish were used in order to prevent possible behavioral fluctuations in males or females as a result of female ovarian cycles.

A convict cichlid (Amatitlania nigrofasciata) was used as a predator fish in the predator task for this study (Figure 2). Although the convict cichlid is not a natural predator of the zebrafish, they have been commonly used as predators in other prey species (e.g., guppies; Budaev, 1997) and a number of visually similar cichlid predators are sympatric with zebrafish populations. The convict cichlid was housed singly in a 10-gallon tank (50.8L x 25.5W x 30.5H cm) with its own water system, such that there was no shared water between zebrafish and cichlid. Otherwise, lighting and water conditions
are the same as for zebrafish. In order to minimize predator attacks towards the zebrafish during the predator task, the cichlid was fed (Top Fin Medium Cichlid Pellets, Phoenix, AZ) to satiation 30 minutes to an hour before the experiment. The food pellets were fed a few at a time and the satiation criteria was considered to be met when the cichlid no longer moved towards and consumed the pellets.

Figure 2. Convict cichlid (*Amatitlania nigrofasciata*).

General Tests and Procedures

The six tasks were conducted in the following manner; the small open field and mirror exposure tasks were completed on day one (in a small, 5-gallon tank). One day of rest was given, and on day three fish were exposed to the final four tasks in the following order: emergence task, large open field, novel object and predator exposure tasks (all within the larger, 10-gallon tank). The decision to conduct experiments in this order was largely based on the design described and utilized by Budaev (1997) for examining personality in another small freshwater species, the guppy (see Figure 3 for layout of the later four tasks). It allows for fluid movement from one task to another without excessive handling by the experimenter, which would be a concern if subjects were moved to separate tanks for each task. A maximum of twenty fish were sampled a week and sampling was repeated on the same days the following week. This time frame served to
reduce bias from learning with repeated exposures and to provide substantial time for recovery from handling and potential treatment-related stress induced by behavioral tasks.

*Figure 3.* The testing apparatus for four behavioral tasks. The tank consists of predator (A), open-field and novel-object (B) and home (C) compartments. 1: solid partition with sliding door (2); 3: sliding solid partition, behind which 2 clear Plexiglas partitions keep the predator separate from compartment B; 4: V-like fold; I: tested ZF; II: cichlid predator. Design and figure are modified from that published by Budaev, 1997. Compartment D is blocked with a door and houses a heater and aerator. Without this modification, pilot studies demonstrated a 2-3° temperature drop over the course of the experiments.

Fish were sampled between the hours of approximately 6:30am and 2:00pm. This time period was maintained in order to reduce behavioral fluctuations associated with natural circadian rhythms, since sampling could not be randomized throughout the day. The order for behavioral sampling for individuals was randomized for each testing day. Experimental tanks were filled with water taken from the original housing systems. A water heater and aerator were placed in each tank to maintain ideal conditions (the same as housing conditions) throughout the experiment and to ensure environmental conditions were as similar as possible between experiments. Behavioral task tanks were illuminated from above by two fluorescent strip lights. “White noise” was present from pumps and heaters in large reserve tanks kept in the experimental room. This served to reduce the
influence of any noise coming from adjacent rooms and created conditions similar to the housing room where pump noises are present 24h a day. All tanks were covered on three sides by white shelving paper to eliminate disturbance from movement outside the tank. One of the long sides of each tank was kept clear for filming purposes. However, a wooden box with a curtain was built around them so that cameras could easily be mounted to them, and disruption from outside movement was eliminated. Tank water was removed between each trial session to eliminate the potential effects of olfactory cues on the behavior of subsequent fish.

Individual zebrafish and cichlid predators were transported from the housing room to the test room via 200mL and 1000mL beakers, respectively. The zebrafish were singly lowered into the starting tank and video camera recording started. One webcam (Logitech webcam video) was attached overhead for a top-down view (best for capturing lateral spatial movement and exploration) and the second (identical) was placed along the open side of the tank, allowing for finer-scale details to be examined as was necessary throughout the later coding process. Live feed from the cameras was recorded by two computers making it possible to watch the fish, undisturbed while the experiments were underway. All video recordings were later coded both manually and via JWatcher (Blumstein & Daniel, 2007). After all behavioral experiments, zebrafish were returned to their housing units. Cichlids were transferred to a small holding container (27L x 17W x 12.5H cm) to wait for the next experiment. The same cichlid was used for all experiments and was returned to its housing tank at the end of each experimental day.
Behavioral Tasks

For behavioral tasks, experimental exposure time periods were chosen with the following considerations in mind: (a) the information that was available from the literature; treatment periods are commonly conducted for five (Budaev, 1997; Burns, 2008; Moretz, Martins, & Robison, 2007; Warren & Callaghan, 1975) to 10 minutes (Blaser & Gerlai, 2006; Gerlai, Lahav, Guo & Rosenthal, 2000) each, with acclimation periods varying from 20-60 seconds (Blaser & Gerlai, 2006; Burns, 2008; Gerlai et al., 2000; Shaklee, 1963) and 2-3 minutes (Budaev, 1997; Warren & Callaghan, 1975), and (b) a detailed methodological review by Burns (2008); reliability for measures on some tasks increased with increasing acclimation periods (i.e., with 300 s compared to 120 s).

Table 1 presents a summary of the experimental treatment timeline which is further detailed in the following explanations.

Table 1

<table>
<thead>
<tr>
<th>Task</th>
<th>Acclimation</th>
<th>Duration</th>
<th>Total (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day One</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sm. Open-Field Task</td>
<td>2 min: start box</td>
<td>5 min: (1st subset)</td>
<td>7</td>
</tr>
<tr>
<td>Mirror Task</td>
<td>2 min: start box</td>
<td>5 min: (2nd subset)</td>
<td>7</td>
</tr>
<tr>
<td><strong>Day Three</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergence Task</td>
<td>2 min: start box</td>
<td>5 min: (1st subset)</td>
<td>18 (at most)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0-11 min: latency</td>
<td></td>
</tr>
<tr>
<td>Lg. Open-Field Task</td>
<td>2 min: start box (Only if they did not move through latency task)</td>
<td>5 min: open-field (3rd subset)</td>
<td>7 (at most)</td>
</tr>
<tr>
<td>Novel Object Task</td>
<td>2 min: start box</td>
<td>5 min: novel object exposure (4th subset)</td>
<td>7</td>
</tr>
<tr>
<td>Predator Task</td>
<td>2 min: start box</td>
<td>5 min: predator exposure (5th subset)</td>
<td>7</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td>53 (at most)</td>
</tr>
</tbody>
</table>
Small Open Field Task: Activity & Exploration

The application of an open-field test has been often utilized in fish species and is intended to primarily examine activity and exploration measures. Open-field tasks are those which present an animal with an open, homogenous and novel environment, void of any structural components. It provides a mechanism to measure naturally-relevant behaviors. As Mikheev and Andreev (1993) point out, fish could potentially end up in novel environments, involuntarily, from being swept away by currents or floods or from desperate escape efforts from predators. Thus, fish need to be able to efficiently explore new environments in order to survive. This task has been used for decades as a test environment from which to observe a rich variety of behaviors in animals (Burns, 2008; Csanyi & Gerlai, 1988), and for which rats have been successfully bred for emotionally reactive and non-reactive phenotypes (Warren & Callaghan, 1975). The general idea is that in order to explore the new environment, an individual must move around it such that the amount of movement can be used as an index of exploration (Russell, 1983).

The small open field and mirror tasks were conducted in the same 5-gallon tank with the water level filled to 13.5 cm deep (8L); however, a solid Plexiglas door, lined with white shelving paper was situated directly in front of the mirror and flush with the sides of the tank for the duration of this task. A 19.5 x 19.5 cm, orange, 36-square coordinate grid (3.25 cm each) was placed on the underside of the clear bottom in order to quantify locomotor activity (and remained there for the mirror task). At the start of the experiment, the focal zebrafish was placed into a 3”- diameter, white, bottomless, opaque, PVC pipe (the start box). There is some suggestion that handling experience could evoke behavioral responses similar to predator capture experiences (Russell, 1983).
and handling can be a stressor. Thus, effort was made to minimize the amount of handling and a 2-minute acclimation period was given to minimize confounding behavioral responses associated with netting and transportation (consistent with methodology from Budaev, 1997; Burns, 2008). The start box was then lifted and behavior (first subset) was recorded for five minutes with both top and side cameras.

**Mirror Task: Aggression**

A mirror task creates a situation in which an animal is confronted with its mirror image (termed **mirror image stimulation**, MIS). It has been widely used for decades to measure aggressive behavior in many species, including, but not limited to, a variety of primates, birds and fish species (see Gallup, 1968). Lissman was among the first to note that male Siamese fighting fish (*Betta splendens*) responded to their mirror images with aggressive displays similar to those used in social aggressive interactions, a finding that was subsequently supported by a considerable body of work (as cited in Gallup, 1968, p.783). By the time Tinbergen published his book on animal instincts in 1951, MIS had become a well-established method for studying fish behavior (Tinbergen, 1974). A mirror task offers a convenient measure of aggression for several reasons. Its application allows for high experimental control, minimizes or eliminates any physical damage to individuals from social aggression and the need for excess handling or invasive procedures (Gallup, 1968; Marks, West, Bagatto, & Moore, 2005). Although it has been suggested that MIS may be useful for studying social behavior (Gallup, 1968), a large body of work has supported the notion that the task influences aggression, including support from neurological research (see Adams, Liley, & Gorzalka, 1996). Aggressive behavioral responses to MIS have been shown for Siamese fighting fish (Baenninger,
1670), guppies (Budaev, 1997), lion-headed cichlids (Budaev, Zworykin, & Mochek, 1999), stickleback fish (Tinbergen, 1974) and is one of the more widely used applications in behavioral research with zebrafish (Blaser & Gerlai, 2006; Gerlai et al., 2000; Gerlai, 2003).

Following the small open field, the fish were returned to the start box for two minutes before mirror exposure. This was done to (a) control the distance each fish was from the mirror at the start of the task and to (b) best create an independent task even though the mirror task followed the small open field task. The mirror (15 x 15 cm) was attached to one end of the tank wall with suction cups and with weather stripping along each side, flush with the mirror, to prevent fish from going behind it during the trial. During the acclimation period, the door in front of the mirror was lifted (from a pulley system to minimize disturbance from overhead shadows) to expose the mirror (not visible to the fish while in the start box). Behavioral recoding began again when the start box was lifted and continued for an additional five minutes.

Emergence Task: Boldness

Although methodology and measures have widely varied, there is empirical support for the use of an emergence tasks as a way to capture responses to a novel environment (Burns, 2008; Toms et al., 2010). The idea is to design some “safe” refuge area from which a fish should emerge. This task is expected to measure the propensity of an animal to leave a safe area, commonly through latency measures. The design by Budaev (1997), in Figure 3, has been used in several studies and modified by other researchers (see Brown, Jones, & Braithwaite, 2005). It offers a simple, convenient design and minimizes other potential interacting variables. The small opening to the
larger novel environment may create a more intimidating situation, such that moving through it may be more representative of bold behavior. It also reduces the possibility that the individual accidently moved through without intention (unintentional approaches are a concern with novel object tests).

The bold-shy axis of behavior is categorized under the empirical indices of introversion and extroversion in humans and has been regarded by psychologists as one of the more distinctive, heritable and stable sources of behavioral variation (Coleman & Wilson, 1998; Kagan, Reznick & Snidman, 1998). In humans, individuals vary on a continuum from extreme shyness to extreme boldness, characterized by an individual’s initial reaction to unfamiliar events (Kegan et al., 1998). Therefore, novelty is key for investigating boldness and has set the foundation for how many animal researchers investigate boldness. With discrepancies in definitions and approach, however, two tasks were incorporated in this project specifically aimed at targeting bold behavior; one with respect to a novel situation and the other with a novel object.

The remaining behavioral tasks were conducted on day three in the large 10-gallon experimental tank (see Figure 3) following the design by Budaev (1997), with some modifications to the procedure. Sections were created with white shelving paper-covered Plexiglas, sealed along the edges to ensure the fish cannot squeeze between sections. At the start of the experiment, the zebrafish was again gently placed into the start box in the home compartment (C in Figure 3) (24.5L x 9W cm) of the 10-gallon tank for two minutes for acclimation. Behavioral recordings started when the start box was lifted and again continued for five minutes (directly comparable with other tasks), with the door to compartment B closed. Following five minutes, the Emergence task began.
The small door (2) was opened via a pulley system from above, revealing a 6cm x 6cm opening. Immediately following the door removal, the latency period for the fish to move through the opening was recorded. The subject enters the novel environment (compartment B) when at least half of its body has passed through the opening. The door was immediately lowered behind the fish.

Pilot studies used cutoffs between 4-6 minutes to remain closely comparable to the time given in other tasks. However, only 46% of fish from pilot research entered compartment B in this allotted time. In order to minimize the potential problem of having a large number of individuals that did not enter the novel compartment, it was more important to increase the time allotted to better judge individual differences, than to remain consistent with respect to timing. Therefore cutoffs were decided as a random number between 9-11 minutes. If fish did not enter within the given interval, it was netted into compartment B, and placed again into the start box (placed in the corner of the environment) to re-acclimate for 2 minutes after handling, before the start of the large open field.

*Large Open-Field Task: Activity & Exploration*

Immediately following the subject’s entry to compartment B (22.5L x 24W cm) and the lowering of the door, behavior recording began again for five minutes for the large open field task. In the instance that the subject did not enter, the time for behavioral recording began as soon as the start box was again lifted. For the duration of this and the *Novel Object Task*, the Plexiglas door (3) between this compartment and the predator compartment (A) was in place.
As with the smaller tank used in the mirror task, an orange coordinate grid was positioned on the underside of the clear bottom (in compartment B) in order to quantify locomotor activity. Although the tank sizes are different, the size of each square remained constant \((3.25^2 \text{ cm})\) each in order to ensure direct comparability of activity rates and exploration measures between the two different sized tanks.

**Novel Object Task: Boldness**

Novel object tasks have also been designed to measure response to novelty (thus boldness and shyness) via responses to objects that are placed in the environment that have not been previously observed. The review by Burns (2008) measured *latency to approach object within 4cm and proportion of time spent within 4cm*. He concluded that the novel object task did not display adequate internal or discriminant validity, arguing that movement in general may have propelled fish towards the object instead of movement resulting from some propensity to explore it. However, the author did not indicate whether an effort was made to introduce the object when the fish was positioned away from where it was to be placed. Placement for this project will be the same for all fish, and slightly off-center. This will create a larger open space on one side farthest away from the novel object such that approaching it would not be accidental. Furthermore, one side of the novel object will only have a very narrow area between it and the side wall, which could be easily avoided. Finally, the Burns (2008) review was with guppies and little work has specifically looked at zebrafish on this task. Combined with the *Emergence* task, the results should be able to help sort out any difference between these two tasks. If behaviors are consistent between the two it would support the idea that they may be eliciting similar types of responses from the fish.
Following the large open field task, subjects will be returned again to the start box for two minutes before the novel object is lowered into the tank. This again ensures the same starting position and initial distance of each fish to the object. Behavior will be recorded for five minutes after placing the object in the tank and lifting the start box. The object will be one of three: an artificial rock that resembles something found in nature, a glass aquarium pebble that could resemble a rock but is different in shape and texture, and a conical lead weight which is something the subject would never have seen before. I chose not to use the same object each time in order to reduce the habituation of novelty. However, they were chosen to be similar, in order to reduce the degree to which they may represent something different in the environment, which might minimize the comparability of behavioral responses. For example, I did not choose to create objects with holes (such as PVC pipe) or an artificial plant since these could be perceived as refuges. The choice of object for each experiment was random.

*Predator Exposure Task: Fear*

Predator models have become an increasingly accepted tool for measuring *fear responses* (Speedie & Gerlai, 2008) in many species, in order to better understand anxiety and phobias in humans. There is much debate over what fear is, and how it can be measured. Psychopharmacological studies tend to define *fear* as “a collection of behavioral responses that are elicited by negative stimuli associated with imminent danger such as the presence of a predator” (see Speedie & Gerlai, 2008, p. 169). Fear has also been described by Boissy and Brain as “an emotional and motivational state normally induced by exposures to potentially dangerous objects or situations” (as cited by Budaev, 1997, p. 408). Archer (1979) argues against a unitary concept of fear because of
evidence that various behaviors that are used to measure it don’t often correlate well. Others suggest that fear may be situation specific, and may be hard to identify since, like many behaviors, the behaviors observed may be the result between fear and other conflicting motivation (Russell, 1973, 1983). Russell (1983) suggests that fear behaviors may actually be testable since there are links to autonomic and endocrine correlates, an idea that has been tested and supported since (e.g., Barcellos et al., 2007). Although predatory-response research has only just begun to make headway using zebrafish, within the behavioral neurosciences community it is generally expected that presenting a live predator or predator model elicits a fear response (see Barcellos et al., 2007; Bass & Gerlai, 2008; Blaser & Gerlai, 2006; Gerlai et al., 2000).

Prior to the final task, once again, the subject is returned to the start box for two minutes. During this time, the door (3) to the predator compartment (A; 12.5L x 24W cm) was lifted, revealing a cichlid predator behind a clear Plexiglas partition. Behavioral recording began as soon as the start box was lifted, and continued again for five minutes. The predator was placed into the experimental tank while the zebrafish was in the start box, just before the predator task. Although this may have led to some disturbance of the zebrafish, the choice to keep the predator in the tank the entire time would have led to chemical exposure to the predator during all tasks and potentially confounded behaviors observed in the emergence, large open field, and novel object tasks. This choice also outweighed the choice of keeping the predator in a completely separate tank, only utilizing visual exposure. There was concern that visual exposure may not be a large enough threat and fear responses would be minimal. Placing the predator within the same
tank ensured chemical and visual exposure, thus potentially eliciting the strongest behavioral responses.

Following all behavioral task trials, a small section of the upper or lower corner of the focal fish caudal fin was clipped for identification in further studies. Eventually, all zebrafish were sacrificed.

Behavioral Measures

Zebrafish behaviors have been documented in many studies. Various behaviors have been targeted for the purpose of looking at pharmacological influences on behavior, but a complete ethogram (i.e., list of species-specific motor and posture patterns) has not been established. Therefore, the behavioral measures chosen were decided on from what was reported in the literature combined with what was reasonable with our equipment based on pilot studies conducted prior to this study.
Table 2

*Summary of Behaviors Recorded, Operational Definitions and Measurement Criteria Used*

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Behavior</th>
<th>Definition</th>
<th>Criteria for Measurement</th>
<th>With Respect To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVE&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Percent time of active movement</td>
<td>Fins are engaged and the fish is moving through the water column</td>
<td>Any movement where fins are moving is considered active movement</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>AMB&lt;sub&gt;r&lt;/sub&gt;</td>
<td>Ambulation rate</td>
<td>Average number of squares entered per minute</td>
<td>Mouth, eyes &amp; gills are the only body parts that move</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>FREEZE&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of time freezing</td>
<td>Complete lack of movement in any direction</td>
<td>Peck fins may or may not still be in motion</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>FREEZE&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of freezing</td>
<td>Defined as above</td>
<td>Defined as above</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>DRIFT&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of time drifting</td>
<td>Slow movement through the water column in which the caudal fin stops moving, but the fish has not stopped forward momentum.</td>
<td>Squares were considered to have been entered if the fish moved at least half way into a given square</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>DRIFT&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of drift</td>
<td>Defined as above</td>
<td>Defined as above</td>
<td>OFs, MR, OFlg, NO, PR</td>
</tr>
<tr>
<td>EXP&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of environment explored</td>
<td>Measured as the percent of total number of squares entered</td>
<td>Considered to have inspected if fish body moves half way or more into the column closest to stimulus. Behavior ends when fish moves at least half way out.</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>INSPT&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of time inspecting</td>
<td>Time spent within the column closest to stimulus (3.25cm from stimulus)</td>
<td>Considered to have approached if fish body moves half way or more into the second column closest to stimulus. Behavior ends when fish moves at least half way out.</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>APP&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of time spent approaching</td>
<td>Time spent within the column second closest to stimulus (6.5 cm from stimulus)</td>
<td>Same as above for inspect</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>INSPT&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of inspection</td>
<td>Defined as above</td>
<td>Same as above for inspect</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>LatINSPT</td>
<td>Latency to inspect (sec)</td>
<td>Latency to move within the column closest to stimulus (3.25cm)</td>
<td>Same as above for approach</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>LatAPP</td>
<td>Latency to approach (sec)</td>
<td>Latency to move within the second column closest to stimulus (6.5 cm)</td>
<td>Same as above for approach</td>
<td>MR, NO, PR</td>
</tr>
</tbody>
</table>
Table 2 (continued).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Behavior</th>
<th>Definition</th>
<th>Criteria for Measurement</th>
<th>With Respect To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>LatEMRG</td>
<td>Latency to emerge</td>
<td>Time it takes to move through a narrow opening into a novel environment</td>
<td>Start time recorded when the door to the opening was half way out of the tank. End time recorded when the fish body was at least half way through the open door.</td>
<td>Emergence</td>
</tr>
<tr>
<td>THRASH&lt;sub&gt;p&lt;/sub&gt; (% thrashing)</td>
<td>Proportion of total time thrashing;</td>
<td>Often very rapid and continuous caudal fin movements back and forth along the mirror. Often nose is pressed up against mirror while moving.</td>
<td>A minimum of 1 sec in duration and movement back and forth at least twice is required to be considered “thrashing”. One pass along mirror is insufficient. Behavior ends when fish moves half way out of column closest to mirror.</td>
<td>MR</td>
</tr>
<tr>
<td>THRASH&lt;sub&gt;f&lt;/sub&gt; (frequency thrashing)</td>
<td>Frequency of thrashing</td>
<td>Defined as above</td>
<td>Same as above for thrash</td>
<td>MR</td>
</tr>
<tr>
<td>ERRATIC</td>
<td>Frequency of erratic movement</td>
<td>Seemingly aimless zigzag movement with frequent directional changes</td>
<td>Lasts for &gt; 1 sec (different from DART). Is unpredictable movement.</td>
<td>OFs, OFlg</td>
</tr>
<tr>
<td>DART</td>
<td>Frequency of Darting</td>
<td>Very brief bouts of obvious rapid movement</td>
<td>Lasts for &lt; 1 sec; Not limited to straight movement</td>
<td>OFs, OFlg</td>
</tr>
</tbody>
</table>

*Note*: The tasks listed under the “with respect to” column indicate which task(s) each behavior was measured in. Measurements across tasks allowed for comparisons across contexts, but not all behaviors were collected across all tasks.
The measures employed were a mixture of frequencies and durations of behavior. Table 2 presents a summary of the behaviors examined and definitions and criteria used to gain reliability for coding. Behaviors were coded using either using an Excel spreadsheet or via Jwatcher (Blumstein & Daniel, 2007). The Jwatcher program is designed for the user to be able to assign key strokes to each behavior of interest. One can design the codes to be entered in such a way that codes are mutually exclusive, effectively turning each other on and off in order to obtain duration information. Single key strokes can also be analyzed as events. In this way, the program was set up to automatically calculate durations and events of desired behaviors once they were recorded, in sequential order. The program also provided a summary of basic descriptive statistics and offered a Kappa reliability calculation feature which was utilized for all behaviors captured in this program. The behaviors presented here (and in Table 2) represent the final collection of behaviors examined; they were chosen as the most important for the experimental tasks and as those with the most empirical support available from the literature.

**Activity and Exploration**

*Percent activity* and *ambulation* scores are commonly used in open-field tasks as indicators of exploration. *Percent activity* (MOVE<sub>p</sub>) is defined as the percentage of total time spent engaged in active movement, propelling the fish through the water column in any direction. Therefore, the opposite of this is *inactivity* or *freezing* (FREEZE<sub>p</sub>): the proportion of time spent with complete lack of movement in any direction with the exception to the eyes, gills and mouth (e.g., also defined in Blaser & Gerlai, 2006; Egan et al., 2009). Freezing behaviors are a common response by many species in the face of
predators (Burns, 2008) and so may be representative of a fear response. Another category of movement was determined from pilot studies in which the caudal fin stopped moving (stopped actively propelling the fish forward) yet forward drifting continued for a few seconds before stopping movement completely. This seemed similar to results reported by Blaser and Gerlai (2006) who described a *creeping* movement defined as < 1cm/sec speed. Unfortunately, this behavior was determined with tracking equipment that was unavailable. Therefore, a measure of proportion of drifting (DRIFT<sub>p</sub>) was added to account for this type of movement. Frequencies for drifting (DRIFT<sub>f</sub>) and freezing (FREEZE<sub>f</sub>) were also measured. These behaviors were all recorded across all the main five tasks (OFs, MR, OFlg, NO, and PR) and were coded using Jwatcher (note; reference to these tasks will continue to as short-hand abbreviations: OFs, MR, OFlg, NO and PR). Also, these are continually referred to as the five main tasks, because the emergence task only measured one behavior (latency to emerge) whereas the rest of the tasks allowed for multiple behaviors to be recorded).

It is important to not only capture percent of total activity, but also the rate of activity, or an *ambulation score* (AMB<sub>r</sub>). Commonly, a grid is marked out beneath the tank and counts of line-crossings in a given treatment are recorded from a top-down perspective in order to best quantify exploratory behaviors (Budaev, 1997; Burns, 2008; Russell, 1983). This was modified slightly so that two pieces of information could be obtained at the same time. Assigning each box in the grid a number, and keeping track of each box that the fish enters, enabled the collection of the total amount of the environment the fish had utilized or explored. Therefore *exploration* (EXP) was defined
as the percent of total boxes entered in a treatment period. AMB and EXP were measured across all five of the main tasks and coded using Excel.

**Aggression**

Gerlai et al. (2000) describe and define an *aggressive display* as “a posture during which the fish erects its dorsal, caudal, pectoral and anal fins” (p. 775). They further describe this behavior as commonly “associated with undulating body movements or small slaps carried out by the caudal fin” (p. 775). It has been difficult to see fin movements or erections with our camera equipment, however undulating body movements have been seen throughout pilot experiments. Regardless of fin movement though, this undulating behavior is consistent with definition of *thrashing* (forceful back and forth swimming against a wall of the tank), commonly recorded in zebrafish behavioral observations for this task (Blaser & Gerlai, 2006; Gerlai, 2003; Gerlai et al., 2000). Therefore this behavior was incorporated and labeled *thrash* to be consistent with the literature. Furthermore, this behavior is characterized by considerable duration so that both duration (THRASH\textsubscript{p} for proportion of time) and frequency (THRASH\textsubscript{f}) were recorded.

**Boldness and Fear**

Unlike concepts such as aggression or exploration, which have been studied extensively, *boldness* is a less-accepted and less-widely used construct in research and is much less straightforward to define and measure in animals. However, if present in a species, bold and shy traits have the potential to influence ecologically important behaviors over an individual’s lifespan and thus should be included when examining personality traits. In animal research, boldness has been commonly associated or even
equated with *fearlessness, exploration* or *activity*, which may be very difficult to tease apart. This project aims to keep novelty as the primary focus for investigating boldness. It is expected and generally accepted by animal researchers that shy individuals will respond to unfamiliar objects or situations by fleeing, retreating, becoming cautious, quiet or inactive. Bold individuals on the other hand do not show these responses or show the opposite behavior (i.e. moving towards, becoming active, exploring and investigating) with the same novel object or in the same novel situation (Toms et al., 2010; Wilson, Coleman, Clark, & Biederman, 1993). Thus, these are the criteria that are used to identify boldness in zebrafish for this project. As described earlier, both the emergence and novel object task are expected to capture bold and shy behaviors and positive correlations of behavior across these contexts would support this.

*Erratic* movement is a behavior often included in fish behavioral research and has been described as one of the behavioral responses to ataxia (stress) (Ross & Ross, 2008). It has been described for zebrafish by Blaser and Gerlai (2006) as “fast and seemingly aimless zig-zagging with frequent changes in direction of swimming” (p. 459). They found this behavior to be present across open field, social preference, aggression, and predator tasks; however, it was most prevalent in open-field tasks. Egan et al. (2009) also described this behavior as “sharp changes in direction or velocity and repeated rapid darting movements” (p. 39). Erratic movement (ERRATIC) has been identified in pilot studies but has only been seen as a continuous movement of rapidly changing direction. For this project it is defined in the same way as Blaser and Gerlai (2006), but for simplicity and due to the relatively short duration of the behavior, it is quantified only as a frequency instead of a duration as they did. Pilot studies demonstrated that a difference
exists between this behavior and *darting* (DART) which is very brief in duration and is usually, although not limited to, a movement in a single direction. Both of these behaviors were originally going to be coded across the five main tasks (OFs, MR, OFlg, NO, and PR). However, these were the two hardest behaviors to get reliable data on, and coding them as a part of a task that involved any other potential influences on behavior (e.g., predator or object) proved too challenging to detect. Therefore, these behaviors were collected only for the small and large open field conditions.

*Distance-from-stimulus* or *latency-to-approach* measures are commonly used in experiments to determine how drugs or other treatment effects influence zebrafish behavior. Measuring the *latency-to-approach* a mirror has been used as an index of aggression (Blaser & Gerlai, 2006; Gerlai, 2003). Similarly, researchers have reported using proportion of time fish spent close to a predator as to examine fear responses (e.g., Blaser & Gerlai, 2006; Gerlai et al., 2006). Zebrafish have been shown to exhibit horizontal preferences for the side of the tank opposite to the introduction of alarm substances (Waldman, 1982). On the other hand, many species of fish actually approach predators instead of avoiding them, a behavior that seems maladaptive (Walling, Dawnay, Kazem, & Wright, 2004), but may represent differences in individuals. This type of measurement can also be utilized in novel object task to discern between bold and shy individuals, in which case this allows for a direct comparison with the *latency-to-emerge* measure from the emergence task.

Therefore *proportion-of-time* spent close to stimuli, *frequency-of-approach* and *latency-to-approach* was measured for the mirror, novel object and predator tasks. The grids placed at the bottom of the tank for measuring ambulation provide a convenient
way to measure these. The line bordering the column of squares closest to any stimuli is 3.25 cm distance from the stimuli for any of the tasks. The column second closest to any given stimuli is at 6.25 cm distance. Pilot work demonstrated that fish often go through a series of approaches to the mirror and predator before getting too close, and fish often inspect the mirror many times before thrashing. The predator task demonstrated that some fish may never even approach the predator within the closest column at any point in the experiment, making it necessary to incorporate the second square into the observations. These observations led to classification of approaches to the second column possibly representing quicker approaches and compared to movement within the first column as more like inspection behavior that may not happen as quickly as approaches but might be indicative of more bold or fearless fish. This was an idea as a way to easily distinguish between the two, as opposed to calling them approach-to-column-1 and approach-to-column-2. Therefore a fish was considered to have approached a stimulus if it came within the second column next to the stimuli (see Figure 4) and proportion of time spent and latency to approach were measured for each (APP, and LatAPP, respectively). A fish was considered to have inspected if it came within the column closest to the stimuli (see Figure 4) and proportion of time, latency to inspect and frequency were recorded (INSPT, LatINSPT and INSP, respectfully). Frequency to approach (column 2) was not recorded due to limitations of the software setup for that behavioral coding scheme.

The above latency measures will be useful for comparing with the Emergence task. The only measurement for this task is latency-to-emerge (LatEMRG) into a novel environment as defined in Table 2.
Figure 4. Explanation of INSPT and APP behavior. (A) Mirror task in 5-gallon tank; (B) Novel object task with pebble in 10-gallon tank; (C) Predator exposure task in 10-gallon tank. In the mirror and predator tasks, column 1 (INSPT behavior) was the column closest to the respective stimulus. Column 2 (APP behavior) was the column second-closest to the respective stimulus. For the novel object task, INSPT behavior was recorded when the fish cross into squares 30, 31, 37 & 38 (within a 1-square-distance) and APP behavior was recorded when fish crossed into squared 22-25, 29, 32, 36, 39, and 40-46 (within a 2-square-distance).
CHAPTER III
RESULTS

Reliability

Ambulation and Exploration

Ambulation and exploration were collected in an excel database. Activity was recorded as a running list of squares crossed, broken down per minute. The excel sheet was designed to then automatically calculate the percentage of squares entered in order to obtain exploration information (i.e., if a zebrafish went into a square at least once, it was calculated as “present”). Therefore, reliable ambulation recordings would automatically indicate that exploration data were also reliable. All of the data were coded by one of four coders. In order to determine if coding was reliable, twenty percent of the data were coded by pairs of raters, with attempts made to have every pair code an equal number of times. Average percent agreement across pairs of coders was 93%. Additionally, in order to ensure high agreement remained across all four coders, of this twenty percent, a small subset of data was examined by all four coders. Average percent agreement across all coders on all tasks was 89%. Due to the incredibly high agreement we achieved across coders, an official reliability statistic was unnecessary. In fact, percent agreement ended up being a somewhat conservative reliability approach, because comparisons across raters were made line by line for the all the squares entered for a given five minute task.

All Other Behavioral Data Recorded Using Jwatcher

Twenty percent of all remaining behaviors were collected and analyzed for inter-rater reliability using Jwatcher, which has a built-in function to calculate a Kappa reliability statistic. The program was designed up to conduct reliability between any two
coded files. A single coded file was a collection of all the behaviors recorded for a given task, in order, for the duration of that task by a single rater. In cases where there were numerous recordings of a given behavior, the Jwatcher Kappa statistic was a very conservative measure of reliability, conducting a line-by-line comparison of any recorded key-press. The following behaviors were included in these interrater reliability calculations: INSPT<sub>p</sub>, APP<sub>p</sub>, INSPT<sub>f</sub>, LatINSPT, and LatAPP. For the mirror task, THRASH<sub>p</sub> and THRASH<sub>f</sub> were also included in the interrater reliability calculations, and Kappa = .95, averaged across each compared file, with an average of 96% agreement. For the novel object task, Kappa = .949, averaged across each compared file, with an average of 96.42% agreement. For the predator exposure task, Kappa = .949, averaged across each compared file, with an average of 98.63% agreement. All of these are considered to be in the “almost perfect” range (Landis & Koch, 1977).

There were several behaviors for which there were very few recordings in any given file, such that the Kappa reported above was not representative of these behaviors. In these cases, behaviors were grouped by behavior instead of by behavioral task, and a Kappa was conducted in SPSS for the behaviors across all tasks for which it was coded. The interrater reliability for MOVE<sub>p</sub>, DRIFT<sub>p</sub>, DRIFT<sub>f</sub>, FREEZE<sub>p</sub>, and FREEZE<sub>f</sub> was found to be Kappa = .793 (p < .001), 95% CI (0.68, 0.91), a “substantial” amount of agreement. Unfortunately, DART and ERRATIC behaviors were not reliable (Kappa = -.002, p = .975).

It is also important to note that Jwatcher calculates durations of behavior from start and stop key entries on the computer. Therefore, the reliability for any duration measure was calculated based on reliable start and end times, with the guideline that any
two recordings of durations were considered the same if they were entered within one second of each other at the start and end of the behavior (e.g., for THRASHp, both raters had to hit the start key “t” within one second of each other and the end key “E” within one second of each other to consider that behavior as coded the same at a given point in the task).

**Double Checking Data**

All data were double checked for errors. Ambulation (and therefore exploration) data were double checked using a Matlab program created for the purpose of double checking the excel database for any illogical entries. Jwatcher data were double checked through a built in feature of the program, based on the rules set up for each respective behavioral coding scheme.

**Exploratory Data Analysis**

All data were initially examined via exploratory data analyses and tested for normality via the Kolmogorov-Smirnov test. Non-normal variables were considered for transformation. Square root, logarithmic (base 10), or inverse transformations were made depending on severity of skew and kurtosis and subsequent re-evaluation of normality. In cases of extreme skew (e.g., MOVEp) the option to dichotomized variables was explored (e.g., those active 100% of the time vs. those that were active for less that 100% of the time). Due to the large number of behavioral variables, the transformation that best suited any one variable did not necessarily suite another, making comparisons challenging. For example, trial one of a behavior may have required a square root transformation, but trial two was already normal. Therefore in order to compare these, one either had to conduct a non-parametric test on the raw data or transform the already normal variable into
something that was thrown away from normality, which may still have required a non-parametric test. In the end, transformations were used when appropriate, but largely the raw data were utilized, double checking results with transformed data for comparison. There results were rarely much different so the results from the raw data were are largely reported.

There were very few occurrences of freeze and drift behaviors within a given task or trial. Because of this, these variables do not lend themselves to statistical analysis. If only the individuals that exhibited these behaviors at least once ever were included, this resulted in a sample size of only $n = 13$. Even if a rank ordered test were utilized, or variables were dichotomized, there were more tied cases where the behavior did not occur across trials or treatments, making statistics useless for this data set. Therefore, freezing and drifting behaviors were not included in further analyses, however, they are still important to examine descriptively (see Table 3 and Figures 5-6).
Table 3

Raw Frequency and Proportion Data for Freezing and Drift Behaviors

<table>
<thead>
<tr>
<th></th>
<th>Sm. Open Field</th>
<th>Mirror Task</th>
<th>Lg. Open Field</th>
<th>Novel Object</th>
<th>Predator Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>TOT</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>FREEZE&lt;sup&gt;a&lt;/sup&gt; (frequency)</td>
<td>1/3</td>
<td>3/3</td>
<td>4/6</td>
<td>2/3</td>
<td>3/4</td>
</tr>
<tr>
<td>FREEZE&lt;sup&gt;c&lt;/sup&gt; (% time freeze)</td>
<td>17.25</td>
<td>5.09-94.18</td>
<td>AVG&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.43-23.62</td>
<td>1.3-77.56</td>
</tr>
<tr>
<td>AVG&lt;sup&gt;d&lt;/sup&gt;</td>
<td>17.25</td>
<td>57.54</td>
<td>13.53</td>
<td>42.99</td>
<td>58.36</td>
</tr>
<tr>
<td>DRIFT&lt;sup&gt;c&lt;/sup&gt; (% time drift)</td>
<td>1.07-</td>
<td>2.32-</td>
<td>28.13</td>
<td>87.97</td>
<td>16.75</td>
</tr>
<tr>
<td>AVG&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2.62</td>
<td>8.98</td>
<td>2.56</td>
<td>8.02</td>
<td>1.15</td>
</tr>
</tbody>
</table>

<sup>Note. </sup><sup>a</sup> The first number is the number of fish that exhibited the behavior in that trial for that task. The second number is the total number of times that behavior was observed in that trial for that task;<br><sup>b</sup> AVG: Average proportion of time across trials for given task;<br><sup>c</sup> TOT: Total number of fish that were observed in the behavior across trials & total number of times that behavior was observed across trials;<br><sup>d</sup> Raw data presented as the range of percent time spent in the given behavior for that trial; All values calculated only for fish that actually conducted the behavior at least once in a given trial.
Figure 5. (A) Average percent time fish spent freezing in trials 1 and 2 for each treatment condition; (B) Average time fish spent drifting in trials 1 and 2 each treatment condition; Data presented here are raw data for fish that were ever observed exhibiting these behaviors at least once (n = 13).
Figure 6. Average percent time fish spent drifting and freezing in trial 1 for each treatment condition. Data presented here are raw data for fish that were ever observed exhibiting these behaviors at least once ($n = 13$).

Table 3 presents the frequencies, range of percentages and summaries across trials for both freezing and darting behaviors for the fish that ever exhibited any of these behaviors at any point in the experiments. For each trial, the frequencies are broken down into number of fish that exhibited the behavior and the number of times that behavior occurred across fish within that trial for the given task. The totals are similarly presented as the total number of fish that were observed for that behavior across both trials and the total number of times the behavior occurred across the two trials. The proportions are presented as the range found for that behavior in a given trial, for a given condition. Averages across trials and averages for each trial, in a given task, are also reported.
Of the thirty fish examined, only thirteen drifted or froze at any point throughout experiments, leaving seventeen that were active 100% of the time, across all tasks, across both trials. Of those that froze or drifted, there were only two instances where a fish exhibited freezing or drifting only once ever. The rest were a mixture of freezing and drifts across contexts and trials. However, there was one instance of a fish that drifted and froze in every trial across every task (except trial 1 of the OFs task). The data does suggest that freezing may occur less frequency but for longer durations than drifting. Also interesting to note is a general trend for an increase in the time spent either drifting or freezing over time (from trial 1 to trial 2), potentially indicating an exposure effect (also see Figure 5).

For further analyses, behaviors were kept separate for each task for any comparisons. For example, there is no certainty that the latency to approach a mirror is a similar behavioral response representing the same thing as the latency to approach a predator (in fact, it would be expected to be unrelated). This is true with any behavior examined in this project, therefore behaviors were not collapsed (e.g., exploration data were not collapsed across contexts and then comparing across time, but instead examined across time for each context).

For the mirror task treatment condition, proportion of time inspecting (i.e., the time spent in column one closest to the mirror; INSPTₚ) also included the proportion of time spent thrashing (THRASHₚ). This was due to the nature of the design where rapid thrashing behavior occurred right up against the mirror which was positioned along column one. In order to better compare this variable to the similar measures in other tasks (i.e., INSPTₚ for predator and novel object tasks) the proportion of time thrashing was
subtracted from the proportion of time inspecting, leaving me with just the time fish spent closest to mirror when NOT thrashing. One can see in Figure 7, that this change resulted in quite a large decrease in magnitude of the average proportion of time spent inspecting for the mirror task (levels that were more similar to the proportions observed in other tasks), but the relationship between the average proportion if time inspecting versus approaching did not change; that is, the average time inspecting was still higher than that for approaching. The frequency variable (INSPT<sub>f</sub>) could not be separated out this way since for any single event where a fish moved close to the mirror, there were potentially several thrashing events. The differences in which INSPT<sub>p</sub> variable (with or without thrashing included) was used in the mirror task are noted as required throughout the remainder of the text.

The remaining data analyses were conducted in three major steps: (a) to examine consistency across time, (b) to examine consistency across context, and (c) to examine the relationship between behaviors and examine evidence for the constructs of interest.

**Examining Relationships across Time**

**Consistency across Time**

*Statistical approach.* Most behaviors were examined for consistency across time by conducting correlations between trial 1 and trial 2. Although, a Pearson product-moment correlation is the parametric and most common correlation statistic, it is sensitive to changes in magnitude between two conditions. Because it uses ranked data, a Spearman rank order correlation comparison is better suited for investigating whether or not the relative standing of individuals remains consistent across trials, regardless of changes in magnitude. This happens to also be a non-parametric test, the results of which
remained the same regardless of whether my data were normal or not. However, because Pearson’s is more widely accepted, this test was conducted only if data were originally normal, or for comparison on transformed normalized data. Table 4 presents the results for Spearman rank order and Pearson correlations.

Table 4

**Consistency across Time (Trials 1 and 2) for Each Behavior on Each Task; Data Prior to Correction for Multiple Tests**

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Sm. Open Field</th>
<th>Mirror Task</th>
<th>Lg. Open Field</th>
<th>Novel Object</th>
<th>Predator Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVE&lt;sub&gt;p&lt;/sub&gt;</td>
<td>K = 0.286</td>
<td>K = 0.143</td>
<td>K = 0.634&lt;sup&gt;***&lt;/sup&gt;</td>
<td>K = 0.423&lt;sup&gt;*&lt;/sup&gt;</td>
<td>K = 0.392&lt;sup&gt;′&lt;/sup&gt;</td>
</tr>
<tr>
<td>AMB&lt;sub&gt;r&lt;/sub&gt;</td>
<td>.415&lt;sup&gt;′&lt;/sup&gt;/ .280</td>
<td>.363&lt;sup&gt;′&lt;/sup&gt;/ .214</td>
<td>.823&lt;sup&gt;<em><strong>&lt;/sup&gt;/ .736&lt;sup&gt;</strong></em>&lt;/sup&gt;</td>
<td>.579&lt;sup&gt;<em><strong>&lt;/sup&gt;/ .678&lt;sup&gt;</strong></em>&lt;/sup&gt;</td>
<td>.727&lt;sup&gt;<em><strong>&lt;/sup&gt;/ .646&lt;sup&gt;</strong></em>&lt;/sup&gt;</td>
</tr>
<tr>
<td>EXP&lt;sub&gt;p&lt;/sub&gt;</td>
<td>.306&lt;sup&gt;′&lt;/sup&gt;</td>
<td>.170/ .095</td>
<td>.404&lt;sup&gt;′&lt;/sup&gt;</td>
<td>.695&lt;sup&gt;***&lt;/sup&gt;</td>
<td>.713&lt;sup&gt;<em><strong>&lt;/sup&gt;/ .723&lt;sup&gt;</strong></em>&lt;/sup&gt;</td>
</tr>
<tr>
<td>INSPT&lt;sub&gt;p&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n/a</td>
<td>.446&lt;sup&gt;***&lt;/sup&gt;/ .407&lt;sup&gt;′&lt;/sup&gt;</td>
<td>n/a</td>
<td>.603&lt;sup&gt;***&lt;/sup&gt;</td>
<td>.609&lt;sup&gt;***&lt;/sup&gt;</td>
</tr>
<tr>
<td>APP&lt;sub&gt;p&lt;/sub&gt;</td>
<td>n/a</td>
<td>.252</td>
<td>n/a</td>
<td>-.269/ -.387&lt;sup&gt;′&lt;/sup&gt;</td>
<td>.447&lt;sup&gt;*<strong>&lt;/sup&gt;/ .554&lt;sup&gt;</strong>&lt;/sup&gt;</td>
</tr>
<tr>
<td>INSPT&lt;sub&gt;f&lt;/sub&gt;</td>
<td>n/a</td>
<td>.340&lt;sup&gt;′&lt;/sup&gt;/ .274</td>
<td>n/a</td>
<td>.506&lt;sup&gt;<strong>&lt;/sup&gt;/ .470&lt;sup&gt;</strong>&lt;/sup&gt;</td>
<td>.591&lt;sup&gt;***&lt;/sup&gt;</td>
</tr>
<tr>
<td>LatINSPT</td>
<td>n/a</td>
<td>.331&lt;sup&gt;′&lt;/sup&gt;/ .351&lt;sup&gt;′&lt;/sup&gt;</td>
<td>n/a</td>
<td>.383&lt;sup&gt;′&lt;/sup&gt;/ .345&lt;sup&gt;′&lt;/sup&gt;</td>
<td>.347&lt;sup&gt;′&lt;/sup&gt;</td>
</tr>
<tr>
<td>LatAPP</td>
<td>n/a</td>
<td>.299</td>
<td>n/a</td>
<td>-.030</td>
<td>.027/.013</td>
</tr>
<tr>
<td>DART</td>
<td>.005/.032</td>
<td>n/a</td>
<td>.334&lt;sup&gt;′&lt;/sup&gt;/ .322&lt;sup&gt;′&lt;/sup&gt;</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ERRATIC</td>
<td>.126</td>
<td>n/a</td>
<td>.706&lt;sup&gt;***&lt;/sup&gt;</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>THARSH&lt;sub&gt;p&lt;/sub&gt;</td>
<td>n/a</td>
<td>.602&lt;sup&gt;***&lt;/sup&gt;</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>THRASH&lt;sub&gt;f&lt;/sub&gt;</td>
<td>n/a</td>
<td>.355&lt;sup&gt;′&lt;/sup&gt;</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>LatEMRG&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>.325&lt;sup&gt;′&lt;/sup&gt;/ .329&lt;sup&gt;′&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Note. <sup>a</sup> Reported for data once thrashing behavior was removed from the data; <sup>b</sup> LatEMRG was only measured for emergence task; All coefficients presented here are prior to Bonferroni correction (N = 30). The number above any given line represents a Spearman rank order correlation coefficient. Anytime normality was not violated with raw or transformed data, a Pearson correlation coefficient was also tested for comparison, and presented here under the diagonal line for any given variable. If data were not normal, even after transformation, Pearson correlations were not tested for comparison due to violated assumptions; * p < .05, ** p < .01, *** p < .001.

The percent activity (MOVE<sub>p</sub>) variables were highly skewed, with most zebrafish observed at one hundred percent activity (as discussed, less than half of the fish ever
demonstrated freezing or drifting at any point). The lack of variability in this behavior limited my ability to conduct a correlation across time. However, for those fish that did drift or freeze (i.e., were not active the whole time) it’s is still of interest to determine if this was consistent. Therefore, these variables were re-coded into dichotomous variables (100% activity vs. those at less than 100% activity, which collapsed DRIFT and FREEZE together). A Kappa statistic was calculated for trial 1 compared to trial 2. Although conventionally designed for examining agreement between raters, a Kappa can be used to examine “agreement across trials,” treating each trial as a rater. These values are also reported in Tables 4 and 5.

Because of the large number of statistical procedures required to examine all behaviors across all tasks, correlational data were corrected for multiple comparisons using a sequential Bonferronni correction (see Rice, 1989). Briefly, this process requires ranking the $p$-values of the tests conducted from least to greatest. Starting with the smallest value, one calculates the corrected significance level as $\alpha/k$, where $k$ is the number of tests. If the test is still significant, one moves to the next smallest $p$-value and recalculates the corrected significance based on how many tests are still left. This continues until one discovers a non-significant case, after which all tests are considered non-significant. Table 5 reports only the Spearman rank order correlation coefficient that remained significant after the corrections were applied (cases that changed in significance are indicated in the table).

Results. In examining Table 5, one can see that most behaviors were consistent over time in each respective treatment condition. Of the thirty-seven comparisons across trials, twenty two were significant and fifteen were non-significant across time. Of the
fifteen that were non-significant, five of these changed from significant to non-significant after the Bonferroni correction (EXP in the small open field task, DART in the large open field, and all three LatINSPT measures).

A surprising result is shown for APP_p and LatAPP in the novel object task. Although insignificant, the correlations across time were negative, indicating a slightly inverse relationship. This may be reflecting differences in responses due to the type of novel object used, which was further explored. There were no differences in the proportion of time inspecting (INSPT_p) or the latency to inspect (LatINSPT) between the three novel objects used (for either trial one or trial two). However, there were notable differences when considering approach behavior (fish coming with the second column closest to the object). There was a difference in the proportion of time in approaching (APP_p) between the three objects used (trial 1: \( H(2) = 12.122, p = .002 \); trial 2: \( H(2) = 7.197, p = .027, \) 2-tailed). Mann-Whitney (2-tailed) tests were used as a posthoc with a Bonferroni corrected significance level of .0167 for trial 1 and 0.025 for trial 2 (only two posthoc comparisons for trial 2). Fish spent less time approaching when the rock was used compared to the weight (trial 1: \( U = 8.00, p = .001, r = -.575 \); trial 2: \( U = 18.5, p = .0162, r = -.43 \)) and compared to the pebble (trial 1: \( U = 10.00, p = .001, r = -.522 \); trial 2: \( U = 27.00, p = .049, r = -.36 \)), but there was no difference when comparing the weight and pebble (trial 1: \( U = 49.00, p = .705 \)).
Based on these findings, the rock seemed to have created a difference in reaction compared to the weight and the pebble. To investigate this difference a little farther the relationships of LatAPP and INSPT\textsubscript{f} were also examined. Only trial one was examined, since these are the data that were further used in analyses. Zebrafish did not inspect the rock nearly as frequently as the other two objects (frequencies = 32\textsubscript{R}, 78\textsubscript{w} and 90\textsubscript{p}). Finally, fish took a significantly longer time to approach (LatAPP) the rock than the weight (\(U = 15.50, p = .008, r = -.472\)) but not the pebble (after the Bonferroni correction; \(U = 20.500, p = .045, r = -.365; (H(2) = 8.361, p = .015)\). Since fish were

### Table 5

**Consistency across Time (Trials 1 and 2) for Each Behavior on Each Task; Final Results after Correction for Multiple Tests**

<table>
<thead>
<tr>
<th>Task</th>
<th>Sm. Open Field</th>
<th>Mirror Task</th>
<th>Lg. Open Field</th>
<th>Novel Object Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVE\textsubscript{p}</td>
<td>K = 0.286</td>
<td>K = 0.143</td>
<td>K = 0.634***</td>
<td>K = 0.423*</td>
</tr>
<tr>
<td>AMB\textsubscript{r}</td>
<td>.415*</td>
<td>.363*</td>
<td>.823***</td>
<td>.579***</td>
</tr>
<tr>
<td>EXP</td>
<td>.306</td>
<td>.170</td>
<td>.404*</td>
<td>.695***</td>
</tr>
<tr>
<td>INSPT\textsubscript{p}</td>
<td>n/a</td>
<td>.446***</td>
<td>n/a</td>
<td>.603***</td>
</tr>
<tr>
<td>APP\textsubscript{p}</td>
<td>n/a</td>
<td>.252</td>
<td>n/a</td>
<td>-.269</td>
</tr>
<tr>
<td>INSPT\textsubscript{f}</td>
<td>n/a</td>
<td>.340*</td>
<td>n/a</td>
<td>.506**</td>
</tr>
<tr>
<td>LatINSPT</td>
<td>n/a</td>
<td>.331</td>
<td>n/a</td>
<td>.383</td>
</tr>
<tr>
<td>LatAPP</td>
<td>n/a</td>
<td>.299</td>
<td>n/a</td>
<td>-.030</td>
</tr>
<tr>
<td>DART</td>
<td>.005</td>
<td>n/a</td>
<td>.334</td>
<td>n/a</td>
</tr>
<tr>
<td>ERRATIC</td>
<td>.126</td>
<td>n/a</td>
<td>.706***</td>
<td>n/a</td>
</tr>
<tr>
<td>THRASH\textsubscript{p}</td>
<td>n/a</td>
<td>.602***</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>THRASH\textsubscript{f}</td>
<td>n/a</td>
<td>.355*</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>LatEMRG</td>
<td>.325*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *Indicates a change in significance after Bonferroni correction; Only Spearman rank order correlation coefficients are reported here; \(N = 30; * p < .05, ** p < .01, *** p < .001.\)
exposed to different objects in trials one and two, these findings may help to explain why LatINSPT and LatAPP variables were not found to be significantly correlated across time (Table 5).

Although MOVE was consistent over time for three of the five treatment conditions (OFlg, NO, & PR), there was such little variability within this behavior, such that it is unlikely to capture information useful for examining personality. Therefore it was removed from further analyses. This was further justified by the fact that ambulation (AMB) rate was normally distributed across all tasks, providing a better, more variable behavioral measure to examine activity with.

Differences in Magnitude across Time

Statistical approach. Before examining consistency across contexts it was important to address whether or not there were changes in the magnitude of behavioral responses across time that might indicate exposure effects. Table 6 presents a summary of the results from these examinations. In order to reduce the number of paired comparisons made for all behaviors, data were first examined visually to determine if a significance test was appropriate. In many cases it was visually obvious from bar graphs and boxplots that a test was not needed to determine that there was no difference between trials. These instances are indicated by “n.s.” in the table. For data that were normal originally (i.e., AMB variables- across all tasks and EXP for predator exposure only), $t$-tests were used for pairwise comparisons. In cases where transformations did result in normal data, $t$-tests were conducted on the transformed data and Wilcoxon signed rank tests (nonparametric) were used on raw data for comparison. There was no instance where a $t$-test was
significant for transformed, normalized data but a Wilcoxon test was not significant with raw data. Therefore, Wilcoxon tests are reported for raw, non-normal data.

Table 6

Results from Examining Differences in Magnitude across Time (Trials 1 and 2) for Each Behavior on Each Task

<table>
<thead>
<tr>
<th>Task</th>
<th>Sm. Open Field</th>
<th>Mirror Task</th>
<th>Lg. Open Field</th>
<th>Novel Object</th>
<th>Predator Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMB, a</td>
<td>1.067</td>
<td>1.257</td>
<td>3.382** (.53)</td>
<td>1.984</td>
<td>3.522** (.55)</td>
</tr>
<tr>
<td>EXP</td>
<td>n.s. c</td>
<td>-1.244b</td>
<td>n.s. c</td>
<td>-1.007b</td>
<td>.853a</td>
</tr>
<tr>
<td>INSPT, p, a, d</td>
<td>n/a</td>
<td>-1.176</td>
<td>n/a</td>
<td>n.s. c</td>
<td>n.s. c</td>
</tr>
<tr>
<td>APP, b</td>
<td>n/a</td>
<td>n.s. c</td>
<td>n/a</td>
<td>-1.841</td>
<td>n.s. c</td>
</tr>
<tr>
<td>INSPT, f</td>
<td>n/a</td>
<td>-.442</td>
<td>n/a</td>
<td>n.s. c</td>
<td>n.s. c</td>
</tr>
<tr>
<td>LatINSPT, b</td>
<td>n/a</td>
<td>-.319</td>
<td>n/a</td>
<td>-.913</td>
<td>-.848</td>
</tr>
<tr>
<td>LatAPP, b</td>
<td>n/a</td>
<td>n.s. c</td>
<td>n/a</td>
<td>-1.503</td>
<td>-.031</td>
</tr>
<tr>
<td>DART, b</td>
<td>n.s. c</td>
<td>n/a</td>
<td>-1.553</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ERRATIC, b</td>
<td>-1.240</td>
<td>n/a</td>
<td>-1.553</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>THRASH, p, b</td>
<td>n/a</td>
<td>-1.799</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>THRASH, f, b</td>
<td>n/a</td>
<td>-.733</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>LatEMRG, b</td>
<td>-.551</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Final results corrected for multiple comparisons using sequential Bonferroni corrections. All tests are 2-tailed; a t-test on originally normal data. Values in parentheses are effect sizes (r), calculated based on recommendations of Fields (2005); b Wilcoxon Signed Ranks tests used on raw, non-normal data (z-score, exact test). c Data were not examined statistically. See text for further explanation (n.s.: not-significant); d Data reported for proportion of inspecting behavior with proportion of thrashing behavior removed from the variable. See text for more detail; n/a: not available since these behaviors were not recorded for the respective task; * p < .05, ** p < .01, *** p < .001.

Similar to before, correlation data were corrected for multiple comparisons using a sequential Bonferroni correction (Rice, 1989). For the mirror task treatment condition, differences across time are examined and reported in Table 6 using the INSPT, p.
variable where the proportion of thrashing was removed. For reasons discussed earlier, MOVEP, DRIFT and FREEZE were not included in this analysis.

Results. From the results in Table 6, there were only two cases of significant differences in behavior across trials, after corrections were applied, and both were for ambulation rate (AMB_r). For the large open field task, on average, zebrafish had a lower ambulation rate on trial 2 ($M = 137.55$, $SE = 9.27$) than on trial 1 ($M = 161.30$, $SE = 9.97$, $t(29) = 3.38$, $p = .002$). For the predator exposure task, zebrafish also had a lower ambulation rate on average for trial 2 ($M = 149.39$, $SE = 7.34$) compared to trial 1 ($M = 169.79$, $SE = 6.22$, $t(29) = 3.52$, $p = .001$). Both were also large effects ($r = .53$ and $.55$, respectively), representing a substantive finding regardless of the fact that a number of comparisons table-wide (Rice, 1989).

Despite differences in magnitude for AMB_r variables, these results indicate that on the whole, potential exposure effects were not a concern. Data could be averaged across trials for further comparisons. However, it is still preferable to examine initial responses to these tasks to gain the most accurate picture of behavioral responses. Therefore, trial one was used for further comparisons across tasks and maintained separately from trial two.

Examining Relationships across Contexts

Consistency across Contexts

Statistical approach. Kendall’s coefficient of concordance ($W$) was used to assess the trend of agreement across multiple contexts. Similar to a Kappa, this is typically used as a reliability statistic for agreement across multiple raters, but one can utilize it to look at “agreement” across treatments (Field, 2005). Similar to a Spearman rank order
correlation, this test works with rank ordered data, and Kendall’s W coefficient ranges from 0 to 1, where 1 indicates complete agreement. The Kendall’s W test is designed for looking at agreement across independent raters, whereas my data is dependent. However, since using an independent test on dependent data results in a loss of power to detect differences, this will actually work to increase confident in any significant results found (R. Mohn, personal communication, May 1, 2011).

Results. Table 7 presents a summary of the results obtained from examining consistency across contexts for each behavior of interest. Results indicate that all but two behaviors were consistent across treatment conditions (standard Bonferronni correction for 7 tests results in an α-level criteria of .007. All p – values were less than .001). DART and ERRATIC were the only two behaviors that were not consistent across the treatments they were measured for (OFs and OFlg).

Differences in Magnitude across Contexts

Statistical approach. Behaviors were examined for differences across tasks. A repeated measures ANOVA was used for data that were normal (i.e., AMBr). A nonparametric Freidman’s ANOVA was used for non-normal data. A Bonferronni posthoc test was used (with corrections) for pairwise comparisons after a significant RMANOVA. Wilcoxon signed rank tests were used for pairwise posthoc comparisons following a significant Freidman’s ANOVA (with a sequential Bonferronni correction for multiple tests). The results for the main effects are presented in Table 7. Posthoc pairwise comparisons are summarized in Tables 8-9 and discussed more in the following sections.
Table 7

Results for Examining Main Effects of Differences in Magnitude across Contexts and for Examining Consistency across Contexts

Main Effects for Difference Across Treatments<sup>a</sup> | Consistency Across Contexts<sup>b</sup>
--- | ---
| Behavior | Test Statistic | p-value | Kendall’s W | Test Statistic | p-value |
| c AMB<sub>r</sub> | $F(4) = 18.954$ | <.001<sup>***</sup> | .403 | $\chi^2(4) = 48.41$ | <.000<sup>***</sup> |
| d EXP | $\chi^2(4) = 51.090$ | <.001<sup>***</sup> | .429 | $\chi^2(4) = 51.09$ | <.001<sup>***</sup> |
| e INSPT<sub>p</sub> | $\chi^2(2) = 41.863$ | <.001<sup>***</sup> | .698 | $\chi^2(2) = 41.86$ | <.001<sup>***</sup> |
| e APP<sub>p</sub> | $\chi^2(2) = 23.27$ | <.001<sup>***</sup> | .388 | $\chi^2(2) = 23.27$ | <.001<sup>***</sup> |
| e INSPT<sub>f</sub> | $\chi^2(2) = 20.123$ | <.001<sup>***</sup> | .335 | $\chi^2(2) = 20.123$ | <.001<sup>***</sup> |
| e LatINSPT | $\chi^2(2) = 17.556$ | <.001<sup>***</sup> | .293 | $\chi^2(2) = 17.56$ | <.001<sup>***</sup> |
| e LatAPP | $\chi^2(2) = 19.467$ | <.001<sup>***</sup> | .324 | $\chi^2(2) = 19.47$ | <.001<sup>***</sup> |
| f DART | -.416 | >.05 | — | $r = .068$ | ns |
| f ERRATIC | -1.150 | >.05 | — | $r = -.036$ | ns |

Note. <sup>a</sup>Results of main effects from examining differences in behaviors across treatments (trial 1 only; N = 30). <sup>b</sup>Also reported are Kendall’s W coefficient of concordance and subsequent results from significance test. Significant values indicate consistency in individual responses across treatments, regardless of any changes in magnitude; <sup>c</sup>Main effect results are from RMANOVA across OFs, MR, OFlg, NO & PR; <sup>d</sup>Main effect results are from Friedmann’s ANOVA across OFs, MR, OFlg, NO & PR; <sup>e</sup>Main effect results are from Freidman’s ANOVA across MR, NO & PR; <sup>f</sup>Main effect results are z-scores from Wilcoxon Sign Rank test (2-tailed, exact test). Consistency across time examined via Spearman rank order correlations across OFs & OFlg (1-tailed); ns: not significant.
Table 8

*Posthoc Results from Tests Indicating Significant Differences in Behaviors across Treatments (for Trial 1 only)*

<table>
<thead>
<tr>
<th>Behavior</th>
<th>OFs</th>
<th>MR</th>
<th>OFlg</th>
<th>NO</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambulation (AMBr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR M = 114.58</td>
<td>-46.72**</td>
<td>-23.740</td>
<td>-55.207***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFlg M =161.30</td>
<td>22.98</td>
<td>-8.487</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO M =138.32</td>
<td>-31.467***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR M = 169.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Exploration (EXP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFs Mdn = 97.00</td>
<td>-4.387*** (-.36)</td>
<td>- .887</td>
<td>-1.901</td>
<td>-4.165*** (-.34)</td>
<td></td>
</tr>
<tr>
<td>MR Mdn = 78.00</td>
<td>-4.034*** (-.33)</td>
<td>-2.273</td>
<td>-1.451</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFlg Mdn = 100.00</td>
<td>-2.592</td>
<td>-4.026*** (-.33)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO Mdn = 95.00</td>
<td></td>
<td>-2.455</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR Mdn = 87.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. a Bonferroni posthoc comparisons (with correction) after significant RMANOVA. Mean values for each task reported along the diagonal; b Wilcoxon Signed Ranks post hoc tests (z-scores, 2tailed, exact test) after significant Friedman’s ANOVA. Significance was corrected for multiple comparisons via a Bonferroni correction (10 comparisons). Effect sizes (r) are presented in parentheses, calculated from the Wilcoxon tests as \( r = z / \sqrt{n} \), where n is the number of observations (Fields, 2005). Median values for each task are presented along the diagonal. * p < .05, ** p < .01, *** p < .001.*
Table 9

Results of Pairwise Posthoc Comparisons after Significant Friedman’s ANOVA

<table>
<thead>
<tr>
<th>Behavior</th>
<th>MR</th>
<th>NO</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of time within column 1 (INSPT&lt;sub&gt;p&lt;/sub&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>$Mdn = 80.13$</td>
<td>-4.762*** (-.50)</td>
<td>-4.703*** (-.50)</td>
</tr>
<tr>
<td>NO</td>
<td>$Mdn = 0.31$</td>
<td>-1.571</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$Mdn = 0.35$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of time within column 2 (APP&lt;sub&gt;p&lt;/sub&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>$Mdn = 4.94$</td>
<td>-3.024** (-.32)</td>
<td>-2.067 (-.22)</td>
</tr>
<tr>
<td>NO</td>
<td>$Mdn = 16.92$</td>
<td>-4.227*** (-.45)</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$Mdn = 1.97$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of approach to column 1 (INSPT&lt;sub&gt;i&lt;/sub&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>$Mdn = 25$</td>
<td>-4.043*** (-.43)</td>
<td>-3.450*** (-.36)</td>
</tr>
<tr>
<td>NO</td>
<td>$Mdn = 3$</td>
<td>-3.97</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$Mdn = 3.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latency to approach column 1 (LatINSPT&lt;sup&gt;a&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>$Mdn = 5.33$</td>
<td>-3.260** (.34)</td>
<td>-3.860*** (.34)</td>
</tr>
<tr>
<td>NO</td>
<td>$Mdn = 59.65$</td>
<td>.091</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$Mdn = 55.34$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latency to approach column 2 (LatAPP&lt;sup&gt;a&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>$Mdn = .91$</td>
<td>-2.232* (-.24)</td>
<td>-3.137** (-.33)</td>
</tr>
<tr>
<td>NO</td>
<td>$Mdn = 3.26$</td>
<td>-3.589*** (-.38)</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>$Mdn = 15.65$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * Measured in seconds; Reporting Wilcoxon Signed Ranks tests with z-scores for significant differences in behaviors across treatments (2tailed, exact test, trial 1 only). Median values for each task are presented along the diagonal for comparison between treatments. Significance was corrected for multiple comparisons via a Bonferroni correction (3 comparisons). Effect sizes ($r$) are presented in parentheses, calculated from the Wilcoxon tests as $r = z/\sqrt{n}$, where $n$ is the number of observations (Fields, 2005); * $p < .05$, ** $p < .01$, *** $p < .001$. 
Relationships Between Behaviors and Underlying Constructs

The results from this research were expected to not only demonstrate consistency across time and contexts as a way to support the idea that personality traits might be identifiable for this species, but there were also three umbrella constructs for exploration that might represent traits of interest (and thus were the foundation for how this project was designed): aggression, fear and boldness.

Statistical Approach

Due to a small sample size, a principle component analysis was not an acceptable approach for looking at the relationships between behaviors, nor was a factor analysis for examining underlying constructs. However, this research began with specific expectations about how the behaviors and tasks might be related to each other based on previous work and what the tasks were designed to measure. Therefore, one can examine the relationships that were expected between behaviors and across tasks, examining converging and discriminant validity, and if expected relationships hold, it would help to support the constructs of interest.

Instead of examining a large correlational matrix to examine the relationships between behaviors of interest, a Cronbach’s α was utilized to examine how behaviors “hung together.” It is used both as a measure of reliability and as a measure of unidimensionality (i.e., the extent to which a scale measures one underlying factor or construct; Fields, 2005). An acceptable range of α is .7-.8, or higher (V. Zeigler-Hill, personal communication, April 5, 2011; Fields, 2005). With fewer items, and when dealing with Psychological constructs, lower values have been shown to be acceptable, but values are definitely not accepted below 0.6. Cronbach’s α has a convenient feature
that tells you whether or not Cronbach’s $\alpha$ would be influenced if any of the items in the test were removed. Therefore, if Cronbach’s $\alpha$ would be increased with the removal of a given behavior, this would indicate that this behavior “didn’t fit” with the rest. More specific relationships can then be explored based on this initial information, avoiding an accumulation of a large number of tests. Furthermore, when using Cronbach’s $\alpha$ in this way, one wants to examine the “Corrected Item-Total Correlation” output to ensure that all values are above 0.3 (Fields, 2005). Anything below that value for a given behavior indicates fairly low internal consistency, identifying a potential problem. In summary, there are three criteria for concluding that a cluster of variables may represent an underlying construct using Cronbach’s $\alpha$: (a) an acceptable value of .7 or better (with a little flexibility for values not much less than this), (b) a “Corrected Item-Total Correlation” value of .3 or better for each variable in the cluster analyzed, and (c) no major changes to $\alpha$ (e.g., a large increase) if any behaviors were to be removed from the cluster (found under “Cronbach’s Alpha if Item Deleted” column in the output). The following sections explore the results from this approach.

Results

Aggression. Thrashing in front of a mirror was the primary behavior that has been empirically supported as an indication of aggression in zebrafish (see Blaser & Gerlai, 2006; Gerlai et al., 2000; Gerlai, 2003). This behavior was shown to be consistent over time for both proportion and frequency variables (Table 5; THRASH$_p$ and THRASH$_f$). Due to the design of the mirror task (refer to Figure 4A) I would expect that those that spend a large proportion of time thrashing should also spend a large proportion of time inspecting (within column one while not thrashing; INSPT$_p$) and approaching (in column
two; APP_p). Aggressive individuals would also be expected to approach the mirror more often (i.e., higher INSPT_f) and more quickly (i.e., lower LatINSPT; LatAPP). Depending on whether or not the frequency and proportion of time thrashing are positively related, those that are aggressive might also thrash more frequently (THRASH_f). Therefore, Cronbach’s $\alpha$ was expected to be high for this collection of variables (THRASH_p, THRASH_f, INSPT_p (thrashing subtracted), INSPT_f, APP_p, LatINSPT, LatAPP). Surprisingly, this was not the case. Cronbach’s $\alpha$ was initially 0.613 (lower that typically accepted), several behaviors with corrected item-total correlations were less than .3, and most surprisingly, results showed that a large increase in $\alpha$ (.713) would be obtained if THRASH_p was removed. In subsequent investigation, it turns out that all three criteria previously discussed were not met until both THRASH_p and THRASH_f were both removed from the cluster, and once removed, $\alpha = .760$. Further examination demonstrated that THRASH_f and THRASH_p were actually not significantly correlated with each other ($r_s = .302$, $p$ (two-tailed) = .105).

**Fear.** A collection of several behaviors were also expected to be indicators of fear responses to a predator, related in the following ways: individuals exhibiting fear responses were expected to show a low proportion of time spent close to the predator (INSPT_p and APP_p) and low frequency of approaches closest to the predator (INSPT_f), but have longer latencies to approach close to the predator (LatINSPT and LatAPP). These expectations were further supported from evidence regarding the differences in these behaviors across treatments (see Table 9 and Figures 7-9). Proportion of time approaching (APP_p) in the predator condition ($Mdn = 1.97$) was significantly lower than both the novel object ($Mdn = 16.92$, Wilcoxon signed ranks $z = -4.227$, $p < .001$, $r = -.45$)
and the mirror task ($Mdn = 4.94$, Wilcoxon signed ranks $z = -2.067$, $p < .05$, $r = -.22$).

This makes sense as it is expected that fish would not spend much time close to the predator. The proportion of time fish spent inspecting (i.e., closest to the predator; INSPT$_p$) was even lower ($Mdn = .35$; Figure 7). Although not significantly different from the novel object task ($Mdn = .31$, Wilcoxon signed ranks $z = -1.571$, $p > .05$), it was much lower than the mirror task ($Mdn = .80.13$, Wilcoxon signed ranks $z = -4.703$, $p < .001$, $r = -.50$). INSPT$_f$ was also significantly lower in the predator task ($Mdn = 3.5$; Table 9, Figure 8) compared to the mirror task ($Mdn = 25$, Wilcoxon signed ranks $z = -3.450$, $p < .001$, $r = -.36$), but not compared to the novel object task ($Mdn = 3$, Wilcoxon signed ranks $z = -.397$, $p > .05$).

Finally, the latency to inspect behavior was much longer in the predator task than for the mirror task (LatINSPT: $Mdn = 55.34$, 5.33, respectively; Wilcoxon signed ranks $z = -3.860$, $p < .001$, $r = -.34$), although, once again, comparable to the novel object task ($Mdn = 55.34$ and 59.65, respectively). The latency to approach the predator was the longest in duration compared to any task (LatAPP$_{PR}$: $Mdn = 15.65$; LatAPP$_{MR}$: $Mdn = 0.91$; Wilcoxon signed ranks $z = -3.137$, $p < .01$, $r = -.33$; LatAPP$_{NO}$: $Mdn = 3.26$, Wilcoxon signed ranks $z = -2.232$, $p < .05$, $r = -.24$).
Figure 7. Comparing the proportion of time inspecting and approaching stimuli across tasks (INSPT, APP; trial 1); (A) Showing inspecting behavior for mirror task before the proportion of thrashing has been removed from the variable; (B) Showing inspecting behavior once proportion of thrashing has been removed from the variable, better reflecting the proportion of time inspecting in this task. Means and 95% C.I. are shown.
Figure 8. Comparison of the frequency to inspect for each task (INSPT, trial 1). Means and 95% C.I. are shown. Zebrasfish inspect the mirror much more frequently than either novel objects or the predator.

Figure 9. Comparison of latency measures (LatINSPT, LatAPP and LatEMRG) across tasks (trial 1). Means and 95% C.I. are shown. Note that the time it took before first inspection (cross within the closest column of the stimuli) was much longer than the time it took to approach (within two columns of a stimuli) in each relevant case.
Cronbach’s $\alpha = .865$, with all criteria met, which indicates excellent clustering of these variables (INSPT$_p$, APP$_p$, INSPT$_f$, LatINSPT, LatAPP) within the predator task. This was confirmed by results from a Spearman rank order correlation across each pair variables with corrected alpha levels. Ten comparisons, required a sequential Bonferroni correction with a new $\alpha$-level criteria of $\alpha/10 = .005$ for the first comparison, $\alpha/9 = .006$ for the second comparison and so on. All comparisons remained significant (2-tailed) with a table-wide $p < .05$ (Rice, 1989). For INSPT$_p$, APP$_p$, INSPT$_f$, $r_s$’s = .918 - .969, indicating that those who spent the most time closest to the predator also spent the most time within only the second column closest to the predator and approached the predator more frequently. All three of these behaviors were significantly negatively correlated with LatINSPT and LatAPP ($r_s$’s = -.379 to -.778), indicating that those that spent the most time closest to the predator also were fastest to approach it initially.

**Bold/shyness.** Two tasks targeting bold and shy traits were incorporated into the design of this project. This was largely due to the ongoing debates over how to best measure behavior that is representative of these constructs (see Toms et al., 2010, for review). Convergence between these tasks would indicate that they are both measuring similar responses. In the novel object task, bolder individuals are expected to spend more time closer to the novel object (higher INSPT$_p$ and APP$_p$), visit it more frequently (higher INSPT$_f$) and approach and inspect it more quickly (short LatINSPT and LatAPP) than shy individuals. The latency to emerge behavior is expected to relate well to behaviors in this task, with bolder animals emerging more quickly.

These behaviors were similar in magnitude to the predator task, largely described in the previous section, but there a couple additional notes. Although the average latency
to inspect a stimulus and the proportion of time spent inspecting was similar between predator and novel object tasks (LatINSPT: $Mdn = 55.34 \ & 59.65$ seconds, respectively; INST$_p$: $Mdn = .35$ and .31, respectively; Figure 9), zebrafish were quicker to approach the novel object (LatAPP: $Mdn = 3.26$ and 15.65 seconds, respectively, $z = -3.589, p < .001, r = -.38$) and spent more time approaching the novel object (APP$_p$: $Mdn = 16.92$ and 1.97 seconds, respectively, $z = -4.227, p < .001, r = -.45$; Figure 7).

Cronbach’s $\alpha = .710$, which indicates substantial clustering of these behaviors (INSPT$_p$, APP$_p$, INSPT$_f$, LatINSPT, LatAPP and LatEMRG). However, one item (APP$_p$) did not meet the 0.3 criteria for corrected item-total correlation and would have resulted in a large increase in $\alpha$ if removed (i.e., $\alpha$ jumps to .803 if APP$_p$ is removed). As discussed earlier, this is the variable that was not consistent over time, possible due to the differences in responses to the rock compared to the pebble and the weight. Therefore, cases where zebrafish were exposed to the rock for the APP$_p$ variable were removed and Cronbach’s $\alpha$ was re-examined. This time the alpha jumped up to 0.801, and all criteria were met. All variables in the cluster appeared to fit well this time. As expected, the latency to emerge variable (LatEMRG) from the emergence task also fit this collection of behaviors well. In fact, if it were removed from the cluster, Cronbach’s $\alpha$ would decrease to .794.

Results from the Spearman rank order correlations (conducted for comparison) provided mixed results (again with a sequential Bonferroni correction applied as was done for the predator task, this time for 15 comparisons). INSPT$_p$ was positively correlated with INSPT$_f$ ($r_s = .928$) and negatively correlated with LatINSPT ($r_s = -.632$); INSPT$_f$ was also negatively correlated to LatINSPT ($r_s = -.635$) indicating those that
spent the most time closest to the objects were also the one that inspected most frequently and were quickest to inspect. However, APP with rocks removed was not correlated with any other behavior, and LatAPP was only correlated with LatINSPT ($r_s = .539$). LatEMRG was only correlated with LatINSPT ($r_s = .475$). This further provides evidence that there is a difference between INSPT behaviors and APP behaviors in this task, possibly due to the differences in novel objects used, which is further discussed in the discussion.

Activity and exploration. Activity and exploration behaviors were incorporated in this project largely for exploratory purposes. The small and large open field tasks were included to capture exploratory or possible escape behaviors in a setting that did not contain other stimuli (unlike the novel object or predator, for example). Additionally, they were expected to be two different sizes of the same task, such that measures between the two were expected to be highly correlated. Unfortunately, darting and erratic behaviors were not correlated across these two tasks; however, that might have been largely due to the lack of reliability obtained in measuring these behaviors.
The average percent of environment explored was largest for both the small and open field tasks (Figure 10), compared to the rest of the main tasks, and they were not significantly different from each other (Table 8; \( Mdn = 97 \) and 100 percent, respectively, Wilcoxon signed rank test, \( z = -.887, \) 2-tailed). Interestingly, ambulation was higher in the large open field compared to the small open field (Figure 11, Table 8; \( M =161.3 \) and 129.57 squares/min, respectively, Bonferronni posthoc pairwise comparison = -31.727, \( p < .01 \)). Based on previous work (Burns, 2008; Mikheev & Andreev, 1993) ambulation was expected to be a response to exploratory motivation in these tasks, such that they would be correlated. Surprisingly, ambulation (AMB) in the small open field was not correlated to exploration (EXP) in either the small or the large open field tasks (\( r_s's = -.253 \) & -.296, respectively), and ambulation in the large open field was inversely correlated to exploration in both the small and large open field tasks (\( r_s = -.419 \) and -.409, \( p (1\text{-tailed}) = .011 \) and .012, respectively). It is possible that higher rates of ambulation

**Figure 10.** Average percent of environment explored across tasks (EXP: trial 1); Means and 95% C.I. are shown.
indicate increased anxiety or fear, which would explain why exploration decreases with high ambulation rates. If this is the case, one might expect ambulation in the open field tasks to correlate with potential fear responses behaviors in the predator task. However, neither ambulation in the small nor large open field tasks were correlated with any of the potential fear response behaviors in the predator task (INSPT\textsubscript{p}, APP\textsubscript{p}, INSPT\textsubscript{f}, LatINSPT or LatAPP; \(r_s\)'s = -.117 to .083, 2-tailed).

Figure 11. Average ambulation rate across tasks (average square crossing per minute; AMB\textsubscript{r}; trial 1); Means and 95% C.I. are shown.

Other evidence has resulted in suggestions that swimming rate is a conflict between boldness and exploratory behavior. Burns (2008) demonstrated that ambulation was related to both exploration in an open field and boldness (latency to emerge) in an emergence task. If this were the case ambulation could positively correlate with INSPT\textsubscript{p}, APP\textsubscript{p} and INSPT\textsubscript{f} from the novel object task and show an inverse relationship with
LatINSPT, LatAPP and LatEMRG. There was only one instances of a significant relationship, and it was in the opposite direction from expected. Ambulation in the large open field was positively correlated with latency to inspect the novel object (LatINSPT; $r_s = .527, p = .003$; even after Bonferroni corrected $\alpha$-level of .004), meaning that those that moved faster on average in the open field were more hesitant to approach closely to the novel object. However, this was the only relationship that was even close to being significant.

There is some debate on whether measures of exploration and of boldness are actually representing one and the same underlying motivation (see Toms et al., 2010 for review). Since open fields are novel environments, and the definition of boldness has to do with responses to novelty, it would follow that these behaviors would be related. One way to examine this is to add EXP to the group of variables that were shown to be highly clustered in the novel object and emergence tasks (INSPT$_p$, APP$_p$ (w/o cases exposed to the rock), INSPT$_f$, LatINSPT, LatAPP and LatEMRG) and examine how the relationship changes with the added variable. The Cronbach’s $\alpha$ changed from .801 to .835 when EXP$_NO$ was added to the cluster, indicating that those that spent a large proportion of time close to the novel object, approached it more frequently and were quicker to approach it initially also explored their environment more in this task. Cronbach’s alpha was also high when exploration variables from both the small and large open field tasks were also included in this cluster (.816). For comparison, exploration was added to the cluster of behaviors examined in the predator task as well (INSPT$_p$, APP$_p$, INSPT$_f$, LatINSPT, and LatAPP). If the predator task indicated fear responses, than a high fear response would be expected to be related to a low percent exploration. When EXP$_PR$ was added to the cluster,
Cronbach’s alpha jumped from .865 to .892. Once again, Cronbach’s alpha was also high when exploration variables from both the small and large open field tasks were also included in this cluster (.841).

*Evidence of discriminate validity for constructs of interest.* A high degree of clustering of behaviors within a given task provides convergent evidence for the constructs examined in this project. Without additional discriminant evidence between tasks, however, one cannot assume that the clusterings within each task represent different underlying motivations.

*Aggression.* The results from the mirror task indicate that the collection of behaviors examined may separate into more than one cluster. The collection of behaviors examined in this task without *thrashing* (INSPT<sub>p</sub> (thrashing subtracted), INSPT<sub>f</sub>, APP<sub>p</sub>, LatINSPT, LatAPP) clustered very well together, possibly representing something separate from aggression. If this task was actually capturing a trade-off between different underlying motivations for behavior, this highly clustered sub-group may relate well with clusters from other tasks. If this cluster represents something similar to fear, it might be expected to group well the cluster of variables from the predator task. If it represents something similar to boldness, it would be expected to group well with the novel object task cluster. Unfortunately, neither case turned out to cluster well. The alpha for these variables compared with the predator task (PR: INSPT<sub>p</sub>, APP<sub>p</sub>, INSPT<sub>f</sub>, LatINSPT, LatAPP) was 0.530. The alpha for these variables compared with the novel object task (NO: INSPT<sub>p</sub>, APP<sub>p</sub> (w/o cases of rock exposure), INSPT<sub>f</sub>, LatINSPT, LatAPP) was 0.595.
**Fearlessness/boldness.** There are many that warn that it is challenging to distinguish between fear and bold/shyness or other conflicting motivations (Archer, 1979; Russell, 1983). If the measures of the predator task don’t group well with measures from the novel object and emergence test, this would provide discriminate evidence supporting the separation of these constructs. To examine this, LatEMRG was also included in the Crobach’s $\alpha$ grouping for the predator task that had already demonstrated an $\alpha = .865$. With the LatEMRG variable included, Cronbach’s $\alpha$ dropped to .843 (INSPT$_p$, APP$_p$, INSPT$_r$, LatINSPT, LatAPP and LatEMRG), however, this value is still considered to be a very good indication of clustering.

The behaviors that were shown to exhibit a high degree of clustering in the novel object task (INSPT$_p$, APP$_p$ (w/o cases of rock exposure), INSPT$_r$, LatINSPT, LatAPP) were compared to those that showed a high degree of clustering in the predator task (INSPT$_p$, APP$_p$, INSPT$_r$, LatINSPT, LatAPP ) via a collective Cronbach’s $\alpha$ test. If these all represent the same underlying construct, one would expect a high degree of clustering between all variables. Any substantial decrease in clustering would provide discriminate validity that these are capturing separate constructs. Consequently, Cronbach’s $\alpha$ dropped to .473. Any one item deletion would only have brought this value up to a max of .568. This does indicate a potential separation of constructs between these two tasks.

**Comparisons of the same type of measurement across tasks.** Related behaviors were examined across contexts, with the assumption that if these tasks were capturing different constructs, the behaviors would not cluster well via Cronbach’s $\alpha$. If the proportion of time spent inspecting and approaching across mirror, novel object and predator tasks all represented something similar in each context, these behaviors should
cluster together well across tasks. Cronbach’s $\alpha = .403$ \( (\text{INSPT}_p \& \text{APP}_p \text{ for each of three task; 6 variables total}) \), indicating a difference in constructs across these three tasks. Interestingly, when \text{INSPT}_p \text{ and } \text{APP}_p \text{ were removed for the mirror task, alpha jumped up to } .767, \text{ with all other criteria met.}

Similarly, if latency variables were all representing similar motivations for fear or shyness across tasks, these would also be expected to cluster well together \( (\text{LatINSPT}, \text{LatAPP} \text{ for each of the three tasks and LatEMRG; 7 variables total}) \). At first look, they appeared to not cluster very well \( (\text{Cronbach’s } \alpha = .679) \). However, once again, when \text{LatINSPT} \text{ and } \text{LatAPP} \text{ were removed for the mirror tasks, alpha jumped up to a respectable } .736, \text{ with all other criteria met. Finally, the clustering of the frequency to inspect} \ (\text{INSPT}_f) \text{ across all three tasks was examined. Cronbach’s } \alpha = -.145, \text{ however, there were only three variables available to put into this cluster, unlike the six or seven variables explored with the other measures. Therefore, correlations across these three variables were examined. INSPT}_f \text{ was not correlated between any pair of the three tasks examined} \ (\text{MR, NO, & PR; } r_s \text{ ranged from } -.309 \text{ to } .355, p’s = .054 - .123).
CHAPTER IV
DISCUSSION

In this study, behavioral relationships were examined for six behavioral tasks designed to capture information that would provide a comprehensive picture of zebrafish behavior and potentially help to provide insight on personality traits present in this species. In order to begin examining implications for personality a few requirements must be met. A high reliability of measures needs to be demonstrated for justification of using them (Gosling, 2001) and behaviors that are expected to represent underlying personality traits are expected to be consistent across time and across contexts (Budaev, 1997; Dall, et al., 2004; Gosling, 2001; Pervin, 1984; Vazire & Gosling, 2004). Finally, construct validity needs to be addressed in order for one to confidently conclude that any given behavior, or collection of behaviors, represent a particular personality trait.

Behaviors

A high standard of reliability was met for all behaviors except two: ERRATIC and DART. These were much harder to detect than any of the other behaviors included in this project. The highest confusion came from distinguishing between these and a general increase in swimming speed at any given point. Furthermore, the frequency at which they occurred in any given task was low (DART ranged from 0-18 and ERRATIC ranged from 0-11, across trials), resulting in variables that did not exhibit much variability in themselves. Higher variability may have been achieved if duration had been recorded instead of frequency; however, much more time would need to be spent working to obtain reliable identification of the start and ends of this behavior.
The proportion of time spent in active movement (MOVE\textsubscript{p}) was another variable with very little variation (although high reliability). Most individuals were active one hundred percent of the time, which also meant that freezing and drifting variables also demonstrated very little variability. In re-reviewing previous work with these behaviors in zebrafish, it appears that often freezing and erratic behavior tends to occur in low frequency for very low proportions of time with little variability (e.g., Blaser & Gerlai, 2006; Egan et al., 2009; Gerlai et al., 2006). It is possible that these behaviors reflect bimodal distributions of freezers and non-freezers (or erratic responses and non-erratic responses) instead of a normal continuous distribution. Budaev (1997) explored this potential in guppies and a cluster analysis revealed two distinct clusters (freezers and non-freezers) indicating a bimodal distribution. Additionally, Burns (2008) suggests that in a predator task, freezing may function to decrease attention conflicts because it allows the animal to focus on predator detection and to avoid being detected. If this was true, it is possible that these behaviors may be more situation-specific and may not be good behaviors to quantify for conclusions about personality traits.

Consistency

Most behaviors demonstrated high consistency across time in most contexts. The lack of consistency in latency variables and APP\textsubscript{p} for the novel object task may have been attributed to the difference in responses to the rock compared to either the weight or the pebble. Thrashing behavioral variables, the main indicator of aggressive behavior, were consistent across time in the mirror task. The lack of consistency in other behaviors for this task may be a result of a trade-off of an animal’s desire to assess dominance and
exhibit aggression in the face of a conspecific vs. a desire for caution in a relatively new environment.

Consistency across tasks for trial one was demonstrated for every behavior accept DART and ERRATIC (likely for reasons already discussed). This combined with consistency across time supports the original hypotheses for consistent individual differences in behavior, and provides the supporting foundation required to suggest that some of these behaviors may represent something analogous to personality traits. The bulk of the remaining work was spent examining which personality traits may have been identified from the tasks used.

Examining Constructs

Aggression

It was surprising that the cluster examined for aggression only met conditions when both of the main behavioral variables expected to measure aggression were removed from the cluster (THRASH_p and THRASH_f). A more detailed look at pairwise correlations showed that THRASH_f was not correlated to INSPT_p (w/o thrashing), APP_p, nor INSPT_f, but THRASH_p was significantly negatively correlated to each (r = -.452, -.725, & -.657, respectively). Furthermore, the correlation coefficient for THRASH_f was low (r = .355, accounting for only 12.6% of the variance across time) compared to THRASH_p (r = .602, accounting for 36% of the variance across time; Table 5). It is possible that THRASH_f may not be a good indicator of aggression. This could be further teased apart if more aggressive behaviors (e.g., charging and hitting) could be included in the future to help determine if these would better cluster together in this task.
It is peculiar that not only were \( \text{THRASH}_p \) and \( \text{THRASH}_f \) unrelated, but that the remaining variables in this task (\( \text{INSPT}_p \) (thrashing subtracted), \( \text{INSPT}_f \), \( \text{APP}_p \), \( \text{LatINSPT} \), \( \text{LatAPP} \)) clustered well together once these two were eliminated from the Cronbach statistic. Furthermore, the remaining cluster did not group well with either the bold or the fear cluster from the other tasks, so there’s no support that the mirror task was capturing a trade-off between motivational systems indicating the propensity to be aggressive vs. shy, or the propensity to be aggressive vs. fearful, in a relatively novel environment. It is possible that \( \text{THRASH}_p \) and \( \text{THRASH}_f \) were not capturing aggression at all, however, this seems unlikely due to the overwhelming support of this behavior relating to aggression from literature (Siamese fighting fish: Baenninger, 1970; guppies: Budaev, 1997; lion-headed cichlids: Budaev et al., 1999; stickleback fish: Tinbergen, 1974; zebrafish: Blaser & Gerlai, 2006; Gerlai et al., 2000; Gerlai, 2003). What may be a reasonable explanation is that the non-thrashing behaviors might represent something related to sociability, and thus this task may capture a tradeoff between the propensity to be aggressive and the propensity to be social. It has been suggested that MIS may be useful for studying social behavior (Gallup, 1968). In the case of schooling fish species, tendency to spend time near the mirror image may suggest a schooling response indicative of social motivation or intent to interact with a social partner.

One final suggestion is that those individuals that spend a large amount of time thrashing may, by design, not spend much time doing much of anything else. The average time spent thrashing at the mirror for trial one was 52.5%, and for trial two was 44.7%. This would explain why thrashing behavior did not cluster with other behaviors measured
in this task, and was negatively correlated to each of the remaining behaviors examined in this task.

*Fear vs. Boldness*

There was evidence that the behaviors measured in the predator, novel object and emergence tasks were related in ways based on the constructs they were expected to capture. However, when looking at both converging and discriminant evidence, there were mixed results for whether or not the predator task captured something different from the novel object and emergence tasks.

For both the predator and novel object tasks, there were strong clusters of behaviors within each task separately (in ways that made sense), which would be expected if these tasks captured distinct constructs. There was also some discriminate evidence to support this. If the tasks were representing different constructs, than the behaviors that clustered well within a given task, should not cluster well when combined together. Results provided evidence of this since Cronbach’s alpha dropped far below acceptable levels when all behaviors were combined. Furthermore, when the LatEMRG variable was added to the cluster in the predator task, it appeared to reduce the clustering relationship, whereas when included in the novel object task, it improved the resulting strength of the cluster. Surprisingly though, the resulting alpha for either case when LatEMRG was included was very high, and in the predator task, even higher than that for the novel object task with LatEMRG included (.843 and .801, respectively).

Discriminate evidence was also evaluated by examining the same types of variables across context. The frequency to *inspect* a stimulus was not correlated between the mirror, novel object or predator tasks, which again provides discriminate evidence for
the tasks representing different constructs. However, when considering proportions and latencies to inspect and approach stimuli, there were once again conflicting results. There appeared to be a lack of clustering in both variable types across tasks initially. However, when the variables from mirror task were removed, the proportion and latency variables (including that for the emergence task) were highly clustered across the novel object and predator tasks.

Finally, the variable measured as the percent of environment explored (EXP) offers one last piece of evidence in the comparison between fear and bold constructs. As previously mentioned, boldness is defined with respect to behavioral responses to novelty. When EXP from the novel object and small and large open field tasks were added to group of variables already determined to show a strong “clustering” relationship in the novel object and emergence tasks, the degree of clustering between variables was improved, suggesting that the propensity to explore might also be related to boldness. Surprisingly, when EXP was added to the cluster in the predator task, the clustering relationship remained very high, and was higher than that found in the collective novel object/emergence task cluster.

The surprising similarity between tasks expected to measure boldness versus fear, may boil down to discrepancies in the novel object task itself. As previously described, there were differences in the responses of fish to the rock compared to the pebble and weight. The weight and pebble were fairly comparable in size, but the rock was more than twice the size (rock: 3.3W x 6.8L x 2H cm; weight: 1 x 1 x 2.8 cm; pebble: 1.8 x 1.8 x 0.8 cm). It is possible that due to its larger size, the rock was a more threatening stimulus, eliciting something more comparable to a fear response. On the other hand, the
pebble and weight were very small, and may have been perceived as non-threatening, in which case the responses to them were more indicative of boldness. As a result, the novel object task may have been capturing both fear and boldness depending on the object used. This is further evident by the fact that although there were no differences in behavioral responses between the objects used when examining either the proportion of time inspecting or the latency of time before inspecting (both related to being within only a one column-distance from the novel object), there were significant differences depending on the type of object when considering the proportion of time approaching or the latency of time before the first approach (relating to being within a 2-column distance from the object). It is possible that there is an initial “investigation” time where fish are cautious against getting too close to a novel object, which may depend on the perceived threat of the object, but once they are they have acclimated to its presence, investigations closer to it reflect a boldness or curiosity to novelty regardless of the size of the object. It would be important to examine this potential in the future by looking at the expression of these behaviors over time instead of as an average over the treatment period, potentially examining whether or not frequencies of approaches and inspections increased over time.

Visual perception in general may also influence responses to a novel object. Research has shown that zebrafish can discriminate between different types of light (e.g., white light vs. monochromatic; Bilotta, Risner, Davis, & Haggbloom, 2005), and even possibly between colors (Colwill, Raymond, Ferreira, & Escudero, 2005). There is some discrepancy on whether the research done demonstrates an actual wavelength (i.e., color) preference or if results reflect dark versus light preferences (Bilotta et al., 2005); however, either way, this might bias results in a novel object task situation. Responses to
a novel object may not be reflecting boldness or differences based on size, but based on color or perceived darkness. This is an empirical question that remains to be examined and should be considered for any novel object task. Ideally, research needs to be done to examine preferences/avoidance responses between many different types of objects (different sizes, shapes, colors, and textures) before understanding how any given object may be used to capture “bold” responses.

Activity and Exploration

Ambulation measures (rate of activity) did not correlate with any of the expected constructs examined in this project. It has been suggested that activity rates habituate over time in zebrafish (Gerlai, 2003; Mikheev & Andreev, 1993), which is supported by work in other species (e.g., guppies: Burns, 2008; blind cave fish: Mikheev & Andreev, 1993). It is possible that a habituation response is the reason ambulation didn’t relate to other variables, since the ambulation was averaged across a five minute treatment exposure in this project. It might be more appropriate to break down all behaviors minute by minute to be able to examine responses over time. The fact that ambulation in the large open field was negatively correlated with exploration in the large and small open field tasks may indicate ambulation is actually representing something similar to anxiety. A larger environment with more open space may be more stressful, leading to an increase in swim rate. A correlation between those exhibiting erratic behavior (an indication of anxiety or escape) and high rates of ambulation would help support this. A quick glance shows that in the small open field, ambulation is correlated to erratic behavior ($r_s = .468, p = .009$). The same is true for the large open field ($r = .501, p = .005$). One must be careful in drawing conclusions from these results since erratic behavior was not shown to
be reliable in this project; however there is preliminary evidence that this should be further explored.

Considerations and Caveats

In addition to some of the concerns already addressed, there are additional considerations for interpreting the results from this project. There were some limitations with not having enough behaviors in a given task or for a given construct to be able to better cluster or discriminate between constructs. For example, including more aggressive behaviors (or another task examining aggressive responses between two live fish compared to a mirror task), could have helped to identify an aggressive construct. Furthermore, latency-to-emerge was the only behavior captured in the emergence task. Other behaviors could be recorded for comparison. For example one could record the frequency of passing back and forth across the opening before passing through, frequency of moving within close proximity to opening without passing through, etc, in order to examine not only if these behaviors clustered within the emergence task but also how they compared to other tasks.

Most behavioral variables were non-normal and skewed variables such that non-parametric tests were widely used throughout the analytical process. Furthermore, there were a very large number of statistical tests run, which does inflate the probably of committing a type I error, and the samples size was not large enough to conduct more sophisticated statistics that would more neatly explore this data set. Despite these potential concerns, corrections were made whenever possible throughout the project to best control for type I error rate inflation. Furthermore, although nonparametric statistical tests are considered to have less power to detect true differences, this is only typically the
case when comparing to parametric tests on *normal* data (Fields, 2005). When data are non-normal, nonparametric tests are very appropriate, and in some cases in this project, actually preferred (such as using a Spearman rank order correlation to capture consistency in behaviors across time).

One must also consider the limitations of using Cronbach’s $\alpha$ for examining constructs. The value of $\alpha$ does depend on the number of items used. However, because of the ability to measure the same types of behaviors across tasks, the number of items included in the cluster often only differed by one or two variables. Interestingly, $\alpha$ should be inflated with increasing number of items, and in cases where there were more than seven variables (i.e., when combining the predator task cluster and novel object cluster all together), $\alpha$ was very weak, not high, which instills confidence in the results of these groupings. Another caveat to using Cronbach’s $\alpha$ is although it does measure the degree of “unidimensionality” (or extent that a group measures one underlying factor or construct) when there is only one underlying factor, it does not identify cases where the variables actually represent more than one underlying factor (Fields, 2005). This was taken into consideration as results were explored and efforts to examine the data in multiple ways were taken in order to best understand the behaviors in this research.

There is some concern that creating an experimental design that minimized handling and increased fluidity between tasks resulted in potentially creating non-independent tasks. The small open field and mirror task were independent from all of the other four tasks, but not necessarily from each other since they were conducted back to back. The same concern exists for the emergence, large open field, novel object and predator tasks. This problem was addressed by implementing a start box before every
task, such that the fish was allowed to re-acclimate for two minutes before starting the new task. It is possible, that the start box created some stress instead of promoting acclimation. However, this procedure has been implemented in fish studies previously (e.g., Budaev, 1997; Burn, 2008). In the future, tasks should be conducted in random order to reduce the concern for independence between tasks.

Predator responses were not controlled for in this project. It is possible that individual responses could be better teased apart if fine scale behaviors could also be included for the predator. It did not seem ideal to add more complexity to an already highly complex design and it would have required a substantial amount of time to have created operational definitions for important behaviors and obtain reasonable reliability on predator behaviors as well. However, major changes in normal response to the zebrafish were noted (e.g., any unusual decrease in activity or unusually aggressive reaction towards the zebrafish). Out of 30 fish with two trails each, there was only one case where the predator appeared to exhibit low levels of aggression.

Finally, one must consider that zebrafish are a shoaling species, meaning that they are social and prefer to be in groups. Fish were tested individually, without any other conspecifics, thereby potentially increasing anxiety or stress due to separation alone. However, fish were singly housed for the duration of the testing period (a little over four weeks for each fish) and allowed a minimum of 48 hours to acclimate to being alone before the first experiments. Furthermore, all individuals were exposed in exactly the same way and housed in the same sized units, such that variations in behavioral responses would still be expected to reflect individual differences in behavior.
Conclusions and Broader Implications

Individual level variation was demonstrated to be consistent across time and across context which was the basic criteria needed to begin to make conclusions about the potential presence of personality traits. Further investigations of the constructs of interest provided mixed results, unfortunately limiting my ability to make solid conclusions about which traits may have been captured. However, regardless of the specific construct captured, results demonstrated that individuals responded consistently across time and across emergence, novel object and predator tasks. Behaviors clustered very well within the novel object and predator tasks, which was further supported by high Spearman correlations across almost all behaviors in the predator task. The novel object task exhibited less internal consistency across behaviors, even though clustering was high. However, as already discussed, this may be attributed to the differences in responses to the novel object. Furthermore, the emergence task clustered well with both predator and novel object tasks.

Collectively, after examining all the evidence for the initial constructs of interest (aggression, boldness and fear), conclusions can be summarized as follows. Results from the mirror exposure task were internally inconsistent, requiring more work before an “aggressive trait” can be specified. There was not a clear distinction between boldness and fearlessness. Data actually supported the idea that emergence, novel object and predator tasks were capturing the same underlying construct or capturing a conflict or tradeoff between boldness and fear. Exploration also seemed linked to both these constructs, further supporting the idea that there may not have been a clear distinction between them.
This is the first comprehensive study of zebrafish behavior that has examined how the same fish respond in multiple situations, measuring and comparing multiple behaviors in each task used. The implications from this research span across three broad fields: psychology (specifically with regards to personality), behavioral ecology and behavioral neuroscience.

More recently, there has been an increasing divide between psychologists who tend to focus on examining personality structure but are criticized for not exploring potential ecological importance of such findings, and behavioral ecologists (and related fields) who focus on ecological significance without first gaining a clear understanding of personality structure in a given species. One of the initial interests in creating this project was not only to explore individual differences and potential personality traits in zebrafish but also to examine the validity of the tests chosen. Over and over again, animal personality researchers make selections of behaviors that are supposed to be representative of particular personality traits, without ever actually exploring evidence supporting the constructs they are expected to represent. For example, it is not uncommon for a researcher to hand select three behaviors that are expected to represent three traits of interest measured via three tasks (e.g., selecting latency to emerge, frequency of thrashing and proportion of time spent next to a predator as measures boldness, aggression and fearlessness, respectively). Without any further examination of what other behaviors may be involved in the selected tasks, how they are related and without providing much evidence for the use of the tasks, behaviors are then compared to reproductive success, foraging ability, mate selection, etc, and conclusions about ecological implications are made with regards to “personality traits” (reviewed in Toms
et al., 2010). For those that do acknowledge the importance of understanding personality structure and providing enough evidence to support the constructs of interest, often researchers conduct sophisticated analyses, such as a principle component or factor analysis on data with sample sizes that may be too small to warrant the use of those statistical approaches (e.g., Bell, 2005; Budaev, 1997; Huntingford, 1976). Ideally, the best approach would be a combination of these approaches. Personality differences may influence how populations respond to change, cope with stress, or deal with anthropogenic impacts, ultimately influencing a species ability to survive and procreate. Similar to cross-cultural studies in humans, cross-population comparative studies are equally important in animals. It may be possible to gain insight in the evolutionary basis of individual differences, and the biological roots of personality in animals.

Burns (2008) recognized these concerns in research examining personality in guppies, and, as a result, conducted a thorough examination of open field, novel object and emergence tasks, looking at reliability of commonly used measures, consistency in behaviors, internal validity and convergent and discriminant evidence of construct validity. He did find evidence in support of the use of these tasks, but also expressed many concerns. Ultimately, he laid out a foundation for future work in guppies. Although there are many reviews of personality research in animals providing many important considerations when designing personality research (e.g., a thorough review by Gosling, 2001), and a larger body of work available for work with primates, I have only come across one article that has done this and set up a framework for how to best explore personality traits in fish. Additionally, this was only done for one species (the guppy) and
one cannot assume that personality traits in another species could be captured in exactly the same way.

The fact that there were no clear-cut differences between the emergence/novel object tasks compared to the predator task should help to demonstrate how important it is to first explore the framework of personality traits and how behaviors are expressed before jumping to conclusions about what a given behavior is an indication of. For example, much work in behavioral neuroscience has attributed predator tasks to fear responses as a way to compare these responses to anxiety and phobias in humans (Barcellos et al., 2007; Bass & Gerlai, 2008; Blaser & Gerlai, 2006; Gerlai et al., 2000; Speedie & Gerlai, 2008). However, in the behavioral ecology field, predator approaches have been attributed to boldness (Moretz et al., 2007). Another example from this study is evident from the differences in responses to different novel objects used. Many researchers do not select more than one object for comparison and some select objects that might even be perceived as a refuge (e.g., PVC pipe or other structure with holes and hiding places). To add even more variety, some choose to combine novel objects with a food source (see Toms et al., 2010 for a review).

On the whole however, animal personality research is a new and burgeoning field. It has only been recently (i.e., the past decade or two) that individual-level investigation has even been considered important (see a brief review in Wilson, 1998) in animal behavior research. Classically, animal behavior was expected to be highly plastic, meaning that individuals could alter their behavioral responses to cope adaptively with environmental change (Briffa, Rundle & Fryer, 2008; Dingemanse, Kazem, Reale, & Wright, 2009). It was assumed that individual variation in a given behavior was non-
adaptive variation around an adaptive mean. While having some level of behavioral plasticity is still clearly beneficial (e.g., aggressive fish may be more likely to get eaten if responding just as aggressively in the face of a predator), individual behavior is not as plastic as once thought, and unlimited plasticity might actually be costly to produce or maintain (Briffa et al., 2008). These considerations have led to an increasing acceptance that individual differences in behavior are likely an adaptive trade-off between an individual’s ability to modulate its own behavior appropriately within different situations (plasticity), while maintaining some general consistent level of response relative to responses of other individuals (consistency) (Briffa et al., 2008; Dingemanse et al., 2009). This suggests that consistency in behavioral response is not expected for a given individual across every single situation, but that there is a general pattern of consistency, or better yet, on average, individuals tend to respond in consistent ways across time and context. This pattern was evident in this research project. Although those that spent considerable time approaching the novel object were the same individuals that tended to spend more approaching a predator, the amount of time spent in each case was very different and individuals demonstrated a difference in the time to initially approach an object versus a predator as well. This is likely a highly adaptive difference, as approaching a predator without hesitation would likely result in death in any natural situation.

The implications from this research with regards to behavioral neuroscience may be less directly obvious, however, equally as important. Although zebrafish have not yet been capitalized on by psychologists and behavioral ecologists, they are becoming increasingly recognized as an important model organism in behavioral neuroscience.
Much of what we do know about their behavior thus far has come from behavioral neuroscience research. Furthermore, the tasks used in this research have been widely used in zebrafish. However, zebrafish behavior is often measured in groups, with behaviors recorded only if a majority of the group conducts a given behavior (e.g., with a group of eight fish, freezing might only be considered to have occurred if five or more froze at any given time), which fails to consider individual-level responses. Those that do examine individuals then average behaviors across individuals, again ignoring individual differences in behavioral responses. Model organisms are incredibly valuable for furthering our understanding of human behavior/systems. However, humans are complex individuals, with unique personalities. If we want to maximize the utility of and best generalize from animal models, we need to examine individual-level behavioral information, a source of variation that is fundamental to human behavior.

In addition to helping to improve our general understanding of human behavior, considering individual differences can greatly improve the development of pharmaceuticals and the efficacy of treatment. Individual personalities may interact with disorders and treatments in important ways. If we continue to only examine averaged behavioral responses when examining the effects of new drugs, we are likely missing a wealth of information that could help to improve the effectiveness of those drugs. It would be widely beneficial to be able to pre-screen individuals for certain personality traits in order to more specifically target treatment. For example, the results from project showed that individuals respond to a predator in consistent ways, and of the tasks examined, the predator task was the most internally consistent. In fact, it has been suggested that individual differences become more stable and predictable under mild
stress (Gerlai & Csanyi, 1990). One could select a few of the behaviors determined from this study to be highly correlated and use this task to prescreen individuals for highly anxious/fearful fish (keep in mind, a little more preliminary work does need to be done to better determine which trait this task captures, however). One could then administer drugs more specifically to certain “types” of individuals and examine differences in responses between them. This could be highly useful for maximizing our ability to create effective pharmaceuticals. Furthermore, work has already shown that drugs of abuse induce functional changes in the brain that can be detected at the behavioral level (for example, see Echevarria, Toms, & Jouandot, 2011 for a review of alcohol induced effects on behavior). Imagine the strength of understanding that could be obtained if we started taking individual differences into account.

This pre-screening process is not limited to testing various drugs and pharmaceuticals. One of the huge benefits to utilizing zebrafish in neuroscience research is the fact that they are considered to have high throughput. A large number of individuals can be screened quickly, housed easily and efficiently compared to other species commonly used in neuroscience (e.g., rats and mice). A single female zebrafish produces 200 eggs per spawning, can spawn several times a week and achieve sexual maturity within two months (Blaser & Gerlai, 2006). Therefore breeding is very quick and simple and one can begin to look at individual differences and respective differences in treatments across developmental stages very quickly. One can even compare individual level differences between strains and mutations, which could help further our understanding of genetic versus environmental basis of personality traits.
In conclusion, whether one is interested in further studies examining personality structure, determining how individual personalities interact with fitness and survival, or improving our ability to understand human behavior and systems and better tailor treatments, the application of exploring individual-level behavior in such a powerful model organism is infinite. The current project provides a much-needed foundation for exploring individual differences and personality traits in zebrafish from which there are endless opportunities for expansion.
APPENDIX

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE APPROVAL FORM

The proposal noted below was reviewed and approved by The University of Southern Mississippi Institutional Animal Care and Use Committee (IACUC) in accordance with regulations by the United States Department of Agriculture and the Public Health Service Office of Laboratory Animal Welfare. The project expiration date is noted below. If for some reason the project is not completed by the end of the three year approval period, your protocol must be reactivated (a new protocol must be submitted and approved) before further work involving the use of animals can be done.

Any significant changes (see attached) should be brought to the attention of the committee at the earliest possible time. If you should have any questions, please contact me.

PROTOCOL NUMBER: 09051402
PROJECT TITLE: Individual Differences and Stress Influences on Group Behavior
PROPOSED PROJECT DATES: 06/01/09 to 08/31/2010
PROJECT TYPE: New Project
PRINCIPAL INVESTIGATOR(S): David J. Echevarria, Ph.D.
COLLEGE/DIVISION: College of Education & Psychology
DEPARTMENT: Psychology
FUNDING AGENCY/SPONSOR: Departmental
IACUC COMMITTEE ACTION: Full Committee Review Approval
PROTOCOL EXPIRATION DATE: 09/30/2011

Robert C. Bateman, Jr., Ph.D.
IACUC Chair

Date
REFERENCES


Table 2

**Summary of Behaviors Recorded, Operational Definitions and Measurement Criteria Used**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Behavior</th>
<th>Definition</th>
<th>Criteria for Measurement</th>
<th>With Respect To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVE_p (% activity)</td>
<td>Percent activity</td>
<td>Percentage of total time spent in active movement</td>
<td>Fins are engaged and the fish is moving through the water column. Any movement where fins are moving is considered active movement</td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>AMB_r (Ambulation rate)</td>
<td>Ambulation (i.e., rate of movement)</td>
<td>Average number of squares entered per minute</td>
<td>Mouth, eyes &amp; gills are the only body parts that move</td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>FREEZE_p (% time freeze)</td>
<td>Proportion of time freezing</td>
<td>Complete lack of movement in any direction</td>
<td></td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>FREEZE_f (Frequency)</td>
<td>Frequency of freezing</td>
<td>Defined as above</td>
<td></td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>DRIFT_p (% time drifting)</td>
<td>Proportion of time drifting</td>
<td>Slow movement through the water column in which the caudal fin stops moving, but the fish has not stopped forward momentum.</td>
<td>Peck fins may or may not still be in motion</td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>DRIFT_f (frequency drift)</td>
<td>Frequency of drift</td>
<td>Defined as above</td>
<td></td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>EXP_p (% exploration)</td>
<td>Proportion of environment explored</td>
<td>Measured as the percent of total number of squares entered</td>
<td>Considered to have been entered if the fish moved at least half way into a given square</td>
<td>OFs, MR, OFFlg, NO, PR</td>
</tr>
<tr>
<td>INSPT_p (% time w/in 1 column)</td>
<td>Proportion of time inspecting</td>
<td>Time spent within the column closest to stimulus (3.25 cm from stimulus)</td>
<td>Considered to have inspected if fish body moves half way or more into the column closest to stimulus. Behavior ends when fish moves at least half way out.</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>APP_p (% time w/in 2 columns)</td>
<td>Proportion of time spent approaching</td>
<td>Time spent within the column second closest to stimulus (6.5 cm from stimulus)</td>
<td>Considered to have approached if fish body moves half way or more into the second column closest to stimulus. Behavior ends when fish moves at least half way out.</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>INSPT_f (frequency w/in 1 column)</td>
<td>Frequency of inspection</td>
<td>Defined as above</td>
<td>Same as above for inspect</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>LatINSPT (latency to 1 column)</td>
<td>Latency to inspect (sec)</td>
<td>Latency to move within the column closest to stimulus (3.25 cm)</td>
<td>Same as above for inspect</td>
<td>MR, NO, PR</td>
</tr>
<tr>
<td>LatAPP (latency to 2 columns)</td>
<td>Latency to approach (sec)</td>
<td>Latency to move within the second column closest to stimulus (6.5 cm)</td>
<td>Same as above for approach</td>
<td>MR, NO, PR</td>
</tr>
</tbody>
</table>
Table 2 (continued).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Behavior</th>
<th>Definition</th>
<th>Criteria for Measurement</th>
<th>With Respect To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>LatEMRG</td>
<td>Latency to emerge</td>
<td>Time it takes to move through a narrow opening into a novel environment</td>
<td>Start time recorded when the door to the opening was half way out of the tank. End time recorded when the fish body was at least half way through the open door</td>
<td>Emergence</td>
</tr>
<tr>
<td>THRASH&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Proportion of total time thrashing;</td>
<td>Often very rapid and continuous caudal fin movements back and forth along the mirror. Often nose is pressed up against mirror while moving.</td>
<td>A minimum of 1 sec in duration and movement back and forth at least twice is required to be considered “thrashing”. One pass along mirror is insufficient. Behavior ends when fish moves half way out of column closest to mirror.</td>
<td>MR</td>
</tr>
<tr>
<td>THRASH&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of thrashing</td>
<td>Defined as above</td>
<td>Same as above for thrash</td>
<td>MR</td>
</tr>
<tr>
<td>ERRATIC&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of erratic movement</td>
<td>Seemingly aimless zigzag movement with frequent directional changes</td>
<td>Lasts for &gt; 1 sec (different from DART). Is unpredictable movement.</td>
<td>OFs, OFlg</td>
</tr>
<tr>
<td>DART&lt;sub&gt;f&lt;/sub&gt;</td>
<td>Frequency of DARTing</td>
<td>Very brief bouts of obvious rapid movement.</td>
<td>Lasts for &lt; 1 sec; Not limited to straight movement</td>
<td>OFs, OFlg</td>
</tr>
</tbody>
</table>

Note. Measurements across tasks allowed for comparisons across contexts, but not all behaviors were collected across all tasks. The subscripts “p” and “f” indicate proportion of time versus frequency variables, respectively, and “r” indicates a rate of behavior. The tasks listed under the “with respect to” column indicate which task(s) each behavior was measured in; OFs: Open field small; OFlg: Open field large; MR: Mirror exposure task; NO: Novel object task; PR: Predator task.
Table 3

*Raw Frequency and Proportion Data for Freezing and Drift Behaviors*

<table>
<thead>
<tr>
<th></th>
<th>Sm. Open Field</th>
<th>Mirror Task</th>
<th>Lg. Open Field</th>
<th>Novel Object</th>
<th>Predator Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>TOT</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td><strong>FREEZE</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(frequency)</td>
<td>1/3</td>
<td>3/3</td>
<td>4/6</td>
<td>2/3</td>
<td>3/4</td>
</tr>
<tr>
<td>(frequency)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FREEZE</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(% time freeze)</td>
<td>17.25</td>
<td>5.09-94.18</td>
<td>47.47</td>
<td>3.43-23.62</td>
<td>1.3-77.56</td>
</tr>
<tr>
<td><strong>AVG</strong>&lt;sup&gt;d&lt;/sup&gt;</td>
<td>17.25</td>
<td>57.54</td>
<td></td>
<td>13.53</td>
<td>42.99</td>
</tr>
<tr>
<td><strong>DRIFT</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.07-28.13</td>
<td>87.97</td>
<td>16.75</td>
<td>0.85-4.97</td>
<td>0.32-15.45</td>
</tr>
<tr>
<td><strong>AVG</strong>&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2.62</td>
<td>8.98</td>
<td></td>
<td>2.56</td>
<td>8.02</td>
</tr>
</tbody>
</table>

*Note.* <sup>a</sup> The first number is the *number of fish* that exhibited the behavior in that trial for that task. The second number is the *total number of times* that behavior was observed in that trial for that task; <sup>b</sup> Total number of fish that were observed in the behavior *across trials* & total number of times that behavior was observed *across trials*; <sup>c</sup> Average proportion of time across trials for given task; <sup>d</sup> Raw data presented as the range of percent time spent in the given behavior for that trial; <sup>e</sup> Average proportion of time spent conducting the behavior within a given trial; All values calculated only for fish that actually conducted the behavior at least once in a given trial.
Figure 5. (A) Average percent time fish spent freezing in trials 1 and 2 for each treatment condition; (B) Average time fish spent drifting in trials 1 and 2 each treatment condition; Data presented here are raw data for fish that were ever observed exhibiting these behaviors at least once (n = 13).
Figure 7. Comparing the proportion of time inspecting and approaching stimuli across tasks ($\text{INSPT}_p$, $\text{APP}_p$; trial 1); (A) Showing inspecting behavior for mirror task before the proportion of thrashing has been removed from the variable; (B) Showing inspecting behavior once proportion of thrashing has been removed from the variable, better reflecting the proportion of time inspecting in this task. Means and 95% C.I. are shown.