DIPLOMA THESIS

Assessment and evaluation of temperament traits in carp (Cyprinus carpio L.), with contrasts between mirror and scaled morphological phenotypes

By
Philip Rebensburg
Matriculation number: 3798019
Free University of Berlin

Supervisor: Prof. Dr. Robert Arlinghaus
Humboldt-University of Berlin, Faculty of Agriculture and
Horticulture, Institute of Animal Sciences
and
Leibniz-Institute of Freshwater Ecology and Inland Fisheries,
Department of Biology and Ecology of Fishes, Berlin

Co-Supervisor: Prof. Constance Scharff, Ph.D.
Free University of Berlin,
Department of Biology, Chemistry and Pharmacy,
Institute of Biology,
Animal Behavior

Berlin, July 2010

Ehrenwörtliche Erklärung

Hiermit erkläre ich an Eides statt, die vorliegende Diplomarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Berlin, den 15.07.2010

Philip Rebensburg

Abstract

Individual variation in behavioral traits of animals has, if at all, been recognized only reluctantly as temperament, defined as consistency in individual behavioral variation over time and / or across situations. Temperament traits like boldness, exploration, activity, aggression and sociability have been identified in a wide variety of taxa, including various fish species. It was only recently discovered that behavioral traits can correlate with one another to form behavioral syndromes and that these correlations across contexts can be stable over time. So far, little is known about temperament traits, behavioral syndromes and repeatability of temperament traits in the carp (Cyprinus carpio L.). Furthermore, little is known about the effects of domestication on temperament and the repeatability of individual behavior as a consequence of artificial selection in captive environments. To address these issues, a laboratory study was conducted to assess several behavioral variables in carp with the goal of extracting temperament traits and behavioral syndromes respectively, focusing on boldness, exploration and sociability. To maximize behavioral variability and to examine possible effects of domestication on behavioral performance, two genotypes of carp, exhibiting a different degree of domestication (scaled and mirror carp), were used. The experimental setup involved four behavioral contexts. These contexts included exploration behavior in a novel environment, investigation of a novel object, feeding under simulated predation and tendency to associate with a group of conspecifics. All behavioral measurements were repeated after six weeks to test for repeatability and consistency of behavior over time. This study revealed that individual carp differ in their behavioral responses, despite habituation effects over the study period and potential individual differences in habituation degree. Several behaviors are consistent over time and / or situations, i.e. temperament traits respectively behavioral syndromes exist in carp. Two distinct temperament dimensions were identified as the temperament traits boldness and exploration. In the case of exploration, a context-independent temperament trait was found. Regarding domestication effects, only minor behavioral differences between mirror carp and scaled carp of the "wild type" phenotype were found under standardized laboratory conditions, possibly indicating context-dependency of behaviors. This study provides evidence for the existence of the temperament traits boldness and exploration in carp.

Table of Contents

List of Figures

List of Tables

1. Introduction/ Literature Review

The scientific study of animal behavior is receiving increasing attention in natural sciences and comprises the investigation of an animal's interaction with its physical environment or with other animals of the same or different species. These interactions include - among others - resource exploitation, predator avoidance, mate choice, reproduction and brood care (Alcock 2005).

Behavior is considered as part of an animal's phenotype and is loosely defined as everything an animal does and how it does it, provided it can be observed and measured (Campbell and Reece 2005). Within a given population it is often observed that individuals behave differently from one another (Clark and Ehlinger 1987; Magurran 1993; Wilson 1998), which can be explained by a combination of genetic, epigenetic (i.e., developmental) and environmental effects (Goldsmith, Buss et al. 1987; McDougall, Réale et al. 2006). These inter-individual differences can be described, quantified and comparatively analyzed (Blumstein and Daniel 2007) and are often interpreted as differences in personality, temperament, individuality or character (Gosling 2001; Sih, Bell et al. 2004; Réale, Reader et al. 2007). Distinctions between the definitions of these terms are rather vague as all of the terms describe similar measurements of behavior (Réale, Reader et al. 2007).

However, most authors (e.g. Budaev 1997; Gosling 2001; Dall, Houston et al. 2004), independent of terminology, conclude that the phenomenon is described as that individual differences in behavior are consistent over time and / or across situations (Réale, Reader et al. 2007).

Following Réale, Reader et al. (2007), the term "temperament" will be used in the present study for this phenomenon. Temperament is defined here as describing consistent behavioral differences amongst individuals over time and / or situations. In this definition of temperament, the attribute of consistency refers to relative behavioral differences between individuals and does not exclude changes in individual behavioral characteristics through changing environmental conditions or physical maturation. Moreover, the attribute of consistency permits the distinction between short-term motivational or emotional conditions referred to as states and

the long-term characteristics of an organism shared by all or some of the individuals of a species that can vary, although not necessarily, among these individuals (Réale, Reader et al. 2007), referred to as traits.

Traits are considered as a tendency to react in a given way and are therefore, in contrast to states, not observable and can only be measured indirectly by quantifying behavior patterns shown in a relevant situation which activates the individuals tendency to react in a certain manner (Strelau 2001). Thus, it is important to bear in mind that traits are abstract and dispositional constructs. Temperament itself is a hypothetical construct, that is, a notionally constructed assembly on the structures and dynamics of living beings (Strelau 2001).

Perhaps due to its hypothetical nature, the concept of temperament, originally being a domain of human psychology research, has only reluctantly been adapted in the field of behavioral ecology (Wilson, Clark et al. 1994). This seems surprising as the concept of temperament has yielded many benefits in other fields of research (see Réale, Reader et al. 2007). However, recent research has revealed that temperament traits play a role in important ecological processes such as dispersal (Cote, Fogarty et al. 2010), social organization (Dyer, Croft et al. 2009) and niche expansion (Dall, Houston et al. 2004) and are linked to life – history productivity (Biro and Stamps 2008). For example, boldness is linked to food intake and growth-rate in fishes (Ward, Thomas et al. 2004). Thus, the concept of temperament is beginning to be perceived as beneficial to the understanding of ecological and evolutionary processes in animals (Gosling 2001; Sih, Bell et al. 2004; Bell 2007; Réale, Reader et al. 2007).

Meanwhile, a multitude of temperament related behavioral traits and constructs like boldness, aggressiveness, reactivity, risk-taking, activity levels and coping-styles have been identified in almost all taxa (see Gosling 2001; Bell 2007) including fish (Budaev and Zworykin 2002; Brown, Jones et al. 2005; Reddon and Hurd 2009; Wilson and Godin 2009). However, a coherency of construct-labeling is missing in temperament research (Block 1995). For instance, the same trait has frequently been given different labels, e.g. boldness has been termed docility or fearfulness in other studies (Réale, Reader et al. 2007). Another example is the

various labeling of behavioral reactions quantified with the "open field test" (Hall 1934), which have been variously termed fearfulness, extraversion, exploration or activity (Gosling 2001).

The lack of a general framework in animal personality research, as illustrated by Gosling (2001), encouraged Réale, Reader et al (2007), to offer a supporting theoretical and methodological structure for the ecological study of temperament. This framework provides a simplified terminology as a standardized working tool for researchers aiming at integrating animal temperament research within ecological theory. According to Réale, Reader et al. (2007) the multitude of existing temperament constructs can be broken down to five relevant temperament traits with regards to the ecological and evolutionary study of animals. These traits are (1) shyness-boldness, (2) exploration-avoidance, (3) activity, (4) aggression and (5) sociability and will be outlined in further detail as follows.

(1)   **Shyness-boldness** (hereafter "boldness") is defined as an individ*ual's* reaction to any risky situation, but not novel situations (Réale, Reader et al. 2007), whereby individuals vary from being extremely bold (e.g. reacting to novel stimuli with pronounced exploratory behavior) to extremely shy or timid (e.g. reacting to novel stimuli with retreat or vigilance). Boldness is a temperament trait of great ecological importance, underlined by its broad taxonomical distribution (see Gosling 2001; Smith and Blumstein 2008) and the range of associations with ecologically important behaviors and fitness-related traits (Dingemanse and Reale 2005; Stamps 2007), e.g. anti-predator behavior (Brown, Jones et al. 2005), general activity and habitat use (Wilson and McLaughlin 2007), exploration (Wilson and Godin 2009), foraging (Wilson and Stevens 2005), aggression (Johnson and Sih 2005), mate selection (Godin and Dugatkin 1996), invasiveness (Wilson, Clark et al. 1994; Budaev 2003; Brown, Jones et al. 2005; Wilson and Stevens 2005) and dispersal (Fraser, Gilliam et al. 2001; Cote, Fogarty et al. 2010). Furthermore, boldness has been shown to be heritable (Brown, Burgess et al. 2007; Biro and Post 2008) and may be linked to life-history productivity (Biro and Stamps 2008).

(2)   **Exploration-avoidance**, henceforth "exploration", is defined as an *individual's reaction to a new situation, which includes behavior towards a new* habitat, new food or novel objects (Réale, Reader et al. 2007). Of importance for this study is that a new situation can also be considered risky and therefore be linked to boldness if, for example, a new object may represent a potential predator (Réale, Reader et al. 2007). As boldness, exploration is linked with fitness and has been described to be a major target of selection (Réale, Reader et al. 2007; Smith and Blumstein 2008). Exploration has moreover been shown to be linked with activity (Budaev 1997).

(3)   **Activity** is defined as the general level of activity in an individual (Réale, Reader et al. 2007). Activity is linked to boldness and exploration (Wilson and Godin 2009). It can interfere with the measurement of these temperament traits, therefore activity should be assessed in a non-risky and non-novel environment as proposed by Renner (1990).

(4)   **Aggressiveness** is defined as *an individual's agonistic behavio*r towards conspecifics (Réale, Reader et al. 2007). Aggressiveness is linked to boldness and activity (Biro, Beckmann et al. 2009; Huntingford, Andrew et al. 2010) and is also relevant regarding fitness and survival (Dall, Houston et al. 2004).

(5)   **Sociability** is defined as *an individual's reaction to the presence or absence* of conspecifics, under the exclusion of aggressive behavior (Réale, Reader et al. 2007). Sociability is linked to boldness in fishes (Ward, Thomas et al. 2004; Cote, Fogarty et al. 2010) and has been shown to be fitness-relevant (Krause and Ruxton 2002; Dall, Houston et al. 2004).

Parallel to the concept of temperament, another approach has been currently adapted by more and more researchers, termed the behavioral syndromes approach (Sih, Bell et al. 2004 a, b). A behavioral syndrome is a property of a population, referring to inter-individual correlations between rank-order differences across time and / or situations *in contrast to a "behavioral type", which refers to a* particular intra-individual configuration of behaviors (Bell 2007). Behavioral

syndromes occur when individual differences in behavior are consistent across contexts *and are analogous to "personality" or "temperament"* (Sih, Bell et al. 2004 a, b). In other words, a behavioral syndrome is a suite of correlated behaviors across multiple (at least two) observations (Sih and Bell 2008). The definition of a behavioral syndrome therefore is based on behavioral correlations across contexts. This includes different behaviors in the same situation and the same behavior in different situational or temporal contexts.

In the most simple case, a behavioral syndrome can be given when one correlation over two observations is found (Sih and Bell 2008). The definitions of both behavioral syndrome and temperament incorporate individual differences in behaviors that are consistent over time and / or situations (Sih, Bell et al. 2004; Bell 2007; Réale, Reader et al. 2007; Sih and Bell 2008). The concepts of temperament and behavioral syndromes are mostly analogue. The main difference between the two concepts is that the behavioral syndrome is more broadly defined than most available definitions of temperament. As opposed to many of the definitions of temperament (see Réale, Reader et al. 2007), a behavioral syndrome does not necessarily involve a genetic basis (Sih and Bell 2008). Therefore, a temperament trait can be considered a behavioral syndrome but a behavioral syndrome is not always a temperament trait. However, most authors use the term behavioral syndrome when correlations between distinct temperament traits are found (e.g. Kortet and Hedrick 2007; Biro and Stamps 2008; Wilson and Godin 2009), which shall also be done in this study.

Thus, the behavioral syndrome approach emphasizes cross-context carryovers and therefore highlights a more holistic view on behavior in the context of ecological study (Sih, Bell et al. 2004 a, b), with special regards to the dominant model in behavioral biology, the optimality approach. The optimality approach assumes that natural selection favors behavior which yields the highest fitness gain in each and every different context (Potochnik 2009), leading to a view that different behaviors shown in different contexts are independent of each other. It also suggests a disregard of behavioral variability on the level of individuals, which has been put aside as maladaptive "noise" (Dall, Houston et al. 2004) or even "statistical noise" (Wilson 1998) around an adaptive mean in the population.

However, for a variety of species, behavioral syndromes research has shown that behaviors are correlated across contexts (e.g. Johnson and Sih 2005; Dingemanse, Wright et al. 2007; Moretz, Martins et al. 2007; Wilson and McLaughlin 2007; Sih and Bell 2008; Smith and Blumstein 2008; Wilson and Godin 2009). Cross-context spillovers could therefore account for suboptimal behavior in a given situation (Sih and Bell 2008). For instance an individual with an aggressive behavioral type might benefit by aggressively defending food resources against conspecifics but, due to the individual's behavioral type limiting behavioral plasticity, may behave inappropriately bold in the presence of a predator (Sih and Bell 2008). Boldness-aggression syndromes are commonly found in animals (Bell 2005; Kortet and Hedrick 2007; Sih and Bell 2008; Huntingford, Andrew et al. 2010). Where the optimality approach fails to account for limited behavioral plasticity in a given situation, the behavioral syndromes approach can offer explanations why apparently maladaptive behavior e.g. precopulatory sexual cannibalism in spiders (Johnson and Sih 2005), is not eradicated by natural selection and remains within a population (Sih, Kats et al. 2003; Bell and Stamps 2004; Sih, Bell et al. 2004; Dingemanse, Wright et al. 2007; Stamps 2007). Behavioral syndromes have been shown to be heritable (van Oers, de Jong et al. 2005) and to vary amongst populations of the same species in different ecological contexts (Bell 2005; Dingemanse, Wright et al. 2007). This suggests that evolution favors trait combinations that should prove to be optimal in a given ecological context rather than optimal behavior in each and every context (Dingemanse, Wright et al. 2007).

For the understanding of the impact of ecological contexts on the evolution of temperament it can be helpful to compare populations of a species that have been subjected to natural selection in the wild with populations that have been subject to selection pressures due to domestication (Verbeek, Iwamoto et al. 2007). Domestication involves a transition from wild to captive status that comes along with changes in social environment, predation and availability of resources such as

food, space or shelter and is defined as that process by which a population of animals becomes adapted to man and to the captive environment by genetic changes occurring over generations and environmentally induced developmental events reoccurring during each generation (Price 1984).

The development of the domestic phenotype is influenced by phenotypic changes resulting from a variety of mechanisms including inbreeding, genetic drift, artificial selection, natural selection in captivity, and relaxed selection (Price 2002). Inbreeding will typically result in a reduction in genetic variability (Price 1984). Genetic drift, like inbreeding, tends to reduce genetic variability within populations by increasing homozygosity. Furthermore, genetic drift increases variability between populations (Dobzhansky and Pavlovsky 1957). The only genetic mechanism unique to the domestication process is artificial selection.

Artificial selection is goal-orientated, i.e. artificial selections takes place in favor of phenotypical properties desired and selected by man (Price 1984). Artificial selection has a by far more rapid impact on phenotypic changes in animals undergoing domestication than natural selection on free living populations. Through intense artificial selection, a wide variety of behavioral traits in populations of animals can be altered in just a few generations (Price 1984). However, apart from artificial selection, natural selection also occurs in captivity, especially during the first few generations in captivity, depending on the degree of preadaptation of a species for the captive environment provided (Price 2002).

Another important mechanism is relaxed selection. Relaxed selection implies the loss of adaptive significance of certain behaviors (e.g. predator avoidance or reproductive isolating mechanisms) in captivity. This may likely result in an increase of both genetic and phenotypic variability for such behavioral traits (Price 1984; 2002) and also lead to a degradation of such behavioral patterns due to response thresholds increasing above normal stimulation level (Price 1999). With respect to the degradation of anti-predator behavior, it has been shown that populations of fish reared in captivity are bolder than their wild counterparts (Sundstrom, Petersson et al. 2004; Millot, Begout et al. 2009).

A species suitable for studying the effects of domestication on temperament is the common carp (Cyprinus carpio L.) The common carp is a species originally native to Eastern Europe and central Asia (Balon 1995) and was the first fish species to be domesticated (Balon 2004). Today common carp can be found in parts of Europe, Africa, North, Central and South America, Australia and Oceania (Jones and Stuart 2009). Domestication of carp was first documented around 2000 BC in China. In Europe, domestication goes back to 1200 AD (Balon 1995). Soon after, carp production took place in pond systems with separate spawning and growing ponds, which resulted in humpbacked and variously scaled or scale-less domesticated phenotypes appearing in most pond systems due to initially unintentional artificial selection (Balon 2004). Therefore, morphological distinct strains of carp exist in Europe nowadays (Kirpichnikov 1999; Müller-Belecke, Füllner et al. 2009). The dominant strain is the mirror carp, which has a more pronounced humpback and has hardly any scales in contrast to the fully scaled "wild-type" common carp which is more torpedo-shaped (Müller-Belecke, Füllner et al. 2009). Domesticated carp has been genetically improved (Kirpichnikov 1999; Balon 2004) for fast growth and low feed cost in comparison to wild-type common carp (Kohinoor, Islam et al. 2002).

Domesticated common carp may escape from production facilities and hybridize with indigenous populations and thus, alter the gene pool of wild populations of common carp (Khalili and Amirkolaie 2010). Therefore, the fully scaled common carp populations that exist in the wild in Germany are likely to be of feral origin (Balon 1995). However, genetic analysis revealed that the genetic variability within the mirror carp type is lower (4.4 alleles / microsatellite) than within the "wild-type" fully scaled carp (8.2 alleles / microsatellite) (Kohlmann, Kersten et al. 2005). This finding can be explained with genetic drift (Chistiakov and Voronova 2009). Therefore a higher degree of domestication can be assumed for the mirror carp (Price 2002; Vandeputte and Launey 2004). The common carp is of great interest in research, especially regarding ecological implications of temperament, due to its invasiveness (Jones and Stuart 2009), its value for recreational and commercial fisheries, and the possibility to study and compare several genotypes of the same species with different degrees of domestication.

Temperament traits like boldness (Godin and Dugatkin 1996; Budaev 1997; Coleman 1998; Ward, Thomas et al. 2004; Brown, Jones et al. 2005; Moretz, Martins et al. 2007; Biro, Beckmann et al. 2009), exploration (Budaev 1997; Fraser, Gilliam et al. 2001; Kobler, Engelen et al. 2009; Millot, Begout et al. 2009; Wilson and Godin 2009) and sociability (Budaev 1997; Budaev 1997; Krause and Ruxton 2002; Ward, Thomas et al. 2004) have been ascertained in a multitude of fish. But until now only one study exists that assesses temperament related behavioral traits in the carp (Huntingford, Andrew et al. 2010). It suggests the existence of a behavioral syndrome in carp involving boldness and aggressiveness. So far, no study has examined the interrelations of multiple behavioral metrics, which have been utilized for other species to measure several different temperament traits, to determine temperament traits in the carp. Furthermore, no study of temperament in carp has included different genotypes.

2. Aim of the study and hypotheses

The aim of this study was to examine the existence of consistent behavioral differences in carp (Cyprinus carpio L.) with a focus on measures for the temperament traits boldness, exploration and sociability, as defined by Réale, Reader et al. (2007). To reflect a wide range of behavioral variability in this species, individuals with different genotypes implying different degrees of domestication were used. It was hypothesized that individual differences in behavioral expressions of carp are consistent i) over time and /or ii) across situations, on the basis of which underlying temperament traits and / or behavioral syndromes can be identified. Furthermore, it was predicted that iii) behavioral measures are correlated with life-history productivity (growth rate) and iv) carp with different degree of domestication differ in their behavioral expressions.

3. Materials and methods

3.1 Experimental fish and their origin

A total of 36 juvenile (young-of-the-year, age class 0) carp (mean weight 30.7 ± 6.6 g, range 20-57g) were used for the experiments. All carp used in this study were raised at a commercial fish hatchery (Fischzucht Wegert, Ostercappeln, Germany) in a common garden pond environment. They were progeny of homozygous and heterozygous common carp (Cyprinus carpio L.) parents with a fully scaled body phenotype that had experienced a different domestication history and were stocked into the same pond for reproduction.

In carp, morphologically different scale phenotypes develop due to the fact that the inheritance of scale pattern in carp is governed by alleles at two autosomal loci (S/s and N/n) with epistatic interaction (Steffens 1980; Hulata 1995; Kirpichnikov 1999). Common carp with the "wild type" fully scaled phenotype (hereafter called common carp for simplicity) carry the gene "S" and can be either homozygous (SSnn) or heterozygous for this allele (Ssnn). Mirror carp are always homozygous recessive (ssnn). Other combinations involving the "N" gene result in two other scale phenotypes (line carp: SsNn or SSNn, and naked carp: ssNn) as well as lethal combinations (SSNN, SsNN and ssNN), due to severe pleiotropic effects of the "N" gene on viability, development and fin-shape (Kirpichnikov 1999).

The brood source used as parents for the experimental fish in the present study encompassed two sources: homozygous common carp descending from a selection line with typical phenotypes of "wild-type" common-carp; the "wild-type" morphological phenotype is characterized through an elongated, torpedo-shaped and fully scaled body, a smaller gape size and smaller head dimensions relative to the mirror carp, which is only partially scaled, has larger scales, and also has a more humpbacked, compact body (Steffens 1980; Balon 1995; Balon 2004).

The second brood source encompassed heterozygous common carp parents descending from a selection line where "wild-type" common carp were crossbred with a mirror carp phenotype from Hungary. This situation is indicated by the fact that in the breeding ponds from where the experimental animals in the present study originated both common carp and mirror carp juveniles were present despite the

existence of morphologically "pure" common carp parents. When heterozygous common carp reproduce, 25% of the offspring encompass mirror carp (Kirpichnikov 1999). Therefore, mirror carp used in this study descended exclusively from the heterozygous and highly domesticated selection line of common carp parents. Common carp in the present experiment, however, were offspring of two selection lines and could be either homozygous or heterozygous for the allele "S". Notwithstanding, it is safe to assume that the mirror-type carp encompassed more domesticated fish because the increasing loss of scale pattern in carp is correlated with the degree of domestication in carp (Balon 2004). Indeed, a genetic analysis of the two distinct morphological phenotypes from the same population of the carp used in this study, utilizing 4 microsatellites following the method of Kohlmann et al. (2005) revealed a greater allele diversity in wild-type fully scaled carp relative to mirror carp, which indicates greater bottlenecks as a result of low effective population sizes, as is common in hatcheries (Kirpichnikov 1999; Huntingford 2004; Chistiakov and Voronova 2009). Therefore, due to the existence of different scale phenotypes it is safe to assume that two genotypes were presented and it is highly likely that the mirror-type individuals had a longer domestication history related to the wild-type phenotypes.

After natural spawning in the commercial pond facility, when the fish reached an age of about 6 months the pond was drained and the fish were transported to IGB (Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany) in November 2009, 6 weeks before the experiments started. The fish were housed indoors in a large tank (0.8 x 0.8 x 0.6 m) at room temperature with fresh water supply and continuous aeration. Feeding took place ad libitum once per day with common carp pellets (Aller Performa 2 mm, 45% protein, 20% fat, 18% N-free extract, 8% ash, 2% fibre; Emsland Aller Aqua, Golßen, Germany). Out of these fish a subsample of 18 individuals from each genotype were randomly selected for the experiments. It was impossible to determine the sex of the juvenile fish. A possible sex effect on behavior was neither studied nor tested for in the present experiment. Within the six weeks before the experiment started and during

the experimental time, all fish were fed with common carp pellets (Aller Performa 2 mm, 45% protein, 20% fat, 18% N-free extract, 8% ash, 2% fibre; Emsland Aller Aqua, Golßen, Germany) with a food amount of 1 % body wet weight once a day.

3.2 Identification of individuals

For individual identification, each fish was equipped with a PIT (Passive Integrated Transponder) tag (Trovan 12 mm, Euro ID Identifikationssysteme GmbH & Co, Weilerswist, Germany), which can be read with a hand scanner. The scanner (Trovan Multi-Chip Handlesegerät, Euro ID Identifikationssysteme GmbH & Co, Weilerswist, Germany) reads the tag's electromagnetic code and displays the tag's number, thus allowing comfortable identification of individuals. The PIT tags were surgically implanted in the abdominal cavity of the fish. For PIT implantation fish were anesthetized by bathing them in 10 ml of a clove oil / ethanol solution (Soto and Burhanuddin 1995) added to 20 l of clear tap water, until the fish lost equilibrium and operculum rate became slow and irregular. A 4-5 mm long vertical incision, approximately 5 mm to the posterior of the left pelvic fin and about 4 mm below the sideline, was made by using a scalpel to insert the PIT tag into the body cavity. For common-type carp and mirror carp with scales at the incision site, 1-2 scales were removed prior to the surgery. The procedure of PIT implantation followed the description by Skov, Brodersen et al. (2005). PIT implantation does not adversely affect behavior (Brännäs, Lundqvist et al. 1994; Knudsen, Johnston et al. 2009) and body condition (Skov, Brodersen et al. 2005) in salmonidae and cyprinidae.

3.3 Experimental holding conditions and general experimental environment

All experiments took place at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB) in Berlin, Germany, between January and March 2010. experiments were conducted in an air conditioned climate chamber (ILKAZELL, inner dimensions: 276 cm x 210 cm x 176 cm; ILKAZELL Isoliertechnik GmbH,

Zwickau, Germany) with a constant temperature of 20 °C, due to recent insights on the severe effects of temperature on the expression of temperament traits in fish (Biro, Beckmann et al. 2009).

The climate chamber was equipped with 6 identical aquaria (130 cm x 40 cm x 40 cm; 208 L), in which the experiments took place. Each aquarium consisted of a central open area compartment (80 cm x 40 cm x 40 cm), henceforth called "arena", bordered by a refuge compartment (25 cm x 40 cm x 40 cm), henceforth called "refuge" and a shoal compartment (same dimensions as the refuge), see figure 1. The shoal compartment was stocked with 4 conspecifics (2 mirror carp and 2 common carp) of similar size as the focal fish (30.75 ± 6.57 g), henceforth referred to as the shoal. The refuge was separated from the arena by an opaque (black) plastic barrier equipped with a remotely operated sliding door (trap door), figure 1. The refuge was used for acclimatization of the focal fish within the aquaria and to shield the fish from experimental setups within the arena and the shoal compartment during non-experimental periods. The sliding door was 20 cm wide, 20 cm high and could be lifted via a remote pulley system from outside the climate chamber, see figure 1. In non-experimental periods the sliding door remained shut.

The shoal compartment was separated from the arena with a transparent PLEXIGLAS® (Röhm GmbH, Darmstadt, Germany) barrier equipped with a one-way window film (Hochleistungsgebäudeglasfolie "15 IS", Werberitter GbR, Berlin, Germany), henceforth called "one-way glass". One-way glass is used for one-sided optical shielding of a space, in this case, the arena, and bears a very low degree of light transmission and a high degree of light reflection. Therefore, it allows for only little light to pass to the other side. Hence the shoal compartment of each aquarium was equipped with extra lighting in form of 60 W spotlights (Sylvana, Danvers, U.S.A.) pointed at the one-way glass (Figure 4).

The extra lighting could be operated independent of the main overhead lighting of the aquaria. The usage of one-way glass in this experimental setup allowed the focal fish to perceive visual cues from the shoal without the shoal being able to interact with the focal fish (for olfactory cues and chemical cues, see below). The one-way glass was covered by a removable opaque (black) plastic barrier in all behavioral assays except for the shoaling tendency (sociability) assay described

below (experimental design; Figure 1-4). The barriers to both sides of the arena contained a 5 mm gap adjacent to the aquarium floor which allowed water to flow freely from the shoal compartment, where the filter outflow was located, through the arena into the refuge where the filter inflow was located. This setup allowed conspecifics' odor cues and chemical cues to diffuse from the shoal through the arena into the refuge. This way, it provided more familiar conditions within the laboratory setup since carp tend to live in groups in the wild and become stressed in the absence of conspecifics (Huntingford, Andrew et al. 2010).

The shoal compartment and the refuge were interchanged in every other aquarium to account for possible side preference effects. The arena was divided into six similar sized grids (3 grids horizontally and 2 grids vertically). It was visualized by using white lines drawn on the front glass plane to facilitate the recording of fish's space- and water column use and numbered 1 – 6 (Figure 1). The water level was by choice 30 cm high. Therefore each of the 6 grids of the arena approximately equaled one fish length in height and two fish lengths in width.

Each aquarium was filled with aerated and dechlorinated tap water. Water temperature was maintained at 20 °C. All aquaria were illuminated overhead with fluorescent tube lighting (Sun-Glo, R.C. Hagen, Montréal, Canada), which simulates natural sunlight spectrum, on a 10:14h light:dark cycle. All aquaria were equipped with external filters (EHEIM Professionel 3 - type 2080, model 350, 1050 l/h output, EHEIM GmbH & Co. KG, Deizisau, Germany). While all behavioral experiments were conducted using one focal fish per aquarium at a time, the other fish were housed in aquaria (90 cm x 50 cm x 50 cm; 225 l) outside the climate chamber. Fish in the holding tanks were manually fed approximately 1% fish biomass once per day. Water temperature outside the climate chamber also remained approximately 20 °C.

All behavioral assays were conducted remotely from outside the climate chamber using a remote pulley system to operate experimental mechanisms (Figure 1 - 4) and by using video cameras to minimize disturbance and observer bias. Every aquarium was equipped with a video camera (OSCAR CCD Camera, 640 x 480 Pixel) positioned opposite of it, allowing to observe and record behaviors in real-time. The video sequences were recorded with a PC equipped with 3 video

grabbing cards (Hauppauge WinTV HVR-1100, Hauppauge Computer Works GmbH, Mönchengladbach, Germany), using the video software VirtualDub version 1.9.8. (http://www.virtualdub.org). Due to space limitation inside the climate chamber, 4 of the 6 aquaria had to be positioned opposite of each other. To prevent visual stimulation of focal fish from other aquaria, remotely operated black curtains were used to cloak all aquaria located vis à vis except the one in which the current experiment was conducted.

3.4 Experimental design

The experimental setup was designed to accommodate a series of behavioral assays. The goal was to assess multiple behavioral variables presumably suitable for the measurement of the temperament traits boldness, exploration and sociability in carp in order to reveal interrelations between behavioral measures allowing insights into temperament structures in carp. For this purpose, experimental aquaria were built which permitted the assessment of behavioral metrics in four different behavioral contexts. The first context served the purpose of quantifying exploratory behavior in the manner of an open field test (Walsh and Cummins 1976), presumably addressing the temperament trait exploration. The assessment of exploration, according to Réale, Reader et al. (2007), is considered to demand a novel environment. Therefore, the experimental context in which the fishes behavioral reactions towards a novel environment was quantified, always took place preceding the other behavioral assessments, which then followed in a randomized order to avoid sequence effects (Díaz-Uriarte 2002). The existence of the temperament trait boldness was investigated by quantifying several behavioral measures which were considered measures for boldness in other studies (e.g. Ward, Thomas et al. 2004; Álvarez and Bell 2007; Wilson and Godin 2009) in two further experimental contexts. One context was based on a novel object test, which quantifies an individual's reaction towards a novel object and is a standard paradigm used to assess boldness in a variety of species (Wilson, Coleman et al. 1993; Carere and van Oers 2004) and the other was a context implying feeding under predation risk. Here, the individuals' latency to begin feeding in a risky

environment and latency to resume feeding after a simulated predation attempt were quantified as presumed measures for boldness (Ward, Thomas et al. 2004; Álvarez and Bell 2007). The last experimental context was designed to quantify sociability as expressed through shoaling tendency (Krause and Ruxton 2002; Cote, Fogarty et al. 2010). As carp are a highly social species (Huntingford, Andrew et al. 2010), all experiments took place under the presence of a group of conspecifics, which was shielded from visual perception but not from olfactory perception in all contexts but the sociability assessment. To maximize behavioral variability, two different genotypes of carp were used.

Focal fish were stocked into the aquaria on Thursdays or Fridays and allowed to recover over the weekend. The selection of focal individuals for each experimental week was randomized by catch order. Six focal fish were netted from the holding tank, scanned for their PIT number, weighed and placed into the refuge compartments of experimental aquaria for acclimatization over the weekend. Behavioral measurements were then conducted on 4 consecutive days using the same fish from Monday until Thursday (henceforth called experimental week). After each experimental day, aquaria were cleaned from debris and food leftovers. Upon the end of one experimental week, focal fish were returned to the housing aquaria outside the climate chamber. Thereafter the water from the experimental aquaria was completely drained, aquaria were thoroughly cleaned, and fresh water was refilled. Then, six new focal fish were caught and handled as described above. The shoal compartments of the aquaria were stocked isochronous with 4 naïve carp each, as described above and thus the shoal and the focal fish remained in the respective experimental aquarium during the whole experimental week.

Testing of focal fish always occurred in a fixed order with respect to aquaria, which were labeled with numbers 1 - 6. During the experimental week, all focal fish in the refuges of the aquaria inside the climate chamber were fed every 24 hours using automatic feeding machines (EHEIM Futterautomat TWIN, EHEIM GmbH & Co. KG, Deizisau, Germany). The feeding machines were programmed so that each focal fish was fed a predefined amount of commercial carp pellets, as mentioned above, equal to 1% of carp body wet weight, 15 hours before the onset of

behavioral assays in order to standardize hunger level. The shoal was manually fed approximately 1% body weight of the same pellets daily, after all experiments of that day were completed and at the same time on non-experimental days.

3.5 Experimental contexts

Each focal fish spent 6-7 successive days in its respective aquarium, and participated in 4 behavioral assays involving 4 experimental contexts, one per day, after an initial 2-3 day acclimatization period. In all experimental contexts, behavior was quantified only in the arena, not in the refuge, following (Wilson and Godin 2009). On the first experimental day, all fish were tested in the exploration context because this test presumably demands a novel environment (Réale, Reader et al. 2007). Experimental context 2, 3 and 4 were randomly selected for day 2, 3 and 4 using dice as shown in table 1.

After all 36 fish completed all 4 behavioral assays, they were anesthetized (as described in section 4.2.), and thereafter weighed and measured. Exact length measurement required anesthetization and, to avoid stress for the fish prior to the experiments, was therefore performed at the end of the experimental trial, i.e. after all fish were tested once in every experimental context. Then, the whole experiment was repeated in the same order of individuals. Thus, every fish was retested exactly 6 weeks after the first run of the experiment. During retesting, assignment of individual test aquaria was randomized. At the end of the retest, all fish were again weighed and measured as described above. The initial experimental trial is henceforth referred to as "trial A" and the retest is referred to as "trial B".

Table 1: Randomization of trial orders: Allocation of behavioral assay orders to dice scores.

Behavioral Assay Order			
Dice score	1st assay (day 1)	2nd assay (day 2)	3rd assay (day 3)	4th assay (day 4)
1	Exploration	Novel Object	Resume Feeding	Shoaling
2	Exploration	Resume Feeding	Shoaling	Novel Object
3	Exploration	Shoaling	Novel Object	Resume Feeding
4	Exploration	Novel Object	Shoaling	Resume Feeding
5	Exploration	Resume Feeding	Novel Object	Shoaling
6	Exploration	Shoaling	Resume Feeding	Novel Object

3.5.1 Experimental context 1: Latency to emerge from a refuge and explore a novel environment (Exploration)

Individual carp were tested for their willingness to explore a novel environment following Wilson and Godin (2009). Upon trial start the trap door separating the refuge compartment from the central open area compartment was remotely lifted from outside the climate chamber using a manually operated pulley system (Figure 1). Thereby, the fish were allowed to enter and explore the arena. After lifting the door each fish was given 60 minutes to leave the refuge compartment and to investigate the arena. Upon entering the arena, which was defined as the first time that the focal fish emerged from the refuge with its full body length, several behavioral measures were quantified for the following 10 minutes. Once lifted, the trap door remained open during the course of the trial and thus the fish was allowed to return to the refuge at any time.

Quantified behavioral measures included the latency of each fish to leave the refuge (minutes), time spent active in the arena and time spent in the upper zone of the water column. The latency to leave refuge was defined as the time span between the opening of the trap door and the subject's complete emergence from the refuge compartment. The latency to leave a refuge is widely used as a measure for exploration (Réale, Reader et al. 2007) or boldness (Brown, Jones et al. 2005; Wilson and Godin 2009). In this study, following Réale, Reader et al. (2007), latency to leave a refuge is presumed to be a measure for exploratory behavior because it is measured in a novel and potentially risky environment (whereas boldness is to be assessed at predation risk in a non-novel environment, see Réale, Reader et al. 2007). Time spent active in the arena was defined as the total amount of time the focal fish was actively moving (in contrast to holding position or resting on the bottom) inside the arena within the 10 minute observation period and is a measure for exploration-activity respectively activity in a risky environment, not to be confused with activity, which has to be assayed in a non-novel environment (Réale, Reader et al. 2007). Time spent in the upper water zone was defined as the

total amount of time that the fish spent in the upper half of the test aquarium. Time spent in the proximity of the water surface has been used as a measure for risk-taking or boldness in other studies (Wilson and Godin 2009).

After all fish were tested, they were gently driven back into their refuge compartments and the trap doors were shut. Thereafter necessary modifications to the experimental setup were made in the arena for the following experimental context on the next day and the shoals were manually fed. The focal fish were then automatically fed as described above and left undisturbed overnight until the onset of the next behavioral assay on the following day.
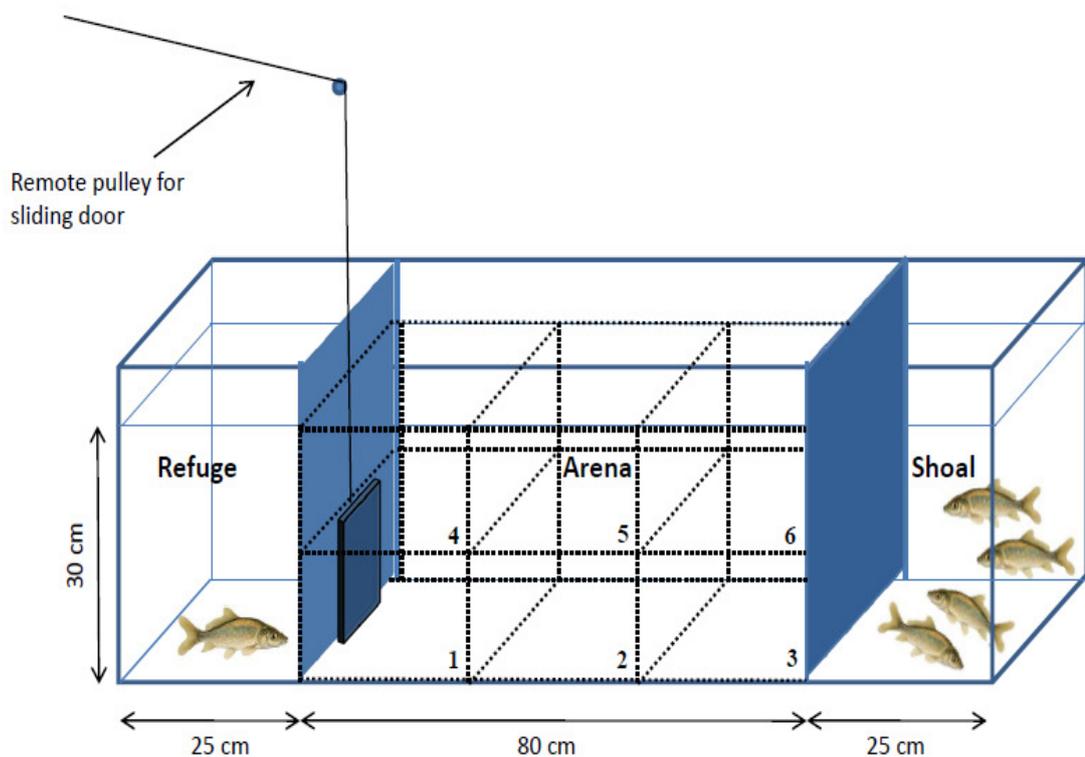
Figure 1: Schematic representation of the experimental aquarium used in experimental context 1 (Exploration) to quantify exploratory behavior in a novel environment. The arena is divided into six grids numbered 1-6 for quantification of locomotor behavior.

3.5.2 Experimental context 2: Novel Object Test (Novel Object)

On the day after completion of the previous experiment, Fish were tested for their willingness to approach a novel object in the arena (Figure 2) following Frost, Winrow-Giffen et al. (2007). This experiment is known to test for boldness (Wilson, Coleman et al. 1993; Wilson, Clark et al. 1994; Frost, Winrow-Giffen et al. 2007). The novel object was placed beforehand in the centre of the middle third of the arena, as to be visible from inside the refuge as soon as the door was opened. The novel object was constructed as a cuboid of 10 cm x 6.5 cm x 3.2 cm from 5 yellow rectangular 8-hole LEGO Duplo bricks (LEGO Group, Billand, Denmark), following Frost, Winrow-Giffen et al. (2007). Hollow space inside the object was filled with sand to allow the object to firmly stay on the bottom of the aquarium. Apart from the presence of the novel object in the centre of the arena, the experimental setup was identical to experimental context 1 (Exploration) described above. Again, onset of the behavioral assay was given by lifting of the trap door, allowing the focal fish to enter the arena. The fish was given 60 minutes to enter the arena and explore the novel object. Upon entry, a 10 minutes observational period began. Proximity to the novel object, defined as the total amount of time spent within one fish length of the novel object, was recorded as a measure for risk-taking behavior. Apart from the measure for risk-taking behavior, the latency to leave refuge as defined in context 1 was additionally recorded.

Figure 2: Schematic representation of the experimental setup used in experimental context 2 (Novel object) to quantify inspection behavior towards a novel object. The experimental setup is identical to that in experimental context 1 but for the addition of a novel object built from yellow Lego Duplo bricks. Dashed lines around the novel object mark the association zone.

3.5.3 Experimental Context 3: Resume feeding after a fright stimulus (Resume feeding)

On the next day after completion of the prior experiment, the fish`s willingness to feed under risk and to resume feeding after a fright stimulus was tested on the basis of Höjesjö, Johnsson et al. (1999), Ward, Thomas et al. (2004) and Álvarez and Bell (2007). The underlying assumption was that individuals differ in risk-taking behavior in terms of feeding under risk (Réale, Reader et al. 2007) and recovery from a fright stimulus (Ward, Thomas et al. 2004). The general setup for this experimental context was analogue to the exploration context described above. In addition, an apparatus was fitted in the center of the transparent coverage of the aquaria, above the arena, containing a mechanism that allowed controlled release of a food pellet, followed by a fright stimulus and simultaneously, another food particle (Figure 3). This apparatus consisted of a 10 cm long plastic tube which was embedded into the coverage. The plastic tube was half cut on two positions,

allowing thin plastic slides, connected to a remote pulley system, to be slid into the cuts, hereby covering the whole diameter of the plastic tube. On top of each plastic slide, a single food pellet was placed. In addition, the second slide also retained a common fishing weight (lead weight 40 g) from falling into the aquaria. By pulling a line from outside the climate chamber, the first plastic slide was removed from the tube and the single food pellet fell into the aquaria. By pulling the second slide, the weight and the pellet were released together, initializing a fright stimulus for the fish (which was found to be effective in a pilot study) and providing more food contemporaneously. To prevent the fishing weight from dropping onto the bottom of the glass-aquaria, it was stopped with a thin line (3 lbs clear fishing line) after plunging 15 cm deep into the water.

Upon the onset of the experiment, the trap door was opened and the first food pellet was simultaneously dropped into the centre of the arena compartment. The fish was then given 60 minutes to exit the refuge. Once the fish was in the arena, it was given 10 minutes to find and consume the food pellet. The time span between the focal fish completely emerging from the refuge and consumption of the food pellet was defined as "latency to feed" and is probably a measure of boldness following Réale, Reader et al. (2007) because of its measurement in a non-novel yet risky environment (open water feeding situation, potential risk of aerial predation). Once the focal fish was observed to feed, the suspended lead weight was dropped together with the second food pellet. The lead plunging into the water caused a noise and a disturbance of the water surface which elicited a considerable fright response. As soon as the falling weight was stopped by the fishing line, it was withdrawn from the aquarium by pulling another line from outside the climate chamber to prevent a persistent stimulation of the focal fish. The time span between triggering of the fright stimulus and consumption of the second pellet was defined as "latency to resume feeding" and was considered a measure for risk-taking or boldness in terms of foraging under predation risk in other studies (Ward, Thomas et al. 2004; Álvarez and Bell 2007). Apart from the two measures for risk-taking behavior, the latency to leave refuge as defined in context 1 was additionally recorded.
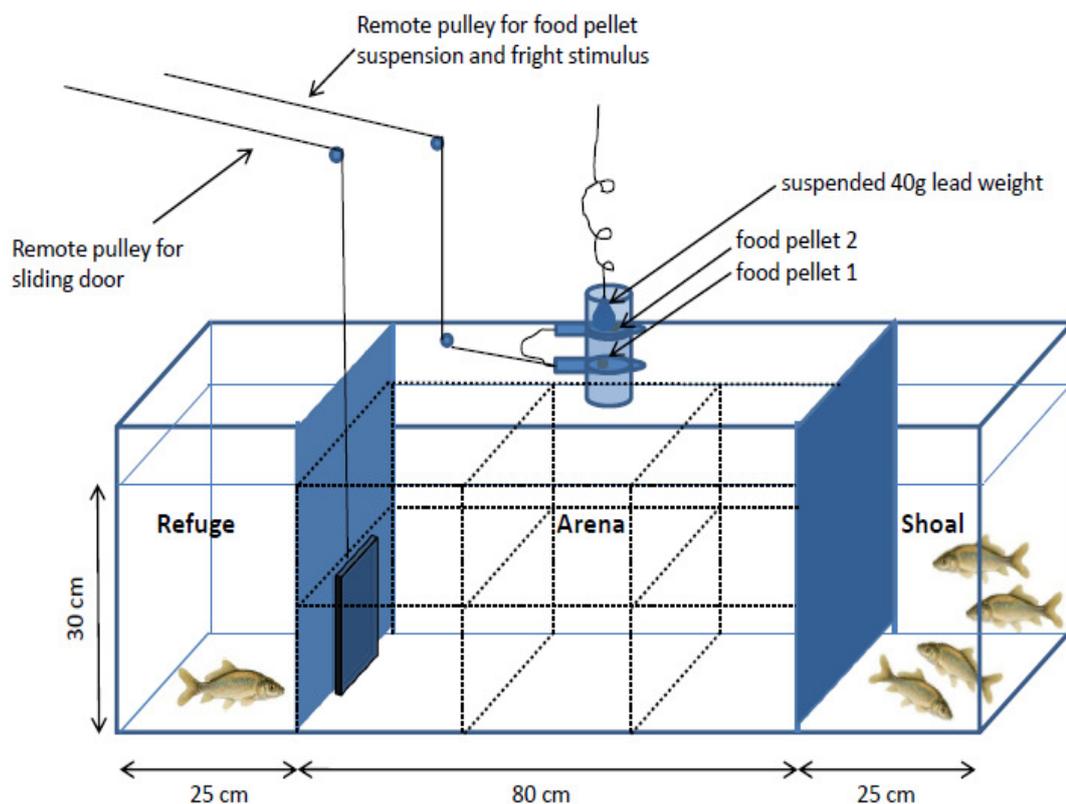
Figure 3: Schematic representation of the experimental setup used in experimental context 3 (Resume feeding) to quantify the latency to feed in a risky environment and the latency to resume feeding after a fright stimulus. The experimental setup is identical to that in experimental context 1 but for the addition of an apparatus allowing suspension of a food pellet and thereafter a further food pellet and the lead weight used as fright stimulus.

3.5.4 Experimental context 4: Shoaling tendency (Sociability)

Following Ward, Thomas et al. (2004) and Krause and Ruxton (2002), individual carp were also tested for their tendency to associate with a stimulus group of conspecifics. For this purpose the coverage of the one-way glass, avoiding visual contacts between the stimulus group and the focal fish in the other experimental contexts was removed antecedent of the overnight acclimatization period between experimental contexts and an additional overhead spotlight facing from the shoal compartment towards the one-way glass separation was switched on preceding the onset of the experiment to provide higher light intensity on the shoal

side of the one-way glass (Figure 4). This is necessary in order for the one-way-glass effect to work properly, which is dependent on a higher light intensity on that side of the one-way glass that is to serve as a visual shield (see section 3.3. above).

Shoaling tendency has been considered a measure for sociability in other studies (Ward, Thomas et al. 2004) and was quantified by recording the amount of time that the focal fish spent within an association zone, defined as within 2 fish lengths of the one-way glass separation (Figure 4). Over a period of 10 minutes, provided the focal fish emerged from refuge within 60 min, sociability of the fish was tested. In nature, members of a shoal are usually less than 3 fish lengths apart from their conspecifics (Pitcher, Magurran et al. 1986; Krause and Ruxton 2002). Apart from shoaling tendency, the latency to emerge from the refuge was recorded as defined in experimental context 1.
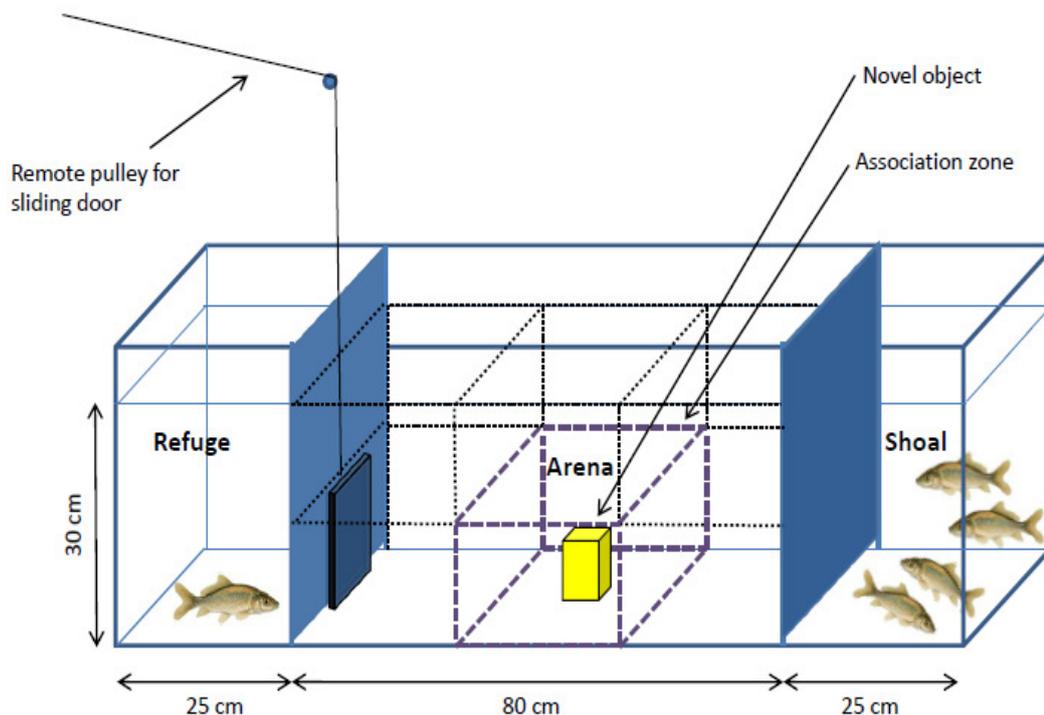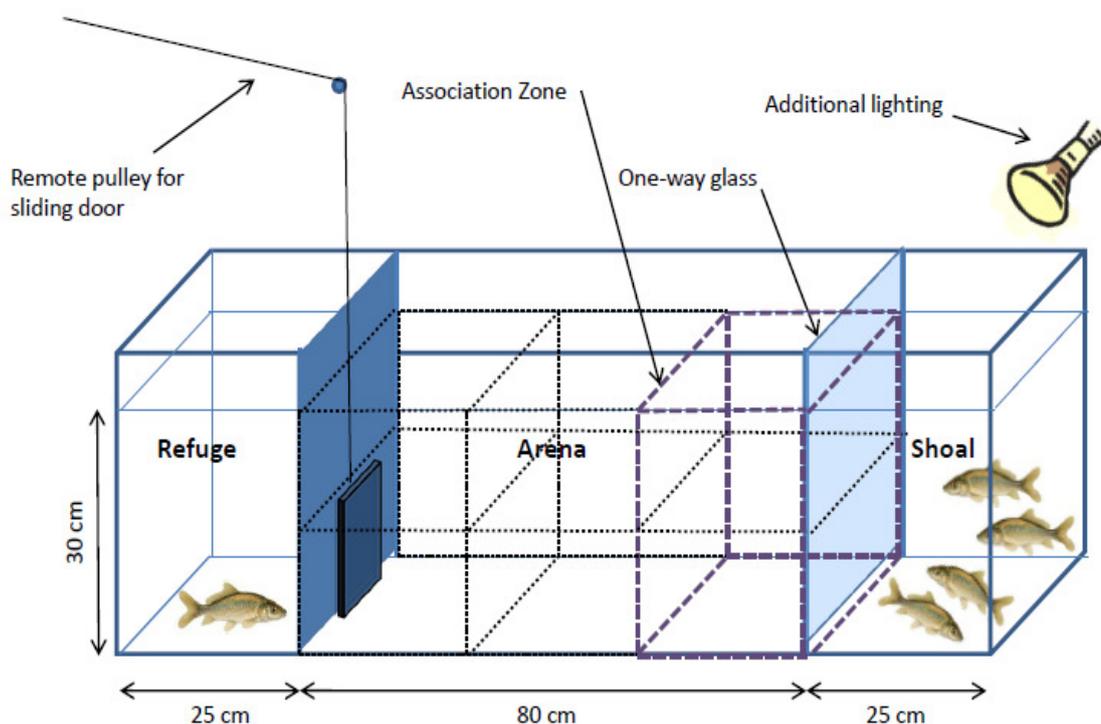

Figure 4: Schematic representation of the experimental setup used in experimental context 4 (Shoaling tendency) to quantify association behavior towards the shoal. The experimental setup is identical to that in experimental context 1 but for the removal of an opaque coverage revealing the one-way glass and an additional lighting. Dashed lines mark the association zone.

3.6 Scoring of behavior

All behaviors and events (Table 2) were quantified from real-time video recordings with the software JWatcher v. 1.0 (http://www.jwatcher.ucla.edu) following Blumstein and Daniel (2007). The JWatcher software created a timeline of key codes pressed during behavioral assessment from which then the duration of behaviors and intervals between events and behaviors as well as the number of their occurrences were calculated. From these event and behavior durations, intervals and occurrences, the variables used in this study (Table 3) were derived.

In detail, the variable "latency to leave refuge" was determined as the time interval between the events "door open" and "enter arena", whereas "Time spent in upper water zone" was calculated by summing up the total time spent in areas 4, 5 and 6. "Time spent within one fish length of a novel object" equaled the time spent in area 2, and "Latency to feed" equaled the time interval between the event "enter arena" and the first recording of the foraging behavior "consume", whereas "Latency to resume feeding" was defined as the time interval between the event "fright stimulus" and the second recording of the foraging behavior "consume". "Time spent within two fish lengths of shoal" equaled the summed-up time spent in the two areas adjacent to the shoal compartment (i.e. area 1 and 4 if the shoal compartment was on the left side or area 3 and 6 if the shoal compartment was on the right side of the respective experimental aquarium) and "Time spent active in the arena" was determined by subtracting the summed time intervals between all "still" and "active" events from the summed time intervals between all "active" and "still" events.

Table 2: Key codings and descriptions of behaviors and events recorded with JWatcher.

Context	Behavior/Event	Key code	Description
Locomotion	Active	"a"	The focal fish uses fins and tail to move in any direction.
	Still	"s"	The focal fish rests on the bottom or holds its position, slight movements possible, but no locomotor activity in any direction using tail fin.
	In area 1-6	"1,2,3,4,5,6"	The focal fish moves into respective area of the arena with more than half of its body length.
Foraging	Consume	"c"	The focal fish completely ingests presented food particle in experimental context 3
Events	Door open	"g"	That moment, in which the sliding door is completely open
	Enter arena	"d"	That moment, in which the focal fish has moved completely outside the refuge
	Back in refuge	"f"	That moment, in which the focal fish has moved completely back inside the refuge
	Fright stimulus	"x"	That moment, in which the dropped fright stimulus has reached the stop position (i.e. is fully suspended)

4. Data handling and statistical data analysis

For orientation purposes all behavioral variables used in this study, their abbreviations, ranges and their special characteristics concerning data handling are listed below (Table 3).

Table 3: Compendium of behavioral variables used, their abbreviations, ranges and special characteristics (specifics).

Behavioral variables

	Abb.	Range (min)	Specifics
Latency to emerge from refuge (Context 1)	LE1	0 - 60	fish that failed to enter the arena within the 60 minute pretrial period were assigned the maximum latency to leave score of 60 minutes
Latency to emerge from refuge (Context 2)	LE2	0 - 60	"
Latency to emerge from refuge (Context 3)	LE3	0 - 60	"
Latency to emerge from refuge (Context 4)	LE4	0 - 60	"
Time spent in upper water zone (Context 1)	UWZ	0 – 10	Fish that failed to enter the arena within 60 minutes were assigned the value "0"
Time spent within one fish length of a novel object (Context 2)	NO	0 – 10	fish that failed to enter the arena within the 60 minute pretrial period were excluded from analysis*
Latency to feed (Context 3)	LF	0 – 10	fish that failed to exit the refuge within 60 minutes or failed to consume the presented food pellet within 10 minutes after entering the arena, were excluded from analysis*
Latency to resume feeding (Context 3)	LRF	0 – 10	fish that failed to consume the second pellet following the fright stimulus within 10 minutes, were assigned a maximum latency to resume feeding score of 10 minutes
Time spent within two fish lengths of shoal (Context 4)	SH	0 – 10	fish that failed to emerge from refuge within 60 minutes were excluded from analysis*
Time spent active in arena (Context 1)	TA	0 – 10	Fish that failed to enter the arena within 60 minutes were assigned the value "0"

* unless stated otherwise

Initially, all ascertained variables were described with sample size (N), mean, median, standard deviation and range, whereby standard deviation and range in particular provide information about variability in behavioral measures. Wilcoxon signed-rank tests were then calculated for all behavioral measures to test for shifts in behavioral responses between trials A and B, indicating habituation effects (Colgan, Nowell et al. 1979; Post and von der Emde 1999).

All statistical tests in this study were calculated with the statistics software package SPSS Statistics v. 17.0 at a type-1-error probability of $\alpha = 0.05$, except Ansari-Bradley tests and permutation tests, which were calculated using the software "R", a language and environment for statistical computing (R Development Core Team 2009).

4.1 Tests of hypothesis i (Behavioral consistency over time)

Concerning hypothesis 1, behavioral consistency over time was assessed by calculating bivariate Spearman rank correlations between the initial (trial A) and the repeated (trial B) measurements of behavior to test for rank-order consistency of individuals over time (6 weeks).

Additionally, repeatability of behavior over time was assessed following Lessels and Boag (1987). For this purpose, one-way ANOVA were computed for each behavioral measure to obtain variance components from which the intraclass correlation coefficients (r) were computed, whereby not normally distributed variables were Ln+1-transformed to approximate normal distributions. The intraclass correlation coefficient "r" estimates repeatability, which is the fraction of behavioral variation that is explained by differences between individuals (see Bell, Hankison et al. 2009). Repeatability was calculated as $r = s^2_A / (s^2 + s^2_A)$, where s^2_A is the inter-individual variance and s^2 is the intra-individual variance over time. A certain behavior or behavioral trait is considered repeatable when the intra-individual variance is relatively low compared to the inter-individual variance (Lessells and Boag 1987; Bell, Hankison et al. 2009). Thus, when individuals show a behavior which is constant over time and there are differences in the expression of this behavior between individuals, then that behavior is considered repeatable.

However, a behavior can be rank-order consistent over time (i.e. a statistical significant Spearman correlation between an earlier and later measurement of the same behavior is given) but not or only low repeatable. A significant Spearman correlation implies that the rank order of individuals' behavioral reactions is largely maintained between two measurements of the same behavior, which will still be the case when e.g. habituation, that is, a steady intra-individual shift of behavioral expression in all individuals (Post and von der Emde 1999), has occurred. The repeatability estimate however, is sensitive to intra-individual shifts in behavioral expression of all individuals (Réale, Reader et al. 2007; Bell, Hankison et al. 2009), as it reflects the proportion of total behavioral variance in a population explained by the inter-individual variance. Thus, a high intra-individual variance over time due to habituation can obscure the inter-individual variance and result in a lower repeatability estimate despite the fact that rank-order consistency over time is given.

In both assessments on consistency of behavioral measures, multiple measurements of the same behavior over different contexts were treated separately (i.e., latency to leave refuge in context 2, 3 and 4), see below.

4.2 Tests of hypothesis ii (Suites of correlated behaviors across situations)

Hypothesis ii was tested by initially calculating bi-variate Spearman rank correlations between all behavioral measures in trial A and between all behavioral measures in trial B to test for correlations in multiple behavioral measures, whereby multiple measurements of the same behavior over different contexts were treated separately (i.e., latency to leave refuge in context 2, 3 and 4), see below. Thereafter, suites of correlated behaviors were identified using principle component analyses (PCA) with varimax rotation for enhanced interpretation of factor structures (Bortz 1999; Quinn and Keough 2006). In this approach, traits correlating with each other will form factors, and the interpretation of which behavioral measures load heavily on a given factor can aid in the identification of underlying (latent) temperament trait dimensions (Quinn and Keough 2006; Cote, Fogarty et al. 2010). Here, initially only variables were considered that concerned the focus of this study and were well-defined in terms of Réale, Reader et al. (2007). Accordingly, variables used for the initial PCA were "Latency to emerge from refuge (Context 1)", "Time spent in

upper water zone (Context 1)", "Time spent within one fish length of a novel object (Context 2)", "Latency to feed (Context 3)", "Latency to resume feeding (Context 3)", and "Time spent within two fish lengths of shoal (Context 4)", all from trial A because at this point it is not clear to what point habituation to novelty (see Réale, Reader et al. 2007) may have occurred between trials A and B, therefore a true novel environment may have been given only in trial A. Behavioral measures with a factor loading value of at least 0.5 were considered meaningful for component interpretation (Peres-Neto, Jackson et al. 2003). Beforehand, all values were z-transformed in order to secure comparability (Daszykowski, Kaczmarek et al. 2007).

Subsequently it was of interest whether factor structures found for trial A were consistent over time respectively similar in trial B. For this purpose a further PCA was conducted using the same variables as above, but from trial B. After sighting of results obtained from the first PCA, further explorative PCA were conducted with different suites of variables from trial A and likewise from trial B to further verify the results obtained from correlations and the first PCA as explained in the results section. To avoid confusion, all factors obtained from PCA were labeled with abbreviations of the variables loading on the respective factor instead of using interpretative labels as done in some other studies (e.g. Cote, Fogarty et. al 2010).

Furthermore, repeatability following Lessells and Boag (1987) and rank order consistency over time (Spearman rank correlations) of factor structures found in PCA from trial A were calculated using summed z-scores of behavioral variables bundled through PCA factors in lieu of PCA factor scores (PC1-scores). Summed z-scores were used because firstly PCA factor scores are not adequate for calculation of repeatability and rank consistency due to weighting diversity of individual parameters and secondly factor structures found in trial B may differ from factor structures found in trial A and thus PC1-scores for a factor in trial A and a factor in trial B are not comparable. Therefore, factor structures which were revealed by PCA in trial A were theoretically reconstructed for both trials by summing z-scores of individual variables shown to load on one factor in PCA from trial A. Summed z-scores were obtained by first assigning missing values that value, which, depending

on the respective variable, equaled the maximum or minimum possible value. This way, an N of 36 fish was obtained for all variables. Then, all variables were transformed in such a manner that all values of variables summed according to factor loadings were equally orientated, i.e. if variables loaded negatively on a factor, they were inverted. Variables were then z-transformed before they were summed according to factor loadings. Hence, one summed variable was obtained per PCA factor, representing the constellation of behavioral variables which the PCA revealed to be associated with one another.

The variables "Latency to emerge from refuge" (Context 1-4) were the only variables that were assessed in all four experimental contexts, therefore, for these variables, consistency over time and contexts as well as repeatability was calculated by first assessing cross-context correlations by conducting a PCA for these four variables in trial A and a second PCA for the same variables measured in trial B. Subsequently, as described above, summed z-scores were obtained from all variables loading on respective factors of PCA from trial A and trial B. Then, using summed z-scores, consistency over time was addressed by calculating spearman correlations between summed z-scores of factor-bundled variables in trial A and trial B and repeatability was determined as described above.

4.3 Tests of hypothesis iii: Interrelations between behavioral metrics and life-history productivity

Since life-history productivity traits such as growth rates of fish can be correlated with temperament traits (Stamps 2007; Biro and Stamps 2008), interrelations were investigated by calculating bi-variate Spearman rank correlations between the relative growth rate for weight (RGR_{weight}), the relative growth rate for standard length ($RGR_{std-length}$) and all behavioral measures and likewise bundled behavioral measures obtained from PCA (summed z-scores of variables co-loading on individual factors). Relative growth rates were computed by using weight and standard length of fish, measured at the end of trial A and at the end of trial B (over a period of 39 days). Relative growth rate (RGR) in % per day for weight was calculated following Hopkins (1992) as $RGR_{weight} = ((W_f - W_i) / W_i \times 100)/t$, where

W_f is the final weight, W_i is the initial weight, and t is the time between W_f and W_i in days. Relative growth rate for standard length was calculated as above with $RGR_{std-length} = ((L_f - L_i) / L_i \times 100)/t$, where L_f is the final length, L_i is the initial length, and t is the time between L_f and L_i in days. Because initial differences in weight and length of test subjects might also influence behavioral performance, bivariate Spearman rank correlations between initial weight of test subjects, length the end of trial A, and behavioral metrics as well as bundles behavioral metrics were also calculated to test for interrelations. Multiple individual measurements of the same behavior over different contexts were not considered (i.e. "Latency to leave refuge" in experimental context 2, 3 and 4) due to redundancy of these behavioral measures.

Differences in relative growth rates for weight and standard length between mirror carp and common carp were assessed by calculating two analyses of covariance (ANCOVA) type general linear model using the variables RGR_{weight} and respectively $RGR_{std-length}$ as dependent variables with genotype as factor and initial standard length as covariate.

4.4 Tests of hypothesis iv: Behavioral differences between common carp and mirror carp

Potential differences between the two genotypes in behavioral expressions regarding behavioral variables assessed in both trials were tested for with permutation tests for differences in mean values of variables were used. A permutation test has several advantages over a simple t-test which would normally be used for this purpose. It can account for small sample sizes, deal with unequal sample sizes and unequal variances of tested samples and does not assume a normal distribution (Ludbrook 1994). To illuminate whether mirror carp exhibit a higher variability in their behaviors than common carp due to domestication effects, Ansari-Bradley tests (Ansari and Bradley 1960) were performed for all behavioral measures in trials A and B respectively.

5. Results

Initially, individual differences were found in all physical and behavioral measures (e.g. in experimental context 1 some individuals left the refuge in less than 1 minute, while other individuals took nearly 60 minutes or never emerged), reflected by standard deviations and ranges for all variables (Tables 4 and 5). In trial B, the mean latency to emerge from refuge (in all contexts) was significantly lower than in trial A (e.g. mean latency to leave the refuge in experimental context 1 was 11.1 minutes versus 30.8 minutes in trial A). Likewise, "Time spent in upper water zone (mean was 0.4 min vs. 0.8 min in trial A)","Latency to feed (mean was 1.1 min vs. 2.8 min in trial A)", "Latency to resume feeding (mean was 3.4 min vs. 7 min in trial A)" and "Time spent within two fish lengths of a shoal (mean was 1.8 min versus 4.3 min in trial A)" were significantly lower. Only the variable "Time spent active in arena" had higher values in trial B (mean was 3.2 min vs. 1.7 min in trial A), although not significantly. "Time spent within one fish length of a novel object" was nearly equal (0.8 minutes respectively, rounded to one decimal) in trials A and B. Wilcoxon signed-rank tests (Table 6) revealed significant differences in the central tendencies of quantified behaviors between trial A and trial B in all analyses but "Time spent active in arena" and "Time spent within one fish length of a novel object".

Table 4: Descriptives (sample size "N", mean, median, standard deviation "SD" and range) of test animals. Relative growth rates were measured over a period of 39 days. All values are rounded to one decimal.

Variable	N	Mean	Median	SD	Range
Initial weight (g)	36	30.8	30	6.57	20 - 51
Weight at the end of trial A (g)	36	35.0	33.5	7.2	21 - 53
Weigh at the end of trial B (g)	36	47.4	46.5	10.1	26 - 76
Standard length at the end of trial A (mm)	36	106.3	105.5	7.18	90 - 122
Standard length at the end of trial B (mm)	36	117.9	117.5	8.0	95 - 138
Relative growth rate for weight (RGR_{weight}) (% / d)	36	0.9	0.8	0.2	0.6 - 1.6
Relative growth rate for standard length ($RGR_{std\text{-}length}$) (% / d)	36	0.3	0.3	0.1	0.1 – 0.4

Table 5: Descriptives (sample size "N", mean, median, standard deviation "SD" and range) of behavioral measures in trial A. All values (except N) are given in metric minutes and rounded to one decimal.

Variable	N	Mean	Median	SD	Range
Latency to emerge from refuge (Context 1 – Trial A)	36	34.4	48.8	27.2	0.0 - 60.0
Latency to emerge from refuge (Context 2 – Trial A)	36	15.7	4.9	21.0	0.0 - 60.0
Latency to emerge from refuge (Context 3 – Trial A)	36	7.9	1.8	16.1	0.1 - 60.0
Latency to emerge from refuge (Context 4 – Trial A)	36	16.2	5.5	21.0	0.2 - 60.0
Time spent active in arena (Context 1 - Trial A)	36	1.7	0.1	2.8	0.0 - 9.3
Time spent in upper water zone (Context 1 - Trial A)	36	0.8	0.0	0.2	0.0 - 0.9
Time within one fish length of novel object (Context 2 - Trial A)	32	0.8	0.6	0.9	0.0 - 4.2
Latency to feed (Context 3 - Trial A)	34	2.8	0.6	3.7	0.0 - 10.0
Latency to resume feeding (Context 3 - Trial A)	28	7.0	9.6	3.6	0.1 - 10.0
Time within 2 fish lengths of shoal (Context 4 - Trial A)	31	4.3	5.5	3.4	0.0 - 8.8
Latency to emerge from refuge (Context 1 - Trial B)	36	11.5	3.3	17.8	0.0 - 60.0
Latency to emerge from refuge (Context 2 - Trial B)	36	7.7	2.2	12.5	0.0 - 60.0
Latency to emerge from refuge (Context 3 - Trial B)	36	1.6	0.7	4.1	0.0 - 25.2
Latency to emerge from refuge (Context 4 - Trial B)	36	3.4	1.2	4.2	0.0 - 20.3
Time spent active in arena (Context 1 - Trial B)	36	3.2	2.2	2.7	0.0 - 8.7
Time spent in upper water zone (Context 1 - Trial B)	36	0.4	0.0	1.03	0.0 - 5.3
Time within one fish length of novel object (Context 2 - Trial B)	35	0.8	0.7	0.8	0.0 - 3.1
Latency to feed (Context 3 - Trial B)	35	1.1	0.3	2.31	0.0 – 10.0
Latency to resume feeding (Context 3 - Trial B)	33	3.6	2.9	3.52	0.0 – 10.0
Time within 2 fish lengths of shoal (Context 4 - Trial B)	36	1.8	1.0	2.03	0.0 – 8.6

Table 6: Wilcoxon signed-rank tests for changes in the central tendencies of behavioral variables between trials A and B.

	N, Z, p
Latency to emerge from refuge (Context 1)	N = 36, Z = -3.708[a], **p = 0.000**
Time active in arena (Context 1)	N = 36, Z = -2.761[b], **p = 0.006**
Time spent in upper water zone (Context 1)	N = 36, Z = -1.373[b], p = 0.170
Time within one fish length of novel object (Context 2)	N = 31, Z = -0.460 [b], p = 0.964
Latency to feed (Context 3)	N = 33, Z = -2.618[a], **p = 0.009**
Latency to resume feeding (Context 3)	N = 27, Z = -3.391[a], **p = 0.001**
Time within two fish lengths of shoal (Context 4)	N = 31, Z = -3.165 [a], **p = 0.002**

[a] based on positive ranks, [b] based on negative ranks

5.1 Behavioral consistency over time (Hypothesis i)

Out of 7 variables tested for rank order consistency over time with bivariate Spearman rank correlations, 3 were found to be consistent over time (Table 7). Individuals were found to be consistent over time in their rank order regarding the behavioral measure "Latency to emerge" (Context 1) (rs = 0.399, P = 0.016, N = 36). Similarly, the behavioral measure "Time spent within one fish length of a novel object" (context 2) significantly correlated between trials A and B (rs = 0.529, p = 0.002, N = 31). Finally, "Time spent within two fish lengths of shoal" (Context 4) was found consistent over time (rs = 0.422, p = 0.018, N = 31).

Behavioral measures which were found to significantly correlate between trial A and B also had the highest repeatability estimates (intraclass correlation coefficients). However, none of the repeatability estimates were found to be significant. Having said this, „Time spent within one fish length of a novel object" (Context 2) and „Time spent within two fish lengths of the shoal" (Context 4) were both found to have intraclass correlation coefficients approaching statistical significance (p = 0.051) despite low sample size.

Table 7: Behavioral consistency (bivariate Spearman rank correlations) and repeatability (r) of behavioral measures (between trial A and trial B).

	Rank consistency	Repeatability
Latency to emerge from refuge (Context 1)	rs = 0.399, **p = 0.016**, N = 36	$F_{35,36}$ = 1.385, p = 0.168, r = 0.162
Time active in arena (Context 1)	rs = 0.174, p = 0.311, N = 36	$F_{35,36}$ = 1.230, p = 0.270, r = 0.099
Time spent in upper water zone (Context 1)	rs = 0.019, p = 0.910, N = 36	$F_{35,36}$ = 0.825, p = 0.725, r = -0.101
Time within one fish length of novel object (Context 2)	rs = 0.529, **p = 0.002**, N = 31	$F_{35,31}$ = 1.785, p = 0.051, r = 0.284
Latency to feed (Context 3)	rs = 0.034, p = 0.851, N = 33	$F_{35,33}$ = 1.184, p = 0.314, r = 0.084
Latency to resume feeding (Context 3)	rs = -0.158, p = 0.430, N = 27	$F_{33,27}$ = 0.480, p = 0.977, r = -0.352
Time within two fishlengths of shoal (Context 4)	rs = 0.422, **p = 0.018**, N = 31	$F_{35,31}$ = 1.793, p = 0.051, r = 0.284

5.2 Suites of correlated behaviors across situations (Hypothesis ii)

In a second step, the correlations between individual behavioral measures were examined separately for trial A and B. Accordingly, in trial A, 20 bi-variate correlations were examined, of which 7 were found to be significantly correlated (Table 8). In particular, „Time spent within one fish length of a novel object" (Context 2) correlated negatively with "Latency to feed" (Context 3), and with "Latency to resume feeding" (Context 3). Furthermore, "Latency to feed" and "Latency to resume feeding" (both Context 3) were correlated. All behavioral measures within experimental context 1, (i.e. "Latency to leave refuge", "Time spent in upper water zone" and „Time spent active in arena") were highly correlated. Moreover, "Time spent within one fish length of a novel object" (Context 2) correlated positively with "Time spent within 2 fish lengths of the shoal" (Context 4).

Table 8: Bivariate Spearman rank correlations between behavioral measures in trial A. Significant p – values are marked in boldface.

	LE1*	UWZ*	NO*	LF*	LRF*	SH*	TA*
Rs							
p							
N							
LE1*		-0.630	-0.015	0.160	0.047	-0.090	-0.897
		0.000	0.935	0.367	0.813	0.632	**0.000**
		36	32	34	28	31	36
UWZ*			-0.026	-0.089	-0.152	0.328	0.742
			0.888	0.617	0.439	0.071	**0.000**
			32	34	28	31	36
NO*				-0.394	-0.542	0.490	0.053
				0.029	**0.004**	**0.007**	0.772
				31	26	29	32
LF*					0.399	-0.295	-0.169
					0.036	0.114	0.340
					28	30	34
LRF*						-0.349	-0.174
						0.074	0.375
						27	28
SH*							0.283
							0.123
							31
TA*							

*Variable names associated with abbreviations can be taken from table 3

In trial B, the same 20 bivariate correlations as above were examined with according variables assessed in trial B. 7 of these variable pairs were found to be significantly correlated (Table 9). In particular, all behavioral measures that were shown to correlate in experimental context 1 of trial A were also correlated in trial B. Moreover, "Time spent active in arena" (Context 1) was also correlated with "Latency to resume feeding" (Context 3) and "Time spent within 2 fish lengths of shoal" (Context 4) in trial B. Furthermore, "Time spent within one fish length of a novel object" (Context 2) and "Time spent within 2 fish lengths of shoal" (Context 4) were correlated, as in trial A. „Time spent within one fish length of a novel object" (Context 2) and "Latency to resume feeding" (Context 3), which correlated

highly significant in trial A, were marginally correlated (p = 0.078) in trial B. In further contrast to the results of trial A, "Latency to feed" (Context 3) and "Latency to resume feeding" (Context 3) are not correlated in trial B.

Table 9: Bivariate Spearman rank correlations between behavioral measures in trial B. Significant p – values are marked in boldface.

rs p N	LE1*	UWZ*	NO*	LF*	LRF*	SH*	TA*
LE1*		-0.401 **0.015** 36	-0.109 0.531 35	0.244 0.159 35	0.487 **0.004** 33	-0.177 0.497 36	-0.370 **0.027** 36
UWZ*			0.033 0.849 35	0.086 0.625 35	-0.165 0.358 33	0.255 0.134 36	0.537 **0.001** 36
NO*				-0.181 0.305 34	-0.316 0.078 32	0.584 **0.000** 35	0.248 0.151 35
LF*					0.007 0.971 33	-0.010 0.954 35	-0.045 0.799 35
LRF*						-0.102 0.572 33	-0.506 **0.003** 33
SH*							0.430 **0.009** 36
TA*							

*Variable names associated with abbreviations can be taken from table 3

All correlations between the measures "Latency to emerge from refuge" (Context 1 - 4) were found to be statistically significant, in trial A (Table 10) as well as in trial B (Table 11).

Table 10: Cross-context correlations (bivariate Spearman rank correlations) between the behavioral measures "Latency to emerge from refuge" in experimental context 1 - 4 (trial A). Significant p – values are marked in boldface.

Rs p N	LE1*	LE2*	LE3*	LE4*
LE1*		0.515 **0.001** 36	0.573 **0.000** 36	0.429 **0.009** 36
LE2*			0.659 **0.000** 36	0.385 **0.020** 36
LE3*				0.657 **0.000** 36
LE4*				

*Variable names associated with abbreviations can be taken from table 3

Table 11: Cross-context correlations (bivariate Spearman rank correlations) between the behavioral measures "Latency to emerge from refuge" in experimental context 1 - 4 (trial B). Significant p – values are marked in boldface.

rs p N	LE1*	LE2*	LE3*	LE4*
LE1*		0.414 **0.012** 36	0.503 **0.002** 36	0.527 **0.001** 36
LE2*			0.603 **0.000** 36	0.396 **0.017** 36
LE3*				0.657 **0.000** 36
LE4*				

*Variable names associated with abbreviations can be taken from table 3

5.3 Further verification of temperament traits and their consistency over time (Hypotheses i & ii)

The first PCA based on z-scores of the seven variables from trial A that concerned the focus of this study and were clearly defined sensu Réale, Reader et al (2007) revealed two factors with eigenvalues greater than 1, composed of three variables each, together explaining 59.5% of total variance (Table 12). Factor 1 incorporated the variables "Latency to emerge from refuge" (Context 1), which loaded negatively on this component, as well as "Time spent in upper water zone" (Context 1), and "Time spent within 2 fish lengths of shoal" (Context 4), which both loaded positively, and was labeled "LE1-UWZ-SH". Factor 2 was composed of the variables „Time spent within one fish length of a novel object" (Context 2), which loaded negatively on this component, and "Latency to feed" (Context 3) as well as "Latency to resume feeding" (Context 3), which both loaded positively and was labeled "NO-LF-LRF".

Table 12: PCA1 - Component loadings of behavioral measures from trial A on two orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 25.

	Principal components	
	Factor 1 **(LE1-UWZ-SH)**	**Factor 2** **(NO-LF-LRF)**
Behavioral measure		
Latency to emerge (Context 1)	-0.752	
Time spent in upper water zone (Context 1)	0.742	
Time within one fish length of Novel object (Context 2)		-0.577
Latency to feed (Context 3)		0.678
Latency to resume feeding (Context 3)		0.825
Time within 2 fish lengths of shoal (Context 4)	0.664	
Eigenvalue	2.201	1.371
Variance explained (%)	30.93	28.61
Total variance explained (%)	59.5	

The PCA based on z-scores for trial B revealed two factors with eigenvalues greater 1, and three (Factor 1) respectively two variables (Factor 2) forming a factor, together explaining 60.4% of variance (Table 13). All variables loaded positively on the respective factors. Factor 1 was composed of the variables "Latency to emerge from refuge" (Context 1), "Latency to feed" (Context 3) and "Latency to resume feeding" (Context 3) and was labeled "LE1-LF-LRF". Factor 2 was composed of the variables „Time spent within 2 fish lengths of a novel object" (Context 2) and „Time spent within 2 fish lengths of shoal" (Context 4) and was labeled "NO-SH". The variable "Time spent in upper water zone" (Context 1) was found to have a factor loading of less than 0.5 and was therefore excluded from PCA in trial B.

Table 13: PCA 1 - Component loadings of behavioral measures from trial B on two orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 32.

	Principal components	
Behavioral measure	**Factor 1 (LE1-LF-LRF)**	**Factor 2 (NO-SH)**
Latency to emerge (Context 1)	0.849	
Time spent in upper water zone (Context 1)		
Time within one fish length of Novel object (Context 2)		0.911
Latency to feed (Context 3)	0.558	
Latency to resume feeding (Context 3)	0.826	
Time within 2 fish lengths of shoal		0.868
Eigenvalue	1.983	1.640
Variance explained (%)	31.81	28.58
Total variance (%)	60.39	

Spearman correlations computed from summed z-scores in trial A and B of variables loading on the respective factors revealed by the PCA in trial A were not found to be significant in the case of factor 1 (LE1-UWZ-SH) and found to be close to statistical significance in the case of factor 2 (NO-LF-LRF) (Table 14). Calculation of repeatability from mean z-scores of variables loading on the

respective factor resulted in a repeatability estimate r = 0.15 for factor 1 (LE1-UWZ-SH), whereas the underlying F-statistics were not significant. Factor 2 (NO-LF-LRF) yielded a repeatability estimate r = 0.26, with close to significant F-statistics (p = 0.06).

Table 14: PCA 1 - Rank consistency over time (bivariate Spearman rank correlations) and repeatability (Intraclass correlation coefficient "r") of bundled variables derived from PCA 1 in trial A.

	Rank consistency	Repeatability
Factor 1 (LE1-UWZ-SH)	rs = 0.204, p = 0.233, N = 36	$F_{35,36}$ = 1.354, p = 0.185, r = 0.150
Factor 2 (NO-LF-LRF)	rs = 0.312, p = 0.064, N = 36	$F_{35,36}$ = 1.694, p = 0.060, r = 0.256

In the second PCA, only variables presumably assessing boldness were used due to possible corruption of presumed exploration measures by reason of a potentially non novel environment in trial B as explained in section 5. For variables from trial A, the PCA revealed two factors with eigenvalues greater than 1 (Table 15), together explaining 74.9% of total variance. Factor 1 was composed of the variables „Time spent within one fish length of a novel object" (Context 2), "Latency to resume feeding" (Context 3) and "Time spent within 2 fish lengths of the shoal" (Context 4) and labeled "NO-LRF-SH". On factor 2, only the variable "Latency to feed" (Context 3) loaded. This factor was labeled "LF".

Table 15: PCA 2 - Component loadings of behavioral measures from trial A on two orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 25.

	Principal components	
Behavioral measure	**Factor 1 (NO-LRF-SH)**	**Factor 2 (LF)**
Time within one fish length of novel object (Context 2)	0.731	
Latency to feed (Context 3)		0.947
Latency to resume feeding (Context 3)	-0.779	
Time within 2 fish lengths of shoal	0.826	
Eigenvalue	1.832	1.163
Variance explained (%)	45.788	29.079
Total variance (%)	74.86	

For variables from trial B, the PCA also revealed two factors with eigenvalues greater than 1 (Table 16), together explaining 74.4% of total variance. Factor 1 was composed of the variables „Time spent within one fish length of a novel object" (Context 2) and „Time spent within 2 fish lengths of the shoal" (Context 4), and was labeled as "NO-SH". In contrast to trial A, the variable "Latency to resume feeding" (Context 3) loaded on factor 2 together with "Latency to feed" (Context 3) and was labeled LF-LRF.

Table 16: PCA 2 - Component loadings of behavioral measures from trial B on two orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 32.

	Principal components	
Behavioral measure	**Factor 1 (NO-SH)**	**Factor 2 (LF-LRF)**
Time within one fish length of novel object (Context 2)	0.921	
Latency to feed (Context 3)		0.819
Latency to resume feeding (Context 3)		0.785
Time within 2 fish lengths of shoal	0.899	
Eigenvalue	1.664	1.313
Variance explained (%)	44.39	30.02
Total variance (%)	74.42	

Spearman correlations computed from summed z-scores in trial A and B of variables loading on the respective factors revealed by the PCA in trial A were neither found to be significant for factor 1 (NO-LRF-SH) or factor 2 (LF) (Table 17). Calculation of repeatability from mean z-scores resulted in a repeatability estimate of r = 0.18 for factor 1 (NO-LRF-SH) with non-significant underlying F-statistics. Factor 2 (LF), yielded a lower repeatability estimate (r = 0.08) than factor 1 with also non-significant underlying F-statistics (Table 17).

Table 17: PCA 2 - Rank consistency over time (bivariate Spearman rank correlations) and repeatability (Intraclass correlation coefficient "r") of bundled variables derived from PCA 2 in trial A.

	Rank consistency	Repeatability
Factor 1 (NO-LRF-SH)	rs = 0.189, p = 0.269, N = 36	$F_{35,36}$ = 1.442, p = 0.140, r = 0.180
Factor 2 (LF)[1]	rs = 0.034, p = 0.851, N = 33	$F_{35,33}$ = 1.184, p = 0.314, r = 0.084

[1]because factor 2 consisted of but one variable ("Latency to feed (Context3)"), rank consistency and repeatability was taken from table 7.

PCA 3 was conducted with all available variables. In trial A, three factors with eigenvalues greater than 1 were revealed (Table 18), together explaining 66.3% of total variance. Factor 1 was composed of the variables "Latency to

emerge from refuge" (Context 1 and 2), "Time spent in upper water zone" (Context 1) and „Time spent active in arena" (Context 1), as well as "Time spent within 2 fish lengths of the shoal" (Context 4) and was labeled "LE1-LE2-UWZ-TA". Factor 2 consisted of the variables "Latency to emerge from refuge" (Context 3 and 4) and was labeled "LE3-LE4". Factor 3 was composed of the variable "Latency to emerge from refuge" (Context 2), which loaded both on factor 1 (but negatively here) and factor 3, as well as the presumed measures for boldness "Latency to feed" (Context 3), "Latency to resume feeding" (Context 3) and "Time spent within one fish length of a novel object" (Context 2) and was labeled "LE2-NO".

Table 18: PCA 3 - Component loadings of all behavioral measures from trial A on three orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 25.

	Principal components		
Behavioral measure	**Factor 1** **(LE1-LE2-UWZ-TA)**	**Factor 2** **(LE3-LE4)**	**Factor 3** **(LE2-NO)**
Latency to emerge from refuge (Context 1)	-0.707		
Latency to emerge from refuge (Context 2	-0.509		0.556
Latency to emerge from refuge (Context 3)		0.869	
Latency to emerge from refuge (Context 4)		0.901	
Time spent in upper water zone (Context 1)	0.810		
Time spent active in arena (Context 1)	0.863		
Time within one fish length of novel object (Context 2)			0.590
Latency to feed (Context 3)			-0.575
Latency to resume feeding (Context 3)			-0.832
Time within 2 fish lengths of shoal (Context 4)	0.511		
Eigenvalue	2.564	2.074	1.993
Variance explained (%)	25.640	20.740	19.934
Total variance (%)	66.314		

In trial B, PCA 3 revealed four factors (Table 19) which were composed differently than in trial A (Table 18). Factor 1 was composed of the variables "Latency to emerge from refuge" (Context 1-3) as well as "Latency to resume feeding" (Context 3) and labeled "LE1-LE2-LE3-LRF". Factor 2 was composed of the variables "Time within one fish length of novel object" (Context 2) and "Time spent within 2 fish lengths of shoal" (Context 4) and labeled "NO-SH". Factor 3 consisted of the variables "Latency to emerge from refuge" (Context 4) and "Latency to feed" (Context 3) and was labeled "LE4-LF". Factor 4 consisted of the variables "Time spent in upper water zone" (Context 1) and "Time spent active in arena" (Context 1) and was labeled "UWZ-TA".

Table 19: PCA 3 - Component loadings of all behavioral measures from trial B on four orthogonally rotated principal components. Only loadings > 0.5 are considered. N = 32.

	Principal components			
Behavioral measure	**Factor 1 (LE1-LE2-LE3-LRF)**	**Factor 2 (NO-SH)**	**Factor 3 (LE4-LF)**	**Factor 4 (UWZ-TA)**
Latency to emerge from refuge (Context 1)	0.762			
Latency to emerge from refuge (Context 2	0.803			
Latency to emerge from refuge (Context 3)	0.767			
Latency to emerge from refuge (Context 4)			0.807	
Time spent in upper water zone (Context 1)				0.781
Time spent active in arena (Context 1)				0.893
Time within one fish length of novel object (Context 2)		0.902		
Latency to feed (Context 3)			0.776	
Latency to resume feeding (Context 3)	0.712			
Time within 2 fish lengths of shoal (Context 4)		0.894		
Eigenvalue	2.449	1.814	1.610	1.525
Variance explained (%)	24.486	18.141	16.103	15.250
Total variance (%)	73.980			

Spearman correlations computed from summed z-scores of bundled variables derived from factor loadings from PCA 3 in trial A were found to be close to statistical significance for factor 1 (LE1-LE2-LE3-TA) and statistically significant for factor 2 (LE3-LE4) whereas factor 3 (LE2-NO) did not correlate significantly (Table 20). Calculation of repeatability resulted in a repeatability estimate of r = 0.34 with significant underlying F-statistics in the case of factor 2 (LE3-LE4) and in a repeatability estimate of r = 0.22 in the case of factor 1(LE1-LE2-LE3-UWZ-TA) with marginally significant underlying F-statistics (Table 20).

Table 20: PCA 3 - Rank consistency over time (bivariate Spearman rank correlations) and repeatability (Intraclass correlation coefficient "r") of bundled variables derived from PCA 3 in trial A.

	Rank consistency	Repeatability
Factor 1 (LE1-LE2-LE3-UWZ-TA)	rs = 0.293, p = 0.082, N = 36	$F_{35,36}$ = 1.564, p = 0.093, r = 0.220
Factor 2 (LE3-LE4)	rs = 0.446, **p = 0.006**, N = 36	$F_{35,36}$= 2.011, **p = 0.020** r = 0.336
Factor 3 (LE2-NO)	rs = 0.042, p = 0.809, N = 36	$F_{35,36}$= 1.011, p = 0.486, r = 0.006

PCA 4 which was computed from the "Latency to leave refuge" variables from all 4 contexts for trial A and trial B respectively, revealed one factor with eigenvalue greater than 1, with all variables loading within, in trial A (Table 21) and likewise in trial B (Table 22). The factor was labeled "LE1-4". Total variance explained was 55.5% in trial A and 53.8% in trial B.

Table 21: PCA 4 - factor loadings of PCA from variables "Latency to emerge from refuge" (Context 1-4). Only loadings > 0.5 are considered. N = 25 (trial A) respectively N = 32 (trial B).

Behavioral measure	Principal components (PCA for trial A) Factor 1 (LE1-4)	Principal components (PCA for trial B) Factor 1 (LE1-4)
Latency to emerge from refuge (Context 1)	0.719	0.801
Latency to emerge from refuge (Context 2	0.663	0.842
Latency to emerge from refuge (Context 3)	0.824	0.679
Latency to emerge from refuge (Context 4)	0.764	0.582
Eigenvalue	2.220	2.151
Variance explained (%)	55.50	53.79
Total variance (%)	55.50	53.79

The bivariate Spearman rank correlation computed from summed z-scores in trial A and B of variables loading on the factor LE 1-4 revealed by PCA 4 in trial A was found to be highly significant (p = 0.007). Calculation of repeatability for the factor exploration revealed a repeatability estimate of r = 0.45 with significant underlying F-statistics (Table 22).

Table 22: PCA 4 - Rank consistency over time (bivariate Spearman rank correlations) and repeatability (Intraclass correlation coefficient "r") of bundled variables derived from PCA 4 in trial A.

	Rank consistency	Repeatability
Factor 1 (LE1 – 4)	rs = 0.440, **p = 0.007**, N = 36	$F_{35,36}$ = 2.642, **p = 0.020**, r = 0.449

5.4 Interrelations between behavioral metrics and life-history productivity (Hypothesis iii)

Significant correlations between the weight of test subjects on onset of the experiments (initial weight) and all behavioral variables used in this study could not be found, neither in trial A nor in trial B. Standard length, ascertained at the end of

trial A, was found to significantly correlate with behavioral variables in only one case (Table 23).

Table 23: Bivariate Spearman rank correlations between initial weight of fish, standard length (at the end of trial A) and behavioral measures in trial A and trial B.

	Behavioral measure	Trial A			Trial B		
		rs	p	N	rs	p	N
Initial weight (g)	Latency to emerge from refuge (Context 1)	-0.274	0.106	36	0.238	0.163	36
Initial weight (g)	Time spent within one fish length of a novel object (Context 2)	-0.066	0.719	32	-0.118	0.500	35
Initial weight (g)	Time active in arena (Context 1)	-0.204	0.233	36	0.037	0.829	36
Initial weight (g)	Time in upper water zone (Context 1)	-0.211	0.217	36	-0.022	0.900	36
Initial weight (g)	Latency to feed (Context 3)	0.160	0.367	34	-0.230	0.183	35
Initial weight (g)	Latency to resume feeding (Context 3)	-0.027	0.891	28	0.008	0.967	33
Initial weight (g)	Time within two fish lengths of shoal (Context 4)	-0.089	0.634	31	-0.168	0.326	36
Standard length (mm)	Latency to emerge from refuge (Context 1)	0.343	**0.041**	36	0.251	0.140	36
Standard length (mm)	Time spent within one fish length of a novel object (Context 2)	0.011	0.952	32	0.155	0.373	35
Standard length (mm)	Time active in arena (Context 1)	-0.280	0.098	36	0.152	0.377	36
Standard length (mm)	Latency to feed (Context 3)	0.101	0.569	34	-0.192	0.270	35
Standard length (mm)	Latency to resume feeding (Context 3)	0.012	0.952	28	-0.128	0.479	33
Standard length (mm)	Time within two fish lengths of shoal (Context 4)	0.058	0.757	31	0.025	0.884	36
Standard length (mm)	Time in upper water zone (Context 1)	-0.200	0.241	36	-0.060	0.729	36

Of the correlations between relative growth rates for weight (RGR_{weight}) and standard length ($RGR_{std-length}$) with behavioral measures in trial A and B, only "Latency to feed" (Context 3) in trial B was significantly correlated with $RGR_{std-length}$ (Table 24). However, "Time spent within one fish length of a novel object (Context 2)" marginally was correlated (p = 0.065) with RGR_{weight} in trial A.

Table 24: Bivariate Spearman rank correlations between relative growth rates for weight and standard length, and behavioral measures.

		Trial A			Trial B		
	Behavioral measure	**rs**	**p**	**N**	**rs**	**p**	**N**
RGR_{weight}	Latency to emerge from refuge (Context 1)	-0.112	0.516	36	-0.113	0.510	36
RGR_{weight}	Time spent active in arena (Context 1)	0.196	0.251	36	-0.104	0.546	36
RGR_{weight}	Time spent in upper water zone (Context 1)	0.600	0.729	36	0.060	0.729	36
RGR_{weight}	Time spent within one fish length of a novel object (Context 2)	-0.145	0.429	32	-0.316	0.065	35
RGR_{weight}	Latency to feed (Context 3)	0.008	0.966	34	-0.251	0.145	35
RGR_{weight}	Latency to resume feeding (Context 3)	-0.116	0.556	28	0.099	0.583	33
RGR_{weight}	Time within two fish lengths of shoal (Context 4)	-0.019	0.920	31	0.089	0.606	36
$RGR_{std-length}$	Latency to leave (Context 1)	-0.151	0.378	36	-0.172	0.317	36
$RGR_{std-length}$	Time active in arena (Context 1)	0.232	0.174	36	-0.006	0.973	36
$RGR_{std-length}$	Time in upper water zone (Context 1)	0.112	0.516	36	0.107	0.534	36
$RGR_{std-length}$	Time spent within one fish length of a novel object (Context 2)	0.013	0.942	36	-0.052	0.765	35
$RGR_{std-length}$	Latency to feed (Context 3)	-0.156	0.378	34	-0.479	**0.004**	35
$RGR_{std-length}$	Latency to resume feeding (Context 3)	-0.104	0.599	28	-0.054	0.765	33
$RGR_{std-length}$	Time within two fish lengths of shoal (Context 4)	0.003	0.987	31	0.073	0.673	36

Significant correlations between growth rates for weight (RGR_{weight}) and standard length ($RGR_{std-length}$) and composite behavioral variables derived from factor loadings were found in two cases. Firstly, the factor "LF" from PCA 2 was significantly correlated with $RGR_{std-length}$. However, this is not a true composite behavioral variable because this factor from PCA2 consists but of the variable "Latency to feed" (Context 3). In the second case, factor 3 from PCA 3, composed of "Time spent within 2 fish lengths of a novel object" (Context 2) and "Latency to emerge from refuge" (Context 2), is correlated with $RGR_{std-length}$.

Table 25: Bivariate Spearman rank correlations between relative growth rates for weight and standard length, and factors revealed from PCA 1-4, taken from trial A. Significant p - values are marked in boldface.

		RGR_{weight}			$RGR_{std-length}$	
Factor	**N**	**rs**	**p**	**N**	**rs**	**p**
PCA1 – factor 1 (LE1-UWZ-SH) – trial A	36	-0.081	0.640	36	-0.078	0.652
PCA1 – factor 1 (LE1-UWZ-SH) – trial B	36	0.038	0.826	36	0.179	0.296
PCA1 – factor 2 (NO – LF – LRF) – trial A	36	0.030	0.863	36	0.198	0.246
PCA1 – factor 2 (NO – LF – LRF) – trial B	36	-0.089	0.608	36	0.221	0.195
PCA2 – factor 1 (No – LRF – SH) – trial A	36	0.084	0.625	36	0.179	0.297
PCA2 – factor 1 (No – LRF – SH) – trial B	36	-0.159	0.354	36	0.171	0.320
PCA2 – factor 2 (LF) - trial A	36	0.024	0.889	36	0.195	0.255
PCA2 – factor 2 (LF) - trial B	36	0.190	0.267	36	0.503	**0.002**
PCA3 - factor 1 (LE1 – LE2 – UWZ – TA) - trial A	36	0.132	0.441	36	0.185	0.280
PCA3 - factor 1 (LE1 – Le2 – UWZ – TA) - trial B	36	-0.023	0.896	36	0.208	0.224
PCA3 - factor 2 (LE3 – LE4) - trial A	36	0.233	0.172	36	0.213	0.210
PCA3 - factor 2 (LE3 – LE4) - trial B	36	0.031	0.859	36	0.231	0.176
PCA3 - factor 3 (LE2 - NO) - trial A	36	0.117	0.496	36	0.229	0.178
PCA3 - factor 3 (LE2 - NO) - trial B	36	-0.069	0.688	36	0.339	**0.043**
PCA4 - factor 1 (LE1 – LE2 – LE3 – LE4) – trial A	36	-0.221	0.195	36	-0.214	0.210
PCA4 - factor 1 (LE1 – LE2 – LE3 – LE4) – trial B	36	-0.105	0.543	36	0.288	0.089

Mirror carp and wild-type fully scaled carp were found to differ significantly in relative growth rate for weight (RGR_{weight}) in the ANCOVA type general linear model (Table 26). Initial length as a covariate had no significant effect. Furthermore, the positive coefficient for genotype indicated a higher growth rate in mirror carp than in common carp, which is visualized in Figure 5. The general linear model calculated to assess differences in relative growth rate for standard length ($RGR_{std-length}$) (Table 27) only revealed close to significant differences in genotype ($p = 0.056$). Here too, mirror carp were found to have a higher growth rate than common carp (Figure 5), as indicated by the positive coefficient for genotype. Initial standard length as a covariate did not have a significant effect.
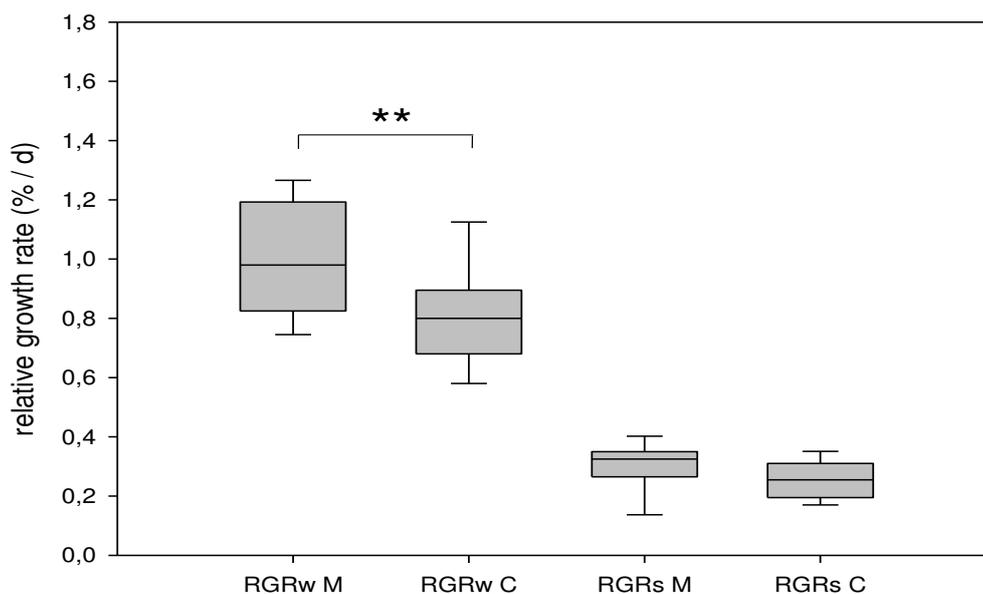

Figure 5: Box plots showing relative growth rates for weight (RGRw) and standard length (RGRs) for mirror carp (M) and common carp (C). Significant differences, tested for with ANCOVA type general linear model, are marked with asterisks.

Table 26: ANCOVA type general linear model: differences in relative growth rate (calculated from weight) between mirror carp (N = 18) and wild-type fully scaled carp (N = 18). Significant p-values are marked in boldface.

Model term	Coefficient	DF	F	*p*
Intercept (RGR$_{weight}$: initial standard length)		2	4.119	**0.025**
Constant term	0.795 (0.512)	1	3.037	0.091
Initial standard length	0.000 (0.005)	1	0.001	0.978
Genotype (mirror carp)	0.195 (0.068)	1	8.200	**0.007**

Table 27: ANCOVA type general linear model: differences in relative growth rate (calculated from standard length) between mirror carp (N = 18) and wild-type fully scaled carp (N = 18). Significant p-values are marked in boldface.

Model term	Coefficient	DF	F	*p*
Intercept (RGR$_{std-length}$: initial standard length)		2	2.517	0.096
Constant term	0.473 (0.187)	1	7.093	0.012
Initial standard length	-0.002 (0.002)	1	1.351	0.253
Genotype (mirror carp)	0.049 (0.025)	1	3.937	0.056

5.5 Behavioral differences between common carp and mirror carp (Hypothesis iv)

The permutation tests revealed no consistent differences in all tested behavioral measures between mirror carp and common carp in trial A, though significant differences were found in trial B for the measures „Time spent active in arena" and "Latency to feed" (Table 28; Figure 6). Common carp spent significantly more time active in the arena (4.1 ± 2.9 minutes) than mirror carp (2.3 ± 2.2 minutes), whereas mirror carp were significantly faster to engage in feeding after leaving the refuge than common carp (0.4 ± 0.6 minutes versus 1.8 ± 3.2 minutes).

Table 28: Permutation tests for mean differences in behavioral measures from trial A and trial B between mirror carp and common carp.

| Tested variable	p (trial A)	p (trial B)
Latency to emerge from refuge (Context 1)	0.823	0.435
Time spent within one fish length of a novel object (Context 2)	0.422	0.861
Time active in arena (Context 1)	0.368	**0.022**
Latency to feed (Context 3)	0.848	**0.030**
Latency to resume feeding (Context 3)	0.124	0.849
Time spent within two fish lengths of shoal (Context 4)	0.361	0.199
Time spent in upper water zone (Context 1)	0.432	0.636
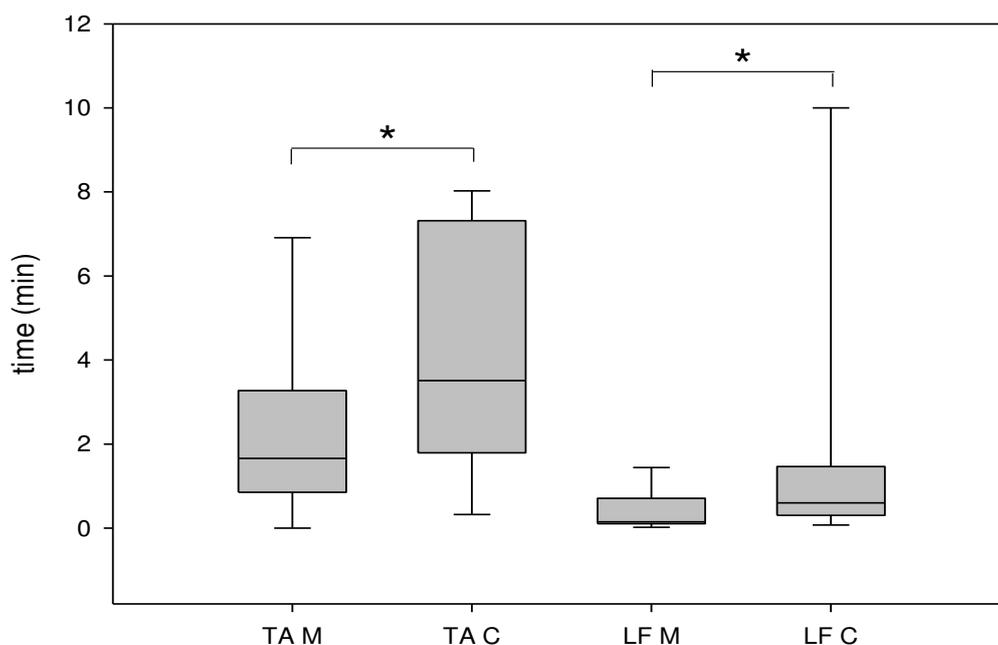
Figure 6: Box plots for the behavioral measures "Time spent active in arena (Context 1)" (TA) and „Latency to feed (Context 3)" (LF) for mirror carp (M) and common carp (C) in trial B. Significant differences tested for with permutation tests are marked with asterisks.

Variability for all behavioral measures was not found to be significantly higher in mirror carp than in wild-type fully scaled carp, in both, trial A and B (Tables 29 and 30). However, there was a trend for variability being higher in mirror carp in the variable "Latency to feed" from trial A.

Table 29: Ansari-Bradley-test testing greater behavioral variability in the mirror carp (trial A).

Tested variable	AB	p- value
Latency to emerge from refuge (Context 1)	145.5	0.160
Time spent within one fish length of a novel object (Context 2)	157	0.830
Time spent active in arena (Context 1)	179	0.736
Latency to feed (Context 3)	143	0.092
Latency to resume feeding (Context 3)	105.5	0.526
Time spent within two fish lengths of shoal (Context 4)	130.5	0.702
Time spent in upper water zone (Context 1)	204.5	0.937

Table 30: Ansari-Bradley-test testing greater behavioral variability in the mirror carp (Trial B).

Tested variable	AB	p-value
Latency to emerge from refuge (Context 1)	165	0.352
Time spent within one fish length of a novel object (Context 2)	162	0.380
Time spent active in arena (Context 1)	181	0.737
Latency to feed (Context 3)	152	0.473
Latency to resume feeding (Context 3)	169	0.796
Time spent within two fish lengths of shoal (Context 4)	181	0.737
Time spent in upper water zone (Context 1)	182.5	0.837

6. Discussion

Ecologically and evolutionary relevant temperament traits like boldness, exploration, activity, aggression and sociability have been identified in numerous species (Dingemanse and Reale 2005). The present study is the first study that simultaneously examined more than two temperament traits in the carp. A methodological assessment of several behavioral variables for boldness, exploration and sociability was used with the goal of extracting temperament traits and behavioral syndromes respectively. Additionally, the effect of domestication on behavioral metrics was tested for by comparing a more domesticated genotype with a less domesticated genotype. The impact of domestication on the temperament of carp has not been examined to date. Therefore the present study may contribute to the further understanding of ecological and evolutionary implication of temperament traits and their contextual bases (Réale, Reader et al. 2007).

6.1 Behavioral differences among individuals

Individual differences in behavioral responses of fish, as they have been found in various other studies (Wilson, Coleman et al. 1993; Budaev 1997; Coleman 1998; Budaev and Zworykin 2002; Bell and Stamps 2004; Ward, Thomas et al. 2004; Brown, Jones et al. 2005; Bleakley, Martell et al. 2006; Álvarez and Bell 2007; Frost, Winrow-Giffen et al. 2007; Moretz, Martins et al. 2007; Biro, Beckmann et al. 2009; Kobler, Engelen et al. 2009; Millot, Begout et al. 2009; Cote, Fogarty et al. 2010; Huntingford, Andrew et al. 2010), could also be observed for carp in this study as clearly reflected by the standard deviations and ranges of behavioral variables. Individual differences in behavioral responses as they were found here are preconditions for temperament assessment (Réale, Reader et al. 2007).

6.2 Behavioral consistency over time (Hypothesis i)

Three of seven behavioral measures were found to be rank order consistent over a time interval of six weeks, i.e. between initial tests (trial A) and retests (trial B). Regarding repeatability, two of these behavioral metrics, „Time spent within one fish length of a novel object" and „Time spent within two fish lengths of the shoal", had the highest repeatability estimates of behavioral variables used in this study. They were found to be close to statistically significant regarding F-statistics. Considering the small sample size in this study, these measures can clearly be regarded as consistent over time.

For the third variable correlating over time, "Latency to emerge from refuge", the repeatability estimate was below the mean value of repeatability assessments from other studies (reviewed by Bell, Hankison et al. 2009) and was not found to be significant regarding F-statistics. Repeatability is the proportion of total behavioral variance in a given population explained by differences between individuals (as opposed to intra-individual behavioral variance). If a mean-level change in behavior occurs between measurements, causing more within-individual than between-individual change, the resulting repeatability will be low (Bell, Hankison et al. 2009). The detected statistically significant rank-order consistency and likewise comparatively low repeatability of these three behavioral measures could therefore be explained by a steady shift in behavioral expression of all individuals between test (trial A) and retest (trial B) due to equal habituation of all fish to laboratory conditions (Biro, Beckmann et al. 2009). The assumption of habituation is supported by significant differences found in the central tendency of this variable between trial A and B. This can be explained with habituation (Colgan, Nowell et al. 1979; Post and von der Emde 1999). In consequence, it can be assumed, that the variable "Latency to leave refuge" is consistent over time despite a possible habituation effect biasing the estimation of repeatability. Due to the given consistency over time regarding inter-individual differences in behaviors reflected by these variables, they are considered as measures of temperament traits and are of particular interest for further interpretation.

Regarding the four remaining behavioral measures, which were neither rank-order consistent over time or highly repeatable, one reason for the lack of consistent differences between individuals could be traced to a possible measurement error associated with each behavioral observation. Measurement error can possibly bias the data set if some behavioral measures have more measurement error associated with them than others (Bell, Hankison et al. 2009).

Another reason could be that uncontrolled motivational variability might have played a role. In this case particularly, individual differences in feeding motivation could have biased the assessment of the variables "Latency to feed" and "Latency to resume feeding, despite the fact that care was taken to normalize feeding conditions in non-experimental periods by standardizing given food amount to body weight. Uncontrolled motivational variability can arise even under the most standardized laboratory conditions as stated by Cote (2010). A further, more likely explanation could be found in individual differences in habituation degree. Another study suggests that individuals can differ in their degree of habituation towards certain stimuli (Martin and Réale 2008). This would account for the lack of not only repeatability but also rank order consistency over time. This assumption is underlined by the finding that in three of the four variables, including "Latency to feed" and "Latency to resume feeding", significant differences in the central tendencies were found over time. This indicates habituation to laboratory conditions (Colgan, Nowell et al. 1979; Post and von der Emde 1999). Ultimately, these four behavioral measures may simply not be adequate measures of temperament traits in the carp.

6.3 Suites of correlated behaviors across situations (Hypothesis ii)

The latency to resume feeding following a fright stimulus and the time spent exploring a novel object were found to be significantly correlated in trial A, and close to statistically significantly correlated in trial B. Consistent with Ward et al (2004), the correlations were negative. Proximity to a novel object, because it implies a potential risk as it may represent a predator (Wright, Ward et al. 2006; Réale, Reader et al. 2007), is regarded as a measure for boldness in several studies

(Wilson, Coleman et al. 1993; Carere and van Oers 2004; Frost, Winrow-Giffen et al. 2007). The latency to resume feeding after a simulated predation attempt is most likely a measure for boldness (Ward, Thomas et al. 2004; Álvarez and Bell 2007). Because both measures are considered measures for boldness, a bold individual can be expected to spend much time in proximity of a novel object and to be quick to resume feeding after a disturbing fright stimulus, as it was shown here. Accordingly, these two measures can be considered as measures for the temperament trait boldness in carp.

Apart from correlating with "Latency to resume feeding", "Time spent within one fish length of a novel object" also correlated positively with "Time spent within two fish lengths of the shoal" in trial A and also in trial B. This result is not consistent with previous findings in other species of fish, whereupon a negative correlation between the two measures could be expected. Proximity to a novel object is a risky behavior due to potential predation risk as the object may resemble a predator (Wright, Ward et al. 2006). Swimming in a shoal dilutes the predation-risk for each individual and hence not swimming in a shoal represents a tradeoff between running a higher predation risk and achieving a better foraging success due to less competition, therefore indicating higher risk-taking tendency of individuals (Krause and Ruxton 2002; Ward, Thomas et al. 2004; Wright, Ward et al. 2006). It has been shown that bolder fish tend to spend less time in proximity of conspecifics (Wilson, Coleman et al. 1993; Ward, Thomas et al. 2004), therefore it was expected here that carp spending more time inspecting a novel object (and hence showing a higher willingness of risk-taking) would spend less time associating with a group of conspecifics. The unexpected positive correlation might be explained by the given experimental setup in experimental context 4 (shoaling context). Here the visual perception of shoal compartment with the shoal inside may elicit the same behavioral response from the focal fish as towards a novel object, because the focal fish never experienced visual contact with the shoal compartment before. Thus the time spent in proximity of the shoal compartment might be a measure for risk taking tendency and not a measure for sociability in this case. Another explanation may be found again in the experimental setup. Here a fish, in order to associate with the stimulus group of conspecifics, had to first emerge from refuge and then cross an

open water space to reach the shoal. Individuals, which were not explorative enough to leave the refuge and not bold enough to cross the open water space never reached the association zone. This might have distorted the results. However, the existence of a correlation between the two measures, regardless of its orientation, suggests a behavioral syndrome incorporating risk-taking and shoaling tendency.

The discovery that, in trial A, "Latency to emerge from refuge" (Context 1) was neither correlated with "Latency to feed", "Latency to resume feeding" or "Time spent within two fish lengths of shoal" is quite interesting. These correlations between "Latency to leave refuge" (Context 1) and the other measures mentioned above, with the exception of "Latency to resume feeding", were also not found in trial B. This is the first indication that, at least in trial A, "Latency to emerge from refuge" (Context 1) diverges in substance from the four other variables that are associated with the temperament trait boldness or the behavioral syndrome boldness/shoal. Accordingly, "Latency to leave refuge" (Context 1) should be considered rather a measure for the temperament trait exploration (sensu Réale, Reader et al. 2007) than a measure for the temperament trait boldness (Wilson and Godin 2009). The correlation found between "Latency to resume feeding" and "Latency to emerge from refuge" (Context 1) in trial B cannot question this assumption, because "Latency to resume feeding" is not consistent over time. The correlation could, however, give a hint that the variable "Latency to emerge from refuge" may have to be interpreted differently in trial B than in trial A.

The finding that "Latency to emerge from refuge" is correlated with "Latency to resume feeding" in trial B, but not in trial A, might be explained by the following. The assessment of exploration in trial B with the variable "Latency to emerge from refuge" could be strongly biased by the fact that, after a period of six weeks, carp may still remember the experimental setup. Therefore, trial B possibly did not offer a novel situation. "Latency to leave refuge" in trial B may not have been an option for a true measure of exploration, but rather a measure of risk-taking. A risky but not novel situation (Réale, Reader et al. 2007) may have been given here, provided carp indeed can remember the experimental setup. This assumption is substantiated regarding behavioral shifts that occurred, most likely due to habituation (see above), between trials A and B. However, this assumption,

that „Latency to emerge from refuge" in trial B is no true measure for exploration was not further clarified by other examinations following below and remains without further support.

All behavioral metrics from experimental context 1 are correlated with each other in both trials. On the first glance this finding indicates an interrelation between measures of boldness, exploration and exploration-activity because "Time spent in upper water zone" is considered a measure for boldness (Wilson and Godin 2009), "Latency to emerge from refuge", after the above interpretation, at least in trial A, a measure for exploration (Réale, Reader et al. 2007), and "Time spent active in arena" a measure for exploration-activity (Réale, Reader et al. 2007). However, the latency to emerge from a refuge did not correlate consistently with other measures of boldness, in particular with the time spent in proximity of a novel object, which was also shown to significantly correlate over time.

The causation must be sought in the assumption that these correlations between context 1 measures are experimental artifacts, originated by multivariate ascertainment in the same experimental context (Díaz-Uriarte 2001). The 16 fish in trial A that failed to exit the refuge and got assigned a maximum "Latency to leave refuge" score of 60 minutes were 0 minutes in the upper water zone and also 0 minutes active in the arena, therefore explaining found correlations. These three variables should be ascertained univariate (Johnson and Wichern 1998; Díaz-Uriarte 2001) in separate experimental contexts in a future investigation.

Further statistically significant correlations that were similar in trials A and B could not be found. Nevertheless, one correlation that was only found in one trial is of interest. The latency to feed correlated with the latency to resume feeding after a disturbing fright stimulus in trial A but not in trial B. The correlation found in trial A would be consistent with the assumption that both behavioral metrics are measures of the temperament trait boldness (Ward, Thomas et al. 2004; Álvarez and Bell 2007). However a correlation would have needed to be found in trial B too, to clearly verify this assumption. Both variables are not time consistent and have significantly different central tendencies between trials A and B, indicating habituation (Post and von der Emde 1999). Ergo, no clear statement concerning these two variables is possible at this point.

Moreover, as expected, the variable "latency to emerge from refuge" correlated across all four experimental contexts in trial A and trial B. Fish were consistent in their rank order of leaving the refuge in every single experimental context. Following the definition of Réale, Reader et al. (2007), the temperament trait exploration implies "an *individual's* reaction to a novel situation". A true novel environment in the sense of a novel habitat is only given in experimental context 1, where the focal fish explore the arena for the first time. But, in the sense of Réale, Reader et al (2007), a novel situation is not restricted to a novel habitat. In consequence, due to the presence of some form of novelty (i.e. a novel object in context 2, a novel feeding situation in context 3, and the shoal compartment in context 4) in every experimental context, "Latency to leave refuge" in context 2, 3 and 4 can still be regarded as measures of the temperament trait exploration, despite the fact that the fish had already explored the arena in experimental context 1 and a novel habitat was not given anymore. The found correlations across contexts, in combination with the detected time consistency in context 1, strongly militate in favor of the fact that the behavioral metric "Latency to emerge from refuge" measured the same temperament trait, namely exploration, in all experimental contexts, likewise in trial A and trial B. Consequently it can even be assumed that a context independent temperament trait "exploration" exists in carp.

6.4 Further verification of temperament traits and their consistency over time (Hypotheses i & ii)

Almost every PCA conducted with different assemblies of behavioral measures revealed different factor structures for trials A and B, which leads to the conclusion that factor structures per se are not consistent over time. This finding can be explained through the fact that not all variables used in the PCA were consistent over time (see consistency of behavioral measures over time, above).

Strikingly, "Latency to leave refuge", to this point interpreted as a measure for exploration and "Time spent within one fish length of a novel object", interpreted as a measure for boldness, are never found to co-load on one factor,

when measured in different experimental contexts. This finding is consistent with the fact that these two time consistent variables are neither correlated with one another in trial A or in trial B. Therefore the assumption (which was postulated already in the analysis of cross-context correlations) is substantiated, that these two measures assess different temperament traits. Hence, "Latency to leave refuge" can be considered as a measure for the temperament trait exploration (Réale, Reader et al. 2007) and "Time spent within one fish length of a novel object" as a measure for the temperament trait boldness (Frost, Winrow-Giffen et al. 2007).

"Latency to feed" and "Latency to resume feeding" were not found to load consistently with the same variables on the same factor. Nevertheless, in all PCA (PCA1, PCA2, PCA3) conducted under inclusion of these variables, in trial A, they loaded together on one factor with "Time spent within one fish length of a novel object", almost without exception, and did not co-load with measures that assess the latency to emerge from the refuge. This finding confirms the assumptions postulated above concerning interrelations of these variables in trial A. In consequence it can be stated that the latency to emerge from a refuge in trial A is a measure for the temperament trait exploration. It can also be assumed that the behavioral measures concerning the latency to feed and resume feeding following a fright stimulus (simulated predation attempt) as well as the time spent in proximity of a novel object can be linked with the temperament trait boldness. Accordingly, concerning trial A, factor 1 (LE1-UWZ-SH) in PCA1 as well as factor 1 (LE1-LE2-UWZ-TA) and factor 2 (LE3-LE4) in PCA3 can alternatively be labeled as "exploration" and factor 2 (NO-LF-LRF) in PCA1, factor 1 (NO-LRF-SH) in PCA2 and factor 3 (LE2-NO) in PCA3 can alternatively be labeled as boldness.

In the case of trial B, PCA did not reveal such clear findings as in trial A. Here, in these cases in which "Latency to emerge from a refuge" was included in PCA (PCA1 and PCA3), "Latency to feed" and "Latency to resume feeding" loaded together with this variable. As in the interpretation of the cross-context correlation between "Latency to emerge from a refuge" and "Latency to resume feeding" in trial B (see above), one would assume that "Latency to emerge from refuge" in trial B is substantially divergent from the same variable in trial A and may be a measure for boldness due to a potentially non novel environment rather than a measure for

exploration in trial B. However, contrary to this assumption, "Time spent within one fish length of a novel object", which has to this point been interpreted as a measure for boldness, not in any case also co-loaded on these factors. Under the assumption that „Latency to emerge from refuge" is substantially divergent in trial A and B, one would have to interpret that „Time spent within one fish length of a novel object" also diverges in substance between trial A and trial B (because it loads together with the variables "Latency to feed" and "Latency to resume feeding" in trial A but not in trial B). Considering the results that this variable is stable over time and no differences in the central tendency were found between trial A and trial B (thus, no habituation took place), this seems highly unlikely. The assumption that the latency to emerge from a refuge is also a measure for exploration in trial B seems more plausible, especially considering the findings that the variables "Latency to feed" and "Latency to resume feeding" are not consistent over time and their central tendencies are divergent between trials A and B, indicating habituation (see above). Altogether, the factor structures cannot be safely interpreted in trial B and no clear labeling is possible.

The variable „Time spent within two fish lengths of the shoal" was found to co-load on one factor with "Time spent within one fish length of a novel object" in four of six cases in those PCA conducted under inclusion of these two behavioral metrics for trial A and trial B (PCA1, PCA2, PCA3). This finding is consistent with the assumption postulated in the discussion of correlations across contexts (see above) that a boldness/shoal syndrome exists in carp. In two cases, "Time spent within two fish lengths of the shoal" did not co-load on the same factor with "Time spent within one fish length of a novel object". That is not consistent with the found cross-context correlation between the two measures discussed above. This is puzzling, as a PCA classifies variables in groups according to their correlative interrelations. However, if sample size is low and not at least four variables loading on one factor have loadings over 0.6 (which is given here), a random structure of loadings can occur (Bortz 1999). This must be considered at this point. Therefore the existence of a boldness/shoal syndrome as indicated by the cross-context correlation is not questioned by the two of six contrary PCA results.

For the variable "Latency to emerge from refuge", which was assessed in all of the four experimental contexts, it was not certain if this measure is a true measure of exploration in all other contexts than the exploration context (Context 1), which was the only experimental context that involved a novel habitat. Similarly, it was unclear if these measures, especially "Latency to emerge from refuge" (Context 1), can be considered measures of exploration in trial B, as discussed above.

A further PCA (PCA4) was conducted to clarify the assumption that all four measures of "Latency to emerge from refuge" are measures of the temperament trait exploration (which was postulated already after the interpretation of given time consistency and given cross-context correlations between all of these measures). All four variables were found to load on a single factor, in trial A as well as in trial B. This clearly speaks for a strong consistency across contexts, underlining that these measures assess the same temperament trait in every different context. Thus, the results from PCA4 in trial A and trial B again underline the existence of a context-independent temperament trait exploration in carp.

The ascertainment of time consistency in factor structures (assessed by theoretically reconstructing factor structures for trial B that were found in trial A, because not in all cases factor structures found in trial B were identical or even similar to those found in trial A) revealed similar results regarding rank order consistency and repeatability. Of the four factors that were labeled as exploration in above PCA analysis, three were regarded to be consistent over time (statistically significant or marginally significant). The single factor revealed by PCA4, consisting of all four "Latency to emerge from refuge" variables, was found to significantly correlate between trials A and B. It was also repeatable with significant underlying F-statistics. Furthermore the repeatability estimate found was well above average of repeatability estimates in other studies (reviewed in Bell, Hankison et al. 2009). In factor structures that were composed of fewer measures of exploration and more other behavioral metrics, correlations and repeatability estimates were weaker (in two cases still significant or close to statistical

significance). Accordingly it can be concluded that the variable "Latency to leave refuge" measures the same temperament trait (exploration) in trial A and trial B. Furthermore, the temperament trait exploration is consistent over time.

Concerning boldness, results were not as clear as those found regarding exploration, which can be explained with the fact that the only truly consistent measure of boldness, „Time spent within one fish length of a novel object", co-loads with inconsistent behavioral measures in every given case. Nonetheless, a close to significant rank correlation and a repeatability estimate (if below average in other studies) with marginally significant underlying F-statistics was found in the case of factor 2 (NO-LF-LRF) from PCA1, which was composed of „Time spent within one fish length of a novel object" in combination with "Latency to feed" and "Latency to resume feeding". However, low repeatability estimate can be biased by steady habituation (Bell, Hankison et al. 2009), which was shown to occur in the variables "Latency to feed" and "Latency to resume feeding", and statistical marginality may likely be accounted for by a larger sample size. Since this factor was the only factor that was composed solely of presumed measures of boldness, the temperament trait boldness is here assumed to be consistent over time.

6.5 Interrelations between behavioral metrics and life-history productivity (Hypothesis iii)

No significant correlations between the initial weight of the fish and behavioral performance were found in this study, which is consistent with findings in another study (Krause, Loader et al. 1999). However, in one case, the length of fish at the end of trial A was correlated with "Latency to emerge from refuge" in trial A, but this could not be reproduced in trial B. Thus it can be concluded that consistent effects of initial differences in size of the fish on behavioral performance were not present.

Concerning correlations between relative growth rates calculated from weight and standard length and behavioral measures, only one significant correlation was found between "Latency to feed" and the relative growth rate

calculated from standard length. This correlation was only found in trial B and was not found between "Latency to feed" and relative growth rate calculated from weight.

The correlations calculated from relative growth rates for weight and standard length and bundled variables derived from factor loadings in all conducted PCA were only found to be significant in two cases. First, factor 2 from PCA 2 (LF) correlated with the relative growth rate for standard length in trial B. However, this factor consists of only one variable, which is "Latency to feed", therefore this correlation is redundant. The second correlation was found between Factor 3 from PCA 3 (LE2 – NO) and the relative growth rate for standard length, also only in trial B, not in trial A. All other correlations were not even close to statistical significance.

Thus, no consistent interrelations between behavioral measures and relative growth rates were found, which is puzzling, as Biro and Stamps (2008), after reviewing several studies, conclude that temperament traits are linked to life history productivity (growth rate). Possibly, here no consistent interrelations were found due to laboratory conditions as Biro and Stamps (2008) state that a correlation between e.g. boldness and growth rate may likely be dependent on the degree of predation, which the fish used in this study have never experienced before. The authors also write that not all temperament traits correlate with life-history traits. Possibly, the behavioral measures used in this study do not address temperament traits that would correlate with life-history traits such as growth rate. Therefore, the search for behavioral measures of temperament traits in carp correlating with life – history traits may be worthwhile addressing in future research.

6.6 Behavioral differences between common carp and mirror carp (Hypothesis iv)

Further aim of the study was to assess behavioral differences between mirror carp and common carp. It was assumed that a different background in domestication history would lead to behavioral differences of the two genotypes, in particular it was expected that mirror carp are bolder than common carp. Price (1984) argues

that due to relaxed selection pressures in captivity, more phenotypes will survive here than in the wild. Therefore a greater spectrum of behavioral variability will be present in domesticated populations. However, behavioral metrics used in this study hardly revealed any differences between mirror carp and common carp at all. Mirror carp do not consistently exhibit a greater variability in behavioral measures than common carp as it could have been expected following Price (1984). Nevertheless, indications were found in trial A, that mirror carp may exhibit a greater behavioral variability in the measure "Latency to feed", but this marginally significant result could not be reproduced in trial B.

Correspondingly the mean values of behavioral variables presumably measuring temperament traits did not differ significantly between mirror carp and common carp in all measures except "Latency to feed" (common carp took longer to consume the food item upon entry of the arena) and „Time spent active in arena" (common carp were active for a longer period of time in the arena), both only in trial B, not in trial A. Regarding the measure that was found to be a measure of boldness in this study, "Time spent within one fish length of a novel object", no significant difference was found between the tested genotypes. It could not be confirmed in this study that mirror carp are bolder than their counterparts of the "wild-type" phenotype. This is contradictory to the results of other studies, which found more domesticated strains in other species of fish, such as Danio rerio, the zebra fish (Moretz, Martins et al. 2007), Salmo trutta, the brown trout (Sundstrom, Petersson et al. 2004) or Dicentrarchus labrax, the sea bass (Millot, Begout et al. 2009) to be bolder than less domesticated or wild strains. no consistent differences in this case were possibly found because common "wild-type" carp were also reared in captivity and never experienced natural predation risk before. Environmental influences may have had an impact on the development of temperament as suggested by some studies (Metcalfe, Valdimarsson et al. 2003; Brown, Burgess et al. 2007; Frost, Winrow-Giffen et al. 2007). It could also be possible that the two used carp strains have been domesticated for a similar time span and thus consistent behavioral differences are lacking.

However, differences in genetic variability between mirror carp and common carp, found by Klefoth et. al (2010, unpublished data) for the population from which focal fish in this study were taken and found by Kohlmann, Kersten et al. (2005) for other populations, contradict this assumption as genetic variability is linked to domestication degree in carp (Vandeputte and Launey 2004). Another reason might be that the parents of the mirror carp used in this study were all common carp and despite morphological differences, might have been siblings. Also the common carp used in this study, depending on which fish were the parents, are partly homozygous and partly heterozygous, possibly resulting in a behavioral continuum between mirror carp and common carp rather than two clearly distinct behavioral phenotypes. In future research, truly pure-bred homozygous genotypes should be used. Nonetheless, it was not possible to obtain mirror carp and common carp for this study that a) were raised under natural conditions in the same pond (common garden environment) and b) of which the common carp were also truly homozygous. Since both genotypes were reared in a common garden environment, behavioral differences, if they would have occurred, would have most likely occurred due to gene-environment interactions (Aday, Wahl et al. 2003). It might be likely, that the lack of behavioral differences was linked to the context-dependency of behaviors.

It has been shown in several other studies that behavioral responses of fishes are context-dependent (Coleman 1998; Bell 2005; Figueira and Lyman 2007; Biro, Beckmann et al. 2009). Hence it is possible that expected differences in the variability in measures of risk-taking, exploration and sociability that were used in this study, did not occur as an effect of laboratory conditions. In all probability, differences could have been found under more natural conditions. Indeed, using carp from the same origin, Klefoth et al. found significant behavioral differences between mirror carp and common carp in a semi-natural pond setup, but differences vanished under laboratory conditions (Klefoth 2010, unpublished data). But when inducing artificial predation risk by angling in the laboratory, Klefoth et al. found a behavioral shift of the fish with significantly less sheltering and more active mirror carp compared to their less domesticated counterparts.

This finding is supported by a multitude of other studies that found differences between wild and domesticated populations of fish (Johnsson, Höjesjö et al. 2001; Metcalfe, Valdimarsson et al. 2003; Wright, Ward et al. 2006; Matsuzaki, Mabuchi et al. 2009; Millot, Begout et al. 2009). Considering the results of Klefoth et al., common carp and mirror carp used in this study would have been expected to show behavioral differences when artificial predation risk was introduced in the resume feeding trial. However, the fright stimulus was only introduced after the fish consumed the first food pellet (which is considered risk-taking behavior in this study, ascertained with the variable "Latency to feed"), potentially leading to a situation where only those fish exhibiting a high risk-taking tendency actually experienced the fright stimulus. Thus fish with a low risk-taking tendency did not ingest the first food item. They were not disturbed at all, thereby biasing the result.

Another explanation could be the stimulus itself. Perhaps, the fish did not associate the fright stimulus with a predation attempt, as intended, even though a considerable fright reaction of most individuals was observed in this study. The perception of a fishing weight dropping into the water might be different to artificial predation stimuli used in other studies, like artificial birds (Krause, Loader et al. 1998). Nonetheless, the dropping of a weight into an experimental aquarium has proven to be a suitable stimulus in other species like Gasterosteus aculeatus, the three-spined stickleback (Ward, Thomas et al. 2004). Based on these considerations, standardized laboratory conditions most likely account for only minor and inconsistent behavioral differences between the two genotypes of carp.

Despite the fact that no noteworthy behavioral differences between mirror carp and common carp could be found, a higher growth rate was found in the mirror carp. This is consistent with findings in other studies that mirror carp have a longer domestication background than common carp with respect to artificial selection for higher growth rates (Balon 1995; Hulata 1995; Kirpichnikov 1999; Balon 2004). Nevertheless, this result needs to be interpreted with caution due to only approximate standardization of feeding conditions in non-experimental periods.

For all results obtained in this study one has to bear in mind that laboratory conditions are, even if designed lifelike, always deficient reproductions of natural conditions (Chapanis 1967). This can dramatically interfere with natural behavior (Naguib 2006). Bell (2008) for instance found that repeatability of behavior was greater in the field than under laboratory conditions. On the other hand, another study found that temperament traits measured in captivity are reflected in the wild. Furthermore, in this study merely 36 fish were analyzed. With this sample size, only indications can be obtained, but not the certainty of representativeness (Naguib 2006). Even though these results offer clear indications, following experiments should be conducted with a larger sample size as done by Huntingford, Andrew et al. (2010) to gain sure evidence. Huntingford, Andrew et al. (2010) also found that isolated carp show high levels of stress likely to obscure results from behavioral tests and therefore tested carp in groups.

This study tried to consider stress effects by allowing the focal fish to receive olfactory and chemical cues from conspecifics while in the experimental aquarium. But still, stress artifacts may have occurred. This could be accounted for by tracking the fishes physiological stress response or, following Huntingford, Andrew et al. (2010), testing fish in groups.

For future research, the experimental framework as it was used in this study could be used in principle. But in retrospective some improvements can be suggested. First, as mentioned above, truly distinct pure-bred genotypes should be used. Also, it would make sense to choose a time period greater than 6 weeks between test and retest to better account for habituation effects. Furthermore, a larger sample size would definitely improve interpretation.

Additionally a clear context-separation with univariate ascertainment seems substantial for a safe interpretation of temperament trait dimensions generated by PCA. Also, other behavioral measures than the ones used in this study may offer better results. For example a measure quantifying the actual fright reaction (e.g. the amount of time spent scattering or freezing after a fright stimulus) of focal fish in context 3 (resume feeding) could be a better indicator for boldness. This could not be tested due to lack of time and should definitely be included in a future investigation.

Last but not least, sex-effects on behavioral variation were ignored in this study because it is impossible to reliably sex juvenile carp based on morphology. Thus, the influence of sex on the expression of temperament traits in adult carp may prove an interesting follow up of this study. Sex-effects on temperament have been identified in other species of fish (Brown, Jones et al. 2005). On the other hand, other previous studies have suggested that sex does not have an influence on temperament related behaviors of fish (Wilson, Coleman et al. 1993; Fraser, Gilliam et al. 2001; Moretz, Martins et al. 2007).

6.7 Conclusions

In the present study, behavioral differences among individuals of carp were investigated by testing the consistency of behavior over time and across situations with the aim of revealing temperament traits and behavioral syndromes in carp. For this purpose, several behavioral measures were used. These were considered to be measures of boldness, exploration and sociability in other studies, to test for the existence of temperament traits or behavioral syndromes in carp under standardized laboratory conditions. By using two different genotypes (common and mirror carp) with different degree of domestication, the effect of domestication on behavioral expressions of carp was additionally tested.

The present study demonstrated that individual carp differ in their behavioral expressions and that several behaviors in carp are consistent over time and / or situations. Therefore it can be concluded that temperament traits respectively behavioral syndromes exist in carp. In this context, two distinct temperament dimensions were identified as boldness and exploration. In the case of exploration, a context-independent temperament trait was found. The present study is the first study to show the existence of the temperament trait exploration in carp. However, strong effects of habituation over the period of study and potential individual differences in habituation degree might have influenced the results and should be considered in future studies of temperament in carp. Additionally, it was discovered that mirror carp have higher growth rates than common carp of the "wild-type"

phenotype, but only minor behavioral differences between mirror carp and common carp of the "wild type" phenotype were found under standardized laboratory conditions. This indicates that behavioral differences as frequently found in comparative studies of animals with different degrees of domestication are potentially context dependent.

7. Acknowledgements

8. References

Aday, D. D., D. H. Wahl, et al. (2003). Assessing Population-specific and Environmental Influences on Bluegill Life Histories: A Common Garden Approach. Ecology **84**(12): 3370-3375.

Alcock, J. (2005). Animal Behavior, 8th ed. Sunderland, MA, Sinauer Associates, Inc.

Álvarez, D. and A. M. Bell (2007). Sticklebacks from streams are more bold than sticklebacks from ponds. Behavioural Processes **76**(3): 215-217.

Ansari, A. R. and R. A. Bradley (1960). Rank-sum Test for Dispersions. Annals of Mathematical Statistics **31**(4): 1174-1189.

Balon, E. K. (1995). Origin and domestication of the wild carp, Cyprinus carpio: from Roman gourmets to the swimming flowers. Aquaculture **129**(1-4): 3-48.

Balon, E. K. (2004). About the oldest domesticates among fishes. Journal of Fish Biology **65**(s1): 1-27.

Bell, A. M. (2005). Behavioural differences between individuals and two populations of stickleback (Gasterosteus aculeatus). Journal of Evolutionary Biology **18**(2): 464-473.

Bell, A. M. (2007). Evolutionary biology: Animal personalities. Nature **447**(7144): 539-540.

Bell, A. M. (2007). Future directions in behavioural syndromes research. Proceedings of the Royal Society B: Biological Sciences **274**(1611): 755-761.

Bell, A. M., S. J. Hankison, et al. (2009). The repeatability of behaviour: a meta-analysis. Animal Behaviour **77**(4): 771-783.

Bell, A. M. and J. A. Stamps (2004). Development of behavioural differences between individuals and populations of sticklebacks, Gasterosteus aculeatus. Animal Behaviour **68**: 1339-1348.

Biro, P. and J. Post (2008). Rapid depletion of genotypes with fast growth and bold personality traits from harvested fish populations. Proceedings of the National Academy of Sciences **105**(8): 2919-2922.

Biro, P. A., C. Beckmann, et al. (2009). Small within-day increases in temperature affects boldness and alters personality in coral reef fish. Proceedings of the Royal Society B-Biological Sciences **277**(1678): 71-77.

Biro, P. A. and J. A. Stamps (2008). Are animal personality traits linked to life-history productivity? Trends in Ecology & Evolution **23**(7): 361-368.

Bleakley, B., C. Martell, et al. (2006). Variation in Anti-Predator Behavior Among Five Strains of Inbred Guppies, Poecilia reticulata. Behavior Genetics **36**(5): 783-791.

Block, J. (1995). A Contrarian View of the 5-Factor Approach to Personality Description. Psychological bulletin **117**(2): 187-215.

Blumstein, D. T. and J. C. Daniel (2007). Quantifying Behavior the JWatcher Way. Sunderland, Sinauer Associates, Inc.

Bortz, J. (1999). Statistik für Sozialwissenschaftler. Berlin - Heidelberg - New York, Springer Verlag.

Brännäs, E., H. Lundqvist, et al. (1994). Use of the Passive Integrated Transponder (PIT) in a Fish Identification and Monitoring System for Fish Behavioral Studies. Transactions of the American Fisheries Society **123**(3): 395-401.

Brown, C., F. Burgess, et al. (2007). Heritable and experiential effects on boldness in a tropical poeciliid. Behavioral Ecology and Sociobiology **62**(2): 237-243.

Brown, C., F. Jones, et al. (2005). In situ examination of boldness-shyness traits in the tropical poeciliid, Brachyraphis episcopi. Animal Behaviour **70**(5): 1003-1009.

Budaev, D. S. (2003). Habituation of Predator Inspection and Boldness in the Guppy (Poecilia reticulata). Cogprints, http://cogprints.org/5878/.

Budaev, D. S. and D. D. Zworykin (2002). Individuality in Fish Behavior: Ecology and Comparative Psychology. Cogprints, http://cogprints.org/5477/.

Budaev, S. V. (1997). Alternative styles in the European wrasse, Symphodus ocellatus: Boldness-related schooling tendency. Environmental Biology of Fishes **49**(1): 71-78.

Budaev, S. V. (1997). "Personality" in the guppy (Poecilia reticulata): A correlational study of exploratory behavior and social tendency. Journal of Comparative Psychology **111**(4): 399-411.

Carere, C. and K. van Oers (2004). Shy and bold great tits (Parus major): body temperature and breath rate in response to handling stress. Physiology & Behavior **82**(5): 905-912.

Chapanis, A. (1967). The Relevance of Laboratory Studies to Practical Situations. Ergonomics **10**(5): 557 - 577.

Chistiakov, D. A. and N. V. Voronova (2009). Genetic evolution and diversity of common carp Cyprinus carpio L. Central European Journal of Biology **4**(3): 304-312.

Clark, A. B. and T. J. Ehlinger (1987). Pattern and adaptation in individual behavioral differences. Perspectives in Ethology, Alternatives 7 P. P. G. Bateson and P. H. Klopfer. New York, Plenum Press**:** 1-47.

Coleman, K. (1998). Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. Animal Behaviour **56**(4): 927-936.

Colgan, P. W., W. A. Nowell, et al. (1979). Aggressive habituation and rim circling in the social organization of bluegill sunfish (Lepomis macrochirus). Environmental Biology of Fishes **4**(1): 29-36.

Cote, J., S. Fogarty, et al. (2010). Personality traits and dispersal tendency in the invasive mosquitofish (Gambusia affinis). Proceedings of the Royal Society B: Biological Sciences **277**(1687): 1571-1579.

Dall, S. R. X., A. I. Houston, et al. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecology Letters **7**(8): 734-739.

Daszykowski, M., K. Kaczmarek, et al. (2007). Robust statistics in data analysis - A review: Basic concepts. Chemometrics and Intelligent Laboratory Systems **85**(2): 203-219.

Díaz-Uriarte, R. (2001). The analysis of cross-over trials in animal behavior experiments: review and guide to the statistical literature. Samizdat Press, http://samizdat.mines.edu/.

Díaz-Uriarte, R. (2002). Incorrect analysis of crossover trials in animal behaviour research. Animal Behaviour **63**: 815-822.

Dingemanse, N. J. and D. Reale (2005). Natural selection and animal personality. Behaviour **142**: 1159-1184.

Dingemanse, N. J., J. Wright, et al. (2007). Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. Journal of Animal Ecology **76**(6): 1128-1138.

Dobzhansky, T. and O. Pavlovsky (1957). An Experimental Study of Interaction between Genetic Drift and Natural Selection. Evolution **11**(3): 311-319.

Dyer, J., D. Croft, et al. (2009). Shoal composition determines foraging success in the guppy. Behav. Ecol. **20**(1): 165-171.

Figueira, W. F. and S. J. Lyman (2007). Context-dependent risk tolerance of the bicolour damselfish: courtship in the presence of fish and egg predators. Animal Behaviour **74**(2): 329-336.

Fraser, D. F., J. F. Gilliam, et al. (2001). Explaining leptokurtic movement distributions: Intrapopulation variation in boldness and exploration. American Naturalist **158**(2): 124-135.

Frost, A. J., A. Winrow-Giffen, et al. (2007). Plasticity in animal personality traits: does prior experience alter the degree of boldness? Proceedings of the Royal Society B: Biological Sciences **274**(1608): 333-339.

Godin, J. G. J. and L. A. Dugatkin (1996). Female mating preference for bold males in the guppy, Poecilia reticulata. Proceedings of the National Academy of Sciences of the United States of America **93**(19): 10262-10267.

Goldsmith, H. H., A. H. Buss, et al. (1987). Round-Table – What is Temperament 4 Approaches. Child Development **58**(2): 505-529.

Gosling, S. D. (2001). From mice to men: what can we learn about personality from animal research? Psychological bulletin **127**(1): 45-86.

Hall, C. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. Journal of Comparative Physiology **18**: 385-403.

Hopkins, K. D. (1992). Reporting Fish Growth: A Review of the Basics. Journal of the World Aquaculture Society **23**(3): 173-179.

Hulata, G. (1995). A review of genetic improvement of the common carp (cyprinus carpio L.) and other cyprinids by crossbreeding, hybridisation and selection. Aquaculture **129**(1-4): 143-155.

Huntingford, F. A. (2004). Implications of domestication and rearing conditions for the behaviour of cultivated fishes. Journal of Fish Biology **65**(s1): 122-142.

Huntingford, F. A., G. Andrew, et al. (2010). Coping strategies in a strongly schooling fish, the common carp Cyprinus carpio. Journal of Fish Biology **76**(7): 1576-1591.

Johnson, J. C. and A. Sih (2005). Precopulatory sexual cannibalism in fishing spiders (Dolomedes triton): a role for behavioral syndromes. Behavioral Ecology and Sociobiology **58**(4): 390-396.

Johnson, R. A. and D. W. Wichern (1998). Applied Multivariate Statistical Analysis. New Jersey, Prentice Hall.

Johnsson, J. I., J. Höjesjö, et al. (2001). Behavioural and heart rate responses to predation risk in wild and domesticated Atlantic salmon. Canadian Journal of Fisheries and Aquatic Sciences **58**(4): 788-794.

Jones, M. J. and I. G. Stuart (2009). Lateral movement of common carp (Cyprinus carpio L.) in a large lowland river and floodplain. Ecology of Freshwater Fish **18**(1): 72-82.

Khalili, K. J. and A. K. Amirkolaie (2010). Comparison of common carp (Cyprinus carpio L.) Morphological and electrophoretic characteristics in the Southern Coast of the Caspian Sea. J. Fish. Aquat. Sci. **5**: 200-207.

Kirpichnikov, V. S. (1999). Genetics and breeding of common carp. Paris, INRA.

Knudsen, C. M., M. V. Johnston, et al. (2009). Effects of Passive Integrated Transponder Tags on Smolt-to-Adult Recruit Survival, Growth, and Behavior of Hatchery Spring Chinook Salmon. North American Journal of Fisheries Management **29**(3): 658-669.

Kobler, A., B. Engelen, et al. (2009). Temperament in bullheads: do laboratory and field explorative behaviour variables correlate? Naturwissenschaften **96**(10): 1229-1233.

Kohinoor, A. H. M., M. S. Islam, et al. (2002). Evalution of different stocks of Chinese carp in Bangladesh: Design and preliminary results. Proceedings of a Workshop on Genetic Management and Improvement Strategies for Exotic Carp in Asia, Dhaka, Bangladesh, Bangladesh Fisheries Research Institute, Mymensingh, Bangladesh.

Kohlmann, K., P. Kersten, et al. (2005). Micro satellite-based genetic variability and differentiation of domesticated, wild and feral common carp (Cyprinus carpio L.) populations. Aquaculture **247**(1-4): 253-266.

Kortet, R. and A. Hedrick (2007). A behavioural syndrome in the field cricket Gryllus integer: intrasexual aggression is correlated with activity in a novel environment. Biological Journal of the Linnean Society **91**(3): 475-482.

Krause, J., S. P. Loader, et al. (1999). Refuge use by fish as a function of body weight changes. Acta ethologica **2**(1): 29-34.

Krause, J., S. P. Loader, et al. (1998). Refuge use by fish as a function of body length related metabolic expenditure and predation risks. Proceedings of the Royal Society of London. Series B: Biological Sciences **265**(1413): 2373-2379.

Krause, J. and G. D. Ruxton (2002). Living in Groups, Oxford University Press.

Lessells, C. M. and P. T. Boag (1987). Unrepeatable Repeatabilities: A Common Mistake. The Auk **104**(1): 116-121.

Ludbrook, J. (1994). Advantages of Permutation (Randomization) Tests in Clinical and Experimental Pharmacology and Physiology. Clinical and Experimental Pharmacology and Physiology **21**(9): 673-686.

Magurran, A. E. (1993). Individual differences and alternative behaviours. The Behaviour of Teleost Fishes. T. J. Pitcher. London, Chapman and Hall**:** 441-477.

Martin, J. G. A. and D. Réale (2008). Temperament, risk assessment and habituation to novelty in eastern chipmunks, Tamias striatus. Animal Behaviour **75**(1): 309-318.

Matsuzaki, S. S., K. Mabuchi, et al. (2009). Behavioural and morphological differences between feral and domesticated strains of common carp Cyprinus carpio. Journal of Fish Biology **75**(6): 1206-1220.

McDougall, P. T., D. Réale, et al. (2006). Wildlife conservation and animal temperament: causes and consequences of evolutionary change for captive, reintroduced, and wild populations. Animal Conservation **9**(1): 39-48.

Metcalfe, N. B., S. K. Valdimarsson, et al. (2003). The relative roles of domestication, rearing environment, prior residence and body size in deciding territorial contests between hatchery and wild juvenile salmon. Journal of Applied Ecology **40**(3): 535-544.

Millot, S., M. L. Begout, et al. (2009). Exploration behaviour and flight response toward a stimulus in three sea bass strains (Dicentrarchus labrax L.). Applied Animal Behaviour Science **119**(1-2): 108-114.

Moretz, J. A., E. P. Martins, et al. (2007). Behavioral syndromes and the evolution of correlated behavior in zebrafish. Behavioral Ecology **18**(3): 556-562.

Müller-Belecke, A., G. Füllner, et al. (2009). Aquatische genetische Ressourcen - Laichfischbestände von Wirtschaftsfischarten in Deutschland. Schriften des Instituts für Binnenfischerei e.V. Potsdam-Sacrow. I. f. B. e.V. Potsdam-Sacrow, Institut für Binnenfischerei e.V. **Bd. 25:** 74 S.

Naguib, M. (2006). Methoden der Verhaltensbiologie. Heidelberg, Springer Verlag.

Peres-Neto, P. R., D. A. Jackson, et al. (2003). Giving Meaningful Interpretation to Ordination Axes: Assessing Loading Significance in Principal Component Analysis. Ecology **84**(9): 2347-2363.

Pitcher, T. J., A. E. Magurran, et al. (1986). Size-Segregative Behavior in Minnow Shoal. Journal of Fish Biology **29**: 83-95.

Post, N. and G. von der Emde (1999). The "novelty response" in an electric fish: response properties and habituation. Physiology & Behavior **68**(1-2): 115-128.

Potochnik, A. (2009). Optimality modeling in a suboptimal world. Biology & Philosophy **24**(2): 183-197.

Price, E. O. (1984). Behavioral Aspects of Animal Domestication. The Quarterly Review of Biology **59**(1): 1-32.

Price, E. O. (1999). Behavioral development in animals undergoing domestication. Applied Animal Behaviour Science **65**(3): 245-271.

Price, E. O. (2002). Animal domestication and behavior. New York, CABI Publishing.

Quinn, G. and M. Keough (2006). Experimental design and data analysis for biologists. Cambridge, U.K., Cambrigde University Press.

R_Development_Core_Team (2009). R: A language and environment for statistical computing. V. R Foundation for Statistical Computing, Austria, http://www.R-project.org.

Réale, D., S. M. Reader, et al. (2007). Integrating animal temperament within ecology and evolution. Biological Reviews **82**(2): 291-318.

Reddon, A. and P. Hurd (2009). Individual differences in cerebral lateralization are associated with shy - bold variation in the convict cichlid. Animal Behaviour **77**(1): 189-193.

Renner, M. J. (1990). Neglected aspects of exploratory and investigative behavior. Psychobiology **18**(1): 16-22.

Sih, A., A. Bell, et al. (2004). Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol. **19**(7): 372-378.

Sih, A. and A. M. Bell (2008). Insights for Behavioral Ecology from Behavioral Syndromes. Advances in the Study of Behavior Vol. 38. H. J. Brockmann, T. J. Roper and M. Naguib. San Diego, CA, Academic Press.

Sih, A., A. M. Bell, et al. (2004). "Behavioral Syndromes: An Integrative Overview. The Quarterly Review of Biology **79**(3): 241-277.

Sih, A., L. B. Kats, et al. (2003). Behavioural correlations across situations and the evolution of antipredator behaviour in a sunfish-salamander system. Animal Behaviour **65**(1): 29-44.

Skov, C., J. Brodersen, et al. (2005). Evaluation of PIT-tagging in cyprinids. Journal of Fish Biology **67**(5): 1195-1201.

Smith, B. R. and D. T. Blumstein (2008). Fitness consequences of personality: a meta-analysis. Behav. Ecol.: arm144.

Soto, C. G. and Burhanuddin (1995). Clove oil as a fish anaesthetic for measuring length and weight of rabbitfish (Siganus lineatus). Aquaculture **136**(1-2): 149-152.

Stamps, J. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. Ecology Letters **10**(5): 355-363.

Steffens, W. (1980). Der Karpfen Cyprinus carpio. Wittenberg Lutherstadt, A. Ziemsen Verlag, DDR.

Strelau, J. (2001). The concept and status of trait in research on temperament. European Journal of Personality **15**(4): 311-325.

Sundstrom, L. F., E. Petersson, et al. (2004). Hatchery selection promotes boldness in newly hatched brown trout (Salmo trutta): implications for dominance." Behav. Ecol. **15**(2): 192-198.

van Oers, K., G. de Jong, et al. (2005). Contribution of genetics to the study of animal personalities: a review of case studies. Behaviour **142**: 1185-1206.

Vandeputte, M. and S. Launey (2004). The genetic management of fish domestication. Productions Animales **17**(3): 237-242.

Verbeek, P., T. Iwamoto, et al. (2007). Differences in aggression between wild-type and domesticated fighting fish are context dependent. Animal Behaviour **73**(1): 75-83.

Walsh, R. N. and R. A. Cummins (1976). Open-Field Test – Critical-Review. Psychological bulletin **83**(3): 482-504.

Ward, A. J. W., P. Thomas, et al. (2004). Correlates of boldness in three-spined sticklebacks (Gasterosteus aculeatus). Behavioral Ecology and Sociobiology **55**(6): 561-568.

Wilson, A. and R. McLaughlin (2007). Behavioural syndromes in brook charr, Salvelinus fontinalis: prey-search in the field corresponds with space use in novel laboratory situations. Animal Behaviour **74**(4): 689-698.

Wilson, A. D. M. and J. G. J. Godin (2009). Boldness and behavioral syndromes in the bluegill sunfish, Lepomis macrochirus. Behavioral Ecology **20**(2): 231-237.

Wilson, A. D. M. and E. D. Stevens (2005). Consistency in context-specific measures of shyness and boldness in rainbow trout, Oncorhynchus mykiss. Ethology **111**(9): 849-862.

Wilson, D. S. (1998). Adaptive individual differences within single populations. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **353**(1366): 199-205.

Wilson, D. S., A. B. Clark, et al. (1994). Shyness and boldness in humans and other animals. Trends in Ecology & Evolution **9**(11): 442-446.

Wilson, D. S., K. Coleman, et al. (1993). Shy bold continuum in pumpkinseed sunfish (Lepomis gibbosus) - an ecological study of a psychological trait. Journal of Comparative Psychology **107**(3): 250-260.

Wright, D., A. J. W. Ward, et al. (2006). Social organization, grouping, and domestication in fish. Zebrafish **3**(2): 141-55.