Behavioural Processes 91 (2012) 133–140

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

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

Quantifying personality in the terrestrial hermit crab: Different measures, different inferences

Noelle M. Watanabe[a], W. David Stahlman[b], Aaron P. Blaisdell[b], Dennis Garlick[b], Cynthia D. Fast[b], Daniel T. Blumstein[a,*]

[a] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA
[b] Department of Psychology, University of California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 11 January 2012
Received in revised form 31 May 2012
Accepted 20 June 2012

Keywords:
Animal personality
Behavioral syndrome
Hermit crabs

ABSTRACT

There is much interest in studying animal personalities but considerable debate as to how to define and evaluate them. We assessed the utility of one proposed framework while studying personality in terrestrial hermit crabs (*Coenobita clypeatus*). We recorded the latency of individuals to emerge from their shells over multiple trials in four unique manipulations. We used the specific testing situations within these manipulations to define two temperament categories (shyness-boldness and exploration-avoidance). Our results identified individual behavioral consistency (i.e., personality) across repeated trials of the same situations, within both categories. Additionally, we found correlations between behaviors across contexts (traits) that suggested that the crabs had behavioral syndromes. While we found some correlations between behaviors that are supposed to measure the same temperament trait, these correlations were not inevitable. Furthermore, a principal component analysis (PCA) of our data revealed new relationships between behaviors and provided the foundation for an alternate interpretation: measured behaviors may be situation-specific, and may not reflect general personality traits at all. These results suggest that more attention must be placed on how we infer personalities from standardized methods, and that we must be careful to not force our data to fit our frameworks.

1. Introduction

There is much interest in studying animal personalities yet there is disagreement as to how they should be defined and studied (Dingemanse and Wolf, 2010; Dingemanse et al., 2010a; Groothuis and Trillmich, 2011; Réale et al., 2010; Stamps and Groothuis, 2010; Uher, 2011; Wolf and Weissing, 2010). In a broad sense, animal personality refers to the existence of consistent differences in behavior between individuals (Réale et al., 2010) and the maintenance of these individual differences both in varying situations and through time (Dingemanse et al., 2010a). The presence of these consistent individual differences can generate a behavioral syndrome, which is seen when there are inter-individual correlations between consistent traits across contexts or time, or between distinct traits, expressed at the population level (Dingemanse et al., 2010a; Réale et al., 2010; Sih et al., 2004; Smith and Blumstein, 2008). But how are these individual differences identified, classified, and connected?

Given the great interest in quantifying personality, it is imperative that there is agreement on how personality is defined. Words commonly used to describe personality traits such as “boldness” and “explorative” may be inconsistently used or defined in inappropriate ways. Some studies have shown these traits to be correlated in species/populations like the great tit (Drent et al., 2003; Van Oers et al., 2003), while other studies (e.g., Yoshida et al., 2005; Wilson and Godin, 2009) emphasized a more causal relationship between the traits.

When boldness is defined by response to novelty, researchers may assume that bolder individuals will be more explorative and vice versa. This idea has been applied in a variety of contexts, for example, testing for boldness by exposing individuals to novel objects (e.g., Wilson et al., 1993; Pronk et al., 2010) or by allowing individuals to emerge and explore in a novel open field test (e.g., Wilson et al., 1994; Brown and Braithwaite, 2004; Yoshida et al., 2005; Brown et al., 2007). However, the terms are often used inconsistently or interchangeably (Réale et al., 2007).

In studies where boldness is defined as a willingness to take risks and be exploratory, exploration tests are used to study boldness. The data produced by these tests are referred to as boldness measures. One such study in sunfish concluded that more-exploratory individuals were more willing to take risks; however, this is redundant with the definition of boldness used

---

* Corresponding author at: Department of Ecology and Evolutionary Biology, 621 Young Drive South, University of California, Los Angeles, CA 90095-1606, USA. Tel.: +1 310 267 4746; fax: +1 310 206 3987.

E-mail addresses: wdstahlman@gmail.com, wstahlma@ucla.edu (W.D. Stahlman), marmots@ucla.edu (D.T. Blumstein).

0376-6357/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.beproc.2012.06.007

(Wilson and Godin, 2009). There are other cases where terms are inconsistently used (e.g., Dzieweczynski and Crovo, 2011). These inconsistencies suggest that it is essential to ensure that researchers consistently define personality dimensions from a set of observed behaviors. While researchers should be free to design unique personality assays to observe behavioral responses, the manner in which the behaviors are classified by trait categories should be clear.

Réale et al. (2007) suggested a framework to classify personality traits and methods for how these traits may be defined. They suggested five distinct categories that were defined by the situation in which they were tested: (1) *shyness-boldness*, which refers to reactions to risky but not new situations, (2) *exploration-avoidance*, which refers to reactions to new situations that may or may not be associated with risk, (3) *activity*, which refers to the level of activity (in non-novel and non-risky situations) often quantified as locomotion, (4) *aggressiveness*, which refers to the level of agonistic behavior to conspecifics, and (5) *sociability*, which refers to reactions (excluding aggression) to the presence or absence of conspecifics.

Since Réale et al. (2007) initially proposed a consistent set of personality dimensions from a set of observed traits, there has been further work using structural equation modeling by Dingemanse et al. (2010b) that permits likely syndrome structures to be defined by testing the fit of a proposed set of relationships between the observed traits to empirical data. Nevertheless, Réale et al.'s (2007) initial framework focuses on inferring traits from qualitatively similar or different observation methods rather than defining their relationships. Various trait categories such as the five suggested by Réale et al. (2007) are used to formulate potential structural equation models. Slight differences in behavior categorization or variation in the testing situation of the measured behavior could lead to different conclusions about the most probable behavioral syndromes. Thus, appropriate testing conditions and categorization of behaviors become critical to properly identify syndrome structure.

Various studies have used Réale et al.'s (2007) framework as a guide to defining animal personality traits (e.g., Archard and Braithwaite, 2011; Eriksson et al., 2010) yet formal tests which both apply and evaluate this methodology are largely lacking. We studied personality and behavioral syndromes in a terrestrial invertebrate and simultaneously assessed the utility of Réale et al.'s (2007) operational definitions of temperament traits. We focused on two of the five traits: shyness-boldness (which involves reactions to situations of risk without any aspect of novelty) and exploration-avoidance (which involves reactions to novel situations that may or may not be risky). Following Réale et al.'s (2007) framework, we should expect that different methods to quantify the same personality trait would be correlated; however, we should not necessarily find correlations between methods that purport to examine different traits.

Aquatic hermit crabs (*Pagurus bernhardus*) have been used in a variety of animal personality studies. For example, Briffa et al. (2008) suggested individual consistency in the differences of startle response of these crabs across two situations (presence and absence of predator cues). A more recent study conducted by Briffa and Twyman (2010) again suggested individual consistency in latency-to-emerge behavior in four situations (varying shell and substrate colors).

We elected to study a terrestrial Caribbean hermit crab (*Coenobita clypeatus*) to evaluate Réale et al.'s (2007) framework. This hermit crab's behaviors are easily observable and quantifiable, but this particular species has not yet been the subject of personality studies. In general, and despite their abundance, concern for the welfare of invertebrates is minimal (Carere et al., 2011). Studies of invertebrate personality may better inform discussions about invertebrate welfare (Mather and Anderson, 1993; Wemelsfelder et al., 2001).

We quantified hermit crab behavior in four separate manipulations. Each manipulation included one or more situations that were categorized according to their corresponding temperament traits using Réale et al.'s (2007) definitions. We quantified the latency-to-emerge in multiple novel situations; although such tests are sometimes used to study boldness (as described above), because of the novelty of each situation, each should quantify exploration-avoidance. We expected to see individual consistency over repeated trials in the same situation. We also expected to find substantial correlations between behaviors within the same temperament trait category, and, if there were syndromes, correlations between behaviors in separate categories.

2. Materials and methods

2.1. Subjects

Our subjects were 24 terrestrial hermit crabs that varied in body mass (mean mass = 6.9 g ± 3.2 SD), shell length (2–5 cm), shell density (2.1 ± 1.1 g/ml), and shell and body colors. Crabs were supplied by two commercial suppliers and were randomly assigned to a home tank (6 crabs per 19 cm × 45 cm × 19.5 cm tank). A total of 9 crabs died or exhibited signs of stress (e.g., limb loss) and were gradually removed from the experimental pool between the manipulations that were conducted over six weeks. The crabs were shipped during the winter and despite a period of acclimatization following their arrival, we suspect that some were cold-stressed during shipping. We recognize this mortality may have been nonrandom. The unexpected loss of subjects also clearly reduced our power. Ultimately, however, we focused on correlations between measures and it is unlikely that this mortality systematically biased our results. We did not account for crab age since we did not know, with certainty, their age. However, we are not aware of any previous hermit crab study that has worked with known-age crabs or has been able to use this as a factor or covariate in analysis. Additionally, prior studies using aquatic hermit crab species found individual differences in startle response to be independent of sex and size (Briffa et al., 2008) so it is not unreasonable to consider our results are independent of these factors as well. For identification, each crab was marked with a different color of non-toxic nail polish painted on both the large claw and the shell.

2.2. Manipulations

Table 2 provides an outline of our basic experimental design. All manipulations were performed in the laboratory and were videotaped. Manipulations were performed in successive order (i.e. inversion; open field; visual predator; shock) to control for individual crab experience level and examine within-manipulation consistency. After completing an entire experimental phase for a *manipulation* we waited between 4 and 14 days before initiating the next one. Since subject order was changed for each trial within a manipulation, each crab had approximately the same mean interval of time between manipulations.

We measured a variety of behaviors and identified them according to their corresponding temperament traits (Table 1). The behaviors *latency-to-hide* or *latency-to-emerge* were measured by: inversion, placement in an open field, exposure to an expanding image of a raptor, and administering an electric shock.

To quantify exploration-avoidance, we timed the initial latency-to-emerge in situations conducted in novel environments: the manual inversion was conducted on cloth; the open field was conducted in a wooden box; the visual predator demonstration

Table 1
Measured behaviors categorized according to their temperament traits as defined by Réale et al. (2007); see text for details.

Behavior measured	Temperament trait
Manual inversion: initial latency-to-emerge	Exploration-avoidance
Open field: initial latency-to-emerge	Exploration-avoidance
Visual predator: initial latency-to-emerge (pre-stimulus)	Exploration-avoidance
Shock: initial latency-to-emerge (pre-shock)	Exploration-avoidance
Visual predator: initial latency-to-hide	Shyness-boldness
Visual predator: initial latency-to-re-emerge (post-predator)	Shyness-boldness
Visual predator: number of trials to habituate	Shyness-boldness
Shock: initial latency-to-re-emerge (post-shock)	Shyness-boldness

involved securing the crab in a novel clamp apparatus in a separate sound-attenuating room; and the shock occurred in a novel chamber. According to Réale et al. (2007), each of these situations should quantify the exploration-avoidance trait because the behavior was measured in a new surrounding environment. To quantify shyness-boldness (which reflects natural risk assessment), we measured both the initial latency-to-hide and the initial latency-to-re-emerge following the presentation of an expanding image of a raptor, as well as determining the number of trials each individual received until it habituated (no longer withdrew into its shell) to the raptor image. Additionally, we timed individual latency-to-re-emerge from the shell following a shock-induced withdrawal response. Réale et al.'s (2007) guidelines suggest that these situations should measure shyness-boldness because behaviors were measured in risky situations.

We first quantified individual consistency within a situation by measuring individual latencies-to-emerge for multiple trials of the manual inversion, open field, and shock manipulations and then calculating an intra-class correlation coefficient, a measure of test–retest reliability (Uher, 2011). We expected to see variation within the population, but also intra-individual consistency across the trials, which would suggest personality. We also expected to see correlations between within-trait behaviors, and possibly between across-trait behaviors as well, which would suggest behavioral syndromes.

2.2.1. Manipulation 1: manual inversion

We followed Briffa et al.'s (2008) protocol to systematically quantify hermit crab hiding behavior via manual inversion. Following this inversion, crabs were placed in a novel cloth and allowed to emerge, which, according to Réale et al.'s (2007) classification, suggests that it quantified exploration-avoidance. It is important to note, however, that a manual inversion may be similar to a natural predatory experience and thus may involve some degree of risk.

We conducted this experiment on a desk surface in a large room (3.7 m × 6.1 m × 2.4 m) under dim light. On day 1 in the morning (starting at 08:30), each individual was manually lifted, flipped, and held in the inverted position for 10 s (see Briffa et al., 2008). This caused the crab to withdraw into its shell and block the aperture with the major claw. Subjects were then placed, aperture up, in the folds of a novel soft washcloth to stabilize the shell and to avoid excess jostling. The latency-to-emerge was measured starting from when the crabs were placed in the washcloth until both pairs of walking legs reappeared outside the aperture. We calculated the time the crab took to emerge, and returned the crabs to their home tanks. Six hours later on the same day starting at 14:30, we repeated the procedure. On the following day, we repeated the inversion two more times (with the same 6-h interval between trials) with subject order reversed to balance any potential temporal effects.

2.2.2. Manipulation 2: open field

Following Réale et al. (2007), we use the initial latency-to-emerge as a measure of exploration-avoidance because the crabs have been placed in a novel environment—the open field.

We conducted this manipulation in the same large room as in *manipulation 1* (Section 2.2.1). An open top wood box (1 m × 1 m × 0.14 m high walls), painted a uniform dark blue, was used as the open field. Each crab was subjected to four trials (1 trial per day for 4 days) of the open field manipulation, where the order in which the crabs were tested was randomized each time. Each trial was recorded with a web camera positioned directly above the open field.

For all trials, each hermit crab was inverted for 10 s to promote withdrawal. The crab was then placed in the center of the open field, aperture facing down. The time from the placement of the crab in the open field to the time of emergence (when both pairs of walking legs emerged) was measured and recorded.

2.2.3. Manipulation 3: simulated visual predator attack

Chan et al. (2010) found that a gradually expanding image of a raptor with spread wings displayed on an LCD monitor successfully caused hermit crabs to withdraw into their shells.

The initial latency-to-emerge before the crab experiences the visual predator is best described as an exploration-avoidance trait because the crab is behaving in a new situation (as described below, it was clamped in front of a screen). The latency-to-hide reflects boldness-shyness because it is a response to the threatening predator. The latency-to-re-emerge after the visual stimulus is

Table 2
An outline of the experimental design.

Manipulation	Number of trials	Measures observed (situations)	Number of times measure was observed	Total number of measures obtained
1: Manual inversion	4	Latency-to-emerge	Once per trial	4
2: Open field	4	Latency-to-emerge	Once per trial	4
3: Visual predator	Equal to number of trials to habituate	Latency-to-emerge	Once	1
		Latency-to-hide	Once per trial	1 less than number of trials to habituate
		Latency-to-re-emerge (post predator)	Once per trial	1 less than number of trials to habituate
		Number of trials to habituate	Once	1
4: Shock	2	Latency-to-emerge	Once per trial	2
		Latency-to-re-emerge (post shock)	Once per trial	2

a response to a risky situation (because the crabs emerges following a simulated predatory attack), and should reflect boldness-shyness. Crabs may habituate to repeated exposure to the predatory stimulus (Rodriguez-Prieto et al., 2009; Stahlman et al., 2011) and stop withdrawing into their shells. Although habituation is a learning process, in this case it is based on the individuals' risk assessment, and thus, the number of trials to habituate reflects boldness-shyness.

We conducted this manipulation in a 2.5 m by 1.5 m sound-attenuating room. Each crab was placed in a C-clamp attached to a wood platform with its aperture facing upward and slightly forward. The clamp held the crab so that its shell touched the platform. An LCD monitor that displayed the visual stimulus (as described above) was placed directly in front of the clamp. The apparatus and background were painted blue to ensure a homogeneous background. A web camera was placed 5 cm in front of the clamp, below the LCD monitor, and a software program recorded crab emergence and hiding times as well as obtain the number of trials to habituate.

The image of the raptor started as one pixel at the center/top of the monitor. Over the next 17 s the raptor descended and enlarged at a constant rate until it reached the bottom of the screen. Its maximum size was 30 cm (900 pixels wide). The time between the visual stimulus appearing to when the hermit crab retracted into its shell (latency-to-hide) was recorded by the detector. Each crab was subjected to repeated trials until the crab habituated (showed no hide response upon predator presentation) to the visual stimulus. The order of testing for individuals was randomized. We tested 4–6 different crabs per day over 4 days. The number of trials it took for each individual to habituate was recorded. Note: because we recorded one single habituation value per individual, it was impossible to calculate repeatability.

2.2.4. Manipulation 4: electric shock

Hermit crabs have been observed to withdraw into their shells in response to receiving an electrical shock (Appel and Elwood, 2009). We conducted this experiment (described below) in a Skinner box. Since the initial latency-to-emerge in the Skinner box was a response in a new environment, we classified this as exploration-avoidance. The latency-to-re-emerge after the shock was a behavior in response to the shock, which is a risk. Thus, the latency-to-re-emerge was classified as a measure of boldness-shyness. Although a shock is a novel situation, latency-to-hide to the shock would measure exploration-avoidance behavior, not the latency-to-re-emerge.

Each subject received an electric shock (30 mA), which caused subjects to withdrawal. The experiment took place in a 22.5 cm × 32 cm × 25 cm Skinner box (Med Associates Inc., model #ENV-008). We designed a circuit in which the crab could receive a shock only when it was emerged from its shell and never while it was hiding. The circuit was created using standard electrical wire between one metal rod from the floor of the chamber and the crab by inserting the exposed leads of one end of the wire into the crab's shell, so that contact was maintained between the wire and the body of the crab. Each crab was held in an elevated position (2 cm) over a thin 6.5 cm × 13.5 cm metal plate using a C-clamp attached to a wooden platform (19 cm × 18 cm) in the box. The metal plate was secured to the wood platform by tape, and was also connected via an alligator clip to a separate rod of the chamber floor because contact with two separate rods in the Skinner box was required to produce a shock. In this arrangement, contact between the crab and the plate closed the circuit and was only achieved if the crab was emerged. Each subject was dipped in a bath of 1% saltwater prior to securing the wires. A few drops of salt water were also dripped on the surface of the metal plate. The shell of the crab was dried to ensure the wires could be securely fastened. After inserting the wires, each crab was manually secured in the C-clamp, aperture facing down. We were able to manually administer a shock to the crab when it was in contact with the metal plate by charging the rods of the Skinner box. All crabs immediately retracted into their shells upon shock initiation. We timed the latency-to-re-emerge. Each crab had one trial of the shock manipulation per day over 2 days. The order of testing for individuals was randomized each day.

2.3. Statistical analysis

Because we found large intra-class correlation coefficients (see Section 3.1), we used each individual's initial response times to calculate correlations between situations and contexts. Large correlation coefficients (r) would indicate individual consistency between both situations and contexts. Based on the Réale et al. (2007) terminology, we expected that the initial latency-to-emerge should be correlated across the manual inversion manipulation, open field manipulation, visual predator manipulation (pre-stimulus), and shock manipulation (pre-shock) since we consider they all relate to the exploration-avoidance temperament trait. Additionally, we expected the initial latency-to-hide in the visual predator attack, the initial latency-to-re-emerge (post-stimulus), the number of trials to habituate to the predator, and the initial latency-to-re-emerge (post-shock) to be correlated since they all presumably relate to the shyness-boldness temperament category. We looked for broader syndromes as evidenced by correlations of behaviors across different traits. In addition to identifying pair-wise correlations, we performed a Principal Component Analysis (PCA) on our data for the situations in the manipulations, excluding the shock manipulation (for which there was insufficient data). We extracted the first two principle components because we expected that if different measures of the same situation were correlated, that they would be load on a single component. Since we were studying exploration and shyness, we expected to see different situations loading uniquely on these two factors. We interpreted Varimax rotated component scores $\geq|0.7|$ as loading on a particular factor.

3. Results

3.1. Are behavioral responses repeatable?

Our measured traits were generally repeatable as measured by their intra-class correlation coefficient (r_{IC}) and they varied substantially among individuals. We found very high levels of repeatability in both the manual inversion latency-to-emerge ($r_{IC} = 0.859$, $p < 0.0001$) and the open field latency-to-emerge ($r_{IC} = 0.891$, $p < 0.0001$) across the 4 trials in each situation. We found generally high repeatability in the initial latency-to-emerge (pre-shock) ($r_{IC} = 0.632$, $p = 0.036$), and the latency-to-emerge post-shock ($r_{IC} = 0.587$, $p = 0.055$) over the 2 trials of the shock manipulation. We did not calculate repeatability in latency-to-emerge or latency-to-hide during the predator presentation habituation trials because inter-individual variation in habituation would render such comparisons uninterpretable. We recognize that variation in habituation rate may influence repeatability. For example, if inter-individual habituation rates were very different, this alone could lead to inconsistencies in inter-individual behavior and therefore affect our conclusions about personality. For the situations in which the intra-class correlation coefficient was calculated, trials were conducted several hours apart to reduce the likelihood that inter-individual differences would result simply from habituation.

Table 3

Correlations among the latency terrestrial hermit crabs took to emerge across situations and contexts. Note that the six correlations above the horizontal line, as well as the six to the right of the vertical line, are expected to be significant given that they were correlations of measures within a temperamental category.

	Manual inversion: initial latency-to-emerge (s)	Open field: initial latency-to-emerge (s)	Visual predator: initial latency-to-emerge (pre-stimulus) (s)	Shock: initial latency-to-emerge (pre-shock) (s)	Visual predator: initial latency-to-hide (s)	Visual predator: initial latency-to-re-emerge (post-stimulus) (s)	Visual predator: number of trials to habituate
Open field: initial latency-to-emerge (s)	$r = 0.311$ $p = 0.139$ $n = 22$						
Visual predator: initial latency-to-emerge (pre-stimulus) (s)	$r = 0.033$ $p = 0.877$ $n = 18$	$r = 0.177$ $p = 0.586$ $n = 18$					
Shock: initial latency-to-emerge (pre-shock) (s)	$r = 0.278$ $p = 0.315$ $n = 15$	**$r = 0.863$** $p < 0.001$ $n = 15$	$r = 0.041$ $p = 0.884$ $n = 15$				
Visual predator: initial latency-to-hide (s)	$r = -0.088$ $p = 0.681$ $n = 18$	$r = 0.139$ $p = 0.516$ $n = 18$	**$r = 0.680$** $p < 0.001$ $n = 18$	$r = -0.484$ $p = 0.067$ $n = 15$			
Visual predator: initial latency-to-re-emerge (post-stimulus) (s)	$r = -0.229$ $p = 0.283$ $n = 18$	**$r = 0.485$** $p < 0.02$ $n = 18$	$r = 0.327$ $p = 0.119$ $n = 18$	**$r = 0.737$** $p < 0.003$ $n = 15$	$r = 0.242$ $p = 0.255$ $n = 18$		
Visual predator: number of trials to habituate	$r = -0.118$ $p = 0.582$ $n = 18$	$r = 0.108$ $p = 0.616$ $n = 18$	**$r = 0.482$** $p < 0.02$ $n = 18$	$r = 0.098$ $p = 0.727$ $n = 15$	$r = 0.390$ $p = 0.060$ $n = 18$	$r = 0.260$ $p = 0.219$ $n = 18$	
Shock: initial latency-to-re-emerge (post-shock) (s)	$r = 0.037$ $p = 0.895$ $n = 15$	$r = 0.514$ $p = 0.050$ $n = 15$	$r = 0.043$ $p = 0.878$ $n = 15$	**$r = 0.809$** $p < 0.001$ $n = 15$	**$r = -0.547$** $p < 0.04$ $n = 15$	**$r = 0.844$** $p < 0.001$ $n = 15$	$r = 0.143$ $p = 0.611$ $n = 15$

Bolded values indicate significant correlation coefficients at $p < 0.05$.

3.2. Are behavioral responses correlated across situations and contexts?

Many of our latency measures were correlated across situations and contexts (Table 3). We found one significant correlation between measures that belonged to the exploration-avoidance temperament trait category. The initial latency-to-emerge in the open field was correlated to the initial latency-to-emerge in the shock manipulation, pre-shock. We found two significant correlations between measures that belonged to the shyness-boldness temperament trait category. The initial latency-to-re-emerge post-shock was correlated with both the initial latency-to-hide in the visual predator manipulation and the initial latency-to-re-emerge in the visual predator manipulation.

We also observed multiple correlations between measures that were categorized according to different temperament traits (Table 3). The open field initial latency-to-emerge (exploration-avoidance) was correlated with the initial latency-to-re-emerge (post-stimulus) in the visual predator manipulation (shyness-boldness). The visual predator manipulation initial latency-to-emerge pre-stimulus (exploration-avoidance) was correlated with both the visual predator initial latency-to-hide and the visual predator number of trials to habituate (both shyness-boldness). The initial latency-to-emerge in the shock manipulation, a measure of exploration-avoidance, was correlated with two measures of shyness-boldness: the visual predator latency-to-re-emerge (post-stimulus) and the latency-to-re-emerge (post-shock).

From our correlation matrix, we found no substantial relationships between the initial latency-to-emerge in the manual inversion and any other behaviors we quantified (Table 3).

Factor 1 of our PCA, included the visual predator initial latency-to-hide and visual predator number of trials to habituate (Table 4),

Table 4

Principal component analysis (PCA) of behaviors from three situations of exploration-avoidance and three situations of shyness-boldness formed into two factors.

	Factor 1	Factor 2
Manual inversion initial latency-to-emerge	−0.223	**0.890**
Open field initial latency-to-emerge	0.096	**0.833**
Visual predator initial latency-to-emerge	**0.868**	0.110
Visual predator initial latency-to-hide	**0.869**	−0.065
Visual predator initial latency-to-re-emerge post predator	0.345	**0.754**
Visual predator number of trials to habituate	**0.706**	0.101

Bolded values indicate significant correlation coefficients at $p < 0.05$.

which we expected since they both measure shyness-boldness. Interestingly, we did not see a significant correlation between these behaviors in the pair-wise correlation coefficients. Additionally, visual predator initial latency-to-emerge was included on this factor, even though it supposedly measured exploration-avoidance, a finding that was also seen by looking at the raw correlation matrix.

Factor 2 of our PCA analysis (Table 4), included the manual inversion initial latency-to-emerge and the open field initial latency-to-emerge, which we expected since they both measure exploration-avoidance. Interestingly, we did not see this relationship in the pair-wise correlations (Table 3). Additionally, visual predator initial latency-to-emerge post-predator was also grouped with Factor 2. From Table 3, we saw that open field latency-to-emerge was correlated with visual predator initial-latency-to-re-emerge post-predator, but from PCA, we now see there is a predictive relationship between all three.

4. Discussion

While there was substantial variation across subjects, individuals responded in consistent ways to the manual inversion manipulation, the open field manipulation, and the shock manipulation. These high levels of repeatability suggest that individual terrestrial hermit crabs vary systematically both in the exploration-avoidance and in the shyness-boldness traits.

Interestingly, we found certain correlations between measures of different temperament traits although we did not detect correlations between all measures within the same temperament trait (Table 3), and the complementary PCA analysis was generally in agreement (Table 4).

4.1. Quantifying exploration-avoidance

While pair-wise correlations did not detect significant correlations between the manual inversion and open field latency-to-emerge, they were grouped together in the PCA. Thus, we conclude that either of these situations would be sufficient to study exploration-avoidance. This is biologically relevant, because naturally falling or rolling crabs may find themselves in a variety of new locations in an inverted or non-inverted position. Thus, the degree to which they explored their new environment should be functionally correlated.

We also found significant correlations between individual latency-to-emerge in the open field manipulation and pre-shock situation of the shock manipulation. We expected to see these relationships since each quantified the exploration-avoidance temperament trait. This suggests that measuring initial latency-to-emerge after inversion, aperture downward in an open field or when secured in a clamp, all measure the same trait—exploration-avoidance. Interestingly, the PCA grouped visual predator latency-to-re-emerge (post predator) with these measures, although we hypothesized it would measure shyness-boldness.

Surprisingly, only the initial latency-to-emerge in the visual predator manipulation was not correlated with any other exploration-avoidance measure. Instead, according to our correlations and PCA, this trait is a better indicator of shyness-boldness.

4.2. Quantifying shyness-boldness

We found both correlational and PCA support for shyness-boldness measures. According to the PCA, and as expected, visual predator initial latency-to-hide and the number of trials to habituate were grouped in the same factor. Additionally, PCA grouped visual predator initial latency-to-emerge with these two shyness-boldness measures. It is possible that this situation is quite risky; for example, if a predator wished to extract a crab from its shell, it might secure the crab in such a way.

We found correlations between the latency-to-re-emerge post-shock and both the latency-to-hide to the visual predator and the latency-to-re-emerge after the visual predator stimulus which we also expected since these behaviors quantify shyness-boldness and relate to how crabs respond in risky situations. Interestingly, visual predator initial latency-to-re-emerge was also correlated with three measures of exploration avoidance, and PCA grouped this measure with the exploration-avoidance trait measures.

4.3. How should temperaments be quantified?

Our findings show that some, but not all, measures suggested by Réale et al. (2007) appropriately assess that which they were suggested to measure. Their framework would have predicted significant correlations among different situations that defined the same temperamental trait. Indeed, testing the same behavior that falls into the same temperament trait, but using two different situations does not inevitably lead to substantial correlations or factor scores. This poses a problem. If we had only tested initial latency-to-emerge in the open-field manipulation, but not in the visual predator manipulation, we might have made erroneous conclusions that the behaviors would be similar in both situations when they were in reality very different. Additionally, our results suggest there are measures that are correlated both within and across traits. It is possible that both novelty and risk are involved. Using one of these more ambiguous measures could lead to improper conclusions when studying personality. The difficulty results from not knowing whether a behavior correlates with both traits unless it was compared to other measures of both traits, but in reality, many studies use only one measure designated to represent each different trait. In such cases, only one representative measured behavior per category may be used as the basis for very general conclusions about personality or behavioral syndromes (e.g., Dingemanse et al., 2002, 2007; Cote et al., 2010; Scales et al., 2011). Interestingly, our study shows that these oversimplified studies may potentially generate erroneous conclusions.

More sophisticated methods to identify behavioral syndromes (e.g., structural equation modeling [SEM]), are potentially vulnerable to this error because models are generally based on the relationships between general behavioral trait categories. If behaviors within the trait categories themselves are not necessarily correlated, the use of different personality tests to measure behaviors that should belong to the same trait could lead to major differences across studies. Differences in experimental design could therefore generate behavioral syndrome variability across populations within species. Thus, until the appropriate assessments about the validity of simplified studies are conducted, it seems prudent to include more than a single testing situation for a particular personality trait (e.g., Carere et al., 2005). Such potentially redundant testing will also ensure that the trait descriptions are highly representative and may help identify higher-level syndromes. In addition, this type of testing could allow for a more comprehensive understanding of a particular personality trait as it may be unlikely that one testing situation alone can precisely define that entire trait category.

It is important to note that a re-analysis of the data in Dingemanse et al. (2007) using structural equation models found additional covariance between traits that suggests a syndrome structure that was not initially observed (Dingemanse et al., 2010b). If simply re-analyzing already existing data sets using SEM yields different results, it may be worthwhile to also investigate how re-categorizing behaviors within trait categories affects syndrome structures when using promising methods like SEM.

4.4. PCA reveals a new interpretation

We initially expected the top three variables in Table 4 to load onto one factor indicating the exploration-avoidance trait, and the bottom three variables to load onto a separate factor indicative of the shyness-boldness trait. While four of the variables fell into the predicted factors, two of them loaded onto the opposite factor than predicted. This suggests that each factor indicates something different than what we presumed.

Let us consider Factor 1. It showed high loading by visual predator initial latency-to-emerge, visual predator latency-to-hide, and visual predator number of trials to habituate. What is common among these three variables? While it seems that the initial response to detecting the visual predator and the number of trials before behavioral habituation to the repeated presentation of the predator are logically connected, the initial latency-to-emerge in this setup also loads on this factor despite the visual predator having not yet appeared. Perhaps the action of being placed into a restraining clamp activates the same behavior class as does detecting an approaching predator. In fact, it may even simulate the handling cues that would be present following being captured and held by a predator. Thus, we might conclude that Factor 1 groups behaviors related to predator detection, rather than a general personality trait like shyness-boldness.

While initial latencies-to-emerge in the manual inversion and open field manipulations load on Factor 2 as expected, initial latency-to-emerge post-predator in the visual predator manipulation unexpectedly loads on this factor. One perspective suggests that what these three variables share in common is a recent encounter with a predator. For the first two variables, the crabs had recently been handled by a human experimenter; for the third variable, the crabs had just encountered a visual predator. After a period of time post-encounter, the hermit crab makes a behavioral decision to investigate whether the predator is no longer present, regardless of its new surroundings.

This reinterpretation based on relabeling the factors provides an important lesson: the experimenter must be cautioned against forcing measures into predefined categories or labels. Animals may perceive situations very differently from us, and thus respond to them in different ways. Dingemanse et al. (2007) found that stickleback behavior in response to the introduction of a perch was correlated with stickleback behavior in a novel environment. Although the researchers initially expected that the perch test should be a test of boldness, the behaviors were concluded to be individuals' responses to the altered environment, not the predator.

While it remains possible that risk and/or novelty were involved in each of our test situations, our results showed that there might be more to correlations in behaviors than a general underlying personality trait. Even though this classification system simplifies personality studies, it is not ideal when results overlook real, complex relationships, and instead generate an artificial personality structure. We must work harder to understand how animals, not we, perceive different situations. Stamps and Groothuis (2010) suggested that rather than designing experiments to test behaviors in specific categories and then looking for correlations, it would be more practical to first identify related behaviors and use them to understand what aspects of personality development control these behaviors. Doing this may enable us to better understand how personality traits play a role in controlling behaviors, rather than strictly define them. Exploratory analyses, such as those provided by PCA, can be useful to check the validity of our measures and offer alternative perspectives from which to view our data sets and the sets of correlations among measures. Such alternative perspectives can then lead to more in-depth investigations of the underlying personality factors.

In conclusion, a well-defined taxonomy of temperamental traits (Réale et al., 2007) may not always generate consistent results. Indeed, tests designed to measure behaviors that indicate these traits may actually reflect much more specific conditions. While there have been suggestions that a more 'ground-up' approach could be useful (Stamps and Groothuis, 2010), we lack direct comparisons of the two approaches. Developing a functional, universal assessment approach which considers not only temperament traits but also groups of related behaviors across those traits will be fundamental to future advances in the study of animal personality.

Acknowledgements

Support for this research was provided by NSF Research Grant BCS-0843027 (to A. P. Blaisdell). We thank Kelsea Ryan, Kirandeep Sumra, Zunera Ghaznavi, Hwee Cheei Lim, Kevin Nguyen, Julien Martin, and Mathew Petelle for assistance and/or comments, and two astute referees for many constructive comments on a previous version.

References

Appel, M., Elwood, R.W., 2009. Motivational trade-offs and potential pain experience in hermit crabs. Appl. Anim. Behav. Sci. 119, 120–124.

Archard, G.A., Braithwaite, V.A., 2011. Variation in aggressive behaviour in the poeciliid fish *Brachyrhaphis episcope*: population and sex differences. Behav. Process. 86, 52–57.

Briffa, M., Rundle, S.D., Fryer, A., 2008. Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab *Pagurus bernhardus*. Proc. R. Soc. Lond. B 275, 1305–1311.

Briffa, M., Twyman, C., 2010. Do I stand out or blend in? Conspicuousness awareness and consistent behavioural differences in hermit crabs. Biol. Lett. 7, 330–332.

Brown, C., Braithwaite, V.A., 2004. Size matters: a test of boldness in eight populations of the poeciliid *Brachyraphis episcopi*. Anim. Behav. 68, 1325–1329.

Brown, C., Burgess, F., Braithwaite, V.A., 2007. Heritable and experiential effects on boldness in a tropical poeciliid. Behav. Ecol. Sociobiol. 62, 237–243.

Carere, C., Drent, P.J., Privitera, L., Koolhaas, J.M., Groothuis, T.G.G., 2005. Personalities in great tits, *Parus major*: stability and consistency. Anim. Behav. 70, 795–805.

Carere, C., Woods, J., Mather, J., 2011. Species differences in captivity: where are the invertebrates? Trends Ecol. Evol. 26, 211.

Chan, A.A.Y.-H., Stahlman, W.D., Garlick, D., Fast, C.D., Blumstein, D.T., Blaisdell, A.P., 2010. Increased amplitude and duration of acoustic stimuli enhance distraction. Anim. Behav. 80, 1075–1079.

Cote, J., Fogarty, S., Weinersmith, K., Brodin, T., Sih, A., 2010. Personality traits and dispersal tendency in the invasive mosquitofish (*Gambusia affinis*). Proc. R. Soc. Lond. B 277, 1571–1579.

Dingemanse, N.J., Both, C., Drent, P.J., van Oers, K., van Noordwijk, A.J., 2002. Repeatability and heritability of exploratory behavior in great tits from the wild. Anim. Behav. 64, 929–938.

Dingemanse, N.J., Wright, J., Kazem, A.J.N., Thomas, D.K., Hickling, R., Dawnay, N., 2007. Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. J. Anim. Ecol. 76, 1128–1138.

Dingemanse, N.J., Kazem, A.J.N., Réale, D., Wright, J., 2010a. Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol. Evol. 25, 81–89.

Dingemanse, N.J., Dochtermann, N., Wright, J., 2010b. A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. Anim. Behav. 79, 439–450.

Dingemanse, N.J., Wolf, M., 2010. Recent models for adaptive personality difference: a review. Philos. Trans. R. Soc. B 365, 3947–3958.

Drent, P.J., van Oers, K., van Noordwijk, A.J., 2003. Realized heritability of personalities in the great tit (*Parus major*). Proc. R. Soc. Lond. B 270, 45–51.

Dzieweczynski, T.L., Crovo, J.A., 2011. Shyness and boldness differences across contexts in juvenile three-spined stickleback *Gasterosteus aculeatus* from an anadromous population. J. Fish Biol. 79, 776–788.

Eriksson, A.C., Booth, D.J., Biro, P.A., 2010. 'Personality' in two species of temperate damselfish. Mar. Ecol. Prog. Ser. 420, 273–276.

Groothuis, T.G.G., Trillmich, F., 2011. Unfolding personalities: the importance of studying ontogeny. Dev. Psychobiol. 53, 641–655.

Mather, J.A., Anderson, R.C., 1993. 'Personalities of octopuses (*Octopus rubescens*). J. Comp. Psychol. 107, 336–430.

Pronk, R., Wilson, D.R., Harcourt, R., 2010. Video playback demonstrates episodic personality in the gloomy octopus. J. Exp. Biol. 213, 1035–1041.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. Camb. Philos. Soc. 82, 291–318.

Réale, D., Dingemanse, N.J., Kazem, A.J.N., Wright, J., 2010. Evolutionary and ecological approaches to the study of personality. Philos. Trans. R. Soc. B 365, 3937–3946.

Rodriguez-Prieto, I., Fernández-Juricic, E., Martín, J., Regis, Y., 2009. Antipredator behaviour in blackbirds: habituation complements risk allocation. Behav. Ecol. 20, 371–377.

Scales, J., Hyman, J., Hughes, M., 2011. Behavioral syndromes break down in urban song sparrow populations. Ethology 117, 887–895.

Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004. Behavioural syndromes: an integrative overview. Q. Rev. Biol. 79, 241–277.

Smith, B.R., Blumstein, D.T., 2008. Fitness consequences of personality: a meta-analysis. Behav. Ecol. 19, 448–455.

Stahlman, W.D., Chan, A.A.Y.-H., Blumstein, D.T., Fast, C.D., Blaisdell, A.P., 2011. Auditory stimulation dishabituates anti-predator escape behaviour in hermit crabs (*Coenobita clypeatus*). Behav. Process. 88, 7–11.

Stamps, J., Groothuis, T.G.G., 2010. The development of animal personality: relevance, concepts and perspectives. Biol. Rev. Camb. Philos. Soc. 85, 301–325.

Uher, J., 2011. Individual behavioral phenotypes: an integrative meta-theoretical framework. Why behavioral syndromes are not analogs of personality. Dev. Psychobiol. 53, 521–548.

Van Oers, K., Drent, P.J., de Goede, P., van Noordwijk, A.J., 2003. Realized heritability and repeatability of risk-taking behavior in relation to avian personalities. Proc. R. Soc. Lond. B 271, 65–73.

Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the ‘whole animal’: a free choice profiling approach. Anim. Behav. 62, 209–220.

Wilson, A.D.M., Godin, J.-G.J., 2009. Boldness and behavioral syndromes in the bluegill sunfish, *Lepomis macrochirus*. Behav. Ecol. 20, 231–237.

Wilson, D.S., Coleman, K., Clark, A.B., Biederman, L., 1993. Shy bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*)—an ecological study of a psychological trait. J. Comp. Psychol. 107, 250–260.

Wilson, D.S., Clark, A.B., Coleman, K., Dearstyne, T., 1994. Shyness and boldness in humans and other animals. Trends Ecol. Evol. 9, 442–446.

Wolf, M., Weissing, F.J., 2010. An explanatory framework for adaptive personality differences. Philos. Trans. R. Soc. B 365, 3959–3968.

Yoshida, M., Nagamine, M., Uematsu, K., 2005. Comparison of behavioral responses to a novel environment between three teleosts, bluegill *Lepomis macrochirus*, crucian carp *Carassius langsdorfil*, and goldfish *Carassius auratus*. Fish. Sci. 71, 314–319.