The evolution of animal “intelligence”: Among-individual differences and the heritable basis of cognitive and personality (co)variation in the Trinidadian guppy (*Poecilia reticulata*)

Submitted by Pamela M Prentice to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences May 2021.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……

i. Abstract

Among-individual variation in cognition is common within populations, and has been demonstrated across a range of animal taxa. From an evolutionary perspective, this variation is a pre-requisite for natural selection and genetic variation – both of which are required for adaptive evolution to occur. Selection has been hypothesised to favour high cognitive performance, however directional selection would be expected to erode genetic (and among-individual) variation over time. Furthermore, as selection does not act on traits in isolation, understanding the extent to which cognitive traits covary with other aspects of phenotype (e.g. personality traits) is an important factor. The question of how among-individual variation is maintained is therefore central to our understanding of the adaptive evolution of cognition in the context of the wider phenotype. The overall aim of my PhD thesis was to study the evolutionary biology of among-individual variation in cognitive and personality traits, and to explore the relationship between them in the Trinidadian guppy (*Poecilia reticulata*). I aimed to characterise among-individual differences in cognitive performance and personality, and investigate the extent to which genetic variation contributes to these, and to relationships between them. Each chapter was intended to obtain novel insights into the mechanisms explaining the existence and maintenance of these two important facets of behaviour across multiple hierarchical levels. I advocate the use of quantitative genetic style modelling approaches throughout, and seek to highlight the value of multivariate approaches to investigating animal cognition and associated behavioural traits. Firstly, relationships between cognition and personality were explored at the among-individual level, using a measure of cognitive performance in a spatial learning task and a stress-related behavioural trait. Secondly, to further scrutinise links between cognition and

personality, the multivariate structure of among-individual variation in cognitive performance across different domains was investigated in addition to variation in personality trait ‘boldness’. Next, among-individual and genetic variation in phenotypic ‘predictability’ (within-individual variation) of a stress-related behavioural trait was analysed using a novel form of ‘double hierarchical’ model. I then explored whether genetic variation contributes to among-individual differences in cognitive performance in a detour reaching task, and further investigated whether an interplay by genotype-by-environment interactions contributed to this variation. And finally, I discuss these results and how they contribute to our understanding of the causes of among-individual variation in cognitive performance, in addition to their evolutionary implications and ideas for future work.

ii. Acknowledgements

First and foremost, I would like to thank my supervisor, Prof Alastair Wilson. His overwhelming expertise, support, positivity and guidance has been a valued source of direction and encouragement, for which I will be eternally grateful. Thanks also to past and present members of the Wilson research group, in addition to the extended 'Quant Gen' family. Many thanks to all the staff at the Penryn Fish Lab – especially for their support during the pandemic – it's been a pleasure to work with such an awesome team. I'd also like to thank Dr Alex Thornton, Dr Erik Postma, Dr Julian Martin and Prof Niclas Kolm for their wisdom, advice and overall stimulating conversations. Lastly, my thanks to my family and friends who have supported me throughout. You know who you are, but maybe not quite how much I have appreciated your love and support.

Table of Contents

iii. List of Tables

iv. List of Figures

vi. Author's Declaration

The work described in this thesis was carried out by P Prentice, unless otherwise stated below.

Chapter 1 was written by P Prentice.

Chapter 2 was based on data collected by MSc student C Mnatzaganian at the University of Exeter, Penryn campus. The data was analysed and manuscript written by P Prentice, with input and support from A Wilson and A Thornton.

Chapter 3 and 5 were based on work designed by P Prentice and A Wilson. Data were collected by P Prentice at the University of Exeter, Penryn campus. Technical staff assisted with pedigree husbandry in **Chapter 5.** The data were analysed and manuscripts written by P Prentice, assisted by A Wilson, and advised by A Thornton and N Kolm (**Chapter 5 only)**.

Chapter 4 was based on work designed by P Prentice, T Houslay and A Wilson. Data was collected by P Prentice and T Houslay at the University of Exeter, Penryn campus, and analysed by P Prentice (assisted by A Wilson), and J Martin. The manuscript was written by P Prentice, assisted by A Wilson, T Houslay and J Martin. **Chapter 4** has undergone peer review.

Chapter 6 was written by P Prentice.

1. Chapter 1: General Introduction

This thesis aims to investigate the causes of variation in cognitive and personality traits in *Poecilia reticulata*, and to explore the relationship between these two important facets of behaviour. I characterise among-individual differences in cognitive performance and personality and investigate the extent to which genetic variation contributes to these and to relationships between them. In so doing I test hypotheses about the mechanisms driving and maintaining behavioural variation at multiple hierarchical levels. Throughout the thesis, I advocate the use of quantitative genetic style modelling approaches, and seek to highlight the value of multivariate approaches to investigating animal cognition and associated behavioural traits.

Cognition in animals

In the broadest sense, cognition refers to adaptive information processing, and so encompasses the set of mechanisms by which animals acquire, process, store and respond to information from the environment (Shettleworth, 2009a). Environmental information is acquired through the senses (e.g., vision, hearing, smell, taste, touch, electroreception) and processed by the brain (Dukas, 2004). Functionally appropriate behaviours are then the result of an animal's decision to act on the information received by the brain (Shettleworth, 2010a; Bräuer et al., 2020). As such, cognitive processes play a major role in driving behaviour and are therefore vital for carrying out the day-to-day activities required for survival and reproduction (Boogert et al., 2011; Maille et al., 2016; Shohet et al., 2009; Smith et al., 2015).

Cognition is often described as including multiple aspects or 'domains' such as memory, attention, learning, behavioural inhibition and self-recognition (to name a few), all of which can themselves be further subdivided (Roitblat, 2008; Shettleworth, 2010a). Research on animals has consequently examined many distinct aspects of cognitive ability across species, for example tool-use in New Caledonian crows (*Corvus moneduloides* (Kenward et al., 2006)), future planning in western scrub-jays (*Aphelocoma californica* (Raby et al., 2007)), numerosity in mosquitofish (*Gambusia holbrooki* (Agrillo et al., 2011)), and social learning in red-footed tortoise (*Geocgelone carbonaria* (Wilkinson et al., 2010)). Furthermore, great progress has been made by comparative psychologists in elucidating the cognitive mechanisms and neural structures underpinning animal behaviour (e.g. by investigating brain size as a predictor of cognitive performance (Kotrschal et al., 2013), or hippocampus volume in food-hoarding animals; (Roth et al., 2010)). Cognitive ability has long been known to vary among species, but it is also becoming clear that some organisms are capable of performing more sophisticated cognitive tasks than was previously thought possible. The cuttlefish (*Sepia officinalis*), for example, is a marine invertebrate that shows episodic-like memory in a foraging task (Jozet-Alves et al., 2013) and displays future planning behaviour when foraging for previously experienced food items (Billard et al., 2020). These abilities are in fact broadly similar to those seen in vertebrate lineages with large complex brains, including birds (Clayton et al., 1998), apes (Martin-Ordas et al., 2010) and even human children (Russell et al., 2011).

Differences in evolutionary history and brain structure suggest that cognitive abilities may have emerged multiple times independently (Roth, 2015). Psychologists have postulated several major adaptive hypotheses to explain both the emergence of cognition, and the presence of among-species differences in

ability. These include the “Ecological Intelligence Hypothesis”, which proposes that complex cognition evolved to meet the challenges associated with finding and processing food (Gibson, 1986; Sayol et al., 2016); the “Social Brain Hypothesis”, which predicts cognitive ability is driven by the demands of living in social groups (Dunbar, 1998; Ashton et al., 2018). Others have postulated cognitive abilities are primarily an adaptation to cope with the challenges of predators (Byrne et al., 2007; Skelhorn et al., 2015; Kotrschal et al., 2017). Support for all these hypotheses, and several others, is claimed across species, but it is important to note they need not be mutually exclusive.

Historically, cognitive research was mainly focused on human intelligence and, as a consequence mammalian species (e.g. mice) used as biomedical models have also been important in experimental work (for a review, see McGonigle et al., 2014; Sternberg, 2002). In this human context, there has been longstanding interest in detecting and understanding differences in cognitive performance among humans (Hunt et al., 1973; Daneman et al., 1980; Deary et al., 2009). In contrast, until recently comparative work on non-human animals has largely addressed differences in cognitive abilities between species and, in some cases, among populations facing distinct environmental challenges (ecological and/or social). In these contexts, behavioural means were compared to address specific questions, while differences among-individuals within groups (e.g. species) were largely ignored, dismissed as noise, or attributed to experimental design inconsistencies. However, in line with the growth of animal personality research (discussed below), studies of cognition in non-human animals have increasingly viewed among-individual variation in cognition as an interesting phenomenon in its own right. In terms of understanding the evolution of traits, this level of variation is incredibly interesting and important, as it is a pre-requisite for

natural selection and genetic variation – both of which are required for adaptive evolution to occur (Wilson et al. 2010).

Among-individual variation in cognition

Biomedical studies, motivated primarily by understanding intelligence in humans, revealed the presence of among-individual variation in cognitive abilities in rodent (Tolman, 1924; Tryon, 1940; Galsworthy et al., 2002; Matzel et al., 2003) and primate models (Banerjee et al., 2009). Rats, for example, show considerable individual differences in the ability to learn to solve mazes, and this variation is at least partially determined by genetic factors (Tolman, 1924; Tryon, 1940). Behavioural ecologists have now started to reveal the extent of this variation across a much wider range of non-model organisms (Sih, Bell, & Johnson, 2004; Sih et al., 2012; Thornton et al., 2014). Many empirical studies show that animal populations can harbour high levels of among-individual variation in cognitive performance (Boogert et al., 2018). This has been seen across a wide taxonomic groups including insects (Li et al., 2017), fish (Lucon-Xiccato & Bisazza, 2017a), birds (Quinn et al., 2016) and mammals (Mazza et al., 2019). For example, individual black-capped chickadees (*Poecile areicapillus*) vary in speed and performance accuracy in an acoustic discrimination task (Guillette et al., 2009, 2010, 2015). Performance in a foraging task varies among individual bank voles (*Myodes glareolus*; (Mazza et al., 2018)), and inhibitory control varies among individual guppies (*Poecillia reticulata*) when presented with a detour reaching task (Macario et al., 2021; **Chapter 5**). Consistent among-individual differences are also found across a range of cognitive domains, such as spatial memory

(Sonnenberg et al., 2019), association learning (Kniel et al., 2020), inhibitory control (Brandão et al., 2019) and problem solving (van Horik et al., 2019).

Efforts to quantify patterns of among-individual variation in cognitive traits are still in relative infancy (Rowe et al., 2014; Thornton et al., 2014; Boogert et al., 2018), and empirical studies are somewhat limited (but see Niemelä et al., 2013; Lucon-Xiccato, 2017a; Ashton et al., 2018; Prentice et al., 2020; **Chapter 2, 3 & 5** for examples). Moreover, progress to date has been limited by experimental and analytical paradigms common to comparative psychology (Rowe et al., 2014; Thornton et al., 2014). Specifically, robustly quantifying (as opposed to simply detecting) among-individual differences requires data structures not readily obtained using classic cognitive experimental paradigms used to estimate phenotypic means in small samples (e.g., tens of individuals and sometimes less). It requires high volume data sets with repeated measures on individuals. These are logistically difficult to collect, and require carefully designed psychometric tests that can be adapted to high throughput phenotyping of cognitive differences across multiple domains (Thornton et al., 2014). Furthermore, there is always a risk that apparent among-individual variation in cognitive performance actually arises from other factors. These can include differences in internal state (e.g. motivation due to satiety, stress, and hormonal or circadian cycles), imperfectly controlled aspects of the external environment (Dohm, 2002), or unknown prior experience (i.e., repeated exposure to cognitive challenges will alter performance via processes such as learning and memory (Griffin et al., 2015)).

While the empirical challenges are not trivial, they are also not unique to the study of cognitive variation. Indeed, many of the same issues have been recognized in the broader field of animal personality (Martin et al., 2008).

Overcoming them requires a continued focus on rigorous experimental design, but this must be coupled with i) a recognition that high throughput data collection is essential and ii) use of analytical methods that are flexible enough to allow (potentially) confounding sources of variation to be jointly modelled. Multivariate linear mixed effects models, developed in evolutionary quantitative genetics and increasingly adopted in behavioural ecology offer a useful strategy in this regard.

Animal personality: what is it and what maintains variation?

Arguably, the recent focus on among-individual differences in cognitive performance owes much to the burgeoning study of 'animal personality'. It is now abundantly clear that species typically harbour high levels of consistent among-individual variation in behavioural traits generally (Dingemanse et al., 2005). Individual differences in (mean) behaviours, commonly referred to as personality, can manifest as, for instance, variation in aggressiveness (or sociability) towards conspecifics, or as differences in response to perceived risk (Réale et al., 2007; Bridger et al., 2015). Evidence from a wide variety of species shows that some individuals within populations are consistently more aggressive, more exploratory, or more 'bold' than others. For example, individual deer mice (*Peromyscus maniculatus*) consistently vary in escalation of aggressive behaviour exhibited when presented with a conspecific rival (Wilson et al., 2009). In this study, differences in aggression were shown to be partly genetic and therefore heritable, a finding that is quite typical of personality traits generally (Dochtermann et al., 2019). For example, risk taking behaviour is heritable in hand reared great tits (*Parsus major;* (Drent et al., 2003)), as is 'shy-bold' type variation characteristic of behavioural stress responses in brown trout (*Salmo*

trutta; (Kortet et al., 2014)) and guppies (*P. reticulata;* (Prentice et al., 2020)). If personality traits are typically heritable, then we also know that they can have fitness consequences and so be under selection (Dingemanse, 2005; Smith et al., 2008). For example, individual wild African striped mice (*Rhabdomys pumilio*) that were quicker to respond to a predator stimuli had increased rates of survival (Maille et al., 2016). The combination of heritable variation and selection sets the stage for adaptive evolution, but also raises the question of what maintains this among-individual variation in behaviour within populations**.** Increasingly, individual personalities are being viewed as components of 'extended' life history strategies, that are expected to be correlated with – and trade-off against - other aspects of the phenotype. Since, strong directional or stabilising selection is usually predicted to erode variation (Roff, 2002), it is widely hypothesised that variation in personality traits is maintained by fitness trade-offs of some kind (Dingemanse et al., 2004; Quinn et al., 2016). Trade-offs may be with other aspects of personality, or with cognition, or perhaps with physiological and/or reproductive traits (Réale et al. 2010; Sih, Bell, & Johnson 2004; Wolf et al. 2007). For example, bolder individuals may be better at acquiring resources to invest in life history traits (e.g., growth, reproduction) but their behaviour may expose them to greater predation risk. If being bold (or shy) has both advantages and costs in this way, selection may maintain variation in a population.

By focussing on among-individual differences the field of animal personality has done much to align the study of behaviour with concepts and methods widely used to investigate the evolution of phenotypic traits more generally (e.g. quantitative genetics; see below). However, the field remains fraught with (largely) semantic arguments about what constitutes personality in general, or how particular aspects of repeatable behaviour should be referred to

(e.g. as boldness, exploration, sociability). While this thesis seeks to avoid such arguments as far as possible, it is necessary to introduce the related concept of 'stress coping style' (SCS). This is a verbal model derived from stress research which predicts that, when challenged by acute stressors in the environment, individuals will vary in response along a proactive/reactive continuum of variation (Coppens et al., 2010; Koolhaas et al., 1999; Sih, Bell, & Johnson 2004). Proactive individuals will tend to express more 'fight or flight' type behavioural responses induced by adrenaline-response to stressors. At the other extreme, reactive coping styles are more behaviourally 'passive' (e.g., freezing or hiding) and show high HPA(I) activity leading to cortisol response (Øverli et al., 2007; Carere et al., 2014). Thus, the model predicts consistent variation in behavioural response to stressors (e.g., isolation in a novel environment (White et al., 2016; Prentice, 2020, 2020) that is broadly analogous to shy-bold type personality variation. Note however that the SCS model explicitly predicts integration of these personality differences with among-individual differences in stress physiology.

Are cognition and personality linked?

Links between cognitive variation and personality are widely hypothesised perhaps in part because trade-offs are so frequently invoked as parallel explanations for both phenomena (for this argument applied to cognitive variation see e.g., Del Giudice et al., 2018). However, there is also empirical evidence of correlation structure between them. For example, great tits (*Parsus major*) that successfully learned to access food during a problem solving task laid more eggs than non-solvers and foraged less for their offspring (Cole et al., 2012). However, problem solvers (with putatively higher cognitive performance) were also less

competitive during foraging and more detrimentally, likely to desert their nest leaving offspring vulnerable to predation and starvation. If fitness trade-offs do shape variation in personality, it seems very possible that they may also drive (co)variation between behavioural and cognitive styles. For example, given the widely postulated cognitive trade-off between speed and accuracy of decision making (Chittka et al., 2009), accurate learners that pay careful attention to stimuli before taking action may generate behavioural profiles characteristic of shy personality types. Conversely, cognitive strategies favouring speed over accuracy may generate bolder, less neophobic and/or more exploratory personalities (Sih et al., 2012).

Consideration of stress responses provides another perspective that has led to hypothesised links between personality and cognitive performance (Raoult et al., 2017; Gibelli et al., 2019). For instance, 'proactive' styles are largely analogous to bold, exploratory and/or risk-taking personalities that, intuitively, may present individuals with more opportunities to learn initially. Conversely, it has been argued that behavioural flexibility, which is associated with 'reactive' stress coping styles (Coppens et al., 2010), is important for cognitive tasks such as reversal learning that require updating information about the environment as conditions change (Koolhaas et al., 1999; Sih et al., 2012; Griffin et al., 2015). More simply, sensitivity to external stressors could also be a source of bias in experiments designed to assay variation in cognitive performance. For instance if more stressed individuals are less motivated to engage in a task and/or are focused on perceived sources of risk, they may perform poorly in assays and so be viewed as having lower cognitive ability.

Variance partitioning and quantitative genetics

The empirical investigations of cognitive and personality variation - and covariation - described in this thesis rely on linear mixed effect models to statistically partition among-individual from within-individual variation. In some analyses, quantitative genetic “animal models” (Wilson et al., 2010) are also employed, in which among –individual variation is further partitioned to estimate genetic variance using a pedigree-based extension of the linear mixed model originally developed for animal breeders. Various other extensions to the basic mixed model are employed including ‘random regression’ (Martin et al., 2017) and ‘double hierarchical models’ (Lee et al., 2006). Detailed explanations of the modelling strategies are presented in the chapters where appropriate but here I give a simple overview of how and why variance partitioning is useful for studying cognitive variation and personality.

Mixed models allow partitioning of trait variation (measured as variance) into within- and among- individual components given data that contains repeated observations of traits on known individuals (Dingemanse et al., 2013; Wilson, 2018). Variance among-individuals necessarily implies some degree of behavioural consistency by those individuals, and is often described in a standardised way by the repeatability (R). This is estimated as an intraclass correlation (Hayes et al., 1997) and can be interpreted as the proportion of phenotypic variation that is due to differences between individuals. Formally, repeatability is

$$R = V_I / V_P \quad (1)$$

where V_I is the variance among individuals and V_P is the total phenotypic variance. The total phenotypic variance comprises

$$V_P = V_I + V_R \quad (2)$$

where V_R = the variance among observations within individuals (often referred to as residual variance). Behaviours having low within-individual variance relative to among-individual variance are more repeatable. In other words, when individuals behave consistently across observations (V_R is low) and when those consistent behaviours differ between individuals (V_I is high) then the behaviour will be highly repeatable. Estimates of repeatability are interesting in their own right, but are also a frequent first step towards studying the genetic basis of behavioural variation as they set an upper bound to heritability (Dohm, 2002). Where pedigree or relatedness data are also available, among-individual variance can be further decomposed to estimate behavioural heritability (h^2 ;Wilson et al., 2010). This is done by further partitioning the variance among individuals (V_I) into variance that is explained by shared genetic information, known as additive genetic variance (V_A), and variance that is due to non-genetic, permanent environment components (V_{PE}):

$$h^2 = V_A/V_P = V_A/ (V_A + V_{PE} + V_R) \quad (3)$$

In fact there are many specific statistical methods for estimating variance components and obtaining heritability estimates. However, unlike older methods such as parent-offspring regression, the 'animal model' (a form of linear mixed effect model) is able to (i) utilise observations made on individuals in any arbitrary pedigree structure, (ii) deal with unbalanced data sets and (iii) be readily extended to include statistical control for other non-genetic sources of similarity between relatives that could otherwise bias estimates of genetic variance (Lynch et al., 1996; Charmantier et al., 2013). Increasing application of animal model analyses in behavioural ecology has revealed low to moderate heritabilities are common for personality traits (Van Oers et al., 2005; Dingemanse et al., 2009; Niemelä et al., 2013; Prentice, Houslay, et al., 2020). However rather few studies

have so far estimated genetic contributions to among-individual variation in animal cognitive traits (but see Hopkins et al., 2014; Langley et al., 2020; Sorato et al., 2018; Vardi et al., 2020).

For a single trait, under (directional) selection, the rate at which the trait evolves is predicted as the product of h^2 and the selection differential S (which provides a measure of how strong selection is (Falconer et al., 1996)). However, as the above discussion of trade-offs makes clear, natural selection rarely operates on single traits in isolation (Roff, 2002). If it is the multivariate phenotype as a whole that determines fitness, and if traits under selection are correlated with each other, then the rate and direction of evolutionary change will depend on the genetic (co)variance matrix (**G**), which provides a multivariate analogue of the single-trait additive genetic variance V_A. For a multivariate phenotype, adaptive potential and evolutionary constraint arise from the alignment (or lack thereof) between **G** and the vector of selection **β** (Lande, 1979; Jones et al., 2004; Walsh et al., 2009), where **β** is the direction of change in multi-trait space that would maximise increase in mean fitness.

Considering a multivariate phenotype gives a much broader and realistic view of how traits are likely to respond to selection and may show ways through which genetic and phenotypic variation in behaviour can be maintained (Wolf et al., 2012). However estimating **G** is very challenging, especially for cognitive and personality traits. If data requirements are already high for quantifying among-individual variance in single cognitive traits, then they become exponentially more so when the target is a multivariate phenotype and among-individual variation is to be partitioned further. Moreover pedigree and/or relatedness information must be obtained meaning either use of breeding experiments or molecular/genomic pedigree analysis techniques (Wilson et al., 2010). A partial solution may lie in

Cheverud's conjecture that patterns of phenotypic variance and covariance could be suitable proxies for the underlying genetic architecture (Cheverud, 1988; Roff, 1996). Accepting this view uncritically implies that the (co)variance structure of the phenotypic matrix **P**, which can be readily estimated in a set of individuals observed once only for each of the target traits, can yield robust insights into the structure of **G**. Although Cheverud's conjecture is (arguably) implicit in most behavioural ecological studies, most quantitative geneticist would view this with scepticism for multivariate behavioural phenotypes characterised by high levels of within-individual variation. However, just as R sets an upper limit for h^2 in the univariate case, estimating the among-individual (co)variance matrix (**ID**) provides a useful intermediate step here (Brommer, 2013). Since genes contribute to among-individual differences, **G** is a component of **ID** and **ID** is a component of **P**. Thus, all else being equal the among-individual covariance matrix will be a better proxy of the genetic matrix than the total phenotypic matrix. (Dochtermann, 2011; Brommer, 2013)

Study species – The Trinidadian guppy (*Poecilia reticulata*)

In this thesis, I take a largely quantitative genetics approach to understanding variation in cognition and personality, relationships between them and genetic underpinnings. I use the Trinidadian guppy (*Poecilia reticulata*) as my model species throughout the thesis to explore these themes.

The guppy is a small, shoaling species from the family *Poeciliidae* that generally inhabits freshwater streams found along the coastal fringes of mainland South America (Magurran, 2005). *P reticulata* is sexually dimorphic for size and growth. Once mature, male growth plateaus with priority switching to

reproduction. Conversely, female guppies exhibit indeterminate growth after maturity in order to maximise fecundity. Males are brightly coloured relative to females, and male colouration is an important factor in mating success (Williams, 1967; Magurran et al., 1990; Nicoletto, 1993; Endler et al., 1995). This species has a promiscuous mating system where females mate with multiple males (Evans et al., 2001). The guppy, like most poeciliids is a livebearer, with mature males using a gonopodium (a modified anal fin) for insemination and internal fertilisation (Wourms, 1981). Females provision the eggs prior to fertilisation and retain them in the ovary cavity until the hatching and 'birth' of offspring (Magurran, 2005). Broods range in size from 1 to 25 fry, with the average brood at around 15 fry. Once released, fry are fully independent and capable of feeding with no active parental care exhibited by either parent.

Guppies live in shoals primarily to reduce predation risk (and potentially increasing foraging efficiency). There is a high frequency of fission-fusion events, with males being the more mobile sex (Croft, et al., 2003), resulting in a dynamic social environments. Males maximise fitness by moving between multiple shoals of females, increasing potential mating encounters (Griffiths et al., 1998; Kelley et al., 1999; Croft, Albanese, et al., 2003; Croft, Arrowsmith, et al., 2003). In females, fitness depends on longevity and fecundity rather than mating opportunities (MaGurran et al., 1994). Therefore, females tend to exhibit stronger shoaling tendencies and higher shoal fidelity to reduce mortality from predation (Griffiths et al., 1998; Magurran et al., 2000; Magurran, 2005; Richards et al., 2010).

The guppy was chosen as a model system for several reasons. First, this species has recently become popular as a model for investigating animal cognition (Kotrschal et al., 2015; Lucon-Xiccato & Bisazza, 2016; Fong et al.,

2019). This means I have been able to build upon the work of others and take advantage of published cognitive testing assay designs. For instance, guppies, have recently been used to investigate learning colour discriminations (Trompf et al., 2014; Buechel et al., 2018), numerical discrimination (Kotrschal et al., 2013; Lucon-Xiccato & Bisazza, 2017b), reversal learning (Buechel et al., 2018), spatial learning (Lucon-Xiccato & Bisazza, 2017c; Prentice, Mnatzaganian, et al., 2020) and inhibitory control (e.g. Lucon-Xiccato & Bisazza, 2016). Second, as an important model system in evolutionary ecology generally, methods for assaying among-individual 'personality' variation are also well established (Burns et al., 2008; White, 2016). One of the most common testing paradigms for quantifying personality is the open field trial (OFT), which provides a measure of exploratory behaviour and/or boldness in a novel, and somewhat risky environment (Burns et al., 2008). Risk is perceived as guppies are a shoaling species and isolation without cover leaves an individual open to increased predation risk.

In fact previous studies on the wild-derived captive population of guppies used throughout my thesis have already highlighted the utility of the OFT for assaying personality variation associated with the behavioural stress response (see e.g., Prentice, Houslay, et al., 2020; Prentice, Mnatzaganian, et al., 2020). Observed behaviours expressed in the OFT are both repeatable and plastic with respect to experimentally-manipulated stressor severity (specifically perceived predation risk) (Houslay et al., 2018). We also know from pedigree-based quantitative genetic studies that individual (mean) behaviours and their predictability (defined as within-individual variance) are heritable (White & Wilson, 2019; White, Houslay, et al., 2019; Prentice, 2020).

Finally, there are pragmatic reasons for the use of guppies as a model system. They are easily maintained in a laboratory environment and it is possible

to breed large numbers over a relatively short space of time. Females are able to reproduce monthly, and can store sperm from a single mating for many months, which makes then highly fecund. This coupled with the (relative) ease with which behavioural data can be collected means this species is an ideal model, not just for investigating the structure of cognitive and personality (co)variation among individuals, but also among genotypes using pedigree-based quantitative genetic methods.

Overview of thesis structure

This thesis aims to investigate the causes of variation in cognition and personality in *Poecilia reticulata*, and to determine whether there are relationship between these two aspects of phenotype. I will test hypotheses about the mechanisms that drive cognitive and behavioural variation at multiple hierarchical levels – among-individuals and among-genotypes. Throughout the thesis I advocate the use of multivariate modelling strategies for quantifying and testing hypotheses about both animal cognition and behaviour. These strategies are commonly used in quantitative genetics but are yet to be widely adopted by behavioural ecologists.

Following this general introduction (**Chapter 1**), **Chapters 2-5** describe a series of four empirical studies of guppy behaviour. In **Chapter 2**, I begin my research by investigating among-individual variation in spatial learning, a cognitive domain assayed using a maze test. I ask whether individuals differ in spatial learning ability, as measured by improvement in performance time across repeated maze trials, but also test whether this itself is repeatable at the individual level across two different maze layouts. I also investigate whether variation in

performance is correlated with, and so potentially explained by, personality variation in the form of behavioural stress response assayed using open field trials.

In **Chapter 3**, I continue to explore the structure of cognition and personality differences at the among-individual level. Here I seek to do this in a more fully multivariate way, by estimating the **ID** matrix among traits used as proxies of cognitive performance across three domains; *association learning* in a colour discrimination task; *motor cognition* in a novel motor task and *cognitive flexibility* in a reversal learning task. I also include a measure of the personality trait of 'boldness'. Broadly the aim is to test hypothesised relationships of cognitive performance across multiple domains, as well as to further scrutinise the links between cognition and personality. More specifically, I ask whether there is evidence of performance trade-offs among cognitive domains, or whether the structure of **ID** is more consistent with variation in an overall, domain-general cognitive performance trait ('general intelligence'; Burkart et al., 2017; Galsworthy et al., 2005; Plomin & Spinath, 2002). Relatively few studies have characterised the among-individual covariance structure between cognitive domains, and support for an underlying general intelligence factor among non-human animals remains limited (but see Galsworthy et al., 2005; Hopkins et al., 2014; Arden et al., 2016).

While **Chapter 2** and **3** characterise among-individual variation only, in **Chapters 4** and **5**, I build on this to investigate the extent to which behavioural differences between fish may be explained by genetic factors. **Chapter 4** does not directly address cognitive traits, but builds on previous findings showing heritable personality variation in (mean) behavioural stress-response in the same population of guppies (White, 2016; Houslay et al., 2019). In particular, using a

novel form of ‘double hierarchical’ model I test for among-individual and genetic variation in not just average (or individual) behaviour, but also in within-individual variation (otherwise known trait ‘predictability’). This utilised a set of fish produced from known crosses (such that pedigree was known) and subjected individuals to repeated OFT assays. The idea of investigating phenotypic ‘predictability’ is relatively new but has attracted increasing attention because, for instance two individuals may be equally ‘bold’ on average (as determined from the individual mean behaviour across multiple observations) but differ in how much variation they exhibit around their individual-level mean phenotypes. Emerging evidence from livestock genetics suggests differences in phenotypic predictability can be due to genetic factors. The evolutionary implications of this can be viewed from two perspectives. Firstly, if predictability is genetically variable, then selection on within-individual variation (e.g. as posited by ‘bet-hedging’ type models) could produce an evolutionary response. Secondly, genetic variance in predictability can also be understood as a form of cryptic genotype-by-environment interaction (GxE). In turn the presence of GxE has two biological interpretations that are equivalent (but not obviously so). GxE means that genetic variance for the behavioural phenotype is environmentally sensitive, but also that behavioural plasticity is heritable (Nussey et al., 2007).

In **Chapter 5**, I return to the central theme of cognitive variation in a study of *inhibitory control*. In this chapter, I repeatedly expose individuals to a ‘detour task’ to test for genetic variance in inhibitory control. I also test for genotype-by-environment interactions (GxE) by testing related fish under alternative experimental treatments that differ in degree of available visual information (using transparent vs semi-transparent barriers in the detour task). The fact that cognition is defined in relation to acquiring, processing and using information in

the environment makes the possibility of GxE interactions very plausible and intuitive. However so far, few animal cognition studies have characterised the importance of GxE, and I aim to address this gap in our knowledge here. I then further seek to validate the assumption that the detour task measures variation in cognitive processes (specifically inhibitory control) that is distinct from among-individual differences in behaviour (be that cognitive or non-cognitive behaviours) expressed during a period of training trials. In this study I used a set of offspring fish produced by setting up small breeding groups with several males and females in each. In order to apply quantitative genetic models it was therefore necessary to first resolve the pedigree structure. This was done using microsatellite genotyping coupled to molecular pedigree analysis.

Finally, in **Chapter 6** I present a short general discussion in which I summarise the main findings from each empirical chapter and end with some final thoughts on improvements and directions for future research in the field of cognitive evolution.

2 Chapter 2: Cognitive performance and stress responsiveness in the Trinidadian Guppy – a multivariate approach

This paper is published as: **Prentice, P.M**., Mnatzaganian, C., Houslay, T.M., Thornton, A. and Wilson, A.J., 2020. Cognitive performance and stress responsiveness in the Trinidadian Guppy – a multivariate approach. *Biorxiv.Org*.

2.1 Abstract

Among-individual variation in cognitive performance has been recently demonstrated across a range of animal taxa. While this variation is a prerequisite for contemporary natural selection, it is also true that selection does not act on traits in isolation. Thus, the extent to which cognitive traits covary with other aspects of phenotype (e.g. personality traits) is expected to be an important factor in shaping evolutionary dynamics. Here we adopt a multivariate approach to test for spatial learning ability in a captive population of male Trinidadian guppies (*Poecilia reticulata*), and ask whether differences in cognitive performance are associated with (repeatable) differences in stress response behaviour. We focus on stress response for two reasons. First, functional links between cognitive traits and 'stress coping style' have been hypothesised. Second, individual-level studies of cognitive performance typically rely on multiple testing paradigms that may themselves be a stressor. Thus, there is a risk that variation in stress responsiveness is itself a cause of apparent, but artefactual variance in cognitive ability. Using a set of fish exposed repeatedly to two distinct spatial learning tasks (maze layouts), and an acute stress response test (open field trial), we find differences among-individuals in task performance that are repeatable within- and across maze layouts. On average performance improves with experience in

the first maze, consistent with spatial learning, but not the second. In both mazes there is among-individual variation in the trajectory of mean performance with trial number suggesting individuals differing in 'learning rate'. Acute stress response behaviour is repeatable but predicts neither average time to solve the maze nor learning rate. We thus find no support for among-individual correlation between acute stress response and cognitive performance. However, we highlight the possibility that cumulative, chronic stress effects may nonetheless cause observed declines in performance across repeats for some individuals (leading to lack of improvement in mean time to solve the second maze). If so, this may represent a pervasive but difficult challenge for our ability to robustly estimate learning rates in studies of animal cognition.

2.2 Introduction

Cognition is defined as the set of mechanisms by which animals acquire, process, store and use information from the environment (Healy et al., 2010; Shettleworth, 2010b), and is vital for carrying out day-to-day behaviours needed for survival and reproduction. While differences in cognitive performance among-species have long been studied in comparative psychology (for a review see Healy 2019), a more recent focus in behavioural ecology has been the characterisation of among-individual variation within populations of non-human animals (Lucon-Xiccato & Bisazza, 2017a; Ashton et al., 2018; Boogert et al., 2018). This among-individual variation is interesting from an evolutionary perspective, as it is a pre-requisite for natural selection and genetic variation – both of which are fundamental for adaptive evolution to occur (Wilson et al. 2010). However, selection does not act on traits in isolation. Functional links between variation in

cognitive performance and other aspects of behaviour (including, for example neophobia, boldness and stress responsiveness) have been hypothesised (Griffin et al. 2015; Medina-García et al. 2017; Quinn et al. 2012; Sweis et al. 2013). Robustly testing these relationships is often challenging, requiring multivariate data collection and analyses to detect and describe patterns of variation between associated traits at the appropriate level (e.g., among-individual and/or among genotype; Dingemanse & Dochtermann, 2013). Nonetheless, such efforts are important if we hope to understand the adaptive evolution of cognition in the context of the wider phenotype (Thornton et al., 2015). Here we address this broad goal in the more specific context of testing hypothesised links between cognitive performance and a stress-response (Øverli et al., 2007; Gibelli et al., 2019) in Trinidadian guppies (*Poecilia reticulata*).

Quantifying patterns of among-individual variation in cognitive traits is still in its infancy (Rowe et al., 2014; Thornton et al., 2014; Boogert et al., 2018), and empirical studies therefore remain somewhat limited (but see Ashton et al. 2018; Tyrone Lucon-Xiccato & Bisazza 2017; Niemelä et al. 2013 for examples). However it is now abundantly clear that populations typically harbour high levels of among-individual variation in behavioural traits more generally (Dingemanse et al., 2005). Individual differences in (mean) behaviours, commonly referred to as personality, can manifest as, for instance, variation in aggressiveness or sociability towards conspecifics, or differences in response when faced with predators or other sources of perceived risk (Bridger et al. 2015; Réale et al. 2007). Since, strong directional or stabilising selection is usually predicted to erode variation (Roff, 2002), it is widely hypothesised that variation in personality traits is maintained by fitness trade-offs of some kind (Dingemanse et al., 2004; Quinn et al., 2016), among other mechanisms. For example, bolder individuals

may be better at acquiring resources to invest in life history traits (e.g., growth, reproduction) but their behaviour may also expose them to greater predation risk. In this way personalities can themselves be viewed as components of life history strategies, leading to an expectation that they will be correlated with – and trade-off against - other aspects of physiological, reproductive, and behavioural phenotype (Réale et al. 2010; Sih, Bell, & Johnson 2004; Wolf et al. 2007). Certainly, arguments that trade-offs can maintain variation in cognitive performance parallel explanations made for widespread presence of personality. These could be trade-offs among cognitive domains, or between, for instance an overall cognitive performance trait ('general intelligence' (Plomin et al., 2002; Galsworthy et al., 2005; Burkart et al., 2017)) or other aspects of phenotype.

Variation in an animal's stress physiology may provide one putative source of among-individual differences in both personality traits and cognitive performance (Raoult et al., 2017; Gibelli et al., 2019). The widely used concept of stress coping style model predicts that individuals will vary - both behaviourally and physiologically- along a proactive/ reactive continuum (Coppens et al., 2010; Koolhaas et al., 1999; Sih, Bell, & Johnson 2004). As originally posited, the model predicts proactive coping styles will express more 'fight or flight' type behavioural responses induced by adrenaline-response to stressors. At the other extreme, reactive coping styles will be more behaviourally 'passive' (e.g., freezing or hiding) and show high HPA(I) activity leading to cortisol response (Øverli et al., 2007; Carere et al., 2014). Various links to cognitive performance variation have been suggested. For instance, proactive styles are broadly thought to be associated with 'bold', exploratory, risk-taking personalities that may present with more opportunities to learn initially. Conversely, greater behavioural flexibility associated with reactive coping styles (Coppens et al., 2010) may be important

for tasks such as reversal learning, that require an ability to acquire (and use) new information under changing environmental conditions (Koolhaas et al., 1999; Sih et al., 2012; Griffin et al., 2015). More generally, sensitivity to external stressors or challenges could impact performance in cognitive assays if more stressed individuals are simply more or less motivated and/or are focused on sources of risk rather than environmental cues of rewards.

Although hypothesised links between stress responsiveness (or coping style) and cognitive performance seem intuitive, empirical evidence is still limited to a small number of studies (Lukowiak et al., 2014; Mesquita et al., 2015; Bebus et al., 2016; Bensky et al., 2017; Brust et al., 2017; Mazza et al., 2018). There are also contrasting studies in which either a weak or no relationship was detected (Cole et al., 2011; Carazo et al., 2014; Guillette et al., 2015). It is also possible that relationships are variable across different aspects of cognition. For instance in sailfin mollies (*Poecilia latipinna*), individual fish displaying less thigmotaxic behaviour (an anxiety related behaviour in fish) performed better in a discrimination learning task than highly anxious individuals, whereas the opposite was found in a reversal learning task (Gibelli et al., 2019). Clearly, there is need for more empirical work before a clear picture of the complex relationship between variation in cognitive performance and stress responsiveness/coping style is understood. Here we address this broad goal by testing the hypothesis that individual differences in cognitive performance and stress responsiveness are correlated in male Trinidadian guppies (*Poecilia reticulata*).

The guppy is a freshwater poeciliid fish that is widely used as a model in behavioural and evolutionary ecology. Methods for assaying among-individual 'personality' variation are well established in this species generally (Burns et al., 2008; S. J. White et al., 2016), while guppies have been used in cognitive studies

that target learning colour discriminations (Trompf et al., 2014; Buechel et al., 2018), numerical discriminations (Kotrschal et al., 2013; Lucon-Xiccato & Bisazza, 2017b), reversal learning (Buechel 2018), spatial learning (Lucon-Xiccato & Bisazza, 2017c; Prentice, Mnatzaganian, et al., 2020) and inhibitory control (e.g. Lucon-Xiccato & Bisazza, 2016). Here, we investigate the relationship between behavioural stress response and performance in a spatial learning task in which male guppies repeatedly navigated a maze to access females as a reward. The cognitive task was repeated using a second, differently structured maze in order that we could assess not just variation in learning within a single spatial context, but also ask whether – for instance – individuals displaying greater performance in trials using the first maze subsequently also performed better in the second. In the wild, male guppies usually utilize large home ranges during mate search and foraging (Croft et al. 2003), and as such spatial learning is expected to be an ecologically relevant trait (Brown et al., 2005). For our measure of stress responsiveness, we utilise 'Open Field Trials' (OFT). Widely used across species as a paradigm for characterising behavioural differences related to exploration, activity, and 'shy-bold' type variation (Gosling, 2001; Bell et al., 2009), previous studies on this captive population of guppies have highlighted its utility for assaying behavioural stress response (see e.g., Prentice et al 2020). Observed behaviours expressed in the OFT are both repeatable and plastic with respect to experimentally-manipulated stressor severity (specifically perceived predation risk) (Houslay et al., 2018). We also know from pedigree-based quantitative genetic studies that individual (mean) behaviours and their predictability (defined as within-individual variance) are heritable (White, 2019, 2019; Prentice, 2020). Furthermore, there is some evidence of genetic integration between OFT behaviour and cortisol expression,

strengthening the view that the OFT provides an appropriate assay of behavioural stress response; (Houslay et al., 2019).

In what follows we: i) test for evidence of learning in naïve guppies repeatedly exposed to a spatial learning task (maze), ii) ask whether individuals differ in cognitive performance across repeated trials and if so; iii) whether performance in the first maze predicts performance in a second spatial context (i.e. reconfigured maze). We predict that time to complete the mazes (our proxy of cognitive performance) will, on average, improve with experience consistent with spatial "learning", but that individuals will consistently differ in cognitive performance within each maze. We also predict that individual performance in the first maze will be positively correlated with performance in the second, consistent with stable differences in cognitive ability, although we acknowledge proactive interference (difficulty inhibiting memory or previously learnt associations; Shettleworth, 2009a) may affect performance in the second maze. Finally, iv) we test the hypothesis that individual differences in cognitive performance will be associated with differences in stress responsiveness. Although empirical evidence suggests potential relationships in both directions between stress responsiveness and cognitive performance, withthe current absence of specific models, we make no *a priori* predictions about the sign of the relationship here.

2.3 Methods

Study site and housing

All behavioural assays were carried out on guppies from a captive population (derived from wild fish collected in the Aripo River, Trinidad in 2008) housed at

the University of Exeter's Penryn campus. Adult males (n = 64) were randomly sampled from the stock population, and housed in groups of 8 in separate home tanks (15 l, 18.5 × 37 × 22 cm) maintained at 23–24°C on a 12:12 light/dark cycle. The tanks shared a recirculating sump water supply which underwent a 25% water change once per week. All fish were fed to satiation twice daily on commercial flake food and live brine shrimp (*Artemia salina*) to control as much as possible for energetic and nutritional states prior to testing. We elected to focus on males only for several reasons. First, pilot studies showed a high occurrence of 'freezing' behaviour in females (relative to males) when introduced to the maze. While freezing can be a component of the behavioural stress response (Houslay et al., 2018), we considered that frequent occurrence during the cognitive assay would complicate data interpretation. Second, males show consistent sexual reproductive motivation towards females (Burns et al., 2008), enabling the use of females as a 'reward' for males solving the maze (Kotrschal et al., 2015). Third, male guppies exhibit distinctive markings and colouration on body and fins. By recording and sketching these for each fish we were able identify individuals within groups without the need to subject individuals to invasive tagging.

Ethics

This work was conducted under the auspices of the Animals (Scientific Procedures Act) out with approval of the University of Exeter research ethics committee, under licence from the Home Office (UK) (Licence Number PPL30/3256). Experimental procedures and behavioural assays were developed in accordance with the principles of the three Rs and ASAB guidelines (Buchanan et al., 2020) for use of animals. All periods of handling and emersion were kept

to a minimum and only fish deemed healthy and exhibiting normal behaviour were used in trials. At the end of the experiment, fish were returned to a designated 'retirement' tank (containing females as well as males) and not used in any further experiments.

Overview of behavioural testing scheme

We used a repeated measures approach to test for among-individual (co)variation in spatial learning performance and stress responsiveness. Spatial learning was first assessed by repeatedly trialling individuals in a maze apparatus (Maze A, Figure 2.1). Each individual fish was tested once per day for 11 consecutive days with reduction in time to complete the maze interpreted as 'learning'. This is consistent with previous studies using either time to complete an objective or to perform a particular task to investigate variation in cognitive performance among-individuals (Guillette et al., 2015; Lucon-Xiccato & Bisazza, 2016; Mazza et al., 2018; Zidar et al., 2018). We acknowledge that this interpretation strictly requires the implicit assumption that the contribution of any other factors to among-individual variation (e.g., motivation, energetic state, experience previous to the experiment; Rowe & Healy 2014) is negligible relative to differential cognitive performance. We attempted to mitigate against other sources of among-individual variation as far as possible using standardised housing and husbandry conditions. Following completion of spatial learning trials using Maze A, individuals were tested for stress responsiveness three times each over a three-week period using Open Field Trials (OFT) with a mean (range) of 4 (1-5) days between successive trials. Finally, fish were retested in a second maze (Maze B) with a different layout, and repeat trials conducted (as before) one per

day for 11 consecutive days. Thus, in total, the design called for all individuals to complete 22 spatial learning trials, 11 on each of two different maze layouts (distributed across two different mazes) and three OFT over a total testing period of 43 days. Note that the sample size declined slightly across the experiment as (i) a few mortalities occurred naturally within the testing period and, (ii) we proactively 'retired' any fish not deemed to be feeding well and behaving normally in their home tanks as a precaution against cumulative adverse effects. Thus 63 fish experienced Maze A, which declined to n=60 at trial 11 and OFT testing. Five fish were then removed prior to experiencing Maze B (n= 55 at trial 1 and n=53 at trial 11).

Spatial Learning Trials

In order to facilitate more rapid data collection, a single aquarium (25 x 45 x 25cm) was divided into two, with each half containing an identical version of maze A (A1, A2). Two replicates of maze B were similarly constructed (Figure 2.1). This allowed two fish to be tested concurrently during trials. Each maze consisted of 6 opaque Perspex panels (8 cm), spaced 5cm apart (Figure 2.1). A visually transparent perforated panel at one end of each maze was used to separate a small holding area (12.5 x 10 x 25 cm) contain two adult females selected randomly from stock. During trials the experimental maze tanks were lit from below by one fluorescent lamp and filled to a depth of 8 cm with room temperature water (approx. 23-24 °C). The water was taken from the same recirculating system used to house the male groups and was changed between each housing group (i.e. after every 4 runs with two fish trialled per run). Stimulus females were also changed at the same time.

At each trial, two males were individually netted from their home tank and quickly identified from natural markings. Each was randomly allocated to one of the two maze replicates and carefully placed within a perforated plastic tube in the 'start' zone (Figure 2.1). They were given 60 s to acclimate before the plastic tubes were removed. A Sunkwang C160 video camera mounted above the tank allowed the fish to be observed without disturbance. Tracking software (http://www.biobserve.com) was then used to determine the *start latency* as the (post-acclimation) time taken before a fish started the maze by leaving the 'start' zone, and *maze time* as the latency from starting to completing the maze (with completion defined as reaching the 'end' zone; Figure 2.1). On reaching the 'end' zone individuals were given 60 s undisturbed visual access to the females before an opaque plastic sheet was inserted to obstruct females from view. Following the 60 s reward period, fish were netted and returned to the home tank. To ensure standardized exposure to the reward stimulus, individuals that did not complete the maze within 480 s post-acclimation period (irrespective of whether they had started) were gently guided through the maze to the end zone using a net behind them and then experienced 60 s visual access to the females. Following the 60 s reward period, fish were netted and returned to the home tank. These fish were assigned a right censored value of 480 second for *maze time*.

Open Field Trial (OFT)

OFTs to characterise stress responsiveness closely followed the protocol described in White et al. (2016). For each trial, a single individual was netted from the home tank, quickly identified and introduced gently into the centre of an open arena (a 30 × 20 cm tank filled to 5 cm water placed on a lightbox). A cardboard

screen was around the tank prevented visual disturbance and a Sunkwang C160 video camera mounted above the arena again allowed movement to be tracked. Following a 30 s acclimation period, individuals' movements were tracked for 4 minutes and 30 s to determine *track length* (total distance swum (cm)) and *area covered* (percent of tank area covered). These two observed behaviours which are known to be repeatable and heritable in this population (Houslay, 2018; White, 2019, 2019), were used to calculate the derived trait of *relative area* following Houslay et al. (2019). *Relative area* is the observed area covered in the trial minus the expected area covered under a simulated 'random swim' of length equal to the observed track length (see Houslay et al. (2019) for further detail on simulations). Low values of *relative area* result from a 'flight type' behavioural stress response in which individuals swim rapidly (yielding a high track length) but exhibit thigmotaxis (staying close to the walls and seeking escape from the arena) and thus cover relatively little of the arena area. In contrast low values of *relative area* correspond to efficient exploration (i.e. a high proportion of the arena covered given distance swum), by putatively less stressed fish.

Statistical Analysis

Data from both types of behavioural assay were analysed using univariate and multivariate linear mixed effect models fitted by REML (restricted maximum likelihood) using ASReml within R (http://www.vsni.com) (Gilmour et al., 2009). By including individual identity as a random effect in these models we test for and characterise among-individual (co)variation. Traits were mean centred and scaled to standard deviation units to ease interpretation of results and facilitate

convergence of multivariate models. For *maze time* we did this using the overall mean and standard deviation of observations from both mazes in order to preserve any meaningful differences in performance between A and B. With traits in standard deviation units (sdu), estimates of among-individual variance (V_{ind}) can be interpreted as repeatabilities (i.e. proportion of the observed phenotypic variance explained by among-individual differences). However, we also calculate estimates of adjusted repeatability (R), the proportion of phenotypic variance explained by consistent among-individual differences, after controlling for fixed effects on the mean (Nakagawa et al., 2010). Thus $R=V_{ind}/(V_{ind}+V_R)$ where V_R is the residual (within-individual) variance estimated from each model. The significance of random effects was tested using likelihood ratio tests (LRT), while fixed effects (included in the various models as described below) were tested using conditional F-statistics. All models assumed Gaussian error structures, an assumption that was deemed acceptable based on visual inspection of the model residuals.

Univariate analyses of maze performance and spatial learning

We use *maze time* as our observed measure of performance. Here we describe in full the univariate analysis of data collected in maze A (subsequently *maze* $time_A$). Identical procedures were then applied to data from maze B. First, we visualised the distribution of *maze* $time_A$ across repeat using box plots and also plotted the proportion of mazes completed as a function of repeat to see if a pattern of increasing average performance (i.e. decreasing *maze time* and/or increasing proportion of successful completion) was immediately apparent. Next a series of three nested models with identical fixed effects but differing random effect structure were fitted to the centred and scaled *maze* $time_A$ data. All models

included a fixed effect of *trial number* (the cumulative number of trials experienced by an individual, treated as a continuous variable), allowing us to test for improvement in the mean (indicative of learning). Additional fixed effects were included as statistical controls for potential sources of variance not relevant to hypotheses being tested here. These included time of day (in minutes after 9 am), maze replicate (as a factor denoting position 1 or 2 in maze tank), and order caught from the home tank. The latter was to account for any cumulative disturbance effect of removing fish sequentially from the home tank and/or build-up of chemical cues in the maze between water changes.

The first model contained no random effects, while the second contained a random intercept of individual identity. Likelihood ratio test (LRT) comparison of these models was conducted to test the hypothesis that individuals differ in their average performance (*maze time*$_A$) across the 11 repeats, and we estimated the (adjusted) repeatability of performance under the second model. For the LRT we assume twice the difference in model log-likelihoods is distributed as a 50:50 mix of X^2_1 and X^2_0 following Stram & Lee (1994). The third model was a first order random regression (i.e. a random slope and intercept model) in which each individual's deviation from the fixed effect mean *maze time* can change as a linear function of *trial number* (1-11). Variation in random slopes means that there is among-individual variation around the mean *maze time*$_A$ - *trial number* relationship. Thus, LRT comparison of the second and third models thus provides a test for among-individual variation in learning rate. This comparison is conducted assuming the test statistics is distributed as X^2_2 since the third model has two extra parameters (a slope variance and a slope-intercept covariance). Note that among-individual variance in slopes cannot be scaled to a repeatability as within individual variance in slope is not estimable (using data from a single

maze; see below). Nor is its magnitude directly comparable to random intercept variance since slopes and intercepts are in different units. However, under the third model, among-individual variance in learning (slope) means that among-individual variance *maze time*$_A$ changes with *trial number* (Supplementary Info Figure S2.1, Appendix 2). Thus, to understand the biological effect size of estimated variance in slopes, we use the third model to predict among-individual variance (V_{ind}) and adjusted repeatability (R) of *maze time*$_A$ at both initial (trial 1) and final (trial 11) performance (following e.g., Nussey et al. (2007); see Supplemental Information Table S2.3 for didactic explanation and corresponding code). We note that among-individual variation at final performance has been used to infer differences in cognitive ability in studies adopting similar repeated measures designs (e.g. Langley et al. 2020) and so also has a useful biological interpretation here.

Univariate analysis of relative area

To verify our expectation that individuals would show consistent differences in stress responsiveness, we fit a simple random intercepts model to (scaled and centred) *relative area*. This model included fixed effects of trial number (1-3), and time of day (in minutes after 9 am in which each trial took place) as well as a random effect of individual identity. Adjusted repeatability (R) of *relative area* was calculated and the significance of among individual variance tested by LRT comparison to a simplified model with no random effect (assuming the test statistic was distributed as a 50:50 mix of X^2_1 and X^2_0 as above).

Multivariate modelling of Maze A, Maze B and OFT data combined

Finally, to test the predicted correlation structure between cognitive performance

and stress responsiveness, we formulated a trivariate mixed model in which the three response variables were *maze time*$_A$, *maze time*$_B$ and *relative area*. Fixed effects were exactly as described above on all three traits. Random effects were also as described above (i.e. individual level random intercepts and slopes for *maze time*$_A$ and *maze time*$_B$ but a random intercept only for *relative area*) but the multivariate formulation allowed us to estimate the full 5x5 among-individual covariance matrix (**ID**) among these effects. Since each observation of a fish provided data on a single trait only, residual covariances among traits were fixed to zero. After fitting the model, we compared it to a simplified fit in which all among-trait covariance elements in **ID** were constrained to zero. This provides a global test of individual covariance between traits. We then scaled estimated pairwise covariances in **ID** to their corresponding correlations for easier interpretation (noting for a pair of effects x,y the correlation $r_{xy} = COV_{xy}/(V_xV_y)^{0.5}$. This allowed us to scrutinise the correlation structure between stress responsiveness and cognitive performance in both mazes A and B, using both final performance and learning rate (i.e. random regression slope) as measures of cognition. Additionally, it allowed us to estimate the individual level correlation in cognitive performance measures (final *maze time* performance, learning) across mazes. These are not strictly equivalent to individual repeatabilities of cognitive performance measures across mazes (as opposed to individual repeatability of *maze time* across trials within mazes) because estimates could be negative. However, they can be readily interpreted in those terms; a strong positive correlation between, for example, individual *learning* in maze A and maze B means this latent variable is highly repeatable across mazes. Conversely, a negative correlation means that individuals learning faster in maze A tend to learn more slowly in maze B (and *vice versa*).

2.4 Results

Performance in Maze A

Plots of the raw data suggest that average time to complete Maze A decreases across trials, and that the success rate (proportion of individuals completing the tasks within the 480 s) tends to increase (Figure 2.2). These patterns are qualitatively consistent with expectations if (average) performance improves as a consequence of learning. The mixed model analysis of *maze time*$_A$ confirms statistical support for this with a significant negative effect of trial repeat number (based on the full random slope and intercept model; coefficient = -0.043 (0.014) sdu, $F_{1,59.8} = 10.140$, P = 0.003. This effect size equates to an estimated decrease of 91.9 seconds in average *maze time* over the 11 trials. Other fixed effects of order caught and maze position were non-significant (see Supplementary Information Table S2.1). Likelihood ratio tests (LRT) confirmed among-individual variation in *maze time*$_A$ (comparison of null and random intercept models; $\chi^2_{0,1} = 155$, P < 0.001). Under the random intercept model, repeatability of *maze time*$_A$ conditional on fixed effects was estimated as $R_A = 0.343$ (0.05).

LRT comparison of the random intercept and first order random regression models showed the latter to be a significantly better fit to the data ($\chi^2_2 = 25.0$, P < 0.001). This comparison provides evidence for among-individual variance in the rate of change of *maze time*$_A$ across repeated trials (interpretable, with caveats discussed below, as variation in rate of learning). Among-individual variance in intercepts (int) and slope (slp) were estimated as $V_{ind_{int}}$ = 0.394 (0.102) and $V_{ind_{slp}}$ = 0.006 (0.002) respectively while the among-individual intercept –slope correlation was estimated as ($r_{ind_{int,}ind_{slp}}$= - 0.489 (0.147)). Biological

interpretation of these parameters is not completely straightforward. Given the scaling of *trial number* in the random effect structure of the model (see Supplementary Information Table S2.3) $V_{ind_{int}}$ is interpretable as among individual variance in *maze time*$_A$ at first trial. While slope variance is in different units and thus not of directly comparable magnitude, variation in slopes actually means that among-individual variance in the observed trait (V_{ind} for *maze time*$_A$) changes with trial repeat number. Here the random regression model predicts values of $V_{ind_{A1}} = 0.394\ (0.102)$, and $V_{ind_{A11}}$ = 0.542 (0.131)) at first and last trial in maze A respectively, suggesting more among individual variation in performance at the end of trials than at the beginning. The corresponding predictions of repeatability at first and last observed trial are $R_{A1} = 0.431\ (0.070)$ and R_{A11} = 0.511 (0.067). The negative intercept-slope correlation ($r_{ind_{A.int,A.slp}}$ = -0.489 (0.1 47), χ^2_2 = 6.182, P = 0.045), means that individuals with higher intercepts (high *maze time*$_A$ at trial 1, tended to have lower slopes (i.e., more negative, indicative of faster learning). These patterns are represented visually in Figure 2.3, which shows the individual reaction norms predicted from the best linear unbiased predictions (BLUPs) of random intercept and slope for each fish (following e.g., Houslay & Wilson (2017)).

Performance in Maze B

In contrast to Maze A, plotting *maze time*$_B$ data reveals no clear increase in performance (i.e. decrease in time) across trials. Furthermore, there is actually a trend towards fewer individuals successfully completing the task (Figure 2.2). However, we note that if the censored data points are excluded to leave only successfully complete trials, there is a decreasing trend in *maze time*$_B$ with trial number. The mixed model analysis (which uses data from all trials) confirms the

lack of improvement in the mean *maze time*$_B$, with a (non-significant) positive estimate of the trial repeat number effect (from random slope and intercept model; coefficient =$0.014\ (0.014)$, $F_{1,538.2} = 1.193,\ P = 0.301$). Effects of order caught and maze position were not significant (Supplementary Information Table S2.1). Likelihood ratio tests (LRT) between the univariate random intercept model and the null model with no random effect, shows the presence of significant among-individual variation for *maze time*$_B$ ($\chi^2_{0,1}$ = 182.041 P < 0.001), with a corresponding repeatability estimate of R_B = 0.401 (0.055). The random slope model was a significantly better fit again (χ^2_2 = 9.995 P = 0.007) providing evidence of among-individual variation in the performance-trial number relationship. Among-individual variance in intercepts (int) and slope (slp) were estimated as 0.472 (0.130) and 0.004 (0.002) respectively. These estimates mean predicted values of $V_{ind_{B1}} = 0.472\ (0.130)$ and $V_{ind_{B11}}$ = 0.635 (0.162)) which correspond to repeatabilities of R_{B1} = $0.439\ (0.074)$ and R_{B11} = 0.512 (0.071). Given that there is no (significant) effect of trial number on mean *maze time*$_B$ the presence of among-individual variance in slope suggest that some individuals are improving (consistent with learning) while for others performance is tending to get worse across repeats in Maze B. Furthermore, the among-individual intercept –slope correlation was non-significant as ($r_{ind_{B.int,B.slp}}$= -0.302 (0.214), χ^2_2 = 1.476, P = 0.478). The predicted patterns are again represented visually by plotting the individual reaction norms (Figure 2.3).

Among-individual differences in OFT behaviour

We found evidence of significant among-individual variation in *relative area*, $\left(repeatability(with\ SE), R = 0.465\ (0.089), \chi^2_{0,1} = 20.421, P < 0.001\right)$. This replicates previous findings in the same population (Prentice, Houslay, et al.,

2020) though the current estimate of repeatability is somewhat higher, likely due to differences in study design (e.g. the current study used a shorter inter-observation interval and was limited to males only). Fixed effects from the OFT behaviour models are presented in the Supplementary Information Table S2.2 for completeness, although are not directly relevant to our hypotheses in this study.

Multivariate model

The full multivariate model (**ID)** of *maze time*$_A$, *maze time*$_B$ and *relative area* provides evidence of some significant among-individual covariance structure between observed traits (comparison of the full model to one in which all among-individual between trait covariances are fixed to zero; $\chi^2_8 = 44.094$, $P < 0.001$). Examination of the estimated covariances and correlations (Table 2.1) suggests this result is largely driven by a strong positive correlation between the individual intercepts for *maze time*$_A$ and *maze time*$_B$ $\left(r_{ind_{A.int,B.int}} = 0.686\ (0.135)\right)$. In other words performance at first trial is positively correlated at the individual level across mazes (since $r_{ind_{A.int,B.int}} = r_{ind_{A1,B1}}$). Using the multivariate random regression model to predict the corresponding correlation at final trial (i.e. trial 11), performances across mazes yields an estimate (SE) of $r_{ind_{A11,B11}} = 0.602\ (0.131)$. Thus, our results suggest strong positive among-individual correlation of performance as measured by maze time across trials and mazes. This is not only the case for first and last performance, but also for intermediate trial numbers as can be shown by transforming the **ID** estimate from the random regression model (as shown in Table 2.1) to a 'character state' correlation matrix among the full set of trials and maze specific observations, and *relative area* (see Supplementary Information Table S2.3 for this matrix and an explanation of the

transformation).

However, returning to a reaction norm interpretation of results, we do not find evidence that reaction norm slopes (i.e. putative rates of learning) are correlated across mazes. While the multivariate model corroborates the presence of among-individual slope variance in mazes A and B, the correlation between them was only weakly positive and non-significant ($r_{ind_{A.slp,B.slp}}$ = 0.216 (0.266); Table 2.1). Nor do we find statistical support for among-individual correlation between maze performance intercepts or slopes (for either maze) and relative area.

2.5 Discussion

Here, we show evidence of among-individual differences in performance – measured as time to complete a maze – in guppies exposed to a spatial learning test paradigm. Performance of individuals is repeatable both within, and across, the two spatial learning tasks (i.e. mazes) presented. However, the question of whether there is robust evidence of learning, on average or by individual fish, is somewhat less clear cut. In particular, in the first maze used (A) we find evidence of improvement in mean performance consistent with learning (on average). We also find among-individual variation in this rate of improvement, and so – putatively their rate of learning. However, the same fish exposed to maze B show (on average), no increase in performance across successive trials. We found among-individual correlation structure between performances (i.e. time in the maze) but not learning (i.e. rate of improvement) across the 2 spatial learning tasks. We did not however find any significant association between individual differences in maze performance (or learning) and repeatable stress responsiveness as measured in the open field trials. In what follows we describe

each of these findings in more detail and discuss them in the wider context of the cognitive literature.

The data from Maze A show that on average, time to complete the maze improves across repeated trials. This improvement suggests that spatial learning is occurring in the guppies, a finding consistent with previous studies of this species (Kotrschal et al., 2015; Lucon-Xiccato & Bisazza, 2017c; Fong et al., 2019). We also see evidence of consistent, repeatable differences among-individuals in performance in Maze A. This is shown in our reaction norm models as significant among-individual variance in intercept, which strictly represents performance at first trial. However, using among-individual variation in intercepts and slope to predict the corresponding variance at, and correlation among-, all trials (see Supplementary Information Table S2.3 for derivation and presentation of these estimates) reveals that in fact individual performance is positively correlated across all trials from 1 to 11. In simple terms, fish that are faster than average at completing Maze A in their first trial, tend to be faster than average across all subsequent trials too. Predicted repeatability of *maze time* is moderately high relative to many behavioural studies (e.g., 43% at trial 1, 51% at trial 11) but broadly comparable to estimates reported from similar assays designed to test cognitive variation; see Cauchoix et al., 2017) for an overview. We note that a contributing factor is likely to be short inter-observation period (here 24 hrs) typical of cognitive studies, since behavioural repeatabilities generally decline as this increases (Boulton et al., 2014).

Accepting that improvement across repeated trials can be interpreted as learning (caveats to this are discussed below), our random regression model also provides evidence for among-individual variation in spatial learning in Maze A. Usefully, our modelling strategy allowed all observations to contribute to

estimating variance in the latent cognitive trait (learning) while avoiding statistically problematic 'two-step' analysis (Houslay et al., 2017). Although this strategy is now widely used in studies of behavioural plasticity, it has not yet been widely adopted by researchers focussing specifically on animal cognition (but see e.g., Langley et al., 2020). In addition to finding variance in slopes (learning), we estimated a negative among-individual intercept-slope correlation using the Maze A data; individuals with higher intercepts (i.e. *maze time* at first trial) tend to have lower (more negative) slopes. While it is therefore the case that those fish performing poorly initially exhibit higher rates of learning, it is also true - as noted above - that individual performance (*maze time*) is positively correlated across trials 1-11. These two results are entirely compatible because differences in learning (slope) are not sufficiently pronounced that initially poor performing (but fast learning) fish will generally 'overtake' initially good performing (but slow learning) individuals in expected time to complete the maze by trial 11. We cannot comment on what fitness consequences, if any, the variation detected here would have in wild fish. Nonetheless, this finding does highlight a danger with any common presumptions that cognitive abilities may be under positive selection. Here, if we assumed that fitness benefits were accrued by rapidly achieving a spatial task (e.g. locating a resource) regardless of mechanism, it would be the slower learners that were advantaged. Thus, while it is tempting to assume fast learners will achieve better outcomes, they may sometimes simply be those with the 'most room for improvement'.

Thus, findings from Maze A are consistent with our initial predictions that time to complete the maze would improve (on average) with experience due to spatial learning, but that individuals would also vary in both performance (*maze time*) and learning (rate of change in performance with experience). We also found that

individuals that were quicker (over all trials) to complete Maze A, tended to be quicker (over all trials) to complete Maze B. While this could be attributable to cognitive differences, there are certainly other possibilities. For instance more explorative and/or less neophobic individuals may be generally faster at solving tasks (Boogert et al., 2006; Bousquet et al., 2015; Zidar et al., 2018). Similarly there could be among-individual variation in perceived cue salience (Meyer et al., 2012), individual physiology (Bókony et al., 2014), or motivation (van Horik et al., 2016). Regardless of these unknowns, an important difference between Maze A and Maze B was that we found no evidence of learning on average in the latter. In fact, for Maze B the mean *maze time* actually increased slightly, though not significantly, across trials. Despite this, patterns of individual variation around the mean trajectory were largely similar to those found in Maze A. Thus, there is among-individual variation in intercept (*maze time* at trial 1) and also in slope. Given that there is no (significant) change in mean performance, but there is significant variation in slopes, we conclude that some individuals are improving (learning) in Maze B while others are getting worse with experience. We also note that, as in Maze A, slope variance is present, but not sufficiently high to break down the positive correlation structure of individual performance (*maze time*) across trials 1-11.

Although we did not formally test for differences in average slope between maze A and B, we note that approximate 95% confidence intervals do not overlap (estimated as coefficient±1.96SE). Several possibilities may explain the finding of spatial learning on average in A but not B. First, the results from maze A may be a false positive (Sterne et al., 2001; Fraser et al., 2018). However coinciding with previous studies which show this species is capable of learning an initial spatial learning task (Kotrschal et al., 2015; Lucon-Xiccato & Bisazza, 2017b;

Fong et al., 2019), we assume this unlikely. Second, it may be that the layout of maze B was more challenging to learn. This could certainly be true if, for instance learning to navigate a new maze following the acquisition of a previously learnt layout poses a more challenging task, for example due to proactive interference (difficulty inhibiting memory; Shettleworth, 2009a)). In this case the second maze may require more trials to detect improvement. There is some evidence for such effects in guppies. For instance, Lucon-Xiccato & Bisazza (2014) found that on average guppies took 14.61 trials to learn a reversed colour cue association, while Fong et al., (2019) found that on average, 15.30 trials were required for guppies to learn a reversed maze layout. A third possible explanation could be that, even if some individuals clearly did perform better over time, mean performance time is confounded by changes in motivation due to trial fatigue (reduced motivation) by the end of the trials in maze B.

Another possible explanation is that learning does lead to gains in maze B performance, but that these are being masked at the level of the sample mean by concurrent changes in aspects of average individual 'state' that reduces cognitive performance and/or motivation. One plausible hypothesis is that chronic stress responses arise cumulatively from repeated capture and handling necessitated by the experimental design (Huntingford et al., 2006; Warren & Callaghan, 1976; Wong et al., 2008). If so, this could negatively impact affected individuals and offset expected improvements in mean performance across trial number. Presently we cannot directly test this possibility, and variation in susceptibility to chronic stress response is not well understood. Nonetheless, our experiment does confirm repeatable among-individual variation ($R = 0.465\ (0.089)$ in *relative area* covered in the OFT, used here as a measure of acute behavioural stress response. This replicates previous results using

independent data sets of fish from the same captive population (White, 2016; Houslay, 2019; Prentice, 2020). Acute stressor exposure has been shown to affect cognitive performance in spatial learning tasks in both mammals and fish (Gaikwad et al., 2011; R. Y. Wong et al., 2019). At the individual level, there is also evidence to suggest short-term measures of acute stress responses can predict longer term organismal performance under chronic and/or repeated stressor exposure (Segerstrom et al., 2004; Salak-Johnson et al., 2007; Øverli et al., 2007).

Here, our modelling approach did not provide compelling statistical support for strong relationships between *relative area* and either initial *maze time*, or learning (i.e. improvement in *maze time*) in either Maze A or B. However, the estimated correlation between *relative area* and *maze time* in Maze B actually rises to r_{ind}=0.336 (0.169) by trial 11 and thus approaching nominal significance at α=0.05 (assuming a lower 95% CI of r_{ind} - 1.96SE). To explore this further we conducted a *post hoc* likelihood ratio test comparison of a bivariate model of *maze time*$_B$ and *relative area*; all effects as described for the trivariate model earlier) to the corresponding model fit where among-individual covariances between *relative area* and *maze time*$_B$ (intercept and slope) were constrained to zero. The LRT did not provide evidence that the trivariate model was not a significantly better fit; χ^2_2 = 3.098, P = 0.212). Thus, we do not find statistical support for the prediction, made under the stress coping style model, that (acute) stress responsiveness will (co)vary with cognitive performance (Coppens et al., 2010; Sih et al., 2012; Griffin et al., 2015). Nonetheless, the possibility that chronic stress negatively impacts apparent learning cannot be completely excluded here. Empiricists rightly seek to minimise the possibility of stress confounding conclusions from cognitive studies. However, we suggest the

assumption that individuals remain (equally) ‘unstressed’ over experimental periods requiring repeated observations (and often repeated capture and or social isolation) is difficult to validate in practice.

In summary, here we have evidence of consistent differences among-individuals in spatial task performance in the guppy *P. reticulata*. Individual performance is repeatable across trials within- and between two different spatial tasks (i.e. maze layouts). This among-individual variation in performance may well be mediated by cognitive factors but differences in ‘personality’ (e.g. neophobia, exploratory tendency) may also contribute. We also find evidence of improved performance with experience, consistent with spatial learning. In both tasks variation around the trajectory of mean performance across trial number was present. While this means individuals can be considered as differing in ‘spatial learning rate’ it is important to note that performance declines for some individuals, especially in the second maze where there was no improvement in average time across 11 trials. We show here that an individual’s (repeatable) behavioural response to an acute stress stimulus does not predict either average performance in the maze or learning rate. However, we suggest the possibility that cumulative, chronic stress effects may contribute to declining performance (or reduced improvement) in our study. If individuals generally differ in susceptibility to chronic stress, this may represent a widespread but currently poorly acknowledged challenge for characterisation of cognitive variation in animal studies.

Table 2.1: Among individual variance–covariance–correlation matrix from the final trivariate model of *maze time*$_A$, *relative area* and *maze time*$_B$. Variances are shown on the diagonal (dark grey shading), with covariances below and correlations above. Light grey shading denotes within trait covariance/correlation estimates (i.e. between reaction norm intercepts and slopes). Standard errors are shown in parentheses and bold font denotes nominally significant pairwise estimates assuming approximate 95% CI of ± 1.96SE).

	***Maze time*$_A$**		***Relative area***	***Maze time*$_B$**	
	intercept$_A$	*slope*$_A$		*intercept*$_B$	*slope*$_B$
***intercept*$_A$**	**0.436 (0.113)**	**-0.489 (0.147)**	0.286 (0.175)	**0.686 (0.135)**	-0.129 (0.254)
***slope*$_A$**	**-0.027 (0.013)**	**0.007 (0.003)**	-0.075 (0.202)	-0.006 (0.209)	0.216 (0.266)
Relative area	0.127 (0.084)	-0.004 (0.011)	**0.451 (0.118)**	0.024 (0.125)	0.377 (0.231)
***intercept*$_B$**	**0.299 (0.091)**	-0.003 (0.012)	0.011 (0.085)	**0.437 (0.119)**	-0.309 (0.212)
***slope*$_B$**	-0.005 (0.010)	0.001 (0.001)	0.016 (0.011)	-0.013 (0.012)	**0.004 (0.002)**

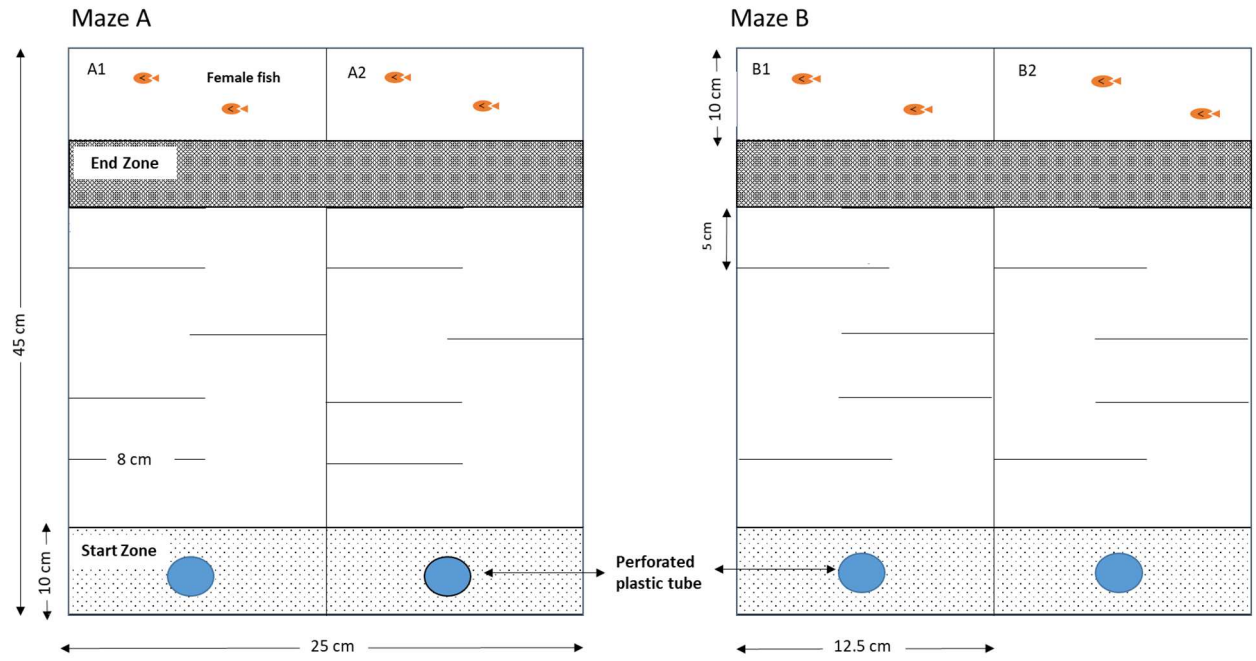


Figure 2.1: Aerial view of the maze designs used in the experiments (A and B), each tank was split into two identical mazes (1 and 2).

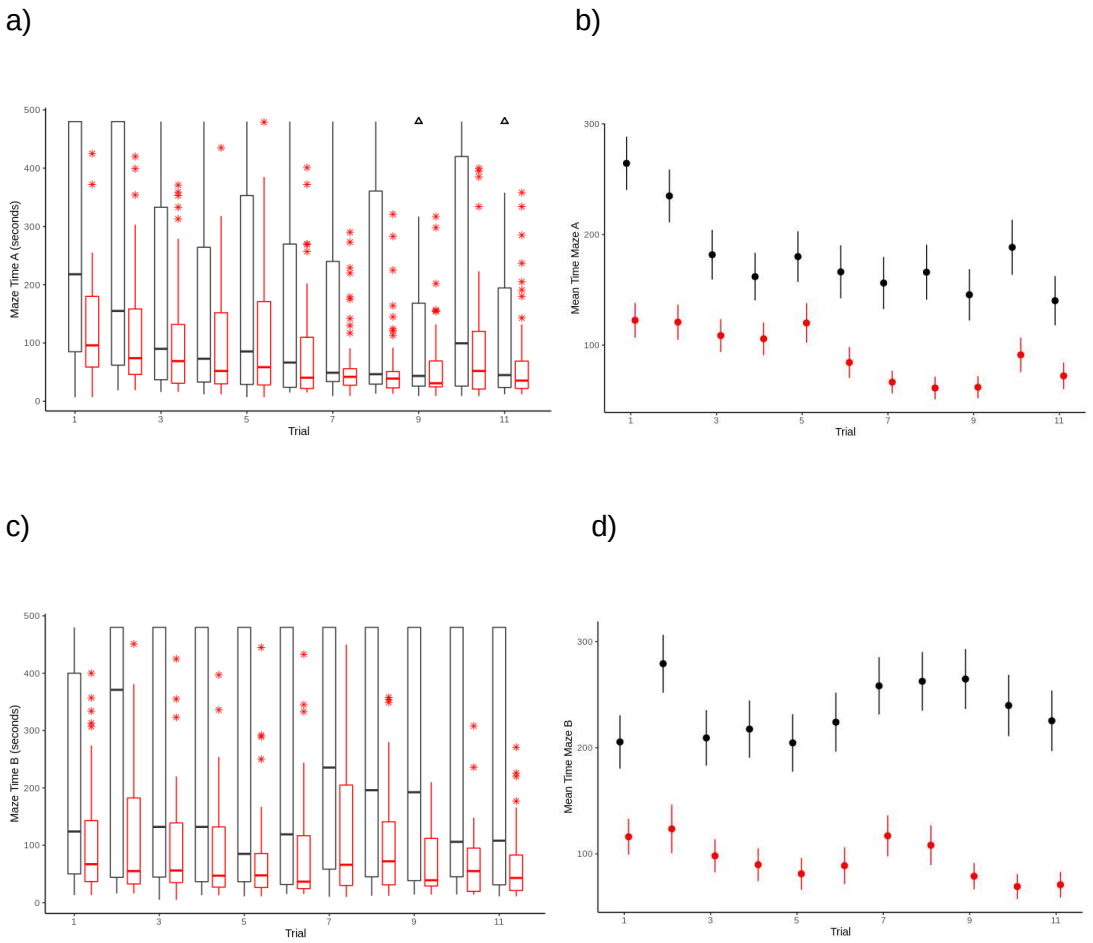
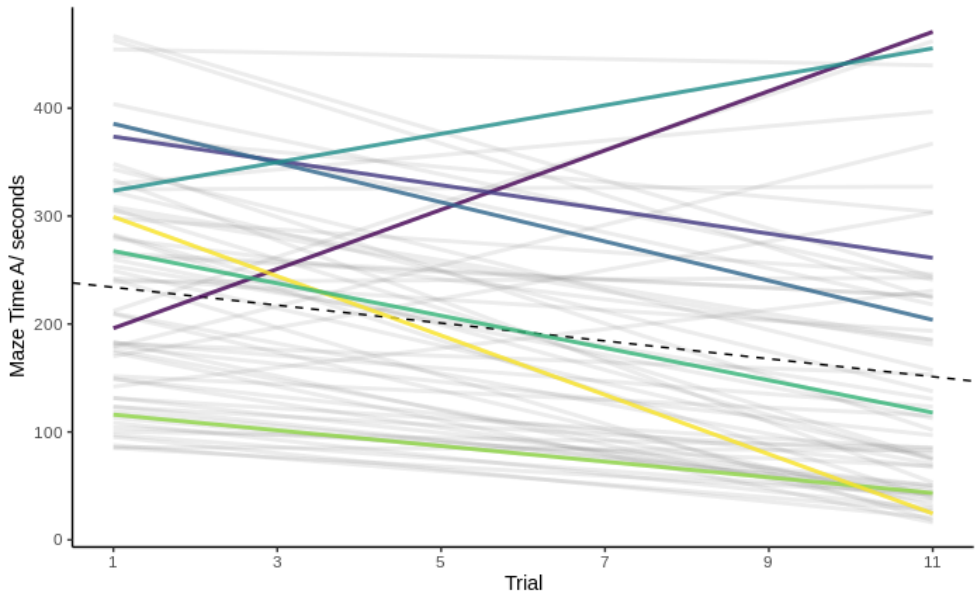
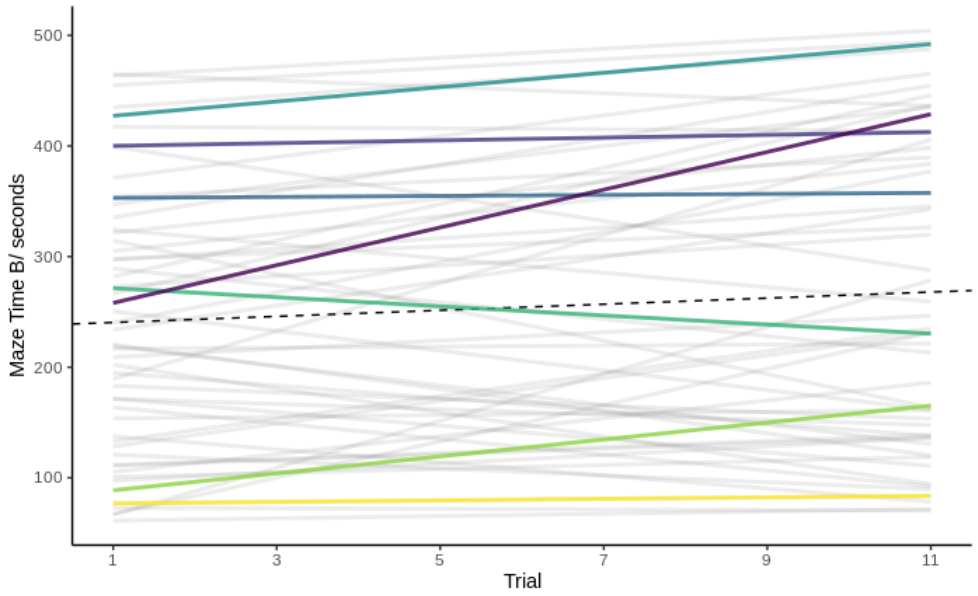


Figure 2.2: Plots of raw data of *maze time* across both maze designs. Boxplots (a) and (c) show the data distributions for time to complete Maze A and Maze B respectively across the 11 trials. Black boxes display data of all individuals and red boxes represent only those individuals that successfully completed the task within 480 s. Horizontal lines within box correspond to behavioural medians, box boundaries correspond to first and third quartiles. When present, whiskers correspond to 10th and 90th percentiles, and points correspond to outliers. Plots (b) and (c) represent mean and standard errors for time to complete Maze A and B respectively. Colours represent the same groups; black error bars represent mean and standard errors of *maze time* for all individuals, and red represent only those individuals that successfully completed the maze in the allocated time.



a)



b)

Figure 2.3: Spatial learning traits across Maze A and Maze B as a function of trial number, *maze time*$_A$ (a), and *maze time*$_B$ (b). Grey lines represent individual predicted reaction norms (BLUPs) from univariate random slope models for each trait. Coloured lines are used to illustrate reaction norms for a small random set of arbitrarily chosen individuals tested in both mazes. Black dashed line represents the trend in fixed effect mean *maze time* across repeat trials

3 Chapter 3: A multivariate view of cognitive differences reveals domain-general correlation structure in the Trinidadian Guppy (*Poecilia reticulata*)

3.1 Abstract

Cognitive variation is common among-individuals within populations, and this variation can be consistent across time and context. From an evolutionary perspective, among-individual variation is important and required for natural selection. Selection has been hypothesised to favour high cognitive performance, however directional selection would be expected to erode variation over time. Additionally, while variation is a prerequisite for natural selection, it is also true that selection does not act on traits in isolation. Thus, the extent to which performance covaries among specific cognitive domains, and other aspects of phenotype (e.g. personality traits) is expected to be an important factor in shaping evolutionary dynamics. Fitness trade-offs could shape patterns of variation in performance across different cognitive domains, however positive correlations between cognitive domains and personality traits are also known to occur. Here we aimed to test this idea using a multivariate approach to characterise and test hypothesised relationships of cognitive performance across multiple domains and personality, in the Trinidadian guppy (*Poecilia reticulata*). We estimate the among-individual correlation matrix (**ID**) in performance across three cognitive domains; association learning in a colour discrimination task; motor cognition in a novel motor task and cognitive flexibility in a reversal learning task, and the personality trait 'boldness' measured as time to emerge. We found no support for trade-offs occurring, but the presence of strong positive domain-general correlations in **ID,** where 57% of the variation is explained by the leading eigen vector. While highlighting caveats of how non-cognitive factors and assay

composition may affect the structure of the **ID**-matrix, we suggest that our findings are consistent with a domain-general axis of cognitive variation in this population, adding to the growing body of support for domain-general variation among-individuals in animal cognitive ability.

3.2 Introduction

Interest in the cognitive mechanisms by which animals acquire, process, store and use information from the environment (Healy et al., 2010; Shettleworth, 2010b), has grown substantially in recent years. Differences in cognitive abilities among-species have long been recognised (Wasserman et al., 2006; Shettleworth, 2009b), but we now know that variation among-individuals that is consistent across time and context is common in non-human animals (for a review, see Cauchoix et al., 2018). From an evolutionary perspective, variation at this level is required for natural selection and is thus central to our understanding of adaptive evolutionary dynamics (Roff, 2002). However, how or why this variation is maintained within populations is not always clear. For instance, while we might intuitively expect selection to favour high cognitive performance, directional selection of this form is generally expected to erode variation over time. One possibility is that fitness trade-offs could shape patterns of variation in cognition, just as they do in other aspects of phenotype (e.g. life history; Stearns, 1992). If so, do trade-offs arise among different cognitive 'domains' or traits, with variation maintained because individuals (or genotypes) that perform better than average in some respects, perform worse in others? Here we test this idea using a multivariate approach to characterise and test hypothesised relationships of cognitive performance across multiple domains in

the Trinidadian guppy (*Poecilia reticulata*). We ask whether there is support for domain-specific correlation structure, where performance is negatively correlated as expected under the trade-off hypothesis. Or whether in fact domain-general correlation structure is present and uniformly positive – a pattern predicted if multivariate cognitive performance variation is explained by a single latent "general intelligence factor" (*g*) (Nisbett et al., 2012).

Many empirical studies have now shown that animal populations can harbour high levels of variation in cognitive performance among individuals (Boogert et al., 2018). This has been shown across taxonomic groups including insects (Li et al., 2017), fish (Lucon-Xiccato & Bisazza, 2017a), birds (Quinn et al., 2016) and mammals (Mazza et al., 2019). Among-individual variation is also found across different cognitive domains (e.g. spatial memory (Sonnenberg et al., 2019); association learning (Kniel et al., 2020), problem solving (van Horik et al., 2019)), though relatively few studies have characterised the among-individual covariance (or correlation) structure between these. This is important to do because trade-offs between domains have been hypothesised to maintain cognitive variation (Del Giudice et al., 2018). This hypothesis leads to an expectation of domain-specific structure of variation, where negative correlation structure is exemplified by predictions of 'speed-accuracy trade-offs' (Biro et al., 2008; Sih et al., 2012). Faster decision making should provide a competitive advantage (e.g. by increasing the potential rate of resource acquisition), but by allowing less time to assess environmental cues, error rates may be increased. In contrast, slower decisions may be more accurate, but being slow to act can mean resources are lost to competitors. A number of empirical studies have provided support for speed-accuracy trade-offs in animals. For example, individual archerfish (*Toxotes chatareus*) that tended to make slower decisions

within a colour discrimination task also had greater accuracy (Jones et al., 2020). A similar result was found in zebrafish (*Danio rerio*) given a visual discrimination task (Wang et al., 2015), and in wild-caught great tits (*Parus major*) given a foraging task (Moiron et al., 2016). However, counter examples also exist. For example, a prior study of guppies found no support for a speed-accuracy trade-off using a shape and a colour discrimination task (Lucon-Xiccato, Dadda, et al., 2016), and three-spined sticklebacks (*Gasterosteus aculeatus*) making fast decisions in a spatial leaning task did not show reduced accuracy (Mamuneas et al., 2015).

Another widely hypothesised cognitive trade-off that might maintain variation is between learning and cognitive 'flexibility' (Del Giudice et al., 2018). Flexibility is broadly defined as the ability to adapt when environmental stimuli or information cues change. It is often tested by reversal learning experimental paradigms in which individuals must be flexible to override old cue-reward associations and form new ones (Bitterman, 1965; Buechel et al., 2018; Kehagia et al., 2010). For example, in Florida scrub-jays (*Aphelocoma coerulescens*) individuals that were quick to learn an initial colour-reward cue (high associative learning performance), were slower to adjust when the cue signal was reversed (Bebus et al., 2016). The implication of such patterns is that some individuals are more 'intrinsically driven' than others; they learn initial associations quickly but, being less sensitive to external stimuli, struggle to adapt when cues are altered. However, again counter-examples can be found. For instance, those individual bumblebees (*Bombus terrestris*) that quickly learned to discriminate between two colours (where one was associated with a floral reward), were also faster to learn a new association when the cues were reversed (Raine et al., 2012).

Thus, while trade-off among components of 'multivariate' cognitive ability are intuitive and evidenced in many cases, they may not be inevitable and positive correlations between specific domains are sometimes found (Guenther et al., 2017; Wallace et al., 2021). If correlation structure among cognitive performance traits is universally positive (with respect to expected fitness consequences), then variation among individuals can be explained by a domain-general structure that invokes a single latent general intelligence factor, sometimes denoted '*g*' (Deaner et al. 2006; Lefebvre and Sol 2008). While there is evidence for this in humans (Deary et al., 2010; Burkart et al., 2017), support for the *g*-model in non-human animals systems remains limited (but see Arden et al., 2016; Galsworthy et al., 2005; Hopkins et al., 2014; Shaw et al., 2015). Note that, under this model, trade-offs may still be important for maintaining cognitive variation, but if so they must operate between general intelligence and other (non-cognitive) aspects of phenotype. This possibility is mirrored elsewhere in the behavioural literature since for example, individual personality can be viewed as a component of an extended 'life history' (Sih, Bell, & Johnson, 2004; Wolf et al., 2007). Indeed, given that personality traits and cognitive performance are frequently correlated (Guenther et al., 2014; Nawroth et al., 2017; White et al., 2017), and both thought to drive variation in resource acquisition, then trade-offs between them are plausible (Sih et al., 2012; Dougherty et al., 2018).

Here, we investigate among-individual (co)variation between cognitive performance across three cognitive domains (associative learning, motor cognition and cognitive flexibility) in a captive population of wild-type guppies (*Poecilia reticulata*). Guppies have been widely used as a model in behavioural and evolutionary ecology, and methods for assaying among-individual variation in cognitive performance are well established (Laland et al., 1999; Miletto

Petrazzini et al., 2016). This species is known to perform well (on average) in tests of associative learning (Lucon-Xiccato & Bisazza, 2016; Kniel et al., 2020), motor cognition (Lucon-Xiccato & Bisazza, 2016; Lucon-Xiccato, Gatto, et al., 2017), and cognitive flexibility (Cauchoix et al., 2018; Fong et al., 2019). However we know little about if and how these traits may covary among individuals within populations. Broadly stated, our goal is to estimate the structure of this multivariate variation to evaluate whether trade-offs among cognitive traits are evident, or whether among-fish variation is consistent with a domain-general intelligence model (*g*).

We note that the extent to which among-trait associations shape, and in the particular context of trade-offs, constrain evolutionary adaptation strictly depends not on the phenotypic correlation structure, but on the genetic contribution to this. In particular, evolutionary constraint arises from the genetic (co)variance structure (**G**) and its alignment (of lack thereof) with selection (Lande, 1979; A. G. Jones et al., 2004; Walsh et al., 2009). Unfortunately estimation of **G** is challenging in general (requiring large volumes of data from related individuals; Wilson et al., 2010), and particularly so for cognitive traits. This is because high-throughput phenotyping of cognitive performance across multiple domains is notoriously laborious and challenging. Consequently, very few studies have estimated **G** for sets of cognitive traits in non-human animals (but see Langley et al., 2020). Thus most multivariate studies of cognition, and behavioural phenotypes more generally, rely on Cheverud's conjecture (Cheverud, 1988) that phenotypic patterns of covariation can be used to infer evolutionarily important relationships between traits. Accepting this view uncritically implies that the (co)variance structure of the phenotypic matrix **P** (historically referred to as the phenotypic gambit by Grafen (1984)), which can

be readily estimated in a set of individuals observed once only for each of the target traits, can yield robust insights into the structure of **G**. However, behavioural traits are typically very plastic and often subject to high levels of measurement error (e.g. relative to morphological traits), in addition to sources of environmental variation. These phenomena lead to high levels of within-individual variation such that a single observation may tell us relatively little about an individual's phenotype (and so genotype) (Brommer, 2013). In other words, the extent to which **P** is a good proxy for **G** will decline if **P** is dominated by within-individual variation (plasticity, measurement error) versus among-individual variation (which includes genetic factors).

In the present study we therefore adopt an intermediate approach in which we estimate the among-individual (co)variance matrix (**ID**) rather than either **G** or **P**. We do this by obtaining repeated measures of performance on individual fish. Since **G** is a component of **ID** while **ID** is a component of **P**, all else being equal we expect the among-individual covariance matrix to be better proxy of the genetic matrix than the total phenotypic matrix. We target multiple domains of cognitive performance allowing us to 1) test whether individuals differ in their ability to discriminate between rewarded and unrewarded colours in an association learning task; 2) ask whether speed accuracy trade- offs mediate variation in performance in the association task; 3) ask if individuals differ in performance in a novel motor task to access a reward; and 4) test whether individuals differ in their ability to learn a reversal learning task when the colour-reward cue from the association task is reversed. We predict that individuals will differ in average performance within all three tasks; the association task, the novel motor task, and the reversal learning task as well as in their rates of improvement with experience (interpretable as learning). Finally, we 5) estimate

the **ID**-matrix among all cognitive performance traits and a shy-bold type personality trait (henceforth referred to as 'boldness') and ask whether it provides evidence for trade-offs among cognitive domains or whether it is dominated by positive correlations consistent with the *g*-model of domain-general intelligence.

3.3 Methods

Ethics

This work was conducted under the auspices of the Animals (Scientific Procedures Act) with approval of the University of Exeter research ethics committee, under licence from the Home Office (UK) (Licence Number PPL30/3256). Experimental procedures and behavioural assays were developed in accordance with the principles of the three R's and ASAB guidelines (Buchanan et al., 2020) for use of animals. All periods of handling and emersion were kept to a minimum and only fish deemed healthy and exhibiting normal behaviour were used in trials. At the end of the experiment, fish were returned to a designated 'retirement' tank (containing females as well as males) and not used in any further experiments.

Husbandry

All behavioural assays were carried out during the months between October 2019 and January 2020. Data was collected from captive-bred guppies bred and housed at the fish laboratory at the University of Exeter's Penryn campus. The population is descended from wild fish caught in February 2017 from the lower Aripo River, Trinidad and has been subsequently maintained with no deliberate

selection or inbreeding. All fish housed in the laboratory were fed to satiation twice daily (0800 – 1000h and again at 1600 – 1800h) using commercial flake food and live *Artemia nauplii*. Water temperature was maintained at 23-24°C in well-aerated closed system tank stacks that undergo 25% water changes each week and with weekly tests for ammonia, nitrate and nitrite levels. Lighting was kept at a 12:12 light/dark cycle.

Experimental apparatus

Adult males (n = 43) and females (n = 37) were sampled from the stock population. Sampling was haphazard but we approximately size matched fish within each sex (e.g. by avoiding very large females). Fish were then housed singly in separate tanks (15 l, 18.5 × 37 × 22cm) for the duration of behavioural testing. The tank set up closely followed that used by Lucon-Xiccato & Bisazza (2014), each being divided equally into a 'home' compartment at the rear of the tank, and a 'test' compartment at the front of the tank, separated into two compartments (using white plastic) by a guillotine door (Figure 3.1). The rear 'home' compartment (20 x 18.5 cm) allowed individuals visual access to fish in neighbouring tanks. Conversely, the 'test' compartment (17 x 18.5 cm) was screened from neighbours (using white plastic) to prevent any possibility of social learning (i.e. by observation of neighbours) influencing cognitive task performance. The 'test' compartment contained a white plastic plate (4 x 10 cm) placed on the gravel substratum, perforated with 2 equally spaced wells (Figure 3.1). A total of 48 experimental tanks were used. These were contained within two 'stacks', each comprising 24 tanks (8 tanks per row, 3 rows high) on a shared recirculating water supply. As the testing protocol (described below) took 18 days

per fish, data were in practice collected in 3 ‘blocks’ over a total testing period spanning 12 weeks. Before cognitive testing all fish were allowed to acclimate to experimental tanks for 48 hours. During this period they were fed twice daily with bloodworm (*Chironomidae* larvae) pipetted into one (randomly chosen) well of the white plastic plate and the guillotine door was left open allowing unrestricted use of both compartments.

Cognitive assay

We used a repeated measures design to test performance across multiple cognitive domains within a single, extended, testing paradigm. Each fish was observed up to 63 times as it emerged from the home compartment to obtain a food item that had been placed in one of the two wells in the test compartment. Individual fish were trained to discriminate between a rewarded and an unrewarded colour cue, indicating which well the reward could be obtained from. In all trials, food items (bloodworms) were actually placed in both wells, but access was restricted by a small plastic disc for the well with the unrewarded colour cue. This was to ensure that olfactory cues would be insufficient to locate available food, instead fish were required to learn to associate a specific colour with reward access. The rewarded colour (either blue or green) was randomly determined for each fish prior to testing, ensuring balance across subjects to control for any innate colour bias. The relative position of the rewarded well (left versus right) for each trial was randomised and we also controlled for any effect this could have in our statistical modelling (described later).

The 63 observations per individual were grouped into 7 ‘sets’ of 9 ‘trials’ (observations). Each set spanned a 2 day period which was followed by a rest

day before the next set. Test condition were altered between trial sets, by adjusting the position of green and blue coloured plastic counters. Specifically, after each set, the association task increased in difficulty as the coloured counter covered the holes to an increasing extent. This also allowed extraction of information on different cognitive traits as we now describe. In set 1, the coloured counters covered 0% of the wells, in set 2 we increased this to 25%, then subsequently this was increased to 50% (set 3), 75% (set 4) and 100% (set 5). Once the counters covered 75% of the wells (set 4), guppies had to dislodge the counters in order to reach the reward in the well underneath. We consider this a novel motor task, as fish had to learn to dislodge the counter by physical manipulation. The difficulty of this task is increased with 100% coverage (set 5). In the final two sets (set 6 and 7) of 9 trials we tested reversal learning task, by reversing the colour cue -food reward association previously learnt in the first 5 sets.

Prior to each trial, fish were guided into the 'home' compartment of the tank with a net, and the guillotine door was closed. The experimenter set up the test plate which was situated in 'test' compartment, by pipetting bloodworm into the wells in the plate (the food in the unrewarded hole was then covered by a small plastic disc), and the two coloured counters placed either to the side of (set 1), or covering the well (partially or fully depending on set number as described). The trial started once the door to the experimental compartment was opened and continued until the fish ate the bloodworm from the rewarded well, or after 10 minutes, whichever occurred first (see the electronic Supplementary Material, video S1, Appendix 1). All trials were recorded with a GoPro Hero 6 camera, mounted in front of the tank and behavioural scoring of videos was quantified using the software BORIS (Friard et al., 2016). We imposed a deliberately weak

learning criterion, removing any fish that failed to achieve a minimum of 5/9 successes (defined as locating and/or eating the food reward) at the end of each set. In fact, we wished to avoid selective removal of fish as far as possible such that, for example, among-individual variation seen in later sets can be interpreted as representative of the initial population. However, correlations between task performance may be confounded if the ability to solve each task is closely associated with the number of learning opportunities (Guez et al., 2016; Shaw et al., 2017). Differences, for example in motivation or neophobia, could mean some individuals engage with and solve the task each trial, while others do not (despite persistently trying to find a solution or because they do not engage with the task at all). In this way, individual participation directly influences the number of learning opportunities. Removal of fish that did not complete the task (by locating and eating the food reward) ensured data was collected in subsequent set of trials, only from individuals that had the same number of learning opportunities, thus reducing the possibility of confounding correlations in performance caused by differences in e.g. motivation or neophobia for example. Furthermore, since additional food (beyond the reward items) was not provided during the behavioural testing, the learning criterion also ensured that individuals not engaging with the task were not deprived of food for any longer than 2 days before being removed from the study. Over the course of the data collection, 16 individuals were removed before the end of set 6.

Trait definition

Data obtained from the trial sets were used to define observed proxies of one personality trait (*boldness*) and 6 measures of cognitive performance that are

collectively informative for 3 cognitive domains (*association learning, motor cognition, and cognitive flexibility*). The specific measures derived are described below and also summarised for reference in Table 3.1.

Boldness

During the first nine trials (set 1), time taken to emerge from the home compartment into the experimental compartment was recorded (*emergence*), as a proxy of consistent differences in personality trait 'boldness' (White et al., 2016; White et al., 2020).

Association learning

We also focused on set 1 trials to test for speed-accuracy trade-offs, expecting that (on average) fish would have effectively learnt the association in later sets making the speed accuracy trade-offs less relevant to observed outcome. We recorded three behavioural variables. The first was whether the first counter explored was the correct (1) or incorrect (0) choice for obtaining the reward ($AL_{accuracy}$). A choice was deemed to have been made if individual fish swam within 1 body length of, and was actively exploring a counter. The second variable was time taken to make this first choice (AL_{speed}), measured in seconds from emergence into the test compartment. To evaluate overall performance in association learning, we used total time taken from emergence to find and eat the food reward. We did this using all observations from all trials in sets 2 and 3 (AL_{time}). During these sets of trials, the counters covered 25% and 50%

respectively, a level of coverage that still allowed fish to access the reward easily without any physical manipulation of the counter required.

Motor Cognition

To assess performance in the novel motor task, we used observations from trials insets 4 and 5 (where the counter covered 75% and 100% of the well area respectively). Note that at 75% coverage fish were required to physically manipulate the counter in order to reach the food reward. We used time taken by each individual in all trials to obtain and eat the food reward (MC_{time}). Note this was measured as time taken to obtain the food reward after they had made the correct choice of counter in each trial.

Cognitive flexibility

This was assessed using observations from all trials in set 6 and 7, during which the reward colour was reversed. Note that counter coverage of the wells remained at 100% so fish needed to dislodge the counter in order to access the reward. For this reversal learning task we recorded the accuracy of each individuals' first choice, ($RL_{accuracy}$) as well as the time taken, from emergence into the test compartment, to locate and eat the food reward (RL_{time}).

Statistical analysis

We used univariate and multivariate linear mixed effect models to characterise among-individual variation in and covariation among the traits defined. For analysis, the observed (censored) time traits from the cognitive tasks were

natural log-transformed (AL_{speed}, AL_{time}, MC_{time} and RL_{time}), and then all continuous time traits were multiplied by -1 (*boldness*, AL_{speed}, AL_{time}, MC_{time} and RL_{time}). Following this, all response variables were mean centred and scaled to standard deviation units. The log transformation was to improve the assumption of Gaussian error structure, multiplication by -1 for ease of interpretation of multivariate analysis (so that larger values equate to higher performance in continuous traits), and scaling to standard deviation units was to ease interpretation of estimated variance components. All models were fit by REML (restricted maximum likelihood) using ASReml-R 4.1 (Gilmour et al., 2009; Butler et al., 2018) within R version 3.6.1 (R Core Team, 2017). We make the standard assumptions that random effects and residuals are normally distributed with means of zero and variances to be estimated. We acknowledge that these assumptions are necessarily violated since the two accuracy traits are recorded as binary variables, while time data are censored (at 10 minutes). Consequently, we consider this strategy justifiable given that Gaussian mixed models are generally robust to violations of distributional assumptions (Schielzeth et al., 2020). Pragmatically, we also note that the obvious alternative of fitting Bayesian multivariate generalised mixed models poses its own challenges.

Univariate analysis

First we fitted separate univariate mixed models to test for and characterise among-individual variation in each of our 7 traits (*boldness*, AL_{speed}, $AL_{accuracy}$, AL_{time}, MC_{time}, RL_{time} and $RL_{accuracy}$). All models included a fixed effect of the mean, as well as fixed factors of *sex*, *stack* (denoting which of two aquaria stacks the fish was tested in), *colour* (denoting the colour of counter individual fish were

trained to associate with the rewarded cue; green or blue), and *reward side* (denoting which side the rewarded counter was placed on; left or right). We also included *trial number* (i.e. the repeat number of trial, fitted as a factor to avoid assuming a linear functional form for any change in the mean across repeats). For each trait, we firstly fit this 'null' model without random effects, and then refitted with *individual identity* as a random intercept. For each trait we compared the full (i.e. with *individual identity* included) and null models by likelihood ratio tests (LRT) to obtain a statistical test of the among-individual variance (V_I). To test a single variance component (which cannot be less than zero) we assume twice the difference in log-likelihood between the full and reduced models is distributed as a 50:50 mix of $\chi^2_{0,1}$ and χ^2_1 as recommended by Visscher (2006). Since transformed traits are analysed in standard deviation units, total observed variance is 1 and V_I can actually be interpreted as an estimate of repeatability. However, we also calculated estimates of the adjusted repeatability (R) which is the proportion of phenotypic variance conditional on fixed effects that is explained by among-individual differences (Nakagawa et al., 2010). Thus $R = V_I / V_P$ where V_I is the among-individual variance and V_P is sum of V_I and V_R, the residual or within-individual variance. Conditional F-statistics were used to determine the significance of fixed effects in each (full) model, and we elected not to perform model simplification, as we wanted repeatability estimates to be conditioned on a common set of fixed effects and thus comparable across abilities.

Multivariate analysis

To test for the presence of, and investigate the structure of, the (co)variance **ID**-matrix, we built a multivariate mixed model among all cognitive performance traits

and ‘boldness’. All response variables were transformed and scaled as per the univariate modelling procedure above, and we note that multiplication by -1 of all the observed (censored) time traits simplified interpretation of the correlation structure (as larger values equate to higher performance in continuous traits). Fixed and random effects were fitted on each trait as specified for the univariate models. Fixed effects estimates are reported in the Supplementary Information Table S3.1, Appendix 3. We specified the among-individual (V_I) covariance structure as an unstructured matrix to be estimated. Note that R partitions observation-level covariances that are not statistically identifiable if traits are not measured at the same time (i.e., all covariances between traits measured in distinct ‘sets’ of trials). Where this was the case we constrained specific covariance terms in the residual (V_R) matrix to equal zero. We tested for overall among-individual covariance among the traits by comparing this model against a reduced one in which **ID** was specified as a diagonal matrix (i.e., among-individual variances are estimated but covariances are assumed to equal zero).

To aid biological interpretation of the **ID**-matrix, we rescaled our estimate to the corresponding correlation matrix in which off diagonal elements are the among-individual correlations (r_I) between each pair of traits. For any pair of traits (x,y), $r_{I(x,y)} = COV_{I(x,y)} / (V_{I(x)}V_{I(y)})^{0.5}$). We then further scrutinised the correlation structure among the cognitive traits only by dropping the row/column corresponding to boldness, and subjecting the resulting 6x6 submatrix, henceforth denoted $\mathbf{ID_{rc}}$ (using rc to denote correlation scale (r) and cognitive traits (c)) to eigen vector decomposition. This simply provides a descriptive view of the major axes (or principle components) of variation in $\mathbf{ID_{rc}}$ allowing us to determine the proportion of among-individual correlations captured by each principal component. Under the *g* model, we would expect uniformly positive

correlation structure in **ID_{rc}** with a single major axis of variation on which all traits loaded with the same signs. Conversely, if trade-offs dominate the structure of among-individual variation in the multivariate cognitive phenotype **ID_{rc}** should contain at least some negative correlations and the one or more important principle component should have loadings that are antagonistic in sign between traits that trade-off with each other. We estimated approximate 95% confidence intervals on the eigen values (scaled to proportions of variation in the matrix) and on the trait loadings associated with each principal component using a parametric bootstrap approach described in Boulton et al (2014).

3.4 Results

Visualisation of the raw data shows a varied effect of trial number on performance across traits and cognitive tasks (Figure 3.2; see Supplemental Information Figure S3.1 *a–g* of raw data across all sets of trials). Average time to *emergence* (boldness) decreased over the first 9 trials but time-based measures of cognitive performance did not uniformly improve over the sets of trials used to assay them, as might intuitively be expected (Figure 3.2). Note however that for AL_{time} and MC_{time} the sets of 18 informative trials include an increase in task difficulty (between trials 9 and 10). There was no increase in task difficulty over the 18 informative trials for the trait RL_{time} anda negative (albeit weak) trend is seen, indicative of improving performance. In the association learning task, AL_{speed} and $AL_{accuracy}$ increased over the first set of trials, similar to $RL_{accuracy}$ in the reversal learning task which also increased with trial number over the set of trials used to assay this trait. The univariate mixed models confirmed the statistical significance of qualitative patterns seen in the raw data with respect to changes across trials

(see Supplementary Information Table S3.1 *a–g)*. Noting that models were run on transformed traits such that positive effects correspond to increasing performance (i.e. decreasing time) and that *trial number* was fitted as a factor to avoid assuming linear trends, we found effects on *emergence* ($F_{8, 555.7}$=9.24, P< 0.001; with coefficients becoming more positive as *trial number* increased; Table S2.1)*,* AL_{time} ($F_{17, 801.9}$=2.246, P=0.002; with coefficients becoming more negative)*,* AL_{speed} ($F_{8, 555.0}$=7.380, P< 0.001; with coefficients becoming more positive)*,* $AL_{accuracy}$ ($F_{8, 555.0}$=4.602, P< 0.001; with coefficients becoming more positive), MC_{time} ($F_{17, 692.6}$=1.174, P< 0.001; coefficients becoming more negative). The effect of *trial number* on (mean) performance was not significant in RL_{time} ($F_{17, 581.2}$=1.565, P= 0.060) and $RL_{accuracy}$ ($F_{17, 584.2}$=1.085, P= 0.271). The significance and magnitude of all other fixed effects included varied across trait performances. These effects are not directly relevant to hypotheses being tested and we do not discuss them further, but they are reported in full in the supplemental information (see Supplementary Information Table S3.1). Likelihood ratio tests (LRT) provided strong statistical support for the presence of among-individual variance in all seven traits tested (LRT of model 1 vs. model 0; all $P < 0.001$; Table 3.2). Repeatabilies (R) estimated conditional on fixed effects (i.e. as $R = V_I/V_P = V_I/(V_I+V_R)$) were low to moderate ranging from 20-46% with a median across the seven traits of 33.5% (Table 3.2).

The multivariate mixed model provided strong support for significant among-individual covariance in the **ID**-matrix. Likelihood ratio comparison of the full model to a reduced fit in which all cross-trait covariances in **ID** were set to zero showed the former was a significantly better to fit to the data ($\chi 2_{21}$=242.756, P<0.001). Among traits, the estimated covariances/correlations were positive between all trait pairs (Table 3.3). Boldness, as captured by *emergence* was

strongly (and significantly; based on assuming approximate 95% CI of r_I ±1.96 SE) correlated with cognitive performance as measured by all traits in the association learning task ((r ± (SE)), AL_{time} = 0.663 (0.095), ALspeed = 0.818 (0.046), ALaccuracy = 0.788 (0.062); Table 3.2), and to decision accuracy within the reversal learning task ($RL_{accuracy}$ = 0.412 (0.171)).

Among the subset of 6 cognitive traits all estimated among-individual correlations (r_I) were positive (Table 3.3). Among-individual correlations were strongly positive among all traits in the association task ($r_{ALtime.\ ALspeed}$ = 0.758 (0.074); $r_{ALtime.\ ALaccuracy}$ = 0.729 (0.092); $r_{ALspeed.\ ALaccuracy}$ = 0.894 (0.039)) providing no support for a speed accuracy trade-off. Similarly both traits in the reversal learning task ($r_{RLtime.\ RLaccuracy}$ = 0.473 (0.144)) were positively correlated, and this same pattern of positive correlations was also found across traits putatively indicative of different cognitive domains. There was thus no evidence of negative correlations as predicted by domain-specific trade-offs and rather the structure of **ID** is consistent with the domain-general intelligence model (Table 3.3).

Eigen decomposition of $\mathbf{ID_{rc}}$ reflects this, with the first major axis (first principle component, PC1, with 95% confidence intervals from 5000 bootstrap replicates) explaining 57.1% (42.7%, 69.8%) of the among-individual variation in multivariate phenotype on a correlation scale. Subsequent vectors necessarily explain sequentially diminishing amounts of variation (PC2 = 25.4% [12.6, 30.9]; PC3 = 12.1% [7.4, 15.6]; PC4 = 3.7% [4.7, 9.9]; PC5 = 1.6% [0.8, 5.5]; PC6 > 0.001% [0, 0]). All 6 cognitive traits load with the same sign on PC1 and these loadings are statistically significant based on bootstrapped confidence intervals not overlapping zero (Figure 3.3; Supplementary Information Table S3.2). Thus the eigen decomposition reiterates the view that the correlation structure is

consistent with a dominant axis of among-individual variation in cognitive phenotype caused by differences in underlying domain-general intelligence *g*. On this axis, individuals at one end of the axis can be considered to have 'higher/better' performance in all tasks (i.e. are faster and more accurate in the association task, faster to solve the motor cognition task, and faster and more accurate in the reversal learning task when the cue is reversed), while individuals at the other end perform relatively poorly in all respects.

3.5 Discussion

In this study we sought to determine whether – and to what extent – there exists among-individual variation for cognitive performance in a captive population of recently wild-derived guppies. Adopting a multivariate approach allowed us to test for covariation among multiple cognitive domains, as well as between cognitive performance and a shy-bold type axis of personality variation. Our study yields three main findings. First, there is among-individual variation underpinning all cognitive traits. Second, there is strong among-individual correlation structure among cognitive traits and between cognition and bold type personality. Thirdly, the structure of **ID**-matrix provides no support for the presence of important performance trade-offs among specific traits or cognitive domains tested. Rather we suggest that it is consistent with a single axis of domain-general variation that causes positively correlated performance across cognitive tasks, such that individuals at one end of the axis can be considered to have relatively low 'general intelligence' compared to those at the other end of the axis. The structure of **ID**-matrix is therefore consistent with expectations under the *g*-model, although it is important to acknowledge that other latent factors could exist that create similar

expectations of covariation among observed behaviours. In what follows, we discuss these findings in the context of understanding variation in, and evolution of, animal cognition, while also drawing attention to important assumptions and caveats underpinning our conclusions.

Among-individual variation was found in each of the behavioural traits and tasks examined. That guppies show consistent among-individual differences in 'boldness', as measured here by time to emerge into the test compartment over the first set of 9 trials (with a conditional repeatability of 43.9%) was expected given earlier studies of this population (e.g. Houslay et al., 2019) and guppies more generally (Brown et al., 2014; O'Neill et al., 2018; Gasparini et al., 2019). Indeed similar findings are common across a wide range of fish models used in behavioural studies (Boulton et al., 2018; White et al., 2020). We also find among-individual variation in performance in all cognitive traits across each task. Thus, we conclude that individuals differ in: ability to discriminate between rewarded and unrewarded colours in an association learning task; ability to learn a novel motor task; and, in cognitive flexibility as tested by a reversal learning task. The estimated repeatabilities of our observed behavioural proxies are moderate relative to behavioural traits generally (Bell et al., 2009), but generally consistent with reports for cognitive assays in animals (Cauchoix et al., (2018)). A contributing factor is likely to be the short inter-observation period used. Here this was usually less than 24hrs and in fact up to 9 observations were made in any single 48hr period. This sampling frequency is quite typical for cognitive studies, and the short intervals between trails may explain a decrease in performance due to reduced motivation. Furthermore, since behavioural repeatabilities generally decline as inter-observation period increases (Boulton et al., 2014), this may also partially explain the strong signal of repeatability.

Our multivariate modelling provides no support for trade-offs among cognitive traits. This conclusion stems from the finding that the correlation structure in **ID** is universally positive. Thus, for example, speed-accuracy trade-offs are not detected in our association learning trials. Speed-accuracy trade-offs are widely predicted to arise where behavioural outcomes depend on cognitive decision making (Briffa, 2013), although this prediction assumes that faster decisions are better (all else being equal). They have been detected in some animal studies, for instance in Carib grackels (*Quiscalus lugubris*) where individuals that solve novel problems faster also make more errors in discrimination learning tasks (Ducatez et al., 2019). Here we actually find the opposite pattern, faster individuals are more accurate, a result that seems counterintuitive but is consistent with reports in song sparrows (*Melospiza melodia*; (Boogert, Anderson, et al., 2011), and bumblebees (*Bombus terrestris;* Raine, 2012). We note that the correlation structure of behavioural **ID**-matrices can be sensitive to environmental context (Houslay et al., 2018) as a consequence of IxE interactions (among-individual differences in plasticity; Nussey et al., 2007). It is therefore possible trade-offs could be apparent only under specific conditions and/or assays. This scenario is suggested by a recent study of archerfish in which a speed-accuracy trade-off was only statistically significant under the more challenging of two cognitive testing treatments applied (Jones et al., 2020). Plastic responses to social context could also be important since, for instance, in less competitive scenarios (including the single housing used here) slow decisions may carry minimal costs removing motivation to be fast (Sih et al., 2012).

More broadly, we also find positive correlations in performance across traits picked as proxies of different cognitive domains. This is counter to

predictions under hypothesised trade-off models for maintaining variation. For instance, it has been hypothesised that individuals more reactive to changes in environmental cues will perform relatively poorly in association learning but – being cognitively flexible – will do well in reversal learning tasks (Sih et al., 2012; Bebus et al., 2016). Here we find a strong positive among-individual correlation between our measures of performance in association and reversal learning (r_I=0.54). Again we note the absence of apparent trade-offs has precedent in the animal cognition literature. For example, positive correlations have been reported between discrimination acquisition and reversal learning in sparrows (Boogert et al., 2011), social learning and innovative problem solving in pigeons (*Columba livia*; Bouchard et al., 2007), motor skill learning and reversal learning in pheasants (van Horik et al., 2018).

If positive correlations in the **ID**-matrix provide no support for trade-offs, then the corollary is that they are consistent with a domain-general axis of cognitive variation among-individuals in this population. Thus, excluding the boldness proxy of emergence time and focusing on cognitive traits only, we found that 57% of the variation in correlation matrix was explained by the leading principle component (or eigen vector). By comparison in tests of the general intelligence model of cognition using human psychometric test batteries, an underlying *g*-factor is typically found to account for about 40–50% of the multivariate variation (Carroll, 1993; Deary, 2001). Under the *g*-model, trait performance in cognitive assays across multiple domains are expected to load strongly (and in the same direction) on a dominant first principle component (Plomin, 2001; Plomin et al., 2002). This is the pattern we found. Our results thus add to a small, but growing body of empirical literature finding support for domain-

general variation among-individuals in animal cognitive ability (Shaw et al., 2015; Ashton et al., 2018).

Our conclusion that (multivariate) cognitive variation in guppies is consistent with an underlying general intelligence factor requires some caveats. These largely arise because latent variables are, by definition, unmeasurable. Thus, all inferences about (latent) cognitive traits from (observed) behavioural performance measures are inevitably subject to assumptions. First, we are limited by the set of traits assayed. Inclusion of more cognitive domains, different assays, and/or different contexts, would provide more robust insight into the biological extent and evolutionary potential of general intelligence (Poirier et al., 2020). Thus, for instance, we detect no evidence of domain-specific cognitive trade-offs, but they could occur among a wider set of traits (or contexts) not examined. Here we targeted the domains of association learning, motor cognition and reversal learning, but did so using a single testing paradigm in which performance measures share a requirement for individuals to discriminate between two coloured cues. This strategy may be predisposed towards finding positive correlation structure. Second, the eigen vector decomposition provides only a statistical description of the estimated correlation structure among observed traits; that **ID** is consistent with an axis of variation in an underlying general intelligence factor does not prove such a factor exists. Third, the positive correlation structure could be explained, at least in part, by shared dependence of the observed traits on some other parameter that varies among individuals such as metabolic state (Biro et al., 2010; Mathot et al., 2013; McKenzie et al., 2016), satiety (Shettleworth, 1972; Ben-Shahar et al., 2001), or other drivers of motivation (van Horik et al., 2016). For example, the increase in mean performance time in the association task could be confounded by changes in

motivation due to trial fatigue (reduced motivation), at least for some individuals, even if some individuals clearly did perform better over time. Experimental conditions can have important consequences for robust interpretation of studies targeting cognitive traits (Rowe et al., 2014; Griffin et al., 2015; Boogert et al., 2018) and while we attempted to minimise 'non-target' sources of among-individual variation (e.g. by standardising food rations, housing and water chemistry) we cannot exclude the possibility that these contribute to the structure of **ID**. Finally, we acknowledge the potential for bias in our findings caused by selective removal of individuals due to our imposed learning criterion. Failure to participate in a cognitive test may result in sampling biases when measuring among-individual variation in cognitive performances (Thornton et al., 2012a; van Horik et al., 2017). Our learning criterion was specifically weak so that all included individuals represented, as far as possible, among-individual variation of the initial population.

We also found that all six cognitive performance traits were positively correlated with 'boldness', measured here as an individual's tendency to consistently emerge rapidly from the home compartment during trial observations. Pairwise correlations were nominally significant in four of the six cases. Our finding that bolder individuals perform better in cognitive tasks mirrors results reported from several other studies. For instance, boldness has been positively linked with associative learning in rainbow trout (*Oncorhynchus mykiss*; Sneddon, 2003), shape discrimination in Cavies (*Cavia aperea*; Guenther et al., 2014a), and reversal learning in the Chimango Caracara, (*Milvago chimango;* Guido et al., 2017). While shy-bold type behavioural variation can have important fitness consequences (A. D. M. Wilson et al., 2010; Ariyomo et al., 2012; Ballew et al., 2017), estimates of contemporary selection on boldness in wild guppies

are lacking. Nonetheless, multiple studies have reported among-population differences in boldness mapping to differences in predation, a link that is consistent with the hypothesis that variable levels of risk could impose balancing selection that maintains variation in boldness. Very speculatively, if this is true, and if correlations in **ID** are recapitulated in **G,** the link between boldness and cognitive performance could help maintain variation in the latter. This is because selection on any one trait (e.g. boldness) has consequences for the evolutionary dynamics of genetically correlated traits (Roff, 2002). We do know that shy-bold type personality variation is heritable in the population of guppies under lab conditions, (White & Wilson, 2019; White, Houslay, et al., 2019; Prentice, Houslay, et al., 2020) but not how selection is acting in the field.

In summary, we find evidence that individual guppies do differ in cognitive performance as measured by a testing protocol that targets association leaning, motor cognition and cognitive flexibility (reversal learning). We find no support for trade-offs occurring either between speed and accuracy (e.g. within the association learning trials) or between overall performance across cognitive domains. Subject to the caveats highlighted above, the absence of domain-specific trade-offs and the presence of strong positive domain-general correlations in **ID** mean that the general intelligence model provides a good description of the structure of multivariate variation. Our results also suggest that, to the extent that directional natural selection favours higher performance in any single cognitive domain, it would lead to positively correlated evolutionary responses across all domains. However, this inference is dependent on **ID** being a reliable proxy for the **G** matrix (Brommer, 2013). Greater efforts to quantify **G** is thus an obvious, if challenging, next step towards understanding the maintenance of variation in and evolutionary dynamics of animal cognition.

Table 3.1: Outline of behaviours used to assess performance within each putative cognitive domain, including details of corresponding set number at which each trait was measured and corresponding acronyms used within the text.

Set Number	Cognitive/ Behavioural Domain	Trait	Acronym
1	**Boldness**	Time to emerge (Time taken to emerge into the test compartment once the compartment door is opened)	*emergence*
1 – 3	**Association learning**		
(*1*)	Speed accuracy trade-offs	Decision speed (After emergence into test compartment, time taken to make first choice, measured as swimming within 1 body length of, and actively exploring correct counter.)	AL_{speed}
		Decision accuracy (Decision of first choice, choice measured as swimming within 1 body length of, and actively exploring correct counter. Either correct or incorrect)	$AL_{accuracy}$
(*2,3*)	Performance	Time to eat (After emergence into test compartment, time taken to eat reward from the correct well)	AL_{time}
4, 5	**Motor cognition**	Time to eat (After correct choice made (i.e. swimming within 1 body length of, and actively exploring correct counter), time taken to dislodge counter and eat the food reward)	MC_{time}
6, 7	**Cognitive flexibility**	Time to eat (After emergence into test compartment, time taken to dislodge disc and eat reward from the correct well)	RL_{time}
		Decision accuracy (Decision of first choice, choice measured as swimming within 1 body length of, and actively exploring correct counter. Either correct or incorrect)	$RL_{accuracy}$

Table 3.2: Estimated among-individual variance (V_I) and adjusted repeatability (R) from univariate models of all traits. Standard errors are shown in parentheses. Also shown are likelihood ratio tests of V_I.

Cognitive Domain	Trait	V_I	R	$X^2_{0,1}$	P
Personality - Boldness	*emergence*	0.415 (0.077)	0.439 (0.049)	206.004	<0.001
Associative Learning	AL_{speed}	0.442 (0.081)	0.460 (0.049)	220.393	<0.001
	$AL_{accuracy}$	0.202 (0.045)	0.264 (0.046)	79.989	<0.001
	AL_{time}	0.341 (0.074)	0.335 (0.050)	186.345	<0.001
Motor Cognition	MC_{time}	0.275 (0.065)	0.329 (0.054)	171.462	<0.001
Reversal Learning	RL_{time}	0.372 (0.088)	0.375 (0.058)	178.426	<0.001
	$RL_{accuracy}$	0.182 (0.050)	0.203 (0.047)	65.327	<0.001

Table 3.3: ID variance–covariance–correlation matrix from the multivariate mixed model. Estimated variances are shown on the diagonal (dark grey shading), with correlations above and covariances below. Standard errors are shown in parentheses, and bold font denotes nominally significant estimates assuming approximate 95% CI of ± 1.96SE.

ID Matrix	*emergence*	AL_{speed}	$AL_{accuracy}$	AL_{time}	MC_{time}	RL_{time}	$RL_{accuracy}$
emergence	**0.426 (0.0773)**	**0.818 (0.046)**	**0.788 (0.062)**	**0.663 (0.095)**	0.239 (0.164)	0.166 (0.176)	**0.412 (0.171)**
AL_{speed}	**0.361 (0.072)**	**0.458 (0.082)**	**0.894 (0.039)**	**0.758 (0.074)**	0.271 (0.155)	0.228 (0.169)	**0.397 (0.168)**
$AL_{accuracy}$	**0.242 (0.052)**	**0.285 (0.056)**	**0.221 (0.046)**	**0.729 (0.092)**	0.124 (0.175)	0.090 (0.184)	0.258 (0.190)
AL_{time}	**0.286 (0.072)**	**0.340 (0.076)**	**0.227 (0.055)**	**0.437 (0.092)**	**0.747 (0.082)**	**0.540 (0.129)**	**0.423 (0.161)**
MC_{time}	0.092 (0.067)	0.108 (0.067)	0.034 (0.049)	**0.291 (0.075)**	**0.346 (0.078)**	**0.817 (0.067)**	**0.340 (0.167)**
RL_{time}	0.074 (0.082)	0.106 (0.083)	0.029 (0.060)	**0.246 (0.085)**	**0.331 (0.083)**	**0.474 (0.112)**	**0.473 (0.144)**
$RL_{accuracy}$	**0.122 (0.060)**	0.122 (0.061)	0.055 (0.043)	**0.127 (0.059)**	0.091 (0.051)	**0.148 (0.062)**	**0.205 (0.056)**

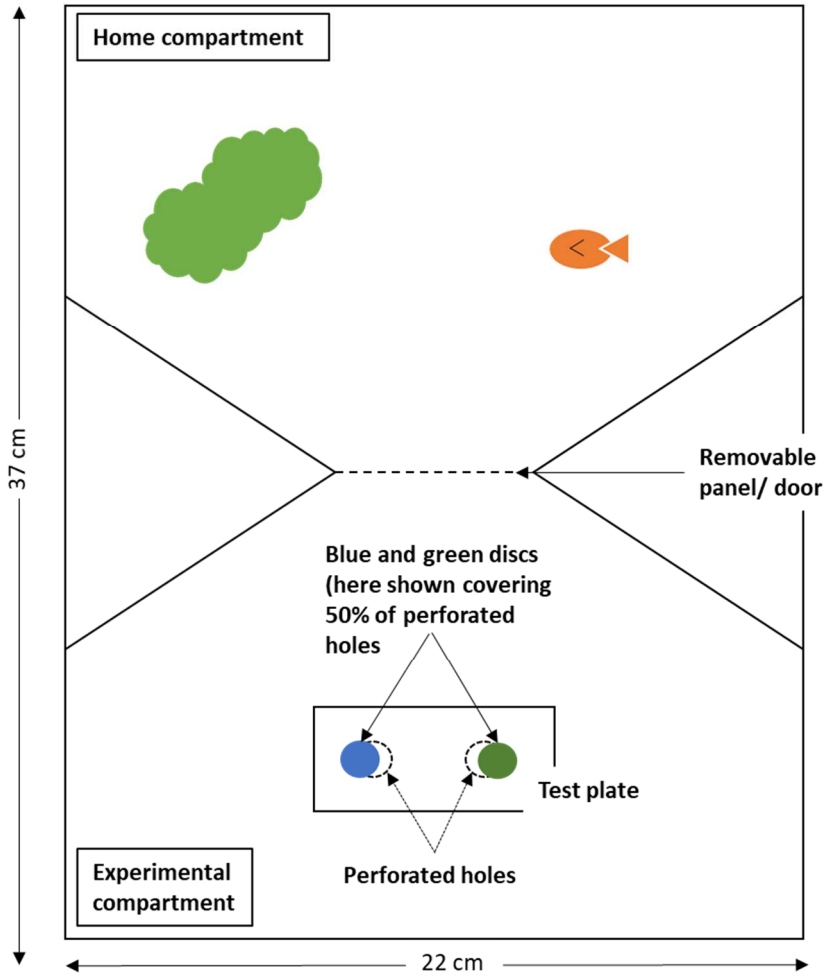


Figure 3.1: Aerial view of the home tank used to house individual fish, showing the experimental compartment and test plate used during each task.

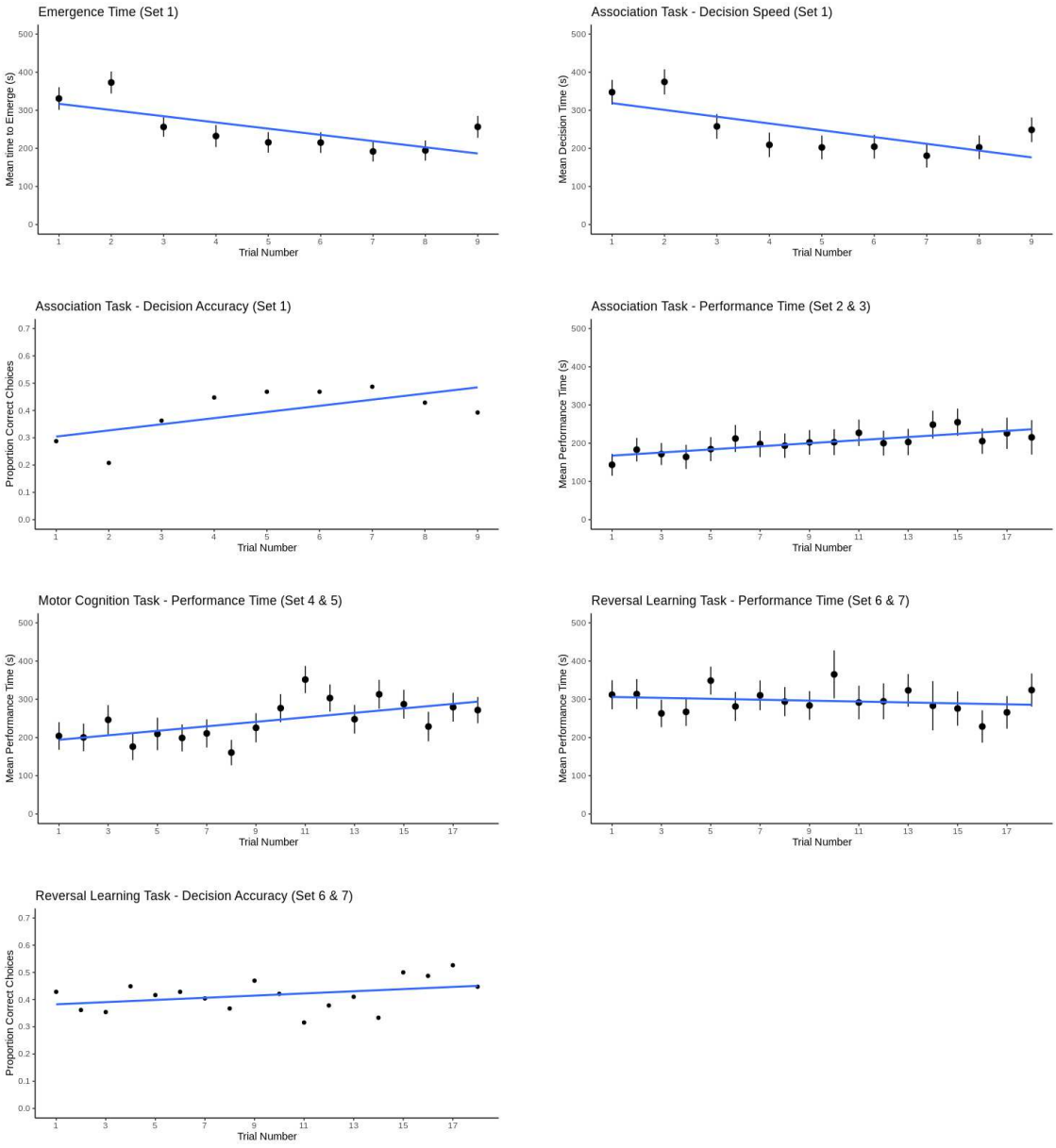


Figure 3.2: Plots of raw data of total performance time across outlined sets and trials used to assay each trait. Emergence time (*emergence)*, association learning task performance time *(*AL_{time}), association learning decision speed (AL_{speed}) and accuracy ($AL_{accuracy}$), motor cognition task performance time (MC_{time}), reversal learning task performance time (RL_{time}) and accuracy ($RL_{accuracy}$). Error bars represent mean and standard errors of performance time

for individuals. Blue line represents regression line through mean performances at each trial.

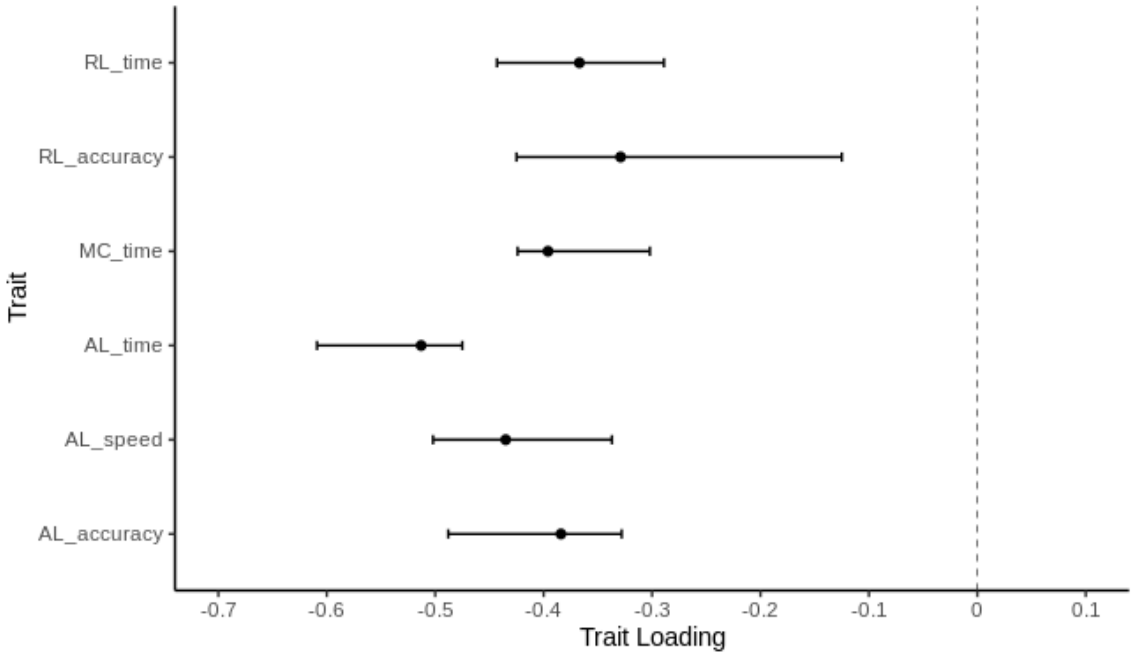


Figure 3.3: Trait loadings from the first eigen vector (principal component, PC1) of the **ID**-matrix. This axis explains 57.1% of the among-individual (co)variation found in the components of the cognitive phenotype in our guppy population. Points show trait loadings from the first eigen vector of our estimate of **ID**, with bars representing 95% confidence intervals on each loading (calculated from 5000 bootstrapped replicates of the model)

4 Chapter 4: Genetic variance for behavioural 'predictability' of stress response

This paper is published as: **Prentice, P.M**., Houslay, T.M., Martin, J.G.A. and Wilson, A.J., 2020. Genetic variance for behavioural 'predictability' of stress response. *33(5*), 642–652.

4.1 Abstract

Genetic factors underpinning phenotypic variation are required if natural selection is to result in adaptive evolution. However, evolutionary and behavioural ecologists typically focus on variation among individuals in their average trait values, and seek to characterise genetic contributions to this. As a result, less attention has been paid to if and how genes could contribute towards within-individual variance, or trait "predictability". In fact, phenotypic 'predictability' can vary among individuals, and emerging evidence from livestock genetics suggests this can be due to genetic factors. Here we test this empirically using repeated measures of a behavioural stress response trait in a pedigreed population of wild-type guppies. We ask (1) whether individuals differ in behavioural predictability, and (2) whether this variation is heritable and so evolvable under selection. Using statistical methodology from the field of quantitative genetics, we find support for both hypotheses and also show evidence of a genetic correlation structure between the behavioural trait mean and individual predictability. We show that investigating sources of variability in trait predictability is statistically tractable, and can yield useful biological interpretation. We conclude that, if widespread,

genetic variance for ‘predictability’ will have major implications for the evolutionary causes and consequences of phenotypic variation.

4.2 Introduction

Among-individual variation in behavioural traits is widely referred to as ‘animal personality’ when individuals display behaviours that are repeatable across time and context (Gosling, 2001; Bell et al., 2009). Individual differences, which can sometimes account for a high proportion of the total observed behavioural variation in a population (Biro et al., 2013)**,** are a pre-requisite for natural selection on behaviour and there is now abundant evidence that personality traits can affect fitness. Since strong directional or stabilising selection is usually predicted to erode variation, it is widely hypothesized that personality variation within populations is maintained by fitness trade-offs (Godin et al., 1995; Réale et al., 2003; Dingemanse et al., 2004; Shackleton et al., 2005). Empirical investigations of this, and related hypotheses, have been facilitated by wide uptake of linear mixed effect models that allow partitioning of among-individual trait variation and estimation of behavioural repeatabilities (R) from data containing repeated observations of known individuals (Dingemanse et al., 2013; Wilson, 2018) . Where pedigree or relatedness data are also available, among-individual variance can be further decomposed to estimate behavioural heritability (Wilson et al., 2010). While the residual, or within-individual, component of variance is normally treated as ‘noise’ arising from plasticity (Nussey et al., 2007) and/or measurement error, some authors have argued that it deserves more attention

as a source of biological insight (Westneat et al., 2015)**.** Here we follow this suggestion in a study of stress-related behaviour in wild-type guppies (*Poecilia reticulata*), and ask whether within-individual variance in behaviour should itself be viewed as a trait that can respond to selection.

A convenient, but rarely scrutinised assumption of typical statistical methods used to characterise personality is that within-individual (or residual) variation in behaviour is homogeneous across individuals. However, this need not be the case. In fact residual variation can itself differ among-individuals (Stamps et al., 2012), a phenomenon variously referred to as among-individual differences in 'within-individual behavioural variance', 'intra-individual variability', or 'consistency' (Müller et al., 2005; Stamps et al., 2012; Biro et al., 2013). Here we refer to this phenomenon as 'predictability' following terminology used by Cleasby et al (2015) and Martin et al (2017). Predictability has been the focus of some studies in human psychology (MacDonald et al., 2006; Hoffman, 2007) but, in recent years, has become a topic of interest in behavioural ecology (Stamps et al., 2012; Westneat et al., 2013). For example, recent empirical studies have provided evidence of variation among individuals in predictability of anti-predatory behaviours (Briffa, 2013) and parental care (Westneat et al., 2013). It is perhaps easy to envisage fitness consequences of within-individual variation (e.g., animals behaving less predictably when fleeing a predator could plausibly have a higher escape probability), although empirical estimates of selection are scarce. Specific hypotheses for the maintenance of among-individual differences in behavioural predictability are also lacking, though trade-offs among associated

life history traits could offer adaptive explanations just as they do for maintenance of variation in behavioural means (Westneat et al., 2013; Bridger et al., 2015; Mulder et al., 2016). ‘Speed-accuracy’ trade-offs have been proposed where observed behavioural outcomes depend on cognitive decision making (Briffa, 2013) and predictability could also be condition-dependent. In the latter case if, for instance, canalising a behavioural response is costly then low predictability may represent phenotypic instability caused by poor individual condition.

Behavioural predictability has also been postulated to have a genetic basis of variation among individuals (Martin et al., 2017). The implication is that if predictability both causes fitness variation and is heritable, it can itself be viewed as a trait that will evolve under natural selection. Though empirical tests of genetic variance for behavioural predictability are scarce more is known for non-behavioural phenotypes. In particular, quantitative genetic methods (Hill, 1984; Hill et al., 2004; Rönnegård et al., 2010) have been increasingly applied to estimate genetic variation for predictability of production traits in livestock, including milk yield (Rönnegård et al., 2013), litter size (Sorensen et al., 2003) and body weight (Sonesson et al., 2013). While increasing the mean of such production traits is a long-standing objective of artificial selection strategies, reducing the level of variation around the means also offers increased efficiency (and profitability) in livestock production and processing. Consequently, the reality that genetic variance ‘for variance’ occurs has prompted development of strategies to select more predictable genotypes, and thus reduce variation in target traits (Hill et al., 2010).

Evolutionary ecologists are now beginning to address the concept of predictability in relation to behavioural and life history traits with the goals of determining whether behavioural predictability consistently varies among-individuals, whether this variation has a genetic basis, and how (if at all) predictability maps to fitness. A hindrance addressing these questions stems from a lack of consensus on how best to quantify and analyse predictability. Most studies to date have taken a two-step approach by, for instance, fitting a linear model to a set of behavioural observations, then calculating an estimate of within-individual variation using model residuals for each individual, which are then used in a subsequent analysis (Stamps et al., 2012; Biro et al., 2013; Highcock et al., 2014). Though intuitive, this approach is statistically problematic for a number of reasons, not least of which is that uncertainty in the predictions of the first model is not accounted for, increasing the risk of type 1 errors and anticonservative hypothesis tests (Houslay & Wilson 2017). Fortunately, a more robust approach to model variation in behavioural predictability is provided by the double hierarchical generalized linear model (DHGLM) developed by Lee and Nelder (2006). This model is an extension of the familiar 'random intercept' mixed model, however instead of only allowing random and fixed effects on the mean trait distribution, it also allows them on the residuals. In other words, it allows us to relax the assumption that residual variance is homogeneous, and ask whether it varies across levels of fixed (e.g. sex) or random (e.g. individual identity) effects (Cleasby et al. 2015; Lee & Nelder 2006).

Applied to repeated measures behavioural data, double-hierarchical models therefore allow simultaneous estimation of 1) among-individual variation in (mean) trait expression (i.e. the normal target of personality studies), 2) variation in predictability of a trait (i.e. differences in within-individual variance) (Lee & Nelder 2006) and 3) the correlation between the mean and the predictability at the individual level. Furthermore, given pedigree data, the DHGLM approach can be combined with the quantitative genetic 'animal model' (in a 'double-hierarchical animal model' DHAM), allowing among-individual variance to be further decomposed into genetic and non-genetic components. To date, only one study has used this approach to test for and estimate the genetic basis of behavioural predictability (Martin et al., 2017). Using a DHGLM, the authors of this study found evidence of among-individual variation in the predictability of docility (the reaction to being trapped and handled), as a repeatable behaviour in marmots (*Marmota flaviventris*). They also showed that individual marmots that were (on average) less docile were also less predictable. Using pedigree information, they went on to show that both (mean) behaviour and its predictability are heritable in this population, and so evolvable under selection.

Here we use a captive population of wild-derived Trinidadian guppies (*Poecilia reticulata)* to test for variation in behavioural predictability and ask whether, if present, it arises in part from genetic differences among individuals. We focus on a putatively stress-related context, specifically the way in which an individual behaves in reaction to isolation in a novel environment, such as an

‘Open Field Trial’ (OFT) arena. The OFT is a widely used paradigm for characterising personality differences related to exploration, activity, and ‘shy-bold’ type variation (Gosling, 2001; Bell et al., 2009). Previous work with *P. reticulata* has demonstrated that behaviours displayed during OFT are associated with exploration, but also risk-taking and stress response (White et al., 2016). The fact that the OFT presents a mild stressor is notable because the widely used concept of ‘stress coping style’ predicts that individuals vary along a proactive/reactive continuum of variation (Coppens et al., 2010; Koolhaas et al., 1999; Sih, Bell, & Johnson 2004), with proactive individuals tending to express more ‘fight or flight’ behaviours on average, but also forming more rigid, stereotyped routines more rapidly (Koolhaas et al., 1999). In other words, the coping style verbal model suggests variation among individuals in not only mean behaviour and behavioural predictability, but also correlation structure between these.

Previous studies of this guppy population have already shown that the behavioural responses to the OFT are repeatable, but also plastic with respect to experimentally-manipulated stressor severity (specifically perceived predation risk) (Houslay et al., 2018). We also now know from pedigree-based analysis that (average) behaviours are heritable (White & Wilson, 2019; White, Houslay, et al., 2019), and that there is genetic correlation structure between OFT behaviour and cortisol expression (strengthening the view that the OFT provides an assay of behavioural stress response; (Houslay et al., 2019)). Here we aim to build on these earlier studies by, firstly confirming the repeatability and heritability of mean

behaviour in an independent sample; secondly, simultaneously estimating among-individual variation in mean behaviour and predictability using a DHGLM; and thirdly, asking whether - if present – variation in predictability is itself heritable using a DHAM. Finally, we test the prediction of the stress coping style model – at both among-individual and genetic levels – that there will be (co)variance between mean behaviour and predictability, with individuals (genotypes) displaying more 'flight' type behavioural stress responses also being more predictable. In the context of a DHGLM, we are thus predicting to have a positive covariance between a 'flight' type behavioural response and its variance at the individual (genotype) level. The 'flight" type behavioural response here was a derived trait called relative area*,* where individuals that have a low relative area, i.e. displaying more flight type response, are expected to have a low within-individual variance (i.e. high predictability).

4.3 Methods

Husbandry

Behavioural assays were carried out on wild-type guppies from a captive population with known pedigree structure housed at the fish laboratory at the University of Exeter's Penryn campus. Data used here have not previously been published, but were collected as part of a larger study for which methods have already been extensively described elsewhere (White et al., 2016; White & Wilson, 2019; White, Houslay, et al., 2019; Houslay et al., 2019). In brief, all fish used were descended from wild guppies caught from the lower Aripo River,

Trinidad in 2008. They were offspring of known parental crosses (as detailed in White & Wilson, 2018, for detailed breeding protocol), that had been raised in families before being tagged at maturity and then allocated to mixed family groups. Groups comprised 16-20 individuals (at 50:50 sex ratio) in 15 l tanks, with 24 tanks within each "stack" having a common sump and shared recirculating water supply. Fish were maintained at 22–24°C on a 12:12 light/dark cycle, with weekly 25% water changes on each stack, and were fed to satiation twice daily on commercial flake food and live brine shrimp (*Artemia salina*). Note these fish were part of a larger pedigree structure containing 1,518 individual fish within a genetic pedigree structure comprised of maternal full-sibships nested within 169 paternal half-sibships (as described in Houslay et al., 2019). Here, we pruned the full pedigree using the prunePed function in the R package MCMCglmm (Hadfield, 2010) to just include the informative individuals. Our final data set contained phenotypic date for 330 individuals from a pedigree with 2113 maternal offspring links, 1654 paternal offspring links, 218 sires and 344 dams, with a maximum depth of 4 generations.

Behavioural data collection and trait definition

Behaviour was assayed using Open Field Trials (OFT), a standardised assay of risk-related behaviours that is widely used in rodent, fish and bird studies (Boulton et al., 2014; White & Wilson, 2019; White, Houslay, et al., 2019). Our assay protocol closely followed that of Boulton et al. (2014) with repeat measures on related individuals providing the data structure needed to estimate among individual and genetic variance in personality and predictability. However, here

we conducted more repeats per individual with a planned maximum of 10 times. In practice some mortality occurred over the course of the data collection period (which was five weeks for each fish). Thus, in total we conducted 2970 behavioural assays on 330 individuals (a mean of 9 per fish) from 23 groups. All experimental data was collected by the same technician, and carried out in two blocks for purely logistical reasons (Batch A; n = 176, Batch B; n= 154). For each block, fish were trialled over five weeks, with data collection occurring in weeks 1, 3 and 5 at not less than 48 hour intervals (weeks 2 and 4 providing 'breaks').

Each OFT comprised a fish being netted from its home tank and placed into an 'arena' comprised of a 30 × 20 cm fish tank filled to 5 cm water depth and lit from below with a light box. (Three identical arena 'set-ups' A, B and C were used concurrently during data collection to facilitate high throughput phenotyping, with fish allocated haphazardly among them). Following a 30 s acclimation period, individuals were tracked for 180 s from a Sunkwang C160 video camera fixed above each tank and the tracking software Viewer II (http://www.biobserve.com). Each fish was then returned to its home tank. Behavioural experiments were conducted under license from the Home Office (UK) and under the auspices of the Animal (Scientific Procedures) 1986 Act, and with local ethical approval from the University of Exeter.

A number of specific variables assayed by OFT have been used to assay 'risk-prone/risk-averse', or 'shy-bold' type personality variation in fishes including guppies (Sih, Bell, & Johnson, 2004; White & Wilson, 2019). Here we extracted two variables from the video - total track length swum (cm) and the area covered

(percent of tank area explored, %). While both are expected a priori to be repeatable and heritable (Houslay et al., 2018; White & Wilson, 2019; White, Houslay, et al., 2019), previous work has failed to detect a strong positive (among-individual) correlation. This is notable since, if fish move randomly in the OFT arena, we expect area covered to increase as a monotonic function of track length. The lack of expected correlation actually arises from variation in how fish respond behaviourally to the stressor stimulus of the OFT. This variation is revealed by calculating the derived trait of relative area –defined as the difference between observed area covered and the predicted area covered given a 'random swim' of the track length actually observed (Houslay et al., 2019). To do this we (i) simulated 'random swims' in the arena across the full range of observed track lengths; (ii) estimated the 'null' relationship between simulated area and simulated track length using a fourth order polynomial regression (which captured 97.85% of the variation); and (iii) used the regression equation to predict area covered given a 'random swim' corresponding to each observed track length. Code and a full description of the simulation approach is provided in Houslay et al. (2019).

Biologically, high values of relative area arise from efficient exploration of the arena by a (putatively) less stressed individual (Figure 4.1 a). In contrast, low values of relative area arise from trials in which fish swim rapidly (yielding a high track length) but also display thigmotaxis (i.e. staying close to the tank wall resulting in a low area covered) (Figure 4.1a). This scenario is commonly observed and is biologically interpretable as a 'flight'-type stress response (i.e.

the fish is seeking escape from the arena).

Statistical analyses

First, we sought to confirm our expectations from previous work on this population that there would be among-individual variation for mean behaviour (relative area). We did this using a simple univariate linear mixed-effects model fit by REML (restricted maximum likelihood) using ASReml-R 3.0 (Butler 2009; Gilmour et al. 2002) within R version 3.4.1 (R Core Team, 2017) in which relative area was modelled with random effects of individual identity and social housing group (Model 1). In addition to the mean, we included fixed effect factors of arena set-up and fish sex, as well as within-group trialling order (as a continuous variable to account for any cumulative disturbance effect of removing fish sequentially from the home tank). Conditional F-statistics were used to determine the significance of fixed effects although we note they were simply included to control statistically for sources of variance not directly relevant to our present goals. Random effects were tested using likelihood ratio tests (LRT), assuming twice the difference in log-likelihood between full and reduced models is distributed as a 50:50 mix of χ^2_0 and χ^2_1 as recommended by Visscher (2006). We make the standard assumptions that random effects and residuals are normally distributed with means of zero and variances to be estimated. Importantly in the current context we also make the standard (but rarely stated) assumption that 'residual' variance is homogeneous across individuals (and fixed effect classes). We also calculated an estimate of the adjusted repeatability (conditional on fixed effects) as the intraclass correlation $R = V_I/V_P$, where V_I is the among-individual variance

and V_P is the total phenotypic variance. V_P is therefore calculated as $V_I + V_{GR} + V_R$, where V_{GR} is the among-group variance (which accounts for environmental and social sources of variation among groups within home tanks) and V_R is the residual (within-individual) variance. The adjusted repeatability R is thus the proportion of phenotypic variance explained by among-individual differences in behavioural mean, after controlling for fixed effects (Nakagawa et al., 2010).

We then extended this model by including the individual genetic merit for (mean) behaviour as an extra random effect (Model 2). This becomes the standard repeated measures animal model of quantitative genetics (with additional fixed and random effects as described above), and allowed us to utilise the pedigree data to partition V_I into additive genetic (V_A) and non-genetic, permanent environment (V_{PE}) components. We tested the significance of V_A by LRT (as described above) and estimated the narrow sense heritability h^2(where h^2=V_A/V_P and V_P is the sum of the variance components and thus conditional on fixed effects).

To estimate among-individual variation of predictability of the behaviour (relative area), we used a double hierarchical generalized linear effect model (DHGLM) (Lee et al., 2006; Cleasby et al., 2015) of relative area (Model 3). The DHGLM allows for the simultaneous analysis of a mean level model and a dispersion level model each including fixed and random effects. We estimated not only the among-individual variation in residual variance (i.e. variation in predictability (V_{I_v})) but also the correlation between the mean behaviour and its predictability at the individual level (Cleasby et al., 2015). To simplify slightly, we

included as fixed effects in the mean model only those variables that were statistically significant in Model 1, while for the dispersion part of the model, we included a fixed effect of sex (i.e. males and females are permitted to differ in average predictability). We included group and individual identity as random effects in both the mean and the dispersion part of the model. We also modelled the covariance (at group and individual levels) between the random means and the predictabilities of relative area.

Finally we extended Model 3 to include random genetic effects on both the mean and the predictability of the behaviour in a double hierarchical animal model (DHAM) (SanCristobal-Gaudy et al., 2009; Rönnegård et al., 2010). The DHAM thus allows us to partition among-individual (co)variance into genetic and non-genetic (permanent environment) components using the pedigree. This DHAM (Model 4) has the same fixed effect structure as Model 3 for both mean and dispersion parts of the model. For the random effects, we included a permanent environment, an additive genetic and a group effect on both the mean and the dispersion models. Thus the double hierarchical models (Model 3 and 4) relate to each other in the same way as the 'normal' mixed models with random effects on the mean behaviour only (Models 1 and 2). In both model 3 and 4, the residual variance is dependent on the fixed and random effects included in the dispersion part of the model. However, it is possible to estimate an average residual variance for DHGLMs, $\overline{V_R}$. Assuming fixed effects in the dispersion part of the model are centred, we can estimate the (average) residual variance in model 4 as follow:

$$\overline{V_R} = exp\left(\eta + \frac{V_{PE_v}}{2} + \frac{V_{A_v}}{2} + \frac{V_{GR_v}}{2}\right)$$

where η is the intercept of the dispersion model and V_{PE_v}, V_{A_v} and V_{GR_v} are the variance components associated with the permanent environment, genetic and group random effects in the dispersion part of the model respectively.

Both DHGLM and DHAM were fitted in a Bayesian framework using Stan (Carpenter et al., 2017) within R version 3.4.1 (R Core Team, 2017) via the package RStan version 2.18.0 (Stan Development Team, 2018). In order to optimize model specification in Stan, the residual variance was modelled on the log-normal scale and the covariance matrices of random effects were estimated as standard deviations and correlation matrices. We used uninformative (or weak) priors on all parameters. For fixed effect priors we used a normal distribution with mean of zero and a variance of 100. We used a half-cauchy distribution (cauchy(0,5)) for standard deviations and for the correlation matrices, we used a LKJ correlation distribution, parametrized in terms of its Cholesky factor (allowing for a uniform distribution between -1 and 1 for the correlation). Model 3 was fitted using 5 Markov chains each including 6,000 iterations, 2,000 burn-in iterations and a thinning interval of 10. Model 4 was fitted using 5 chains each with 43,000 iterations, 3,000 burn-in iterations and a thinning interval of 100. Convergence was first assessed by visually inspecting the trace plots, which were also used to identify an appropriate number of burn-in iterations. We then checked that the Monte Carlo error was less than 1-5% of the posterior standard deviation, that the Brooks-Gelman-Rubin (BGR) diagnostic converged to 1 ± 0.2

and that the autocorrelation was below 0.05 for all parameters (Kass et al., 1997). The mode and 95% Highest Posterior Density Intervals (HPDI) were used to summarise the posterior distributions of the model parameters. For all calculated parameters (e.g., R, h^2), the parameters are calculated at each iteration and we reported their posterior mode and HPDI. Consequently, their estimates might differ slightly from the calculation done directly on the posterior mode of their components.

4.4 Results

Model 1 revealed significant among-individual variation in relative area, $\left(repeatability(with\ SE), R = 0.288\ (0.024), \chi^2_{0,1} = 517.44, P < 0.001\right)$ (Table 4.1). Under this model the social group effect was also significant ($\chi^2_{0,1} = 10.63, P = 0.001$ though it only explained 3.7% (1.9%) of the total variance. Comparison between model 1 and 2 provided strong evidence for significant additive genetic variance in mean behaviour $\left(\chi^2_{0,1} = 10.88, P < 0.001\right)$, with the latter yielding an estimated heritability for relative area of $h^2 = 0.110\ (0.052)$, conditional on fixed effects (Table 4.1). These results are consistent with previously reported estimates using independent data from the same population ($h^2 = 0.080\ (0.003)$) (Houslay et al., 2019), but note that both estimates are conditional on fixed effects that differ slightly. Fixed effects estimated from the current (and subsequent) models are presented in the Supplementary Information (Table S4.1 and S4.2, Appendix 4) for completeness, although are not directly relevant to our hypotheses in this study.

The mean part of model 3 yields very similar point estimates of individual and group level variances to model 1, although the posterior of the latter was not very clearly distinct from zero. The repeatability of (mean) behaviour under Model 3 is the same as that obtained in the standard repeat measures mixed model (Model 1) with R =0.288 (95% CrI , 0.248-0.348). More notably Model 3 provided evidence of among-individual variance in predictability, and also of a strong negative correlation between the individual mean and predictability of behaviour (Table 4.1). While frequentist-type P values are not applicable given the Bayesian inference, the 95% credible interval of the individual level variance in the dispersion part of the mode (V_{I_v}) is clearly distinct from zero. Similarly, the credible interval of the individual mean-predictability correlation is narrow and does not span zero (from which we can conclude statistical 'significance').

These findings are mirrored at the genetic level. Thus Model 4 yields very similar estimates for the heritability of (mean) behaviour. However, the DHAM also shows that both the variation in individual predictability and the mean-predictability correlation estimated in Model 3 have a genetic basis (Table 4.1). More precisely, both the genetic and permanent environment correlations between the mean behaviour and the variation (i.e. predictability) are strongly negative with 95% CrI that do not overlap zero (Table 4.1, Figure 4.2). Thus, individuals - and genotypes – that are more explorative (express high mean relative area) are also more predictable (i.e. less variable) in their behavioural response to the OFT. Point estimates of the corresponding group level correlations are similarly strongly negative, though we reiterate that the amount

of variance in mean behaviour explained by group is low. Fixed effect estimates from all models are not discussed here but are reported in full in the Supplementary Information (Supplementary Information, Tables S4.1 and S4.2).

4.5 Discussion

Here, we show evidence of among-individual variation in stress-related behaviour in the guppy *P. reticulata* and show that variation arises partly through heritable differences among fish. The present data thus provide confirmation of earlier results showing genetic variation for individual mean behaviours expressed during open field trials (Houslay et al., 2018, 2019; White & Wilson, 2019; White, Houslay, et al., 2019). However, while previous analyses were limited to individual means, we now also show that (i) fish differ in behavioural predictability of relative area; (ii) variation in predictability is underpinned by additive genetic effects, and (iii) individual mean behaviour is genetically correlated with predictability. Our results show that the assumption of homogeneous residuals, which is typical to linear (mixed) models applied in personality research (Dingemanse et al., 2013; Brommer, 2013), is violated. Fortunately, this will not generally bias measurement of among-individual or additive genetic variance in mean behaviours. However, it does highlight how standard analytical approaches will necessarily miss interesting and important components of variation among-individuals. In what follows we first discuss our findings in relation to the behavioural stress response in guppies. We then broaden our focus with the aim of highlighting several consequences of (genetic) variance in predictability. We

argue that this phenomenon has interesting implications for the evolution of phenotypes under selection that are more general than the current behavioural context.

First, we found variation in (mean) risk-related behaviour in this population of guppies at both the individual and genetic level. This is consistent with our previous work on the same population (Houslay et al., 2019), other species of wild-type poeciliid (Boulton et al. 2018), and the growing empirical evidence of heritable 'personality' variation across taxa (Dochtermann et al. 2014). Second, and of greater novelty, is the finding that behavioural predictability differs among individuals. Furthermore, our analyses demonstrate correlations between mean and predictability such that individuals expressing low relative area (i.e. more flight-type behavioural responses) are also less predictable (i.e. more variation in response to the OFT). The presence of correlation structure between behavioural mean and predictability is consistent with findings at the phenotypic (among individual) level from several other recent studies (Stamps et al., 2012; Mitchell et al., 2016). For example, a negative phenotypic correlation between mean activity rates and within-individual variation was previously estimated in guppies, where individuals that were more active, were also more predictable (Mitchell et al., 2016). Here, by using pedigree analysis we are also able to show that these individual-level patterns are underpinned by correlated genetic effects on behavioural means and predictabilities. To our knowledge only one previous study has attempted to measure a genetic correlation between mean behaviour and predictability (Martin et al., 2017). This study of docility in marmots estimated

a negative correlation between mean behaviour and predictability, though the genetic correlation was not statistically significant.

Although the stress coping style (SCS) model does propose a relationship between average behavioural response to a stressor stimulus, and the predictability of behaviour, our results do not fully align with its specific predictions. This is because the structural pattern of observed variation in relative area found is not consistent with the proactive-reactive model of SCS (Koolhaas et al., 1999; Sih, Bell, & Johnson, 2004; Coppens et al., 2010) in which proactive individuals are expected to express more 'fight or flight' stress responses on average, but are also expected to be 'bolder' and/or more exploratory than reactive types. In fact, variation in relative area is orthogonal to this expectation, because it discriminates between a (putatively stressed) 'flight' response to the OFT (low relative area) and a (putatively less stressed) exploratory response (high relative area). In other words, relative area is probably better interpreted as measuring the magnitude, rather than 'style' of the behavioural stress response. Thus, while SCS predicts that high (mean) 'flight' behaviour will be linked to high predictability within the proactive coping style, we find it is linked to low predictability instead and likely reflects a high magnitude of stress responsiveness. While this means the stress coping model does not provide a good description of guppy responses to the OFT (Houslay et al., 2019), we nevertheless argue that it provides a useful heuristic framework precisely because it emphasises the need to evaluate integration among stress-response components in a multivariate empirical framework. Here we show links between mean behaviour and predictability, but there is also evidence of genetic

integration between (mean) behaviour and glucocorticoid (GC) physiology (flight type behaviours being associated with higher GC levels; Houslay et al. 2019). It therefore seems likely that predictability will also be genetically correlated with GC responses and their rates of habituation to repeated or chronic stressor exposure (Houslay et al., 2019) though this remains to be confirmed.

Before considering the evolutionary implications of this genetic covariance structure further, it is perhaps worth noting that mean-variance (or predictability) relationships may sometimes be inevitable given trait definitions and distributions (Tatliyer et al., 2019). Here we derived the trait of relative area as a biologically relevant measure of behavioural stress response using observed data on the actual area covered and the distance swum. Specifically, relative area is defined as the difference between observed area covered and the predicted area covered given a 'random swim' as long as the observed track length. It is inevitable that the possible range of observed area covered is restricted for trials of low track lengths (i.e. a fish cannot cover 100% of the area with a very short track length) and, thus it is possible that (genetic) variance in predictability of relative area arises in part from (genetic) variance in mean track length. Although such dependencies might in principle also drive the (genetic) correlation between mean and predictability of relative area, this would lead in the present case to a positive correlation (Tatliyer et al., 2019), not a negative one as estimated here. However, to explore this further we fitted a post hoc DHAM model of the observed area covered (Supplementary Information, Table S4.3). We found the same pattern as reported above with respect to relative area; negative correlations were present between the mean behaviour and predictability at the individual and

genetic levels. Thus, while it is necessarily true that our quantitative results for relative area depend on track length, we do not think there is any sense in which our qualitative conclusions are driven by mathematical artefacts of trait definition. For completeness, we also ran a further DHAM for track length (another biologically relevant trait often used to investigate personality traits such as boldness or activity) (Burns et al., 2008; S. J. White & Wilson, 2019) (Supplementary Information, Table S4.4). We found among-individual differences in the trait mean and within-individual variance, which was in part due to additive genetic effects. This further suggests that the presence of genetic variance in predictability is not a particularly trait-limited phenomenon.

Our results add to the small but emerging set of studies evidencing among-individual and genetic variance for predictability (or intra-individual variation). If widespread, this could have major implications beyond the present focus on stress response and coping strategies. Variance among individuals means that behavioural predictability could be a direct target of selection, and if this does occur, the fact that it is heritable means it could evolve under selection. Furthermore, genetic correlation between the individual mean and the variation around it will allow correlated evolution of predictability in response to selection on 'personality' (individual average behaviour) and vice versa. Clearly our study tells us nothing about the fitness consequences of behavioural predictability in wild guppies. Nonetheless, low predictability can sometimes be selectively advantageous for prey species under specific predation threats (Briffa, 2013; Chang et al., 2017). For instance, in the jumping spider (*Cosmophasis umbratical*) low predictability is advantageous when faced with aggressive

predators (Chang et al., 2017). There is also some evidence for predictability-fitness associations in the pill bug (*Armadillidium vulgare*), where individuals become less predictable in risk-taking behaviour in unfamiliar, rather than familiar environments (Horváth et al., 2019) (but see Richardson et al. 2018 for a counter-example). Predator-mediated direct selection on predictability thus seems at least a plausible hypothesis in guppies (though indirect selection arising from causal effects of the genetically correlated mean behaviour could be more important).

We also note that, while advantages of low predictability do occur in a behavioural context (Briffa, 2013), it seems likely that high within-individual variation may more often be costly. For instance, given a single (constant) phenotypic optimum, an individual predictably expressing this value across multiple selective events will have higher fitness than a less predictable individual with the same mean phenotype. Where canalisation of some continuously distributed trait around the optimum is itself 'expensive', individuals of lower 'quality' or 'condition' may also be less predictable in trait expression and incur costs as a result (Westneat et al., 2015). In other words, low predictability can be a symptom of inability to buffer trait expression against environmental effects. A complementary perspective at the genetic level is gained by recognising that heritable differences in predictability can equally be viewed as 'genetic heterogeneity of environmental variance' (Mulder et al., 2007). Thus, this phenomenon is a manifestation of genotype x environment (GxE) interaction, in which the genotype-phenotype map is sensitive to one or more environmental parameters (Nussey et al., 2007). Although GxE are normally investigated across gradients of some environmental parameter defined a priori, our results show that

some guppy genotypes are more phenotypically plastic (i.e. less predictable) than others in respect to unknown (and uncontrolled) environmental variables. This shows that application of DHAM could be a useful strategy for characterising the potential importance of GxE in scenarios where the most relevant or appropriate descriptor of environmental variation is itself unclear (e.g. wild populations experiencing complex multivariate changes in environmental state). It is notable, for instance, that (linear) reaction norm models applied to wild vertebrates in naturally variable environments have generally detected limited support for GxE (e.g. Hayward et al. 2018), while evidence from experimental studies that manipulate environment conditions is compelling (Pigliucci et al., 1995; Ingleby et al., 2010; Des Marais et al., 2013). One explanation for this might be the (univariate) environmental descriptors used in the former, and/or the assumption that reaction norms are linear, have been inadequate or inappropriate choices. Typically 'extrinsic' variables (e.g. measures of climate) have been used, though some studies used environment specific trait means (Ramakers et al., 2018) or other measures of average 'performance' (e.g., annual mortality; Wilson et al., 2006) as proxies for overall environmental quality. This approach is common in plant studies (following Finlay & Wilkinson, 1963) and may well have wider utility in evolutionary ecology. Regardless, demonstrating the presence of genetic variance for 'predictability' in behaviours or other traits could be a useful starting point for more targeted investigation of which specific environmental factors genotypes are responding to, and of what functional form those responses take.

In conclusion, here we build on previous studies highlighting genetic variation in mean behavioural stress response traits, to show that variance is also

present in predictability of behaviour. This variation among individuals is itself underpinned by additive genetic effects, meaning behavioural predictability can be viewed as a trait with adaptive potential under selection. Furthermore, this is one of the first studies to estimate a genetic correlation between mean behaviour and predictability (i.e. within-individual variation), and so highlights the expectation that these aspects of phenotype will coevolve under selection. We recommend wider application of double hierarchical models, including the DHAM used here, to investigate the presence and causes of among-individual heterogeneity in environmental sensitivity of phenotypes generally (including but not limited to behaviours). By doing this we will gain a more complete picture of how variation is structured within and across hierarchical levels, and consequently a deeper understanding of the evolutionary ecology of labile traits in general.

Table 4.1: Estimated variance components and derived parameters for mean and dispersion parts of the four models of relative area. Subscripts denote residual (*R*), group (*GR*), individual (*I*), permanent environment (*PE*) and additive genetic (*A*) components of variance and the corresponding mean-predictability correlations (*r*). We use the second subscripts (*v*) to denote variance in the dispersion part of the model (applicable to models 3 and 4 only). Also shown are the familiar measures of repeatability (*R*) and heritability (h^2) of mean behaviour (estimated at an average residual variance in models 3 and 4). Values in parentheses indicate approximate standard errors for models 1 and 2, and 95% credible intervals for models 3 and 4.

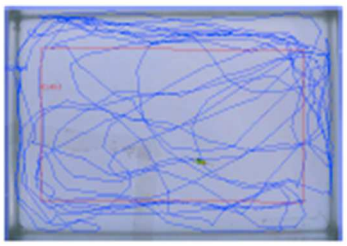
Parameter	Model			
	1 (repeatability model)	2 (animal model)	3 (DHGLM)	4 (DHAM)
V_R	0.640 (0.018)	0.640 (0.018)	0.643 (0.587, 0.716)	0.617 (0.563, 0.74)
Mean				
V_{GR}	0.035 (0.019)	0.034 (0.018)	0.020 (0.004, 0.059)	0.021 (<0.001, 0.056)
V_I	0.273 (0.028)	-	0.234 (0.191, 0.295)	-
V_{PE}	-	0.173 (0.044)	-	0.138 (<0.001, 0.226)
V_A	-	0.105 (0.052)	-	0.052 (0.004, 0.302)
R	0.288 (0.024)	-	0.269 (0.221, 0.309)	-

pe^2	-	0.182 (0.047)	-	0.165 (<0.001, 0.242)
h^2	-	0.110 (0.052)	-	0.092 (0.009, 0.306)
Predictability (dispersion)				
V_{GR_v}	-	-	<0.001 (<0.001, 0.038)	<0.001 (<0.001, 0.038)
V_{I_v}	-	-	0.328 (0.244, 0.421)	-
V_{PE_v}	-	-	-	0.157 (<0.001, 0.264)
V_{A_v}	-	-	-	0.146 (0.034, 0.420)
Mean-predictability correlation				
r_{GR}	-	-	-0.482 (-0.922, 0.486)	-0.603 (-0.951, 0.473)
r_I	-	-	-0.955 (-0.988, -0.858)	-
r_{PE}	-	-	-	-0.956 (-0.998, -0.199)
r_A	-	-	-	-0.921 (-0.987, -0.623)

Figure 4.1: (a) Illustration of contrasting Open Field Trials (OFT), showing a) tracks swum by two individual fish (1, 2) as blue lines, and b) the resulting trait data. Here both fish swim a very similar track length, but individual 1 also covers a high percentage of the tank (Area covered = 65.7%) and displays an exploratory phenotype. By comparison Individual 2 covers much less area and is exhibiting a characteristic stress response of fast swimming along the tank walls. This results in very different values of relative area (RA) (where $RA = AC - \hat{ACTL}$, AC = area covered, = track length, and $\hat{ACTL}$ is the expected area covered in a random swim of observed TL, predicted by a fourth order polynomial regression fitted to simulated data; see Houslay et al 2019 for further details).

a)

1

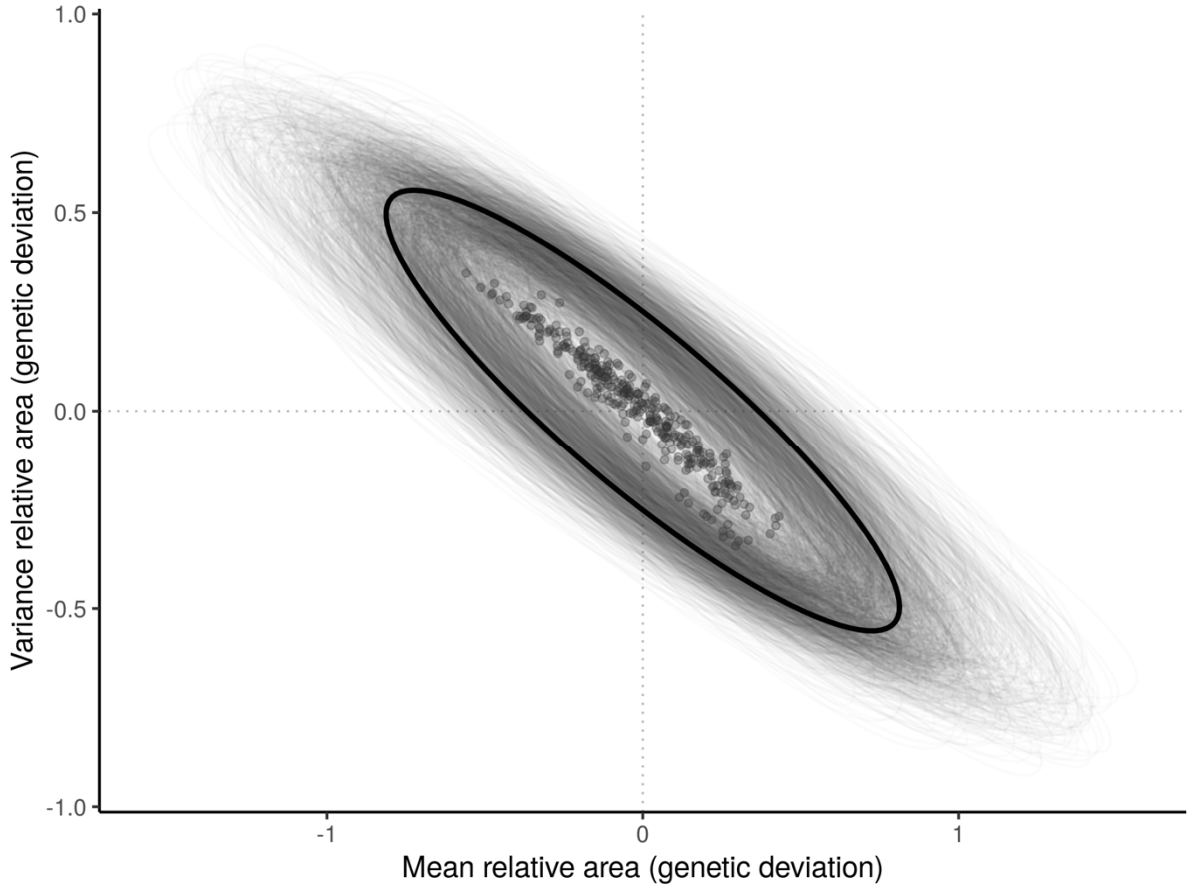


2

b)

Trait / Individual	Track Length (cm)	Area Covered (%)	Relative Area
1	661.1	65.7	15.0
2	633.2	12.4	-37.7

Figure 4.2: Estimated genetic matrix of relative area mean and trait predictability. The black line captures 95% of variance in the genetic correlation between mean and predictability in relative area. The grey ellipses are bootstrapped replicates from model 4, showing uncertainty around the estimated matrix. Individual points are best linear unbiased predictions (BLUPs) of genetic values from individuals in our data set.



5 Chapter 5: Genetic and context-specific effects on inhibitory control performance in the Trinidadian Guppy (*Poecilia reticulata*)

5.1 Abstract

Among-individual variation in cognitive traits, which are widely assumed to have evolved under adaptive processes, is increasingly being demonstrated across animal taxa. As variation among individuals is required for natural selection, characterising the extent of individual differences is an important step towards understanding how cognitive traits evolve within populations. However, adaptive evolution also requires that differences in cognitive performance are heritable. Here we use a quantitative genetic study of wild-type guppies repeatedly exposed to a 'detour task' to test for genetic variance in the cognitive trait of inhibitory control. We also test for genotype-by-environment interactions (GxE) by testing related fish under alternative experimental treatments that differ in degree of available visual information (using transparent vs semi-transparent barriers in the detour task). Finally, by analysing behavioural data collected over a pre-testing training period (in which a cue-food reward association was trained), we seek to validate the assumption that the detour task measures variation in cognitive processes (specifically inhibitory control) that is distinct from among-individual differences apparent during training (e.g., personality, motivation, associative learning ability). We find among-individual variation in detour task performance, consistent with differences in inhibitory control. However, GxE means that heritable variation only contributes to this in one treatment (the fully transparent

cylinder). This suggest that the adaptive evolutionary potential of inhibitory control may be highly sensitive to environmental conditions. Equivalently, this result means that the plastic response of detour task performance to treatment environment is itself genetically variable. We find individual performance in training trials positively predicts performance in the detour task. However, among-individual variation in the former is not a sufficient explanation for the latter. This supports our assumption that detour task performance effectively captures variation in cognitive processes (putatively inhibitory control), that is distinct from variation in other factors (cognitive or non-cognitive) expected to generate among-individual variation in training performance.

5.2 Introduction

Cognitive traits, defined as the set of mechanisms by which animals acquire, process, store and use information from their environment (Shettleworth, 2009a; Healy et al., 2010), are widely assumed to have evolved under adaptive processes. While comparative approaches to cognitive performance have a long history (for a review see Healy, 2019), a more recent aim within behavioural ecology has been to characterise variation in cognitive performance among conspecific individuals within populations (for a review see Boogert et al., 2018). Variation among individuals is required for natural selection and is central to our understanding of adaptive evolutionary dynamics for any aspect of the phenotype (Roff, 2002). Thus, characterising the extent of individual differences in cognitive performance, as well as their consequences for fitness, will shed light into how

cognitive traits evolve within populations, and so diverge among populations and among species. However, adaptive evolution requires not just variation and selection, but also that differences in cognitive performance are heritable (Croston et al., 2015; Thornton et al., 2015). Here we test this premise using a quantitative genetic study of wild-type guppies repeatedly exposed to a 'detour task'. Differences in performance within the detour task are expected to arise from cognitive processes (including, but not limited to, inhibitory control). In order to exclude the possibility that non-cognitive factors (such as motivation, state or personality) or other cognitive factors (such as associative learning ability) could drive individual performance in the detour task, we compared performance with prior training trials to aid discrimination between sources of variation.

Recent studies have shown high levels of variation in cognitive performance among individuals of the same species (for a review see Boogert et al., 2018; Cauchoix et al., 2018), a pattern that is seen across taxa, from insects (Li et al., 2017), fish (Prentice, Mnatzaganian, et al., 2020), to mammals (Nawroth et al., 2017) and birds (Guillette et al., 2015). Furthermore, among-individual differences in cognitive performance extend across many aspects, or 'domains', of cognition (e.g. spatial memory (Sonnenberg et al., 2019); spatial leaning (White et al., 2017), inhibitory control (van Horik et al., 2019)). The fitness consequences of this variation remain somewhat unclear; cognitive studies are often conducted under controlled laboratory conditions which can render meaningful estimation of selection on cognitive performance problematic. Early studies (Raine et al., 2008; Cole et al., 2012) suggest that differences in cognition do have fitness consequences and so are subject to natural selection, and evidence emerging

from more recent field studies further confirm this. . For example, cognitive performance across multiple domains was positively associated with reproductive success in female Australian magpies (*Cracticus tibicen dorsalis*) (Ashton et al., 2018). In New Zealand robins (*Petroica longipes*), males that performed better in a spatial learning task produced more fledglings with higher rates of survival (Shaw et al., 2019). Similarly, in mountain chickadees (*Poecile gambeli*) higher performance in spatial learning and memory tasks was associated with increased probability of overwinter survival (Sonnenberg et al., 2019). While these studies provide evidence for contemporary selection acting on cognitive traits, genomic approaches have also been used to investigate strong selection in the past. For instance, using the fact that genetic variation is reduced in the vicinity of a beneficial mutation that spreads rapidly through a population (i.e. undergoing a 'selective sweep'; Smith & Haigh, 1974), recent studies have detected the signature of positive selection on genes associated with cognition and learning in great tits (*Parsus major)* (Laine et al., 2016), and face recognition in populations of paper wasp (*Polistes fuscatus*) (Miller et al., 2020).

Thus among-individual variation in cognitive performance has been shown in many animals (Guillette et al., 2015; Nawroth et al., 2017; Li et al., 2017; Prentice, Mnatzaganian, et al., 2020) and is likely to have fitness consequences (Ashton et al., 2018; Shaw et al., 2019; Sonnenberg et al., 2019). These phenomena – variation among individuals, and fitness consequences of that variation, represent two of three fundamental requirements for adaptive phenotypic evolution to occur (Wilson et al., 2010). The third requirement is that differences among individuals in a population arise, at least in part from heritable

genetic factors. To date rather few studies have explicitly investigated the extent to which heritable genetic variance contribute to among-individual variation in animal cognition (but see Hopkins et al., 2014; Quinn et al., 2016; Sorato et al., 2018; Vardi et al., 2020; Langley et al., 2020). Furthermore, results from the few studies that have been done are somewhat mixed. For instance, genetic variation did contribute to differences in reversal learning among red junglefowl (*Gallus gallus*), but when individuals were trained to discriminate between a rewarded and an unrewarded cue, no heritable component to discrimination learning was detected (Sorato et al., 2018). Similarly no evidence of genetic variation in spatial learning was found in a recent study of delicate skinks (*Lampropholis delicate*) presented with a Y maze testing paradigm (Vardi et al., 2020). In a recent study of pheasants (*Phasianus colchicus*), Langley et al. (2020) reported quite low estimates of heritability (h^2, the proportion of observed trait variance attributable to additive genetic effects) for some tasks (e.g. spatial learning, $h^2 = 0.09$), but more moderate ones for others (e.g., discrimination learning, $h^2=0.17$; inhibitory control, $h^2 = 0.23$).

We should not be surprised that heritability estimates vary among populations and specific cognitive traits (indeed it would be more surprising if they did not). However, it is clear that further empirical exploration of the genetic underpinnings of animal cognition is required if we hope to gain a better understanding of how variation is structured generally. The very fact that cognition is defined in relation to acquiring, processing and using information in the environment makes the possibility of genotype-by-environment interactions (GxE) very plausible and intuitive. GxE occurs if the genotype-phenotype map is

sensitive to environmental conditions (Via et al., 1985; Nussey et al., 2007) and causes genetic variance within a population to vary with those conditions (Mackay, 1981; Roff, 2012). Thus we might predict that the contribution of genetic variation to among-individual differences in performance when challenged by a cognitive assay, will be highly sensitive to the information context provided (e.g. visual information availability; Pike et al., 2018). Equivalently, GxE means there is genetic variance for plasticity of performance across environmental contexts (Roff & Wilson, 2014).

So far, if few animal cognition studies have characterised the importance of genetic variance within populations, then the contribution of GxE to date has been almost completely unexplored (see Thornton & Boogert, 2019). For other aspects of phenotype, there is abundant and compelling evidence of GxE from experimental studies that manipulate environment conditions (Pigliucci et al., 1995; Ingleby et al., 2010; Des Marais et al., 2013). This work includes many examples of highly labile behavioural traits. For example, 'calling' effort by male crickets has been shown to depend on the interactions of genes with environmental factors such as temperature, diet and social context (e.g. Callander et al., 2013; Hedrick et al., 2002; Kasumovic et al., 2012; Rapkin et al., 2017). Similarly the emergent field of 'animal personality' (among-individual variation in behavioural traits repeatable across time and context; Bell et al., 2009; Gosling, 2001) has moved beyond characterising differences in average behaviour to also consider the implications of among-individual variation in plasticity (IxE). A growing number of studies have used quantitative genetic approaches to demonstrate that genetic variation in plasticity (i.e. GxE)

contributes to these personality traits (Edwards et al., 2017; Prentice et al., 2020; Rudin et al., 2019; Wey et al., 2019).

In this study we seek to address these general gaps in our knowledge about genetic variation underpinning differences in cognitive performance. More specifically, we aim to characterise the importance of genetic variation, including GxE, among Trinidadian guppies (*Poecilia reticulata*) in inhibitory control as measured within a 'detour task'. The detour task is a widely applied cognitive testing paradigm that requires an individual to temporarily move away from a (visible) goal or reward, and detour around a (typically) transparent obstacle in order to reach it. It tests an individual's ability to inhibit pre-potent responses (i.e., attempting to move straight towards a reward) in favour of more effective or appropriate behaviours (i.e., detouring around the obstacle). This ability is thought to depend on several executive cognitive functions, such as working memory, route planning, and object permanence, but particularly on inhibitory control (for a review see Kabadayi et al., 2018). In human studies, inhibitory control positively predicts cognitive abilities across a range of tasks (Shamosh et al., 2008; Diamond, 2013) and academic achievement in children (Duckworth et al., 2012) while being negatively correlated with propensity for antisocial behaviour and drug abuse (Feil et al., 2010; White et al., 1994). We note here that variation in performance in the detour task may also reflect other factors beyond inhibitory control, including motivation and associative learning of the affordances of the task (van Horik, Langley, et al., 2018; van Horik et al., 2019).

For non-human animals, the quantification of natural selection on inhibitory control remains limited (but see Ashton et al., 2018). However, it is well documented that animals can benefit from inhibiting behaviours, such as foraging or parental care when competing conspecifics or predator densities are high (Beran, 2015; Fontaine & Martin, 2006; Soltis et al., 2001). Comparative research suggest this type of executive functioning and self-control is complex, and correlates with brain size (MacLean et al., 2014; although see Jelbert et al., 2016 for an alterntive intepretation), however taxonomically widespread, presenting in mammals (Diamond, 1990; Bobrowicz et al., 2018; Barrera et al., 2019), birds (Boogert, Anderson, et al., 2011; Kabadayi et al., 2016; van Horik et al., 2019) and fish (Lucon-Xiccato, Gatto, et al., 2017; Lucon-Xiccato et al., 2019; Brandão et al., 2019).

Guppies are a freshwater poeciliid fish increasingly used as a model for both behavioural genetics (Prentice et al., 2020) and animal cognition (Kotrschal et al., 2015; Lucon-Xiccato & Bisazza, 2016; Fong et al., 2019). They have recently been used to explore variation in performance in learning colour associations (Trompf et al., 2014; Buechel et al., 2018), reversal learning (Buechel et al., 2018), numerical discriminations (Kotrschal et al., 2013; Lucon-Xiccato & Bisazza, 2017b) and spatial associations (Lucon-Xiccato & Bisazza, 2017b; Prentice et al., 2020). Guppies also exhibit inhibitory control (e.g. Lucon-Xiccato & Bisazza, 2016), and consistent among-individual differences in detour task performance have recently been documented (Macario et al., 2021). There is also some evidence that inhibitory control differs between sexes since, for example, in reversal learning tasks (which themselves require inhibition of

learned behaviours (Lai et al., 1995)) females have been found to inhibit a previously learned colour-reward association faster than males (Lucon-Xiccato et al., 2014; Lucon-Xiccato & Bisazza, 2017b). These studies provide us with a clear expectation that guppies will use inhibitory control to successfully complete a detour task, and that there will be variation among-individuals. However, the extent to which among-individual variation depends on genetic variation (including GxE) remains unknown.

In what follows we address these questions using a quantitative genetic study using guppies bred from a captive colony derived from fish sampled in the Aripo River, Trinidad. Fish in the study were first trained within a colour discrimination paradigm, and trained to feed from a green plastic disc placed in the same consistent location on the tank floor. Fish were then repeatedly assayed in a detour task requiring navigation around a transparent cylindrical barrier in order to reach a visible food reward on the opposite side. We tested for genetic variance in performance, but also for genotype-by-environment interactions (GxE). If present, GxE can be interpreted as genetic variance in plasticity of cognitive performance or (equivalently) as the presence of environment-specific genetic variance in performance. To do this each fish was tested under one of two treatment 'environments'; half undergoing a detour task using a standard transparent cylinder, and half using a striped cylinder that provides additional visual information. The use of 'semi-transparent' barriers has been shown to improve average detour task performance in many cases resulting from an increase in cue salience (e.g., Juszczak & Miller, 2016; Noland, 2008; Santos et al., 1999), but not all. For instance, in a study of three avian species, semi-

transparent barriers improved average performance in herring gulls (*Larus cachinnans*) and quails (*Coturnix x C. Japonica* hybrids) but had no effect in canaries (*Serinus canaria*) (Zucca et al., 2005). Here the objective is to test for an effect of treatment on genetic variance (i.e. GxE) and we make no strong predictions about any directional effect on mean performance. Relative to the standard cylinder, the striped treatment provides additional visual information that may improve average performance, but we also consider it possible the stripes could provide a mildly aversive stimulus, potentially causing a neophobic response, negatively impacting the speed with which fish reach the food reward.

Finally, our experimental design necessarily requires some assumptions to be made about the relationship between latent cognitive traits (inhibitory control) and observed performance in the detour task. We therefore seek – as far as possible – to validate the involvement of cognitive traits by jointly analysing detour task performance with data on time to feed in the training stage of the experiment. Specifically we ask whether among-individual and/or genetic differences in time to feed in the absence of the cylinder (e.g. either due variation in motivation, personality or associative learning), might be sufficient to explain later variation in detour task performance. This allows us to quantify the extent to which the detour task reveals (genetic) variation in cognitive performance (and putatively in inhibitory control), that could not be explained by variation in other factors (e.g., motivation, experience (van Horik, Langley, et al., 2018)).

5.3 Methods

Ethics

This work was conducted under the auspices of the UK Animals (Scientific Procedures) Act (1986) with approval of the University of Exeter research ethics committee, under licence from the Home Office (UK) (Licence Number PPL30/3256). Experimental procedures and behavioural assays were developed in accordance with the principles of the three Rs and ASAB guidelines (Buchanan et al., 2020) for use of animals. All periods of handling and emersion were kept to a minimum and only fish deemed healthy and exhibiting normal behaviour were used in trials.

Fish husbandry and breeding

Fish used in this study were bred from a captive population *P. reticulata* housed at the fish laboratory at the University of Exeter's Penryn campus. The population is descended from wild fish caught in 2017 from the lower Aripo River, Trinidad and has been subsequently maintained with no deliberate selection or inbreeding. All fish were fed to satiation twice daily (0800 – 1000h and again at 1600 – 1800h) using commercial flake food and live *Artemia nauplii*. Water temperature was maintained at 23-24°C in well-aerated closed system tank stacks that undergo 25% water changes each week and with weekly tests for ammonia, nitrate and nitrite levels. Lighting was kept at a 12:12 light/dark cycle.

Quantitative genetic analyses require knowledge of pedigree structure. Here we collected behavioural data an offspring generation comprising 374 guppies (all tested as adults), produced from 6 small breeding groups over a

period of 4 months. The breeding groups were housed in 15L tanks (18.5cm x 37cm x 22cm) and varied somewhat in size and sex ratio (groups containing 2 - 7 males, and 4 - 7 females at any time). Variation was due to some mortalities over the breeding period, which were replaced with new fish. In total 54 adults from the parental generation entered breeding groups; all were sampled at random from the stock tanks but females used were kept isolated from males for 3 months before use, reducing the possibility that they were carrying viable sperm from previous matings. Offspring produced in the breeding tanks were removed on sight and placed in separate 2.8L brood tanks (10cm x 28cm x 15cm). The 'broods' were therefore grouped by (putative) family and date of birth. Note that multiple females in a breeding group may have given birth on the same day and/or paternity may be mixed within broods so putative family structure was subsequently tested using molecular data (described below). Offspring were raised in brood tanks to sexual maturity and all behavioural data was collected in a single period once the total offspring generation of mature fish reached sufficient sample size. This strategy was largely due to closure of the fish lab facility and working restrictions imposed during the COVID-19 pandemic. Consequently all offspring fish subjects exposed to behavioural tests were sexually mature but of varying age. Exact age of individual fish was unknown but ranged from 4 – 15 months from date of birth.

Cognitive testing

We used a repeated measures design to test individual performance in a detour task. Guppies were individually transferred into 15L experimental tanks (18.5cm x 37cm x 22cm) and housed alone for the duration of the behavioural testing. The set-up of these experimental tanks closely followed that used by Lucon-Xiccato & Bisazza (2014), each being divided into two compartments (using white plastic) separated by a guillotine door (Figure 5.1). The rear 'home' compartment (20 x 18.5 cm) allowed individuals visual access to fish in neighbouring tanks. Conversely, the 'test' compartment (17 x 18.5 cm) at the front of the tank was visually screened from neighbours (using white plastic). This was to prevent the possibility of social learning (from observation of neighbours) influencing individual performance in the detour task. Individuals were allowed to acclimate to experimental tanks for 48 hours prior to training and testing. During acclimation they were fed once a day (with 1/3 pipette of live artemia for males, and 1 pipette for females) and the guillotine door was left open to allow unrestricted use of both compartments. Experimental tanks were contained within two 'stacks', each comprising 24 tanks (8 tanks per row, 3 rows high) on a shared recirculating water supply. Thus data could be collected in a block of (up to) 48 fish at once. The testing protocol (described below) took 8 days per fish, so in practice this was done in 11 blocks over a total period spanning 13 weeks.

For each individual in each block, the behavioural data collection protocol included two stages. First we trained naïve guppies to associate the appearance of a green disc placed on the floor of the test compartment with a food reward. Individual guppies were given nine food association training 'trials' (3 per day for 3 successive days). To ensure fish had equal opportunity to learn the food

location, they did not proceed to the detour task if they did not locate food in 5/9 of the training trials (See **Chapter 3** for details of this weak learning criterion). Across all groups, a total of 8 females and 4 males did not proceed to the detour task trails. Prior to each trial, fish were gently guided with a net into the home compartment and the guillotine door was closed. The experimenter then placed a white plastic test plate (4 x 10 cm) with a green disc in the middle (diameter 1.5 cm) on the tank floor in the test compartment. A food reward was then carefully placed on the green disc using a plastic pipette. For males a single (previously frozen) artemia was used, for females (which are much larger than males) we used 3 artemia. The guillotine door was then opened allowing the fish to swim into the test compartment and feed. The time to locate and eat the food reward was recorded for each fish using censored measurements. Specifically, we assessed whether the food item had been consumed at 1, 5, 10, and 20 minutes after opening the door, and (if not consumed) at 20 minute intervals thereafter to a maximum of 140 minutes. This recording strategy allowed a single experimenter to collect data on (up to) 48 individuals simultaneously.

After the 3 day training stage, fish that successfully learnt the association were assayed three times in a detour task (once per day for three successive days). Detour task trials essentially repeated the training trials; fish were guided into the home compartment prior, the guillotine door closed, and the food item placed on the green disk. However, the green disk was now inside a plastic cylinder (Figure 5.2). Thus the fish needed to navigate around a barrier and access the cylinder (from either end) to obtain the food reward. Again we recorded (censored) time to eat by assessing the presence/absence of the food

item at 1, 5, 10, and 20 minutes after opening the door, and thereafter at 20 min intervals for a further 2 hours, then 60 min intervals for a further 5 hours, and a final check at 24 hours after opening the door. Note that due to severe constraints on researcher lab time (arising from the covid-19 pandemic), collection of more detailed behavioural data including, for example, number of redundant attempts to directly acquire the food reward through the barrier (a common trait measured in detour reaching tasks; Kabadayi et al., 2017; Santacà., et al., 2019), were unfortunately not possible.

To allow testing for GxE, each fish was randomly assigned to one of two alternative treatment levels in the detour task; either a transparent (unmarked) cylinder was used (low visual information of the cylinder), or a cylinder with three black horizontal lines was used (high visual information of the cylinder). Thus each individual experiences only one treatment but sibships are represented across both levels. Our naïve expectation was that more visual information should, on average, reduce time to access food in the detour task. However, increased visibility of the cylinder could also lead to neophobic behaviour having the opposite effects. It is important to note however that for the current objective of testing whether genotypes differ in their response to increased visual information, the magnitude and direction of any change in mean performance is incidental (though interesting).

Microsatellite genotyping and pedigree analysis

Molecular pedigree analysis was conducted to fully resolve the pedigree structure needed for quantitative genetic modelling. At the end of the breeding stage, remaining parental generation individuals were euthanised by overdose of buffered MS-222 and individually stored in 70% ethanol at -5°C. All offspring individuals were similarly euthanised at the end of behavioural data collection. For all individuals, DNA was extracted from tail tissue and processed according to the protocol described in Becher et al. (2002). We then genotyped fish at 6 autosomal microsatellite loci (see Supplementary Information Table S5.1, Appendix 5 for details) as described in Becher et al. (2002) and Bergero et al. (2019). Polymerase chain reaction (PCR) reaction conditions were as follows: 8.3 μL GoTaq Mastermix (Promega), 0.2 μL each of fluorescently-labelled forward and reverse primer, 7.8 μL nuclease free-water and 2.5 μL DNA. The PCR programme consisted of 3 minutes at 90 °C, then 30 cycles of 30 seconds at 90 °C, 20 seconds at 55 °C and 30 seconds at 72 °C, before a final extension period at 72 °C for 5 minutes. A random selection of individual amplifications across all 6 primers were visualized on 2% agarose gel to confirm successful amplification. Following this, individual PCR products were separated by capillary electrophoresis in AB 3500 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA).

We genotyped the parental and offspring generations using the Genemapper ® ID-X software (Thermofisher Scientific, Waltham, MA, USA) by scoring individual genotypes across all 6 microsatellite markers using GENEMAPPER 3.7. Population allele frequencies were estimated in CERVUS

3.0 using the genotypes of the offspring generation across the six breeding groups. Individuals were assigned to full- and half-sibships using the program COLONY 2.0.4.5 (http://www.zsl.org/science/software/COLONY), which reconstructs parental genotypes from offspring genotypes using maximum likelihood (Jones &Wang 2009; Wang 2013). Colony runs were programmed so that (half) sibships were assigned for each breeding group separately (recognising that relatedness across groups is assumed to be zero) and that the 'full likelihood structure ' was then obtained for each group and combined across groups to give an overall pedigree structure. The details of COLONY run parameters can be seen in Supplementary Information S5.2.

Statistical analyses

Performance in the detour task

We estimated among-individual variation, heritability and GxE for the performance in the detour task by fitting a series of six univariate linear mixed effect models differing in random effect structure (explained below). All models were fit by REML (restricted maximum likelihood) using ASReml-R 4.1 (Gilmour et al., 2009; Butler et al., 2018) within R version 3.6.1 (R Core Team, 2017). We make the standard assumptions that random effects and residuals are normally distributed with means of zero and variances to be estimated.

We use *time to eat* as our observed measure of performance in the detour task (and in the training trials). For analysis, the observed (censored) time data was tested for normality, natural log-transformed and then mean centred and

scaled to standarddeviation units. The log transformation improved the assumption of Gaussian error structure, while scaling to standard deviation units was just to ease interpretation of estimated variance components. All models included a fixed effect of the mean, as well fixed factors of *sex*, *stack* (denoting which of two aquaria stacks the fish was tested in), and *treatment* (1 = clear cylinder, 2 = stripy cylinder). We also included *trial number* (i.e. the repeat number 1-3, fitted as a factor to avoid assuming a linear functional form for any change in the mean across repeats). Conditional F-statistics were used to determine the significance of fixed effects.

All six models fitted included random effects of *block* (the set of up to 48 fish that were phenotyped simultaneously) and *brood tank* (denoting early life housing environment). Model 1 included these two random effects only. We then extended this model by including an extra random effect of *individual identity*, to model among-individual variance (V_I) in (mean) performance (Model 2). In Model 3 we allowed for the possibility of IxE (among-individual variance in plasticity) by fitting treatment specific among-individual variances. Note that since each fish is observed for one treatment level only, individual plasticity across treatments is not observed. However, a test (albeit of limited statistical power) is possible because IxE leads to an expectation that V_I will differ with environment (here between clear and stripy cylinder treatments). In Model 4 we reverted to the assumption of homogeneous V_I across treatments, but extended the Model 2 to include the individual genetic merit for (mean) performance as an extra random effect. This becomes the standard repeated measures animal model of quantitative genetics, and allowed us to utilise the pedigree data to partition V_I

into additive genetic (V_A) and non-genetic, permanent environment (V_{PE}) components (assumed homogeneous across treatments. In Model 5 we modelled GxE, extending the Model 3 formulation by partitioning treatment specific among-individual variance into treatment specific permanent environment and additive genetic variances, as well as estimating the cross-treatment genetic correlation (r_G). Note that, for the genetic part of the model this represents a 'character state' conceptualisation in which GxE would be manifest as $V_{A1} \neq V_{A2}$ and/or $r_{G12}<+1$ (where subscripts 1 and 2 denote the clear and stripy cylinder treatments respectively). Finally, in Model 6 we allowed V_{PE} to differ between treatments but assumed no GxE (i.e. this model allows IxE but assumes any genetic variation is not treatment specific).

We compared among these six models using the Akaike information criterion (AIC) and likelihood ratio tests (LRT). The latter were used to compare nested models with a standard hypothesis testing framework (e.g. LRT comparison of Models 1 and 2 provides a statistical test of among-individual variance). Where models were compared by LRT to test a single variance component we assuming twice the difference in log-likelihood between full and reduced models is distributed as a 50:50 mix of $\chi^2_{0,1}$ and χ^2_1 as recommended by Visscher (2006). For all other situations we (conservatively) set the degrees of freedom equal to the number of additional (co)variance components in the model complex model.

In order to provide intuitive measures of effect size, we calculated the adjusted repeatability (R), where R is conditional on fixed effects and represents

the proportion of phenotypic variance explained by among-individual differences in behavioural mean (Nakagawa et al., 2010). Thus $R = V_I / V_P$ where V_I is the among-individual variance and V_P is the phenotypic variance conditional on fixed effects (i.e. $V_P = V_I + V_B + V_{GR} + V_R$ where V_B, V_{BT} and V_R are among- block, among-brood tank and residual variances respectively). In Models 2 and 4 we similarly estimated adjusted heritabilities h^2 (where $h^2 = V_A/V_P$, and is conditional on fixed effects). Note that in Models 3 and 5 treatment-specific estimates of V_I, V_{PE} and V_A are used as appropriate to estimate treatment-specific R and/or h^2.

Performance in the training trials

The primary purpose of modelling performance in the training trials was to determine whether differences among individuals and/or genotypes were detectable at this stage of the experiment, and - if so – to consider whether these alone might be sufficient to explain variation seen in detour task performance. We therefore first analysed training trial data in a very similar way to the detour task itself using a set of univariate mixed effects models. The response variable of *time to eat* was treated as described above, and the null model (Model 1) included the same fixed effects used for the detour task (but without *treatment*) and random effects of *block* and *brood tank*. To test for among-individual variance (V_I) in performance we added a random effect of individual identity and compared this model (Model 2) to Model 1 by LRT. We then fitted a repeated measures animal model (Model 3), in which V_I was partitioned into V_A and V_{PE} components.

We compared Models 2 and 3 to test the statistical significance of V_A, and generated conditional estimates of R and h^2 as described above.

Since all fish experienced the same conditions for the training trials, we did not model IxE or GxE across cylinder treatments. However, we did consider the possibility that there could be among-individual and/or genotype variation in the pattern of any change in performance across the 9 repeated training trials. We expect that, on average, time to obtain the food will decline across training trials (as individuals learn the association of the green disc with food) but variation in this process could have implications for determining whether differences in training trial performance are sufficient to explain variation in detour task performance. This would be especially true if individuals (or genotypes) tend to change their ranking of performance over the training period such that, for example, relative performance in early training trials did not predict performance at the end of the training periods. We therefore fitted two additional models; Model 4 extended the Model 2 to a 'random slope' model in which the effect of individual identity was modelled as a first order (linear) function of *trial number* (treated as a continuous covariate). Finally Model 5 was a random regression model in which both additive genetic and permanent environment effects were modelled in this way as linear functions of *trial number*.

Effects of training on performance in the detour task

To examine whether performance in the training trials predicts performance in the detour task, we lastly built bivariate mixed models. This allowed us to test the

among-individual and genetic relationships between performance in training trials and performance in the detour task. Fixed effects on each trait were as specified above. Random effects of *block* and *brood tank* were included and allowed to covary across traits while residual (observation level) covariance across traits is not statistically identifiable so was not estimated. Based on univariate modelling of training trial data we elected to include random intercepts only for this trait (see results for justification), but we did include treatment specific variances for the detour task. This meant the among-individual variation was estimated as a 3x3 variance-covariance-correlation matrix (**ID**) in which the diagonal elements correspond to V_I in: (mean) performance in the training trials; (mean) performance in the detour task for treatment 1; and (mean) performance for treatment 2. All correlations in the matrix are identifiable and so were estimated. We note that the data structure contains very little information for estimating the among-individual correlation (r_{ID}) between detour task performance in the two treatments (since each fish experienced only one). However, constraining this correlation to zero may limit the model to unlikely parameter space (i.e. given 3 variables A, B, C, if, for instance r_{AB} and r_{AC} are strongly positive then it becomes very unlikely that $r_{BC} = 0$). We compared this to a simplified formulation in which the among-trait correlations in **ID** were all constrained to zero, providing a global test of among-individual correlation between traits. We then used an analogous multivariate animal models to estimate the genetic (**G**) and permanent – environment (**PE**) components of the **ID** matrix. We tested similarly conducted an overall tests for genetic correlation structure in **G** by LRT comparison to a

simpler model in which all pairwise genetic correlations (r_G) were constrained to zero.

5.4 Results

Detour task

Visual inspection of the raw data revealed an increase in average performance (i.e. decrease in time) across repeated trials (Figure 5.3). This was true for both sexes irrespective of treatment. Females were faster to eat than males at all trials in both treatments, while both sexes were faster to eat at all trials when experiencing the clear cylinder treatment (Figure 5.3). The univariate models confirmed the statistical significance of these qualitative patterns, revealing significant effects of *sex*, *trial number* and *treatment* on (mean) performance in the detour task. Estimated fixed effects were very similar across all models. Here we present estimates made under Model 2 (see Supplementary Information Table 5.3 for all fixed effect results, including those not directly relevant to current hypotheses). On average, males took longer to locate and eat the food reward (coefficient ± SE=0.617 ± 0.074, $F_{1,1070}$=64.98, P< 0.001; note all effect sizes are in SDU of log transformed times), and fish became, on average, quicker at obtaining the food reward across *trials* (*trial* 1 (-0.411 ± 0.089); *trial* 2 (-0.287 ± 0.047); *trial* 3 (-0.575 ± 0.047) $F_{2,712.7}$=74.05, P< 0.001). Furthermore, performance was slower on average for fish that experienced the striped cylinder treatment (0.613 ± 0.074, $F_{1,1070}$=69.75, P<0.001), compared to the clear cylinder.

This provides evidence of plasticity in (mean) performance across the two treatment environments in the detour task.

Model comparisons based on AIC and LRT provided strong support for the presence of among-individual variation in performance (e.g., LRT Model 2 vs. Model 1; $\chi 2_{0,1}$=214.541, P<0.001; Table 5.1). Under model 2, performance in the detour task was highly repeatable (R=0.467 ± 0.033; Table 5.2). Model 3, did not significantly improve the fit (LRT Model 3 vs. Model 2; $\chi 2_1$=0.124, P=0.725) and yielded treatment specific estimates of among-individual variance that were very similar (Treatment 1 (V_I =0.372 ±0.056); Treatment 2 (V_I=0.343 ±0.052); Table 5.2). Thus there is strong evidence of among-individual variance and population level plasticity across the treatments, but not for IxE.

Statistical support for genetic contributions to variance in detour task performance is somewhat equivocal. On one hand, the animal model was not a better fit than the simple repeated measures mixed model (LRT comparison of Models 4 and 2; Table 5.1) and in fact the V_A estimate was bound to zero in the former. On the other hand, LRT comparison suggested Model 5 was a significant improvement on Model 6 (which allowed no GxE and constrained constant V_A across treatments; $\chi 2_2$=12.452, P=0.002). We therefore fully acknowledge that statistical support for genetic involvement is limited. However, while Model 2 (simple repeatability model) has the lowest AIC, Model 5 (GxE) is next (with a ΔAIC of 1.633) suggesting that, to the extent genes do matter, so does GxE. Model 5 certainly provides our best estimate of genetic parameters and indicates the presence of moderate V_A under treatment 1, while V_A in treatment 2 was

bound to zero (model 5; treatment 1, V_A=0.214 ±0.140, h^2=0.263 ±0.154, Table 5.2). The cross treatment genetic correlation was estimated as negative but also bound to the edge of permissible parameter ($r_{G.treatment1.2}$ =-0.995, no standard error estimated).

Training trials

Plots of the raw data show a strong pattern of improvement in average performance (i.e. decreases in time to obtain food) across the trials (Figure 5.4). Males were slower on average (median) than females. Univariate models confirmed the statistical significance of these qualitative patterns, revealing significant effects of *sex* and *trial number* on (mean) performance. Note that estimated fixed effects were similar across models, and here we present estimates as obtained under Model 2. (Note that all fixed effect results, can be found in Supplementary Information Table S5.3). On average, males took longer to eat the food reward (coefficient ± SE=0.414 ± 0.088, $F_{1,285.1}$=14.94,P< 0.001), and *trial number* had a significant effect on performance $F_{8,2373}$=183.80, P< 0.001). On average fish obtained the food item faster as training experience in increased, coefficients becoming more negative as *trial number* increased (see Table S4.3). In contrast to the detour task itself, we also detected a *stack* had a significant effect of *stack* on performance in training, with fish in stack B significantly slower to obtain food than those in A (0.241 ± 0.088, $F_{1,283.8}$=7.53, P=0.006). Model comparisons provided strong support for the presence of among-individual variation in performance during the training period (LRT model

2 vs. model 1; $\chi 2_{0,1}$=1142.475, P<0.001, Table 5.3) with Model 2 yielding an estimate of R=0.485 ± 0.033. Model 3 (animal model) was not a significant improvement on this (LRT model 3 vs. model 2; $\chi 2_{0,1}$=0.352, P=0.277) so there is no statistical support for genetic variance. Under Model 3 the estimated heritability was very low (h^2= 0.034 ±0.060; Table S4.4).

The random regression modelling provided evidence for among-individual variation in the rate of improvement across training trials (LRT Model 4 vs. Model 2; $\chi 2_{0,2}$=116.760, P<0.001, Table 5.3). Among-individual variance in intercepts (int) and slope (slp) were estimated as V_{I_int}=0.581 ± 0.064 and V_{I_slp}=0.007 ± 0.001 respectively, with an among-individual intercept –slope correlation was estimated as $r_{I\_int\_slp}$=- 0.575 ± 0.057 (Note variance component estimates from all models can been seen in Supplementary Information Table S5.4). Here the negative among-individual intercept –slope correlation means that individuals with higher intercepts (i.e., poor performance time at the beginning of trials) tended to have smaller (i.e. more negative) slopes which correspond to faster improvement across training. Model 5 (GxE model) was not a significant improvement on Model 4 (LRT Model 5 vs. Model 4; $\chi 2_{0,3}$=2.096, P=0.553) providing no statistical support for the presence of genetic variance, or GxE, across repeated *trial number*.

There is thus strong statistical support for IxE (but not GxE) in training trial performance, where E is the 'environmental' axis of experience as measured by *trial number*. Since the consequences of this for current purposes are not immediately apparent, we projected the among-individual intercept-slope

covariance matrix to a ‘character state’ view (following e.g. Nussey et al., 2007; see Supplemental Information Table S5.5 for didactic explanation and corresponding code). This simply transforms the slope-intercept (co)variance structure to an estimate of the among-individual covariance matrix for the nine *trial number* specific performances. This transformation shows that, while IxE is statistically significant, its magnitude is insufficient to cause much change in V_I with *trial number* or to disrupt the uniformly positive among-individual correlation structure (Table S4.5). This means that individual rank ordering of performance largely stays the same across the training periods (i.e. there is not much reaction norm crossing). Due to this result, we elected revert to a random intercepts approach for this trait in the multivariate model (below).

Bivariate analysis

The bivariate model yielded an estimate of the among-individual covariance structure (**ID)** that revealed positive relationships between individual performance in the treatment specific detour tasks and the training trials (comparison of the full model to one in which all among-individual between trait covariances are fixed to zero; χ^2_3 = 33.389, P <0.001). Estimated among-individual correlations (± SE) were 0.491 ± 0.089 in treatment 1 (clear cylinder) and 0.306 ± 0.097 in treatment 2 (stripy cylinder; Table 5.4). Note that a strong positive correlation between individual performance in the two detour task between treatments was also obtained ($r_{ind.treatment1.2}$=0.816 ± 0.610; Table 5.4). Strictly this parameter is actually identifiable in our model however, although sibships are represented

across both levels no individual experienced both treatments, therefore the data here contains very little information for its estimation. Hence the SE is very large and we do not interpret this correlation further.

Extending the bivariate model to partition **ID** into **G** and **PE** added little additional insight, which was not unexpected given the results of univariate analyses. Estimates of V_A for training trial performance and detour task under treatment 1 were very similar to univariate models, while V_A for the detour task under model 2 was again bound to zero. All estimated genetic correlations were bound to (effectively) -1. Thus, with respect to genetics, the multivariate model simply recapitulates the conclusion that there is substantial genetic variance for the detour task in treatment 1 (V_A=0.238 ±0.122; similar to the univariate estimate). Statistical support remains equivocal however. The multivariate animal model is not a significantly better fit than the **ID** model using the expected degrees of freedom (LRT, **G** vs. **ID**; χ^2_6 = 9.356, P=0.155). However, one could make a *post hoc* argument that this is significant given that only 2 additional parameters are estimated in **G** (the other 4 of 6 being set to boundary conditions by the REML algorithm).

5.5 Discussion

In this study we sought to characterise among-individual and genetic variation in inhibitory control, using performance in a detour task as an observable proxy for the ability to inhibit prepotent behaviour. Our analyses show that

individuals vary considerably in detour task performance and thus, subject to important assumptions (discussed below), are likely to differ in inhibitory control. We also find strong evidence of plasticity in average performance across treatment environments. Contrary to predictions, when the visual information available for the detour task was increased by adding stripes to the clear cylinder, fish took longer to obtain the food item. Support for genetic contributions to observed variance was somewhat mixed, although we find some evidence for a GxE interaction, which results in heritable variation in performance only being evident in one treatment (the clear cylinder). By also analysing data collected from the period of training used to establish the cue-reward association needed for the detour task we show that individuals differ in performance during training and that this positively predicts, but is insufficient to fully explain, variation in the detour task itself. In what follows, we discuss the implications of these results in the context of understanding variation in, and evolution of, animal cognition. However, throughout our discussion we also highlight caveats and assumptions arising from the use of performance in the detour task as a proxy for the latent cognitive trait of inhibitory control, and consider alternative explanations for the observed results.

Among-individual and genetic variance in detour task performance

Guppies in our study show consistent among-individual variation in detour task performance measured as the time to navigate a barrier and obtain a food reward. Repeatability of performance is moderately high (e.g., 47% averaged across the

treatments) in comparison to estimates reported across assays designed to test cognitive variation in animals (see Cauchoix et al., 2018 for a review). Accepting that the task provides a valid measure of inhibitory control (an assumption we explore further below), this result is consistent with studies adopting similar experimental paradigms in this species (Santacà, Busatta, Lucon-Xiccato, et al., 2019; Lucon-Xiccato, Bisazza, et al., 2020; Macario et al., 2021). Our observations of apparent learning (i.e. fish become faster at locating the food reward over successive trials) and sexual dimorphism (i.e. females perform better than males) also recapitulate findings from these previous studies. Males are more persistent in trying to pass through the transparent cylinder and are therefore slower in the detour task (Lucon-Xiccato & Bisazza, 2017b). Though speculative, sex differences in inhibitory control might reflect sex-specific selection on behaviour since, for instance, persistent male harassment of females can be rewarded by higher rates of copulation (Endler, 1984; MaGurran et al., 1994). As such, sex specific reward systems may account for the differences in performance we see here between females and males. Based on performance in reversal learning tasks, several studies have also shown that male fish present lower cognitive flexibility than females when a learned response becomes inappropriate (Lucon-Xiccato et al., 2014; Brandão et al., 2019). It seems at least plausible that increased persistence coupled to reduced cognitive flexibility, presents as lower inhibitory control in male guppies relative to females in our experiment. We also note that it is possible that sex differences in performance may arise due to differences in reinforcement of the cue-reward association, rather than differences in learning ability per se.

While among-individual variation was strong, we did not find evidence of a significant genetic contribution to this using the standard animal model approach that assumes the phenotypic value of a genotype is constant across the two treatments. This result contrasts somewhat with reports of moderate, significant h^2 estimates for inhibitory control in humans and other animals (Friedman et al., 2008; Schachar et al., 2011; Cervantes et al., 2013; Gnanadesikan et al., 2020). However, when modelling genotype-by- environment interaction (GxE), our analyses actually suggest that there is moderate heritability (26.3%) but that this is context-specific, and specifically is limited to the clear cylinder treatment. From an evolutionary perspective, the obvious implication of this result is that, if inhibitory control is under directional selection, whether or not it evolves at all may depend on the environmental conditions.

While GxE implies the presence of environment-specific genetic variance, it can equally be understood as genetic variance for, and thus adaptive potential of, phenotypic plasticity (Pigliucci, 2001; Hill et al., 2010). Here, viewing our results from the perspective of phenotypic plasticity presents some complementary insights. We note that while our experimental design precludes directly observing individual-level plastic responses to treatment, differences in average detour task performance between clear and stripy cylinders show that plasticity occurs. Although our directional prediction was tentative, we had expected the stripy cylinder to provide more visual information and result in faster location of the food reward (Santos et al., 1999; Juszczak et al., 2016). In fact the opposite pattern arose, a result that suggests the more visible cylinder presents as an aversive stimulus, potentially by eliciting neophobia and/or predator

avoidance behaviour (Brown et al., 2013). A point that follows from this is that, the detected GxE for observed performance in the detour task need not strictly imply GxE for the latent trait of inhibitory control if observed time to eat the food object also depends on motivation (or neophobia) that is also genetically variable (at least in some environments).

As a more general caveat to our interpretation of results, we suggest that here – as in most cognitive testing studies – it is potentially problematic to view observed performance variation as being attributable to a single latent trait. Time to complete a detour task is widely used as a proxy for inhibitory control (although additional behaviours are also often measured; Lucon-Xiccato et al., 2017; Macario et al., 2021). Collection of more detailed behavioural data including, for example, whether individuals successfully inhibited redundant attempts to directly acquire the food reward through the barrier (Kabadayi et al., 2017; Santacà., et al., 2019), would have been useful to strengthen the assumed link between time to complete the task and inhibitory control. Unfortunately, the requirement for high-throughput phenotyping, coupled with severe constraints on researcher lab time (arising from the covid-19 pandemic) precluded the collection of such data. We acknowledge that completing the detour task may involve additional cognitive processes (e.g., working memory, route planning), behavioural traits (e.g., exploration; Kabadayi et al., 2018) and aspects of 'state' (e.g. prior experience, physiological condition, motivation). We attempted to minimise these sources of 'non-target' variation as much as possible by: standardising food rations, housing and water conditions to minimise differences in motivation or state; avoiding any capture of fish during the training and testing periods (since handling stress

influences observed behaviours Wong et al., 2008)); acclimating all individuals to experimental housing for 48 hours before the onset of training; and feeding only in the 'test' compartment of experimental housing to minimise any aversion to emergence into this area.

Variation in training performance

Several additional conclusions emerge from our analysis of the performance (measured as time to eat the food) in the training portion of the study. As in the subsequent detour test we see sex differences. Females are faster to eat than males, a common finding in guppy studies using food rewards (Laland et al., 1999; Lucon-Xiccato, Gatto, et al., 2020). Fish also become, on average faster at locating the food reward over successive training trials. We interpret this as likely reflecting that the training of the colour signal-food reward association utilised in the detour task was successful. Although, not a specific objective of our study this result adds to the accumulating evidence for associative learning in guppies (Trompf et al., 2014; Lucon-Xiccato & Bisazza, 2017a; Kniel et al., 2020).

We also show that individual fish vary in training performance overall (averaged across trials), and in the rates at which they improve over repeated training opportunities, but we find no statistical support for genetic contributions to this. If we accept an increase in mean performance with repeated training events as evidence of association learning in the population, then statistical support for among-individual variation in random regression slopes (IxE for

training performance where E is *trial number*) demonstrates among-individual differences in association learning (following e.g., Langley et al., 2020). However, some caution is warranted here. First, our analysis shows that the 'fastest learners' (i.e. those with steepest negative slopes) actually tend to be the poorer performers overall. In other words, fast learning fish are actually those with the greatest 'most room for improvement'. Second, the positive correlations across all ages under a 'character state' view of among-individual variation shows that initially poor, but fast learning individuals, remain poorer than average after 9 training trials (i.e. fast learning is not enough to compensate for a poor starting performance). Third, and just as with respect to the detour task data, we caution against an uncritical interpretation of these patterns. The variation among-individuals in time to obtain food in training is consistent with variation in learning ability, but other possibilities cannot be excluded. For instance, in the absence of cognitive variation *per se* some fish may be consistently faster that others to obtain a food if they are more exploratory (or less neophobic; Boogert et al., 2006; Bousquet et al., 2015; Zidar et al., 2018), or more motivated (van Horik & Madden, 2016). Distinguishing between these possibilities is experimentally challenging, and will be particularly so if, for instance, associative learning ability is correlated with aspects of personality (e.g. exploration and boldness) expected to impact time to obtain the reward. Such correlation structure has been reported in a number of fish and bird studies (Guillette et al., 2009; DePasquale et al., 2014; Quinn et al., 2016).

Although we have argued above that variation in training performance is interesting in its own right, the primary rationale for analysing this data was to

validate our assumption that the detour task is informative for cognitive processes, and specifically inhibitory control, that are distinct from those employed during the training period. Statistically, this assumption would be difficult to justify if variation among-individuals (and/or genotypes) in training was sufficient to explain all variation revealed in detour task performance. Biologically, this could arise if, for instance, variation in the detour task performance was completely explained by differences in motivation to feed, neophobia or ability to learn the cue-food association.

We actually find that individual performance in training positively predicts detour task performance in both treatments, but also that (based on approximate 95% CI) the corresponding among-individual correlations are also significantly less than +1. This implies that, factors contributing to among-individual variation in training performance (whether cognitive or otherwise) likely contribute to, but are insufficient to fully explain, detour task performance. Specifically, the estimated covariance structure in **ID** implies that 75% of among-individual variance in the detour task in treatment 1, and 90% in treatment 2, is statistically independent of training performance (based on calculating conditional variances following (Hansen et al., 2008).

This result is consistent with our assertion that detour task performance is informative for inhibitory control, even if observations also depend causally on, for instance, personality or other cognitive variation (such as operant rather than classical conditioning) revealed in the training period. We also note that the correlation structure could arise because of non-causal association of traits that

contribute to training performance (e.g. association learning) with inhibitory control. We cannot distinguish these possibilities from our experimental design but individual-level correlations of inhibitory control with personality and/or other cognitive traits have been reported in guppies (Lucon-Xiccato, Montalbano, et al., 2020), and several bird studies (Ashton et al., 2018; Langley et al., 2020).

Summary

In conclusion, here we find evidence of among-individual variation in inhibitory control in guppies. Our results suggest that GxE occurs such that heritable variation is present in one environment when a standard detour testing paradigm is applied, but not when additional visual information is provided in the form of a striped cylinder. This suggest that the genetic variance in, and so potential for further adaptive evolution of, inhibitory control may be highly sensitive to environmental conditions. An alternative view of the same phenomenon is that plasticity in detour task performance, and so (with some assumptions) inhibitory control is genetically variable. We also find that an individual's performance in training trials used to create a food-cue association positively predicts individual performance in the detour task. However, it is insufficient to explain all among-fish variance in the latter. As such, we argue that the detour task is capturing distinct cognitive processes (i.e., inhibitory control), however other cognitive factors and/or differences in 'personality' are likely contributing to among-individual variation in performance.

Table 5.1: LRT and AIC model comparison – detour task

AIC values and likelihood ratio test comparisons across univariate models for performance in the detour task trials.

Model	AIC	LnL	LRT Comparison	Test for	X2	DF	P
1	799.719	-396.860	-	-	-	-	-
2	587.178	-289.589	2 vs 1	V_I	214.541	0,1	<0.001
3	589.054	-289.527	3 vs 2	IxE	0.124	1	0.725
4	589.178	-289.589	4 vs 2	V_A	0	0,1	0.499
5	588.811	-286.406	-	-	-	-	-
6	591.054	-289.527	6 vs 5	GxE	12.452	2	0.002

Table 5.2: Estimated variance components and derived parameters for the six models of performance in the detour task trials. Subscripts denote block (B), brood tank (BT), residual (R), individual (I), permanent environment (PE) and additive genetic (A) components of variance. Also shown, where applicable are corresponding estimates of repeatability (R), heritability (h^2), the intraclass correlation corresponding to V_{PE} (denoted pe^2) and the cross-treatment genetic correlation (r_G). Parameter estimates specific to treatment 1 (clear cylinder) and 2 (stripy cylinder) are denoted with subscripts and standard errors are provided in parentheses where available. Note that variances were constrained to be positive and correlations between -1 and +1. Where parameters are fixed at boundary conditions no SE is estimated.

Parameter	**Model**					
	1	**2**	**3**	**4**	**5**	**6**
V_B	0.018 (0.013)	0.016 (0.014)	0.015 (0.013)	0.016 (0.014)	0.011 (0.012)	0.015 (0.013)
V_{BT}	0.013 (0.012)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
V_R	0.730 (0.032)	0.392 (0.021)	0.392 (0.021)	0.392 (0.021)	0.392 (0.021)	0.392 (0.021)
V_I	-	0.357 (0.038)	$0.372\ (0.056)_1$	-	-	-
			$0.343\ (0.052)_2$			
V_{PE}	-	-	-	0.357 (0.038)	$0.199\ (0.096)_1$	$0.371\ (0.056)_1$
					$0.338\ (0.052)_2$	$0.343\ (0.051)_2$
V_A	-	-	-	0.000 (-)	$0.214\ (0.140)_1$	0.000 (-)
					$0.000\ (-)_2$	
R	-	0.467 (0.033)	$0.477\ (0.043)_1$	-	-	0.392 (0.021)

			0.458 (0.042)$_2$			
pe^2	-	-	-	0.467 (0.033)	0.244 (0.126)$_1$	0.477 (0.043)$_1$
					0.456 (0.043)$_2$	0.458 (0.042)$_2$
h^2	-	-	-	0.000 (-)	0.263 (0.154)$_1$	0.000 (-)
					0.000 (-)$_2$	
r$_G$	-	-	-	-	-0.995 (-)	-

Table 5.3: LRT and AIC model comparison – training trials

AIC values and likelihood ratio test comparisons across univariate models for performance in the training trials.

Model	AIC	LnL	LRT Comparison	Test for	X2	DF	P
1	1803.996	-898.998	-	-	-	-	-
2	663.521	-327.761	2 vs 1	V_I	1008.849	0,1	<0.001
3	665.170	-327.585	3 vs 2	V_A	0.352	0,1	0.277
4	550.762	-269.381	4 vs 2	IxE	116.760	0,2	<0.001
5	554.666	-268.333	5 vs 4	GxE	2.096	0.3	0.553

Table 5.4: ID variance–covariance–correlation matrix from the bivariate mixed model. Estimated variances are shown on the diagonal (dark grey shading), with correlations above and covariances below. Also shown are the **G** and **PE** variance–correlation matrices estimates from the bivariate animal model (covariances are not shown as back-calculating them becomes problematic with variances bound to zero and/or correlations at ±1). Standard errors are shown in parentheses where available, and bold font denotes nominally significant estimates assuming approximate 95% CI of ± 1.96SE.

	Training Trials	**Detour Task**	
		Treatment 1	**Treatment 2**
ID Matrix			
Training Trials	**0.379 (0.037)**	**0.491 (0.089)**	**0.306 (0.097)**
Detour Task - Treatment 1	**0.187 (0.040)**	**0.376 (0.037)**	0.816 (0.607)
Detour Task - Treatment 2	**0.110 (0.038)**	0.293 (0.220)	**0.343 (0.052)**
G Matrix			
Training Trials	0.004 (0.012)	-0.999 (-)	-0.999 (-)
Detour Task - Treatment 1	-	0.238 (0.122)	-0.999 (-)
Detour Task - Treatment 2	-	-	0.000 (-)
PE Matrix			
Training Trials	**0.377 (0.037)**	**0.770 (0.212)**	**0.309 (0.096)**
Detour Task - Treatment 1	-	**0.190 (0.087)**	0.999 (-)
Detour Task - Treatment 2	-	-	**0.338 (0.051)**

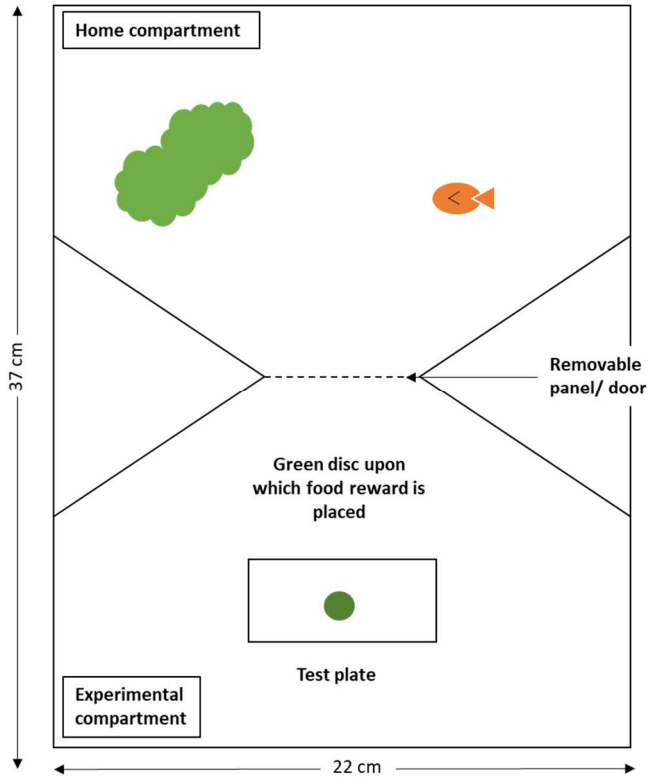


Figure 5.1: Aerial view of the tank set up used for association training trials and detour task trials.

(a) (b)

Figure 5.2: Photographs of the test plate used during the detour trials for a) treatment 1, showing the (unmarked) transparent cylinder which represented low visual information, and for b) treatment 2 which shows the cylinder with three black horizontal lines, used to represent high visual information. The photographs show the open ends of the cylinders in which fish could access to the food reward, in addition to the green plastic disc upon which the food reward was placed.

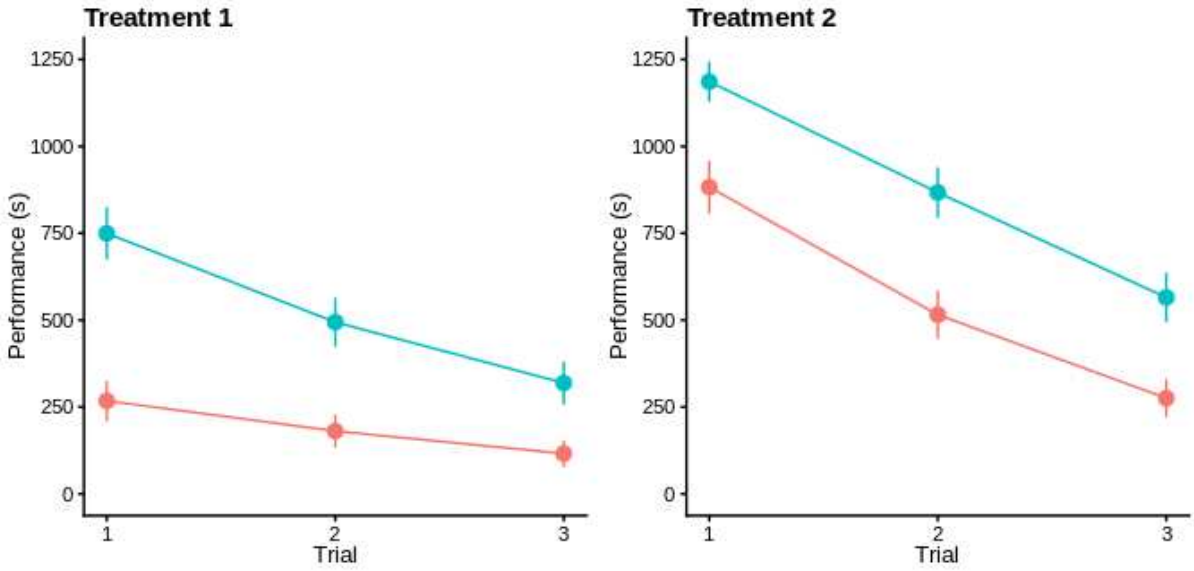


Figure 5.3: Plots of raw data of performance time across trials. Plot represent mean and standard errors for performance time in the detour task in treatment 1 (clear cylinder) and 2 (stripy cylinder), across the 3 trials for males (blue) and females (red). Error bars represent mean and standard errors of performance time for individuals.

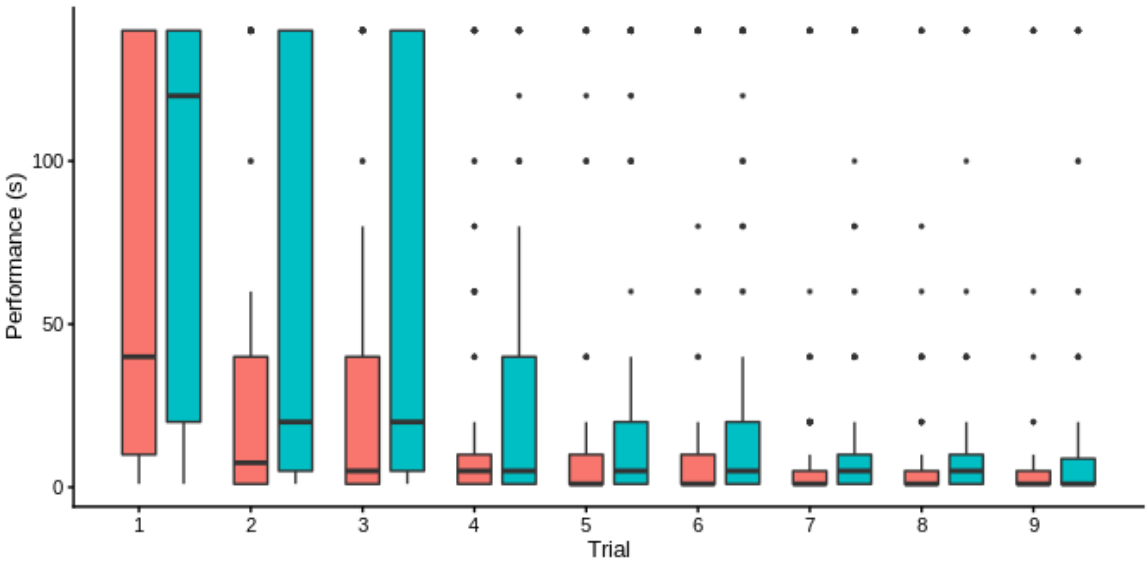


Figure 5.4. Boxplot of data distributions for performance in the 9 training trials for males (blue) and females (red). Horizontal lines within box correspond to behavioural medians, box boundaries correspond to first and third quartiles. When present, whiskers correspond to 10th and 90th percentiles, and points correspond to outliers.

6 Chapter 6: General Discussion

My PhD thesis broadly aimed to investigate the causes of variation in cognitive and personality traits in *Poecilia reticulata*, and investigate the extent to which genetic variation contributes to these and to relationships between them. The question of how natural selection can give rise to repeatable among-individual differences in cognitive and personality traits has attracted much attention from theoreticians and empiricists (Réale et al., 2007; Wolf et al., 2012; Croston et al., 2015; Boogert et al., 2018). However, to date and despite the notable scientific input, the evolutionary origin of adaptive cognitive behaviour is still poorly understood. My thesis chapters were specifically planned to further our understanding of the adaptive causes of individual differences in cognitive and personality traits. Here, I summarise the results of each chapter before suggesting some potential improvements to methodologies used here, and then outline some possible directions of future work to assess the adaptive nature of cognitive performance traits.

In terms of understanding the evolution of traits, among-individual variation is important as it is as pre-requisite for natural selection and genetic variation – both of which are required for adaptive evolution to occur (Wilson et al., 2010). Phenotypic variation is found in almost all labile traits even in the presence of strong natural selection; however strong directional or stabilising selection is usually predicted to erode variation in phenotypes, or at least its genetic component. This presents an evolutionary conundrum. What maintains genetic, and so presumably, among-individual variation?

The most commonly hypothesised ‘resolution’ to this conundrum in evolutionary ecology is that because (i) natural selection does not act on traits in isolation, and (ii) traits are correlated, trade-offs arise that impose constraints on adaptation and maintain variation. If this idea applies to cognition, then the extent to which cognitive traits covary with other aspects of phenotype (e.g. personality traits) is expected to be an important factor in shaping their evolutionary dynamics.

In **Chapter 2,** I investigated hypothesised links between cognitive performance and personality traits. I used a multivariate approach to investigate the presence of (co)variation in cognitive performance in two spatial learning tasks and a behavioural stress response in an open field task. I found among-individual differences in task performance that were repeatable within- and across both maze layouts, where performance in the first maze predicted performance in the second maze. Average performance improved with experience in the first maze, consistent with spatial learning. However, there was no evidence of average improvement within the second maze, potentially owing to cumulative chronic stress effects. Individuals also differed in ‘learning rate’ across both mazes, suggesting that while some individuals improved across trials, performance declines for others, especially in the second maze. Although personality, measured here as (repeatable) stress response behaviour, did not correlate with cognitive performance, it seems both possible and plausible that cumulative, chronic stress effects contribute to declining performance. If so, this study points to the potential importance of considering chronic stress effects in cognitive assays. Chronic stress can be induced by repeated handling and/or

exposure to assays and thus represents a potentially widespread but currently poorly acknowledged challenge for characterisation of cognitive variation in animal studies.

In **Chapter 3**, I continued my investigation of the structure of variation in cognition and personality differences at the among-individual level. I again used multivariate analytical approaches, but here also took a more 'holistic' view of cognitive phenotype by estimating the **ID** matrix among traits used as performance proxies across three domains; *association learning* in a colour discrimination task; *motor cognition* in a novel motor task and *cognitive flexibility* in a reversal learning task, in addition to a measure of the personality trait of 'boldness'. There was no evidence of trade-offs between personality and cognitive performance (both among- and within- cognitive domains). I found a strong positive domain-general correlation structure in the **ID**-matrix among all traits, and (on the correlation scale) 57% of the variation in multivariate cognitive performance was explained by the leading eigen vector. Although there are some important assumptions and caveats, these findings are consistent with variation in an overall, domain-general cognitive performance trait (similar in structure to the 'general intelligence (*g*)' model; Burkart et al., 2017; Galsworthy et al., 2005; Plomin & Spinath, 2002). If the correlation structure in **ID** is recapitulated in **G**, selection on any one trait would cause positively correlated evolutionary responses across all others (Roff, 2002). Shy-bold type behavioural variation can have important fitness consequences (Wilson et al., 2010; Ariyomo et al., 2012; Ballew et al., 2017), and selection on this is therefore likely to impact evolutionary dynamics of cognitive performance (and *vice versa*). Thus **Chapter 3**

demonstrates the importance of considering the multivariate phenotype, which gives a much broader and realistic view of how traits are likely to respond to selection, revealing ways through which genetic and phenotypic variation in behaviour can be maintained (Wolf et al., 2012).

The extent to which among-trait associations shape evolutionary adaptation strictly really depends not on the phenotypic correlation structure *per se*, but on the genetic contribution to this. As such, in **Chapter 4** I built on the among-individual variation characterised in **Chapter 2** and **3** by investigating the extent to which behavioural differences may be explained by genetic factors. In **Chapter 4**, I built on previous work demonstrating genetic variation in personality, measured as (individual mean) behavioural stress-response in open field trials (OFT)**.** With collaborators, I used a novel form of 'double hierarchical' model to test for among-individual and genetic variation in mean behaviour, but also in within-individual variation (otherwise known trait 'predictability'). We used a pedigree of fish and subjected individuals to repeated OFT assays. The idea of investigating phenotypic 'predictability' is relatively new but has attracted increasing attention because, for instance two individuals may be equally 'bold' on average (as determined from the individual mean behaviour across multiple observations) but differ in how much variation they exhibit around their individual-level mean phenotypes.

Here we found individuals differed in behavioural predictability, and that this variation was heritable, meaning behavioural predictability can be viewed as a trait with adaptive potential under selection. Furthermore, we also found

evidence of a genetic correlation structure between the behavioural trait mean and individual predictability, where individuals - and genotypes – that were more explorative were also more predictable (i.e. less variable) in their behavioural response to the OFT. This leads to the expectation that these aspects of phenotype (mean behaviour and predictability) will coevolve under selection. In fact the stress coping style (SCS) model proposes a relationship between average behavioural response to a stressor stimulus, and the predictability of behaviour, where high (mean) 'flight' behaviour will be linked to high trait predictability within the proactive coping style. Our results confirm the association but did not fully align with its specific predictions (we found high (mean) 'flight' behaviour was linked to low predictability). In part this may be because the trait we focussed on to measure behavioural stress response arguably revealed variation in the magnitude, rather than 'style' of the behavioural stress response. Admittedly the distinction between these concepts is somewhat indistinct. Nonetheless, our findings do emphasise the need to evaluate genetic integration among stress-response traits in a multivariate empirical framework. On a more general level, our results show that the common assumption of homogeneous residuals, typical to linear (mixed) models applied in personality research is violated (Dingemanse et al., 2013; Brommer, 2013). While this does not invalidate conclusions from simpler models addressing variation in (individual) mean behaviour only, it does demonstrate that deeper insights can be obtained by also modelling variability in trait predictability.

Finally, in **Chapter 5** I returned to the central theme of cognitive variation in a study of *inhibitory control*. In this chapter, I firstly used a mixed breeding

design, coupled with microsatellite genotyping and pedigree analysis, to reconstruct a pedigree of fish of known relatedness. I then exposed individuals to a 'detour task' to test for genetic variance in inhibitory control. I also tested for genotype-by-environment interactions (GxE) by testing related fish under alternative experimental treatments that differ in degree of available visual information (using transparent vs semi-transparent barriers in the detour task). To date, few animal cognition studies have characterised the importance of GxE. Given that cognition is defined in relation to acquiring, processing and using *information in the environment*, the possibility of GxE interactions are very plausible and intuitive. In terms of phenotypic plasticity, average performance was plastic across treatments, and mean performance in the clear cylinder treatment was faster.

There was also evidence of GxE, which resulted in heritable variation in performance being present in the clear cylinder treatment only. GxE here implies the presence of environment-specific genetic variance, but it can equally be understood as genetic variance for, and thus adaptive potential of, phenotypic plasticity (Pigliucci, 2001; Hill et al., 2010). In terms of adaptive evolutionary potential, inhibitory control may be highly sensitive to environmental conditions, which may provide a mechanism by which additive genetic variation is maintained. Equivalently, this result means that the plastic response of detour task performance to treatment environment is itself genetically variable. This highlights the importance of testing environment in studies investigating genetic variation of cognitive traits. Furthermore in **Chapter 5,** I presented repeatable among-individual differences in performance in the training trials and in learning

rate across trials, where ‘fastest learners’ (i.e. those with the steepest negative slopes) were poorer performers (slower than average by the end of trials). I also demonstrated that, although individual performance in training trials positively predicted performance in the detour task, it was not a sufficient explanation for the latter. This finding suggests that the detour task incorporates cognitive processes (putatively inhibitory control) that are distinct from those employed during the initial training trails (cognitive or non-cognitive). These findings again highlight the importance of multivariate approaches for estimating patterns of (co)variation that might otherwise go undetected in univariate trait-by-trait analysis.

Concluding remarks and future directions

The question of how, if and why, cognition evolves has been pivotal, receiving growing attention in the last few years. Thornton & Wilson's (2015) commentary on cognitive evolution highlighted the necessary conditions for traits to respond to selection, expressed as the “Darwinian Holy Trinity”: 1) phenotypic variation that is 2) heritable and 3) affects fitness. It is clear that high levels of variation in cognitive performance is present among individuals of the same species (Boogert et al., 2018; Cauchoix et al., 2018). However while estimates of among-individual variation are common, estimates of heritable variation are limited in comparison. This is not surprising owing to challenges of collecting large enough data sets of robust measures of cognition required to statistically partition

variance using quantitative genetic analysis (discussed in **Chapter 1)**. The challenges of estimating genetic variation in cognitive traits were also discussed in **Chapter 5**, where genetic variance in detour task performance was detected under one environmental treatment. This chapter highlighted the importance of the environment on the expression of (genetic) variance in (putative) inhibitory control, and suggests that the adaptive evolutionary potential of cognitive performance may be highly sensitive to environmental conditions. GxE interactions are increasingly studied in the animal personality literature. In **Chapter 4,** we used quantitative genetic approaches to demonstrate that GxE interactions contribute to personality traits. However to date, few animal cognition studies have characterised the importance of GxE. Given that cognition is defined in relation to acquiring, processing and using *information in the environment*, the possibility of GxE interactions are very plausible, and represent an important and interesting avenue for future research.

It is commonly assumed (but seldom tested) that elevated cognitive performance provides fitness benefits (Thornton et al., 2019). For traits to evolve, there must be selective consequences of the underlying phenotypic variation present within populations. Throughout my thesis I have presented evidence of among-individual variation in cognitive performance across multiple cognitive domains (**Chapter 2, 3 and 5)**, in addition to evidence of genetic variation contributing to among-individual differences (**Chapter 5**). Fitness consequences arising from among-individual and genetic variance in cognitive performance were not explored in this thesis. However, in order to fully test the adaptive potential of cognitive traits, we require both knowledge of the underlying genetic

architecture, and selection. Estimates of selection on cognitive traits are critical to our understanding of how natural selection may drive the evolution of cognitive abilities (Thornton et al., 2012b, 2014). An approach that targets individual variation in cognition and fitness is key to elucidating contemporary selection in wild populations. Potential avenues for future research could explore experimental assays targeting heritable cognitive performance and fitness measures. For example, an intermediary solution could involve measuring cognitive performance on a pedigree of guppies (of known relatedness) bred in the lab, following which individuals are released into a semi-wild mesocosm, where measures of fitness can be estimated over time (e.g. survival and reproduction).

Throughout my thesis, I have advocated the use of quantitative genetic modelling approaches as providing an ideal framework for investigating among-individual differences in cognitive and behavioural variation. They allow patterns of among-individual (co)variation to be estimated and, through the use of multivariate mixed models and eigen decomposition, direct estimation of axes of among-individual variation to be summarised. In **Chapter 3,** I used a multivariate approach to characterise and test hypothesised relationships of cognitive performance across multiple domains and personality. Although the use of multivariate approaches are limited within the field of animal cognition, recent studies have utilised linear mixed models (e.g. Moiron et al., 2016), and with the inclusion of pedigree information, animal models have been used to investigate multivariate correlation structures in cognitive traits (Sorato et al., 2018; Langley et al., 2020; Poirier et al., 2020). Future research investigating evolutionary

causes and consequences of cognitive traits would benefit from considering the multivariate approaches to give a broader and more realistic view of how traits are likely to respond to selection (Wolf et al., 2012).

Recently mixed model approaches have been advocated due to their ease in controlling for confounding variables, which are common when measuring latent traits such as cognition (Thornton et al., 2015; Morand-Ferron et al., 2016). In contrast to other biological traits, cognition is unobservable and can only be inferred through observation of behavioural or neural responses in carefully designed experiments. This represents a fundamental problem when trying to accurately measure a latent trait such a cognition. As in most cognitive studies, it is potentially problematic to view observed performance variation as being attributable to a single latent trait. General caveats and assumptions must be made when quantifying individual differences in cognitive abilities, and confounding variables must be accounted for. Cognitive performance in many tasks may involve additional cognitive processes and/ or behavioural traits. Furthermore, aspects of 'state' (e.g. prior experience, physiological condition, motivation) will add further complexity to how cognitive abilities are expressed, and each of these factors are also likely to interact. The effect of 'non-target' variation on the expression of cognitive abilities can have implications for quantifying individual differences, and may also obscure repeatability estimates for learning performances (Cauchoix et al., 2018) and general cognitive ability (Burkart et al., 2017). These issues are highlighted throughout this thesis (**Chapters 1, 2, 3 and 5)**, and I have attempted to account for and minimise sources of 'non-target' variation as much as possible. Multivariate approaches

make it easier to control for such confounding variables (Nakagawa et al., 2010). However, although complimentary, multivariate analysis is not a substitute for good experimental design. Combining conceptual and experimental rigor of cognitive science, with the large sample sizes and powerful analytical framework of quantitative genetics is advised. In this way, we can improve our understanding of the genetic causes and fitness consequences of among-individual variation in cognition, in the hope of elucidating the evolutionary processes that have shaped, and continue to shape, animal cognition.

7 Bibliography

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281–287. https://doi.org/10.1016/j.cognition.2010.10.022

Arden, R., & Adams, M. J. (2016). A general intelligence factor in dogs. *Intelligence*, *55*, 79–85. https://doi.org/10.1016/j.intell.2016.01.008

Ariyomo, T. O., & Watt, P. J. (2012). The effect of variation in boldness and aggressiveness on the reproductive success of zebrafish. *Animal Behaviour*, *83*(1), 41–46. https://doi.org/10.1016/j.anbehav.2011.10.004

Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature*, *554*(7692), 364–367. https://doi.org/10.1038/nature25503

Ballew, N. G., Mittelbach, G. G., & Scribner, K. T. (2017). Fitness consequences of boldness in Juvenile and adult largemouth bass. *American Naturalist*, *189*(4), 396–406. https://doi.org/10.1086/690909

Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: Individual differences across cognitive task performance in a new world monkey (Saguinus oedipus). *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005883

Barrera, G., Alterisio, A., Scandurra, A., Bentosela, M., & D'Aniello, B. (2019). Training improves inhibitory control in water rescue dogs. *Animal Cognition*, *22*(1), 127–131. https://doi.org/10.1007/s10071-018-1224-9

Bebus, S. E., Small, T. W., Jones, B. C., Elderbrock, E. K., & Schoech, S. J. (2016). Associative learning is inversely related to reversal learning and varies with nestling corticosterone exposure. *Animal Behaviour*, *111*, 251–260. https://doi.org/10.1016/j.anbehav.2015.10.027

Becher, S. A., Russell, S. T., & Magurran, A. E. (2002). Isolation and characterization of polymorphic microsatellites in the Trinidadian guppy (Poecilia reticulata). *Molecular Ecology Notes*, *2*(4), 456–458. https://doi.org/10.1046/j.1471-8286.2002.00276.x

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. In *Animal Behaviour* (Vol. 77, Issue 4, pp. 771–783). https://doi.org/10.1016/j.anbehav.2008.12.022

Ben-Shahar, Y., & Robinson, G. E. (2001). Satiation differentially affects performance in a learning assay by nurse and forager honey bees. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology*, *187*(11), 891–899. https://doi.org/10.1007/s00359-001-0260-z

Bensky, M. K., Paitz, R., Pereira, L., & Bell, A. M. (2017). Testing the predictions of coping styles theory in threespined sticklebacks. *Behavioural Processes*, *136*, 1–10. https://doi.org/10.1016/j.beproc.2016.12.011

Beran, M. J. (2015). The comparative science of “self-control”: What are we talking about? *Frontiers in Psychology*, *6*(JAN). https://doi.org/10.3389/fpsyg.2015.00051

Bergero, R., Gardner, J., Bader, B., Yong, L., & Charlesworth, D. (2019).

Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(14), 6924–6931. https://doi.org/10.1073/pnas.1818486116

Billard, P., Clayton, N. S., & Jozet-Alves, C. (2020). Cuttlefish retrieve whether they smelt or saw a previously encountered item. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-62335-x

Biro, P. A., & Adriaenssens, B. (2013). Predictability as a Personality Trait: Consistent Differences in Intraindividual Behavioral Variation. *The American Naturalist*, *182*(5), 621–629. https://doi.org/10.1086/673213

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

Biro, P. A., & Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? In *Trends in Ecology and Evolution* (Vol. 25, Issue 11, pp. 653–659). https://doi.org/10.1016/j.tree.2010.08.003

Bitterman, M. E. (1965). Phyletic differences in learning. *The American Psychologist*, *20*, 396–410. https://doi.org/10.1037/h0022328

Bobrowicz, K., & Osvath, M. (2018). Cats parallel great apes and corvids in motor self-regulation - not brain but material size matters. *Frontiers in Psychology*, *9*(OCT). https://doi.org/10.3389/fpsyg.2018.01995

Bókony, V., Lendvai, Á. Z., Vágási, C. I., Pătraş, L., Pap, P. L., Németh, J.,

Vincze, E., Papp, S., Preiszner, B., Seress, G., & Liker, A. (2014). Necessity or capacity? Physiological state predicts problem-solving performance in house sparrows. *Behavioral Ecology*, *25*(1), 124–135. https://doi.org/10.1093/beheco/art094

Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour*, *81*(6), 1209–1216. https://doi.org/10.1016/j.anbehav.2011.03.004

Boogert, N. J., Fawcett, T. W., & Lefebvre, L. (2011). Mate choice for cognitive traits: A review of the evidence in nonhuman vertebrates. In *Behavioral Ecology* (Vol. 22, Issue 3, pp. 447–459). https://doi.org/10.1093/beheco/arq173

Boogert, N. J., Madden, J. R., Morand-Ferron, J., & Thornton, A. (2018). Measuring and understanding individual differences in cognition. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1756, p. 20170280). https://doi.org/10.1098/rstb.2017.0280

Boogert, N. J., Reader, S. M., & Laland, K. N. (2006). The relation between social rank, neophobia and individual learning in starlings. *Animal Behaviour*, *72*(6), 1229–1239. https://doi.org/10.1016/j.anbehav.2006.02.021

Bouchard, J., Goodyer, W., & Lefebvre, L. (2007). Social learning and innovation are positively correlated in pigeons (Columba livia). *Animal*

Cognition, *10*(2), 259–266. https://doi.org/10.1007/s10071-006-0064-1

Boulton, K., Grimmer, A. J., Rosenthal, G. G., Walling, C. A., & Wilson, A. J. (2014). How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheepshead swordtail, Xiphophorus birchmanni. *Behavioral Ecology and Sociobiology*, *68*(5), 791–803. https://doi.org/10.1007/s00265-014-1692-0

Boulton, K., Walling, C. A., Grimmer, A. J., Rosenthal, G. G., & Wilson, A. J. (2018). Phenotypic and genetic integration of personality and growth under competition in the sheepshead swordtail, Xiphophorus birchmanni. *Evolution*, *72*(1), 187–201. https://doi.org/10.1111/evo.13398

Bousquet, C. a. H., Petit, O., Arrivé, M., Robin, J. P., & Sueur, C. (2015). Personality tests predict responses to a spatial-learning task in mallards, Anas platyrhynchos. *Animal Behaviour*, *110*, 145–154. https://doi.org/10.1016/j.anbehav.2015.09.024

Brandão, M. L., Fernandes, A. M. T. de A., & Gonçalves-de-Freitas, E. (2019). Male and female cichlid fish show cognitive inhibitory control ability. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-52384-2

Bräuer, J., Hanus, D., Pika, S., Gray, R., & Uomini, N. (2020). Old and new approaches to animal cognition: There is not “one cognition.” In *Journal of Intelligence* (Vol. 8, Issue 3, pp. 1–25). https://doi.org/10.3390/jintelligence8030028

Bridger, D., Bonner, S. J., & Briffa, M. (2015). Individual quality and personality:

bolder males are less fecund in the hermit crab Pagurus bernhardus. *Proc Biol Sci*, *282*(1803), 20142492. https://doi.org/10.1098/rspb.2014.2492

Briffa, M. (2013). Plastic proteans: Reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, *9*(5), 20130592. https://doi.org/10.1098/rsbl.2013.0592

Brommer, J. E. (2013). On between-individual and residual (co)variances in the study of animal personality: Are you willing to take the “individual gambit”? In *Behavioral Ecology and Sociobiology* (Vol. 67, Issue 6, pp. 1027–1032). https://doi.org/10.1007/s00265-013-1527-4

Brown, C., & Braithwaite, V. A. (2005). Effects of predation pressure on the cognitive ability of the poeciliid Brachyraphis episcopi. *Behavioral Ecology*, *16*(2), 482–487. https://doi.org/10.1093/beheco/ari016

Brown, C., & Irving, E. (2014). Individual personality traits influence group exploration in a feral guppy population. *Behavioral Ecology*, *25*(1), 95–101. https://doi.org/10.1093/beheco/art090

Brown, G. E., Ferrari, M. C. O., Elvidge, C. K., Ramnarine, I., & Chivers, D. P. (2013). Phenotypically plastic neophobia: A response to variable predation risk. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1756). https://doi.org/10.1098/rspb.2012.2712

Brust, V., & Guenther, A. (2017). Stability of the guinea pigs personality – cognition – linkage over time. In *Behavioural Processes* (Vol. 134, pp. 4–11). https://doi.org/10.1016/j.beproc.2016.06.009

Buchanan, K., Perera, T. B. de, … C. C.-A., & 2012, U. (2020). Guidelines for the treatment of animals in behavioural research and teaching. *Animal Behaviour*, *159*, I–XI. https://doi.org/10.1016/j.anbehav.2019.11.002

Buechel, S. D., Boussard, A., Kotrschal, A., van Der Bijl, W., & Kolm, N. (2018). Brain size affects performance in a reversal-learning test. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871). https://doi.org/10.1098/rspb.2017.2031

Burkart, J. M., Schubiger, M. N., & Van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*, e195. https://doi.org/10.1017/S0140525X16000959

Burns, J. G., & Rodd, F. H. (2008). Hastiness, brain size and predation regime affect the performance of wild guppies in a spatial memory task. *Animal Behaviour*, *76*(3), 911–922. https://doi.org/10.1016/j.anbehav.2008.02.017

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2018). ASReml-R Reference Manual Version 4 ASReml estimates variance components under a general linear mixed model by residual maximum likelihood (REML). In *asreml.kb.vsni.co.uk*. http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Byrne, R. W., & Bates, L. A. (2007). Sociality, Evolution and Cognition. In *Current Biology* (Vol. 17, Issue 16). https://doi.org/10.1016/j.cub.2007.05.069

Callander, S., Kahn, A. T., Hunt, J., Backwell, P. R. Y., & Jennions, M. D.

(2013). The effect of competitors on calling effort and life span in male field crickets. *Behavioral Ecology*, *24*(5), 1251–1259. https://doi.org/10.1093/beheco/art059

Carazo, P., Noble, D. W. A., Chandrasoma, D., & Whiting, M. J. (2014). Sex and boldness explain individual differences in spatial learning in a lizard. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1782), 20133275. https://doi.org/10.1098/rspb.2013.3275

Carere, C., Maestripieri, D., Caramaschi, D., Carere, C., Sgoifo, A., & Koolhaas, J. M. (2014). Neuroendocrine and Autonomic Correlates of Animal Personalities. In *Animal Personalities* (pp. 353–377). https://doi.org/10.7208/chicago/9780226922065.003.0013

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Carroll, J. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Cauchoix, M., Chow, P. K. Y., Van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., Bize, P., Boussard, A., Buechel, S. D., Cabirol, A., Cauchard, L., Claidière, N., Dalesman, S., Devaud, J. M., Didic, M., Doligez, B., Fagot, J., Fichtel, C., Henke-Von Der Malsburg, J., … Morand-Ferron, J. (2018). The repeatability of cognitive performance: A meta-

analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756). https://doi.org/10.1098/rstb.2017.0281

Cauchoix, M., Hermer, E., Chaine, A. S., & Morand-Ferron, J. (2017). Cognition in the field: Comparison of reversal learning performance in captive and wild passerines. *Scientific Reports*, *7*(1), 12945. https://doi.org/10.1038/s41598-017-13179-5

Cervantes, M. C., Laughlin, R. E., & Jentsch, J. D. (2013). Cocaine self-administration behavior in inbred mouse lines segregating different capacities for inhibitory control. *Psychopharmacology*, *229*(3), 515–525. https://doi.org/10.1007/s00213-013-3135-4

Chang, C. C., Teo, H. Y., Norma-Rashid, Y., & Li, D. (2017). Predator personality and prey behavioural predictability jointly determine foraging performance. *Scientific Reports*, *7*(1), 40734. https://doi.org/10.1038/srep40734

Charmantier, A., Garant, D., & Kruuk, L. (2013). *Quantitative genetics in the wild*. Oxford University Press. https://books.google.co.uk/books?hl=en&lr=&id=DHVYAwAAQBAJ&oi=fnd&pg=PP1&dq=Quantitative+genetics+in+the+wild&ots=trdFh_SKUL&sig=FBRiAlswrhHxHSNuSht8lsVmdi4

Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution*, *42*(5), 958–968. https://doi.org/10.1111/j.1558-5646.1988.tb02514.x

Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed-accuracy tradeoffs in animal decision making. In *Trends in Ecology and Evolution* (Vol. 24, Issue 7, pp. 400–407). https://doi.org/10.1016/j.tree.2009.02.010

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272–274. https://doi.org/10.1038/26216

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: Statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37. https://doi.org/10.1111/2041-210X.12281

Cole, E. F., Cram, D. L., & Quinn, J. L. (2011). Individual variation in spontaneous problem-solving performance among wild great tits. *Animal Behaviour*, *81*(2), 491–498. https://doi.org/10.1016/j.anbehav.2010.11.025

Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences reproductive life history variation in the wild. *Current Biology*, *22*(19), 1808–1812. https://doi.org/10.1016/j.cub.2012.07.051

Coppens, C. M., de Boer, S. F., & Koolhaas, J. M. (2010). Coping styles and behavioural flexibility: towards underlying mechanisms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/rstb.2010.0217

Croft, D. P., Albanese, B., Arrowsmith, B. J., Botham, M., Webster, M., & Krause, J. (2003). Sex-biased movement in the guppy (Poecilia reticulata).

Oecologia, *137*(1), 62–68. https://doi.org/10.1007/s00442-003-1268-6

Croft, D. P., Arrowsmith, B. J., Bielby, J., Skinner, K., White, E., Couzin, I. D., Magurran, A. E., Ramnarine, I., & Krause, J. (2003). Mechanisms underlying shoal composition in the Trinidadian guppy, Poecilia reticulata. *Oikos*, *100*(3), 429–438. https://doi.org/10.1034/j.1600-0706.2003.12023.x

Croston, R., Branch, C. L., Kozlovsky, D. Y., Dukas, R., & Pravosudov, V. V. (2015). Heritability and the evolution of cognitive traits. In *Behavioral Ecology* (Vol. 26, Issue 6, pp. 1447–1459). https://doi.org/10.1093/beheco/arv088

Daneman, M., & Carpenter, P. A. (1980). Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behvior*, *19*, 450–466.

Deary, I. J. (2001). Human intelligence differences: A recent history. In *Trends in Cognitive Sciences* (Vol. 5, Issue 3, pp. 127–130). Elsevier. https://doi.org/10.1016/S1364-6613(00)01621-1

Deary, I. J., Johnson, W., & Houlihan, L. M. (2009). Genetic foundations of human intelligence. In *Human Genetics* (Vol. 126, Issue 1, pp. 215–232). https://doi.org/10.1007/s00439-009-0655-4

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nat Rev Neurosci*, *11*(3), 201–211. https://doi.org/10.1038/nrn2793

Del Giudice, M., & Crespi, B. J. (2018). Basic functional trade-offs in cognition:

An integrative framework. *Cognition*, *179*, 56–70. https://doi.org/10.1016/j.cognition.2018.06.008

DePasquale, C., Wagner, T., Archard, G. A., Ferguson, B., & Braithwaite, V. A. (2014). Learning rate and temperament in a high predation risk environment. *Oecologia*, *176*(3), 661–667. https://doi.org/10.1007/s00442-014-3099-z

Des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (2013). Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. In *Annual Review of Ecology, Evolution, and Systematics* (Vol. 44, Issue 1, pp. 5–29). https://doi.org/10.1146/annurev-ecolsys-110512-135806

Diamond, A. (1990). Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of, Inhibitory Control in Reaching. *Annals of the New York Academy of Sciences*, *608*(1), 637–676. https://doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond, A. (2013). Executive functions. In *Annual Review of Psychology* (Vol. 64, pp. 135–168). Annual Reviews Inc. https://doi.org/10.1146/annurev-psych-113011-143750

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *271*(1541), 847–852.

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., & Réale, D. (2005). Natural Selection and Animal Personality. *Behaviour*, *142*(9/10), 1159–1184. https://doi.org/10.1163/156853905774539445

Dingemanse, N. J., Van Der Plas, F., Wright, J., Réale, D., Schrama, M., Roff, D. A., Van Der Zee, E., & Barber, I. (2009). Individual experience and evolutionary history of predation affect expression of heritable variation in fish personality and morphology. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1285–1293. https://doi.org/10.1098/rspb.2008.1555

Dochtermann, N. A. (2011). Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution*, *65*(6), 1814–1820. https://doi.org/10.1111/j.1558-5646.2011.01264.x

Dochtermann, N. A., Schwab, T., Anderson Berdal, M., Dalos, J., & Royauté, R. (2019). The Heritability of Behavior: A Meta-analysis. *Journal of Heredity*, *110*(4), 403–410. https://doi.org/10.1093/jhered/esz023

Dochtermann, N. A., Schwab, T., & Sih, A. (2014). The contribution of additive genetic variation to personality variation: Heritability of personality. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798). https://doi.org/10.1098/rspb.2014.2201

Dohm, M. R. (2002). Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*, *16*(2), 273–280. https://doi.org/10.1046/j.1365-2435.2002.00621.x

Dougherty, L. R., & Guillette, L. M. (2018). Linking personality and cognition: A meta-analysis. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1756, p. 20170282). https://doi.org/10.1098/rstb.2017.0282

Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2003). Realized heritability of personalities in the great tit (Parus major). *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1510), 45–51.

Ducatez, S., Audet, J. N., & Lefebvre, L. (2019). Speed–accuracy trade-off, detour reaching and response to PHA in Carib grackles. *Animal Cognition*, *22*(5), 625–633. https://doi.org/10.1007/s10071-019-01258-1

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades. *Journal of Educational Psychology*, *104*(2), 439–451. https://doi.org/10.1037/a0026280

Dukas, R. (2004). Evolutionary biology of animal cognition. In *Annual Review of Ecology, Evolution, and Systematics* (Vol. 35, pp. 347–374). https://doi.org/10.1146/annurev.ecolsys.35.112202.130152

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology*,

6(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Edwards, H. A., Burke, T., & Dugdale, H. L. (2017). Repeatable and heritable behavioural variation in a wild cooperative breeder. *Behavioral Ecology*, *28*(3), 668–676. https://doi.org/10.1093/beheco/arx013

Endler, J. A. (1984). *Natural and sexual selection on color patterns in poeciliid fishes* (pp. 95–111). https://doi.org/10.1007/978-94-015-7682-6_7

Endler, J. A., & Houde, A. E. (1995). Geographic variation in female preferences for male traits in Poecilia reticulata. *Evolution*, *49*(3), 456–468. https://doi.org/10.1111/j.1558-5646.1995.tb02278.x

Evans, J. P., & Magurran, A. E. (2001). Patterns of sperm precedence and predictors of paternity in the Trinidadian guppy. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1468), 719–724. https://doi.org/10.1098/rspb.2000.1577

Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to quantitative genetics. In *Introduction to quantitative genetics*. https://doi.org/10.1016/j.bbabio.2008.04.029

Feil, J., Sheppard, D., Fitzgerald, P. B., Yücel, M., Lubman, D. I., & Bradshaw, J. L. (2010). Addiction, compulsive drug seeking, and the role of frontostriatal mechanisms in regulating inhibitory control. In *Neuroscience and Biobehavioral Reviews* (Vol. 35, Issue 2, pp. 248–275). https://doi.org/10.1016/j.neubiorev.2010.03.001

Fong, S., Buechel, S. D., Boussard, A., Kotrschal, A., & Kolm, N. (2019). Plastic changes in brain morphology in relation to learning and environmental enrichment in the guppy (Poecilia reticulata). *Journal of Experimental Biology*, *222*(10). https://doi.org/10.1242/jeb.200402

Fontaine, J. J., & Martin, T. E. (2006). Parent birds assess nest predation risk and adjust their reproductive strategies. *Ecology Letters*, *9*(4), 428–434. https://doi.org/10.1111/j.1461-0248.2006.00892.x

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS ONE*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gaikwad, S., Stewart, A., Hart, P., Wong, K., Piet, V., Cachat, J., & Kalueff, A. V. (2011). Acute stress disrupts performance of zebrafish in the cued and spatial memory tests: The utility of fish models to study stress-memory interplay. *Behavioural Processes*, *87*(2), 224–230.

https://doi.org/10.1016/j.beproc.2011.04.004

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., Schalkwyk, L. C., & Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior Genetics*, *35*(5), 675–692. https://doi.org/10.1007/s10519-005-3423-9

Galsworthy, M. J., Paya-Cano, J. L., Monleón, S., & Plomin, R. (2002). Evidence for general cognitive ability (g) in heterogeneous stock mice and an analysis of potential confounds. *Genes, Brain and Behavior*, *1*(2), 88–95. https://doi.org/10.1034/j.1601-183X.2002.10204.x

Gasparini, C., Speechley, E. M., & Polverino, G. (2019). The bold and the sperm: Positive association between boldness and sperm number in the guppy. *Royal Society Open Science*, *6*(7). https://doi.org/10.1098/rsos.190474

Gibelli, J., Aubin-Horth, N., & Dubois, F. (2019). Individual differences in anxiety are related to differences in learning performance and cognitive style. *Animal Behaviour*, *157*, 121–128. https://doi.org/10.1016/j.anbehav.2019.09.008

Gibson, K. R. (1986). Cognition, brain size and the extraction of embedded food resources. In *Primate Ontogeny, Cognition, and Social Behavior* (pp. 93–104). Cambridge University Press.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Thompson, R., & Butler, D. (2009).

ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

Gnanadesikan, G. E., Hare, B., Snyder-Mackler, N., & MacLean, E. L. (2020). Estimating the heritability of cognitive traits across dog breeds reveals highly heritable inhibitory control and communication factors. *Animal Cognition*, *23*(5), 953–964. https://doi.org/10.1007/s10071-020-01400-4

Godin, J.-G. J., & Davis, S. A. (1995). Who Dares, Benefits: Predator Approach Behaviour in the Guppy (Poecilia reticulata) Deters Predator Pursuit. *Proceedings of the Royal Society B: Biological Sciences*, *259*(1355), 193–200. https://doi.org/10.1098/rspb.1995.0028

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. https://doi.org/10.1037/0033-2909.127.1.45

Grafen A. (1984). *Natural selection, kin selection and group selection. In Behavioural Ecology: An Evolutionary Approach.* (Krebs J.R & Davies N.B (Eds.); 2nd ed.). Blackwell Scientific, Oxford. https://scholar.google.co.uk/scholar?hl=en&as_sdt=0%2C5&q=Natural+selection%2C+kin+selection+and+group+selection.+&btnG=

Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: An analysis of an emerging field. In *Trends in Ecology and Evolution* (Vol. 30, Issue 4, pp. 207–214). https://doi.org/10.1016/j.tree.2015.01.012

Griffiths, S. W., & Magurran, A. E. (1998). Sex and schooling behaviour in the

Trinidadian guppy. *Animal Behaviour*, *56*(3), 689–693. https://doi.org/10.1006/anbe.1998.0767

Guenther, A., & Brust, V. (2017). Individual consistency in multiple cognitive performance: behavioural versus cognitive syndromes. *Animal Behaviour*, *130*, 119–131. https://doi.org/10.1016/j.anbehav.2017.06.011

Guenther, A., Brust, V., Dersen, M., & Trillmich, F. (2014). Learning and personality types are related in cavies (cavia aperea). *Journal of Comparative Psychology*, *128*(1), 74–81. https://doi.org/10.1037/a0033678

Guez, D., & Griffin, A. S. (2016). Unraveling the key to innovative problem solving: A test of learning versus persistence. *Behavioral Ecology*, *27*(5), 1449–1460. https://doi.org/10.1093/beheco/arw055

Guido, J. M., Biondi, L. M., Vasallo, A. I., & Muzio, R. N. (2017). Neophobia is negatively related to reversal learning ability in females of a generalist bird of prey, the Chimango Caracara, Milvago chimango. *Animal Cognition*, *20*(4), 591–602. https://doi.org/10.1007/s10071-017-1083-9

Guillette, L. M., Hahn, A. H., Hoeschele, M., Przyslupski, A. M., & Sturdy, C. B. (2015). Individual differences in learning speed, performance accuracy and exploratory behaviour in black-capped chickadees. *Animal Cognition*, *18*(1), 165–178. https://doi.org/10.1007/s10071-014-0787-3

Guillette, L. M., Reddon, A. R., Hoeschele, M., & Sturdy, C. B. (2010). Sometimes slower is better: slow-exploring birds are more sensitive to changes in a vocal discrimination task. *Proceedings of the Royal Society of*

London B: Biological Sciences, rspb20101669.

Guillette, L. M., Reddon, A. R., Hurd, P. L., & Sturdy, C. B. (2009). Exploration of a novel space is associated with individual differences in learning speed in black-capped chickadees, Poecile atricapillus. *Behavioural Processes*, *82*(3), 265–270. https://doi.org/10.1016/j.beproc.2009.07.005

Hammond, R. L., Bourke, A. F. G., & Bruford, M. W. (2001). Mating frequency and mating system of the polygynous ant, Leptothorax acervorum. *Molecular Ecology*, *10*(11), 2719–2728. https://doi.org/10.1046/j.0962-1083.2001.01394.x

Hansen, T. F., & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, *21*(5), 1201–1219. https://doi.org/10.1111/j.1420-9101.2008.01573.x

Hayes, J. P., & Jenkins, S. H. (1997). Individual variation in mammals. In *Journal of Mammalogy* (Vol. 78, Issue 2, pp. 274–293). https://doi.org/10.2307/1382882

Hayward, A. D., Pemberton, J. M., Berenos, C., Wilson, A. J., Pilkington, J. G., & Kruuk, L. E. B. (2018). Evidence for selection-by-environment but not genotype-by-environment interactions for fitness-related traits in a wild mammal population. *Genetics*, *208*(1), 349–364. https://doi.org/10.1534/genetics.117.300498

Healy, S. D. (2019). The face of animal cognition. In *Integrative Zoology* (Vol. 14, Issue 2, pp. 132–144). John Wiley & Sons, Ltd.

https://doi.org/10.1111/1749-4877.12361

Healy, S. D., & Rowe, C. (2010). Information Processing: The Ecology and Evolution of Cognitive Abilities. In *Evolutionary Behavioral Ecology* (pp. 162–174). https://eprints.ncl.ac.uk/139823

Hedrick, A. V, Perez, D., Lichti, N., & Yew, J. (2002). Temperature preferences of male field crickets (Gryllus integer) alter their mating calls. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *188*(10), 799–805. https://doi.org/10.1007/s00359-002-0368-9

Highcock, L., & Carter, A. J. (2014). Intraindividual variability of boldness is repeatable across contexts in a wild lizard. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095179

Hill, W. G. (1984). On selection among groups with heterogeneous variance. *Animal Production*, *39*(03), 473–477. https://doi.org/10.1017/s0003356100032220

Hill, W. G., & Mulder, H. A. (2010). Genetic analysis of environmental variation. *Genetics Research*, *92*(5–6), 381–395. https://doi.org/10.1017/S0016672310000546

Hill, W. G., & Zhang, X. S. (2004). Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genetical Research*, *83*(2), 121–132. https://doi.org/10.1017/S0016672304006640

Hoffman, L. (2007). Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time. *Multivariate Behavioral Research*, *42*(4), 609–629. https://doi.org/10.1080/00273170701710072

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology*, *24*(14), 1649–1652. https://doi.org/10.1016/j.cub.2014.05.076

Horváth, G., Garamszegi, L. Z., Bereczki, J., Urszán, T. J., Balázs, G., & Herczeg, G. (2019). Roll with the fear: environment and state dependence of pill bug (Armadillidium vulgare) personalities. *Science of Nature*, *106*(3–4). https://doi.org/10.1007/s00114-019-1602-4

Houslay, T. M., Earley, R. L., White, S. J., Lammers, W., Grimmer, A. J., Travers, L. M., Johnson, E. L., Young, A. J., & Wilson, A. J. (2019). Genetic integration of the stress response. *BioRxiv*, 770586. https://doi.org/10.1101/770586

Houslay, T. M., Vierbuchen, M., Grimmer, A. J., Young, A. J., & Wilson, A. J. (2018). Testing the stability of behavioural coping style across stress contexts in the Trinidadian guppy. *Functional Ecology*, *32*(2), 424–438. https://doi.org/10.1111/1365-2435.12981

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse of BLUP in behavioural ecology. *Behavioral Ecology*, *28*(4), 948–952. https://doi.org/10.1093/beheco/arx023

Hunt, E., Frost, N., & Lunneborg, C. (1973). Individual Differences in Cognition:

A New Approach to Intelligence. *Psychology of Learning and Motivation - Advances in Research and Theory*, *7*(C), 87–122. https://doi.org/10.1016/S0079-7421(08)60066-3

Huntingford, F. A., Adams, C., Braithwaite, V. A., Kadri, S., Pottinger, T. G., Sandøe, P., & Turnbull, J. F. (2006). Current issues in fish welfare. In *Journal of Fish Biology* (Vol. 68, Issue 2, pp. 332–372). https://doi.org/10.1111/j.0022-1112.2006.001046.x

Ingleby, F. C., Hunt, J., & Hosken, D. J. (2010). The role of genotype-by-environment interactions in sexual selection. In *Journal of Evolutionary Biology* (Vol. 23, Issue 10, pp. 2031–2045). https://doi.org/10.1111/j.1420-9101.2010.02080.x

Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, *12*(2). https://doi.org/10.1098/rsbl.2015.0871

Jones, A. G., Arnold, S. J., & Bürger, R. (2004). Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution*, *58*(8), 1639–1654. https://doi.org/10.1111/j.0014-3820.2004.tb00450.x

Jones, N., Webster, M., Newport, C., Templeton, C. N., Schuster, S., & Rendell, L. (2020). Cognitive styles: speed–accuracy trade-offs underlie individual differences in archerfish. *Animal Behaviour*, *160*, 1–14. https://doi.org/10.1016/j.anbehav.2019.11.019

Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*(3), 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Jozet-Alves, C., Bertin, M., & Clayton, N. S. (2013). Evidence of episodic-like memory in cuttlefish. In *Current Biology* (Vol. 23, Issue 23, pp. R1033–R1035). Cell Press. https://doi.org/10.1016/j.cub.2013.10.021

Juszczak, G. R., & Miller, M. (2016). Detour behavior of mice trained with transparent, semitransparent and opaque barriers. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162018

Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. In *Animal Cognition* (Vol. 21, Issue 1, pp. 21–35). Springer Verlag. https://doi.org/10.1007/s10071-017-1152-0

Kabadayi, C., Krasheninnikova, A., O'Neill, L., van de Weijer, J., Osvath, M., & von Bayern, A. M. P. (2017). Are parrots poor at motor self-regulation or is the cylinder task poor at measuring it? *Animal Cognition*, *20*(6), 1137–1146. https://doi.org/10.1007/s10071-017-1131-5

Kabadayi, C., Taylor, L. A., von Bayern, A. M. P., & Osvath, M. (2016). Ravens, New Caledonian crows and Jackdaws parallel great apes in motor self-regulation despite smaller brains. *Royal Society Open Science*, *3*(4). https://doi.org/10.1098/rsos.160104

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the

computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*(5), 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kass, R. E., Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1997). Markov Chain Monte Carlo in Practice. *Journal of the American Statistical Association*, *92*(440), 1645. https://doi.org/10.2307/2965438

Kasumovic, M. M., Hall, M. D., & Brooks, R. C. (2012). The juvenile social environment introduces variation in the choice and expression of sexually selected traits. *Ecology and Evolution*, *2*(5), 1036–1047. https://doi.org/10.1002/ece3.230

Kehagia, A. A., Murray, G. K., & Robbins, T. W. (2010). Learning and cognitive flexibility: Frontostriatal function and monoaminergic modulation. In *Current Opinion in Neurobiology* (Vol. 20, Issue 2, pp. 199–204). https://doi.org/10.1016/j.conb.2010.01.007

Kelley, J. L., Graves, J. A., & Magurran, A. E. (1999). Familiarity breeds contempt in guppies. *Nature*, *401*(6754), 661–662. https://doi.org/10.1038/44314

Kenward, B., Rutz, C., Weir, A. A. S., & Kacelnik, A. (2006). Development of tool use in New Caledonian crows: inherited action patterns and social influences. *Animal Behaviour*, *72*(6), 1329–1343. https://doi.org/10.1016/j.anbehav.2006.04.007

Kniel, N., Guenther, A., & Godin, J. G. J. (2020). Individual personality does not

predict learning performance in a foraging context in female guppies, Poecilia reticulata. *Animal Behaviour*, *167*, 3–12. https://doi.org/10.1016/j.anbehav.2020.07.007

Koolhaas, J. M., Korte, S. M., De Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W., & Blokhuis, H. J. (1999). Coping styles in animals: Current status in behavior and stress- physiology. *Neuroscience and Biobehavioral Reviews*, *23*(7), 925–935. https://doi.org/10.1016/S0149-7634(99)00026-3

Kortet, R., Vainikka, A., Janhunen, M., Piironen, J., & Hyvärinen, P. (2014). Behavioral variation shows heritability in juvenile brown trout Salmo trutta. *Behavioral Ecology and Sociobiology*, *68*(6), 927–934. https://doi.org/10.1007/s00265-014-1705-z

Kotrschal, A., Corral-Lopez, A., Amcoff, M., & Kolm, N. (2015). A larger brain confers a benefit in a spatial mate search learning task in male guppies. *Behavioral Ecology*, *26*(2), 527–532. https://doi.org/10.1093/beheco/aru227

Kotrschal, A., Deacon, A. E., Magurran, A. E., & Kolm, N. (2017). Predation pressure shapes brain anatomy in the wild. *Evolutionary Ecology*, *31*(5), 619–633. https://doi.org/10.1007/s10682-017-9901-8

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., Immler, S., Maklakov, A. A., & Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology*, *23*(2), 168–171.

Lai, Z. C., Moss, M. B., Killiany, R. J., Rosene, D. L., & Herndon, J. G. (1995). Executive system dysfunction in the aged monkey: Spatial and object reversal learning. *Neurobiology of Aging*, *16*(6), 947–954. https://doi.org/10.1016/0197-4580(95)02014-4

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., De Jager, V., Megens, H. J., Warren, W. C., Minx, P., Crooijmans, R. P. M. A., Corcoran, P., Adriaensen, F., Belda, E., Bushuev, A., Cichon, M., Charmantier, A., Dingemanse, N., Doligez, B., … Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*(1), 1–9. https://doi.org/10.1038/ncomms10474

Laland, K. N., & Reader, S. M. (1999). Foraging innovation in the guppy. *Animal Behaviour*, *57*(2), 331–340. https://doi.org/10.1006/anbe.1998.0967

Lande, R. (1979). Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. *Evolution*, *33*(1), 402. https://doi.org/10.2307/2407630

Langley, E. J. G., Adams, G., Beardsworth, C. E., Dawson, D. A., Laker, P. R., van Horik, J. O., Whiteside, M. A., Wilson, A. J., & Madden, J. R. (2020). Heritability and correlations among learning and inhibitory control traits. *Behavioral Ecology*. https://doi.org/10.1093/beheco/araa029

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models - Discussion. *Journal of the Royal Statistical Society. Series C: Applied*

Statistics, *55*(2), 139–185. https://doi.org/10.1111/j.1467-9876.2006.00538.x

Li, L., MaBouDi, H. Di, Egertová, M., Elphick, M. R., Chittka, L., & Perry, C. J. (2017). A possible structural correlate of learning performance on a colour discrimination task in the brain of the bumblebee. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1864). https://doi.org/10.1098/rspb.2017.1323

Lucon-Xiccato, T., & Bertolucci, C. (2019). Guppies show rapid and lasting inhibition of foraging behaviour. *Behavioural Processes*, *164*, 91–99. https://doi.org/10.1016/j.beproc.2019.04.011

Lucon-Xiccato, T., & Bisazza, A. (2014). Discrimination reversal learning reveals greater female behavioural flexibility in guppies. *Biology Letters*, *10*(6), 20140206.

Lucon-Xiccato, T., & Bisazza, A. (2016). Male and female guppies differ in speed but not in accuracy in visual discrimination learning. *Animal Cognition*, *19*(4), 733–744. https://doi.org/10.1007/s10071-016-0969-2

Lucon-Xiccato, T., & Bisazza, A. (2017a). Individual differences in cognition among teleost fishes. In *Behavioural Processes* (Vol. 141, pp. 184–195). https://doi.org/10.1016/j.beproc.2017.01.015

Lucon-Xiccato, T., & Bisazza, A. (2017b). Sex differences in spatial abilities and cognitive flexibility in the guppy. *Animal Behaviour*, *123*, 53–60. https://doi.org/10.1016/j.anbehav.2016.10.026

Lucon-Xiccato, T., & Bisazza, A. (2017c). Complex maze learning by fish. *Animal Behaviour*, *125*, 69–75. https://doi.org/10.1016/j.anbehav.2016.12.022

Lucon-Xiccato, T., Bisazza, A., & Bertolucci, C. (2020). Guppies show sex and individual differences in the ability to inhibit behaviour. *Animal Cognition*, *23*(3), 535–543. https://doi.org/10.1007/s10071-020-01357-4

Lucon-Xiccato, T., Dadda, M., & Bisazza, A. (2016). Sex Differences in Discrimination of Shoal Size in the Guppy (Poecilia reticulata). *Ethology*, *122*(6), 481–491. https://doi.org/10.1111/eth.12498

Lucon-Xiccato, T., Gatto, E., & Bisazza, A. (2017). Fish perform like mammals and birds in inhibitory motor control tasks. *Scientific Reports*, *7*, 13144. https://doi.org/10.1101/188359

Lucon-Xiccato, T., Gatto, E., & Bisazza, A. (2020). Male and female guppies differ in problem-solving abilities. *Current Zoology*, *66*(1), 83–90. https://doi.org/10.1093/cz/zoz017

Lucon-Xiccato, T., Montalbano, G., & Bertolucci, C. (2020). Personality traits covary with individual differences in inhibitory abilities in 2 species of fish. *Current Zoology*, *66*(2), 187–195. https://doi.org/10.1093/cz/zoz039

Lukowiak, K., Orr, M., de Caigny, P., Lukowiak, K. S., Rosenegger, D., Han, J. Il, & Dalesman, S. (2014). Ecologically relevant stressors modify long-term memory formation in a model system. *Behavioural Brain Research*, *214*(1), 18–24. https://doi.org/10.1016/j.bbr.2010.05.011

Lynch, M., & Walsh, B. (1996). Genetics and Analysis of Quantitative Traits. In *Journal of Animal Breeding and Genetics*. Sinauer Associates, Sunderland, MA. https://doi.org/10.1046/j.1439-0388.2002.00356.x

Macario, A., Darden, S. K., Verbruggen, F., & Croft, D. P. (2021). Intraspecific variation in inhibitory motor control in guppies, Poecilia reticulata. *Journal of Fish Biology*, *98*(1), 317–328. https://doi.org/10.1111/jfb.14608

MacDonald, S. W. S., Backman, L., & Nyberg, L. (2006). Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, *29*(8), 474–480. https://www.sciencedirect.com/science/article/pii/S0166223606001251

Mackay, T. F. C. (1981). Genetic variation in varying environments. *Genetical Research*, *37*(1), 79–93. https://doi.org/10.1017/S0016672300020036

MacLean, E. L., Hare, B., Nun, C. L., Addess, E., Amic, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlo, J. F., Chek, L. G., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20). https://doi.org/10.1073/pnas.1323533111

Magurran, A. E. (2005). *Evolutionary ecology: the Trinidadian guppy*. Oxford University Press.

Magurran, A. E., & Garcia, C. M. (2000). Sex differences in behaviour as an indirect consequence of mating system. *Journal of Fish Biology*, *57*(4),

839–857. https://doi.org/10.1111/j.1095-8649.2000.tb02196.x

Magurran, A. E., & Seghers, B. H. (1990). Risk Sensitive Courtship in the Guppy (Poecilia Reticulata). *Behaviour*, *112*(3–4), 194–201. https://doi.org/10.1163/156853990X00194

MaGurran, A. E., & Seghers, B. H. (1994). Sexual conflict as a consequence of ecology: Evidence from guppy, Poecilia reticulata, populations in Trinidad. *Proceedings of the Royal Society B: Biological Sciences*, *255*(1342), 31–36. https://doi.org/10.1098/rspb.1994.0005

Maille, A., & Schradin, C. (2016). Survival is linked with reaction time and spatial memory in African striped mice. *Biology Letters*, *12*(8). https://doi.org/10.1098/rsbl.2016.0346

Mamuneas, D., Spence, A. J., Manica, A., & King, A. J. (2015). Bolder stickleback fish make faster decisions, but they are not less accurate. *Behavioral Ecology*, *26*(1), 91–96. https://doi.org/10.1093/beheco/aru160

Martin-Ordas, G., Haun, D., Colmenares, F., & Call, J. (2010). Keeping track of time: Evidence for episodic-like memory in great apes. *Animal Cognition*, *13*(2), 331–340. https://doi.org/10.1007/s10071-009-0282-4

Martin, J. G. A., Pirotta, E., Petelle, M. B., & Blumstein, D. T. (2017). Genetic basis of between-individual and within-individual variance of docility. *Journal of Evolutionary Biology*, *30*(4), 796–805. https://doi.org/10.1111/jeb.13048

Martin, J. G. A., & Réale, D. (2008). Temperament, risk assessment and

habituation to novelty in eastern chipmunks, Tamias striatus. *Animal Behaviour*, *75*(1), 309–318.

Mathot, K. J., & Dall, S. R. X. (2013). Metabolic rates can drive individual differences in information and insurance use under the risk of starvation. *American Naturalist*, *182*(5), 611–620. https://doi.org/10.1086/673300

Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., Specht, S. M., & Gandhi, C. C. (2003). Individual differences in the expression of a “general” learning ability in mice. *Journal of Neuroscience*, *23*(16), 6423–6433. https://doi.org/10.1523/jneurosci.23-16-06423.2003

Mazza, V., Eccard, J. A., Zaccaroni, M., Jacob, J., & Dammhahn, M. (2018). The fast and the flexible: cognitive style drives individual variation in cognition in a small mammal. *Animal Behaviour*, *137*, 119–132. https://doi.org/10.1016/j.anbehav.2018.01.011

Mazza, V., Jacob, J., Dammhahn, M., Zaccaroni, M., & Eccard, J. A. (2019). Individual variation in cognitive style reflects foraging and anti-predator strategies in a small mammal. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-46582-1

McGonigle, P., & Ruggeri, B. (2014). Animal models of human disease: Challenges in enabling translation. In *Biochemical Pharmacology* (Vol. 87, Issue 1, pp. 162–171). https://doi.org/10.1016/j.bcp.2013.08.006

McKenzie, D. J., Belaõ, T. C., Killen, S. S., & Rantin, F. T. (2016). To boldly gulp: Standard metabolic rate and boldness have context-dependent

influences on risk-taking to breathe air in a catfish (Journal of Experimental Biology (2016) 218 (3762-3770)). In *Journal of Experimental Biology* (Vol. 219, Issue 5, p. 767). https://doi.org/10.1242/jeb.138842

Mesquita, F. O., Borcato, F. L., & Huntingford, F. A. (2015). Cue-based and algorithmic learning in common carp: A possible link to stress coping style. *Behavioural Processes*, *115*, 25–29. https://doi.org/10.1016/j.beproc.2015.02.017

Meyer, P. J., Lovic, V., Saunders, B. T., Yager, L. M., Flagel, S. B., Morrow, J. D., & Robinson, T. E. (2012). Quantifying individual variation in the propensity to attribute incentive salience to reward cues. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0038987

Miletto Petrazzini, M. E., & Agrillo, C. (2016). Turning to the larger shoal: are there individual differences in small- and large-quantity discrimination of guppies? *Ethology Ecology and Evolution*, *28*(2), 211–220. https://doi.org/10.1080/03949370.2015.1029011

Miller, S. E., Legan, A. W., Henshaw, M. T., Ostevik, K. L., Samuk, K., Uy, F. M. K., & Sheehan, M. J. (2020). Evolutionary dynamics of recent selection on cognitive abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(6), 3045–3052. https://doi.org/10.1073/pnas.1918592117

Mitchell, D. J., Fanson, B. G., Beckmann, C., & Biro, P. A. (2016). Towards powerful experimental and statistical approaches to study intraindividual

variability in labile traits. *Royal Society Open Science*, *3*(10), 160352. https://doi.org/10.1098/rsos.160352

Moiron, M., Mathot, K. J., & Dingemanse, N. J. (2016). A multi-level approach to quantify speed-accuracy trade-offs in great tits (Parus major). *Behavioral Ecology*, *27*(5), 1539–1546. https://doi.org/10.1093/beheco/arw077

Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biological Reviews*, *91*(2), 367–389. https://doi.org/10.1111/brv.12174

Mulder, H. A., Bijma, P., & Hill, W. G. (2007). Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics*, *175*(4), 1895–1910. https://doi.org/10.1534/genetics.106.063743

Mulder, H. A., Gienapp, P., & Visser, M. E. (2016). Genetic variation in variability: Phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution*, *70*(9), 2004–2016. https://doi.org/10.1111/evo.13008

Müller, R., & Schrader, L. (2005). Behavioural consistency during social separation and personality in dairy cows. *Behaviour*, *142*(9–10), 1289–1306.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. In *Biological Reviews* (Vol. 85, Issue 4, pp. 935–956). https://doi.org/10.1111/j.1469-

185X.2010.00141.x

Nawroth, C., Prentice, P. M., & McElligott, A. G. (2017). Individual personality differences in goats predict their performance in visual learning and non-associative cognitive tasks. In *Behavioural Processes* (Vol. 134, pp. 43–53). https://doi.org/10.1016/j.beproc.2016.08.001

Nicoletto, P. F. (1993). Female sexual response to condition-dependent ornaments in the guppy, Poecilia reticulata. *Animal Behaviour*, *46*(3), 441–450. https://doi.org/10.1006/anbe.1993.1213

Niemelä, P. T., Vainikka, A., Forsman, J. T., Loukola, O. J., & Kortet, R. (2013). How does variation in the environment and individual cognition explain the existence of consistent behavioral differences? *Ecology and Evolution*, *3*(2), 457–464. https://doi.org/10.1002/ece3.451

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New Findings and Theoretical Developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Noland, J. S. (2008). Executive functioning demands of the object retrieval task for 8-month-old infants. *Child Neuropsychology*, *14*(6), 504–509. https://doi.org/10.1080/09297040701770819

Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, *20*(3), 831–844. https://doi.org/10.1111/j.1420-

9101.2007.01300.x

O'Neill, S. J., Williamson, J. E., Tosetto, L., & Brown, C. (2018). Effects of acclimatisation on behavioural repeatability in two behaviour assays of the guppy Poecilia reticulata. *Behavioral Ecology and Sociobiology*, *72*(10). https://doi.org/10.1007/s00265-018-2582-7

Øverli, Ø., Sørensen, C., Pulman, K. G. T., Pottinger, T. G., Korzan, W., Summers, C. H., & Nilsson, G. E. (2007). Evolutionary background for stress-coping styles: Relationships between physiological, behavioral, and cognitive traits in non-mammalian vertebrates. In *Neuroscience and Biobehavioral Reviews* (Vol. 31, Issue 3, pp. 396–412). Pergamon. https://doi.org/10.1016/j.neubiorev.2006.10.006

Pigliucci, M. (2001). *Phenotypic plasticity: beyond nature and nurture*. John Hopkins University Press, Baltimore. https://books.google.com/books?hl=en&lr=&id=JoCFupYhdNQC&oi=fnd&pg=PR9&ots=K6-7MHmfxN&sig=kaLvC61JhDDfGQYcVxeCFlzkjH8

Pigliucci, M., Whitton, J., & Schlichting, C. D. (1995). Reaction norms of Arabidopsis. I. Plasticity of characters and correlations across water, nutrient and light gradients. *Journal of Evolutionary Biology*, *8*(4), 421–438. https://doi.org/10.1046/j.1420-9101.1995.8040421.x

Pike, T. W., Ramsey, M., & Wilkinson, A. (2018). Environmentally induced changes to brain morphology predict cognitive performance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756).

https://doi.org/10.1098/rstb.2017.0287

Plomin, R. (2001). The genetics of g in human and mouse. *Nature Reviews Neuroscience*. http://www.nature.com/nrn/journal/v2/n2/abs/nrn0201_136a.html

Plomin, R., & Spinath, F, M. (2002). Genetics and general cognitive ability (g). In *Trends in Cognitive Sciences* (Vol. 6, Issue 4, pp. 169–176). https://doi.org/10.1016/S1364-6613(00)01853-2

Poirier, M.-A., Kozlovsky, D. Y., Morand-Ferron, J., & Careau, V. (2020). How general is cognitive ability in non-human animals? A meta-analytical and multi-level reanalysis approach. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1940), 20201853. https://doi.org/10.1098/rspb.2020.1853

Prentice, P. M., Houslay, T. M., Martin, J. G. A., & Wilson, A. J. (2020). Genetic variance for behavioural 'predictability' of stress response. *Journal of Evolutionary Biology*, *33*(5), 642–652. https://doi.org/10.1111/jeb.13601

Prentice, P. M., Mnatzaganian, C., Houslay, T. M., Thornton, A., & Wilson, A. J. (2020). Cognitive performance and stress responsiveness in the Trinidadian Guppy – a multivariate approach. *Biorxiv.Org*. https://doi.org/10.1101/2020.05.19.103689

Quinn, J. L., Cole, E. F., Reed, T. E., & Morand-Ferron, J. (2016). Environmental and genetic determinants of innovativeness in a natural population of birds. *Philosophical Transactions of the Royal Society B:*

Biological Sciences, *371*(1690). https://doi.org/10.1098/rstb.2015.0184

R Development Core Team. (2008). *Control*, *55*, 275–286. https://stat.ethz.ch/pipermail/r-help/2008-May/161481.html

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, *445*(7130), 919–921. https://doi.org/10.1038/nature05575

Raine, N. E., & Chittka, L. (2008). The correlation of learning speed and natural foraging success in bumble-bees. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1636), 803–808. https://doi.org/10.1098/rspb.2007.1652

Raine, N. E., & Chittka, L. (2012). No Trade-Off between Learning Speed and Associative Flexibility in Bumblebees: A Reversal Learning Test with Multiple Colonies. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045096

Ramakers, J. J. C., Culina, A., Visser, M. E., & Gienapp, P. (2018). Environmental coupling of heritability and selection is rare and of minor evolutionary significance in wild populations. *Nature Ecology and Evolution*, *2*(7), 1093–1103. https://doi.org/10.1038/s41559-018-0577-4

Raoult, V., Trompf, L., Williamson, J. E., & Brown, C. (2017). Stress profile influences learning approach in a marine fish. *PeerJ*, *2017*(6). https://doi.org/10.7717/peerj.3445

Rapkin, J., Archer, C. R., Grant, C. E., Jensen, K., House, C. M., Wilson, A. J.,

& Hunt, J. (2017). Little evidence for intralocus sexual conflict over the optimal intake of nutrients for life span and reproduction in the black field cricket Teleogryllus commodus. *Evolution*, *71*(9), 2159–2177. https://doi.org/10.1111/evo.13299

Réale, D., & Festa-Bianchet, M. (2003). Predator-induced natural selection on temperament in bighorn ewes. *Animal Behaviour*, *65*(3), 463–470. https://doi.org/10.1006/anbe.2003.2100

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P. O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318.

Richards, E. L., van Oosterhout, C., & Cable, J. (2010). Sex-specific differences in shoaling affect parasite transmission in guppies. *PLoS ONE*, *5*(10). https://doi.org/10.1371/journal.pone.0013285

Richardson, G., Dickinson, P., Burman, O. H. P., & Pike, T. W. (2018). Unpredictable movement as an anti-predator strategy. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1885), 20181112. https://doi.org/10.1098/rspb.2018.1112

Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution*, *50*(4), 1392–1403. https://doi.org/10.1111/j.1558-5646.1996.tb03913.x

Roff, D. A. (2002). *Life history evolution* (p. 527). Sinauer Associates.

Roff, D. A. (2012). *Evolutionary quantitative genetics*. Chapman & Hall. https://doi.org/10.1007/978-1-4615-4080-9

Roff, D. A., & Wilson, A. J. (2014). Quantifying Genotype-by-Environment Interactions in Laboratory Systems. In *Genotype-by-Environment Interactions and Sexual Selection* (pp. 101–136). Wiley Blackwell. https://doi.org/10.1002/9781118912591.ch5

Roitblat, H. L. (2008). Animal Cognition. In *A Companion to Cognitive Science* (pp. 114–120). https://doi.org/10.1002/9781405164535.ch2

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - Estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, *42*(1), 8. https://doi.org/10.1186/1297-9686-42-8

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2013). Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *Journal of Dairy Science*, *96*(4), 2627–2636. https://doi.org/10.3168/jds.2012-6198

Roth, G. (2015). Convergent evolution of complex brains and high intelligence. In *Philosophical Transactions of the Royal Society B: Biological Sciences*

(Vol. 370, Issue 1684). Royal Society of London. https://doi.org/10.1098/rstb.2015.0049

Roth, T. C., Brodin, A., Smulders, T. V., Ladage, L. D., & Pravosudov, V. V. (2010). Is bigger always better? A critical appraisal of the use of volumetric analysis in the study of the hippocampus. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 365, Issue 1542, pp. 915–931). https://doi.org/10.1098/rstb.2009.0208

Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. In *Behavioral Ecology* (Vol. 25, Issue 6, pp. 1287–1292). https://doi.org/10.1093/beheco/aru090

Rudin, F. S., Simmons, L. W., & Tomkins, J. L. (2019). Social cues affect quantitative genetic variation and covariation in animal personality traits. *Evolution*, *73*(3), 540–553. https://doi.org/10.1111/evo.13661

Russell, J., Cheke, L. G., Clayton, N. S., & Meltzoff, A. N. (2011). What can What-When-Where (WWW) binding tasks tell us about young children's episodic foresight? Theory and two experiments. *Cognitive Development*, *26*(4), 356–370. https://doi.org/10.1016/j.cogdev.2011.09.002

Salak-Johnson, J. L., & McGlone, J. J. (2007). Making sense of apparently conflicting data: stress and immunity in swine and cattle. In *Journal of animal science* (Vol. 85, Issue 13 Suppl, pp. 81–88). https://doi.org/10.2527/jas.2006-538

SanCristobal-Gaudy, M., Elsen, J.-M., Bodin, L., & Chevalet, C. (2009).

Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genetics Selection Evolution*, *30*(5), 423. https://doi.org/10.1186/1297-9686-30-5-423

Santacà, M., Busatta, M., Lucon-Xiccato, T., & Bisazza, A. (2019). Sensory differences mediate species variation in detour task performance. *Animal Behaviour*, *155*, 153–162. https://doi.org/10.1016/j.anbehav.2019.05.022

Santacà, M., Busatta, M., Savaşçı, B. B., Lucon-Xiccato, T., & Bisazza, A. (2019). The effect of experience and olfactory cue in an inhibitory control task in guppies, Poecilia reticulata. *Animal Behaviour*, *151*, 1–7. https://doi.org/10.1016/j.anbehav.2019.03.003

Santos, L. R., Ericson, B. N., & Hauser, M. D. (1999). Constraints on problem solving and inhibition: Object retrieval in cotton-top tamarins (Saguinus oedipus oedipus). *Journal of Comparative Psychology*, *113*(2). https://doi.org/10.1037//0735-7036.113.2.186

Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13971

Schachar, R. J., Forget-Dubois, N., Dionne, G., Boivin, M., & Robaey, P. (2011). Heritability of response inhibition in children. *Journal of the International Neuropsychological Society*, *17*(2), 238–247. https://doi.org/10.1017/S1355617710001463

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H.,

Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, *130*(4), 601–630. https://doi.org/10.1037/0033-2909.130.4.601

Shackleton, M. A., Jennions, M. D., & Hunt, J. (2005). Fighting success and attractiveness as predictors of male mating success in the black field cricket, Teleogryllus commodus: The effectiveness of no-choice tests. *Behavioral Ecology and Sociobiology*, *58*(1), 1–8. https://doi.org/10.1007/s00265-004-0907-1

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for “general” cognitive performance in wild New Zealand robins, Petroica longipes. *Animal Behaviour*, *109*, 101–111. https://doi.org/10.1016/j.anbehav.2015.08.001

Shaw, R. C., MacKinlay, R. D., Clayton, N. S., & Burns, K. C. (2019). Memory Performance Influences Male Reproductive Success in a Wild Bird. *Current Biology*, *29*(9), 1498-1502.e3. https://doi.org/10.1016/j.cub.2019.03.027

Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. In *Animal Cognition* (Vol. 20, Issue 6, pp. 1003–1018).

https://doi.org/10.1007/s10071-017-1135-1

Shettleworth, S. J. (1972). Constraints on Learning. *Advances in the Study of Behavior*, *4*(C), 1–68. https://doi.org/10.1016/S0065-3454(08)60006-6

Shettleworth, S. J. (2009a). Cognition, Evolution, and Behavior. *Oxford University Press*.

Shettleworth, S. J. (2009b). The evolution of comparative cognition: Is the snark still a boojum? *Behavioural Processes*, *80*(3), 210–217. https://doi.org/10.1016/j.beproc.2008.09.001

Shettleworth, S. J. (2010a). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481. https://doi.org/10.1016/j.tics.2010.07.002

Shettleworth, S. J. (2010b). Getting Around: Spatial Cognition. In *Cognition, Evolution, and Behavior* (second, pp. 261–312). Oxford University Press.

Shohet, A. J., & Watt, P. J. (2009). Female guppies Poecilia reticulata prefer males that can learn fast. *Journal of Fish Biology*, *75*(6), 1323–1330. https://doi.org/10.1111/j.1095-8649.2009.02366.x

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378.

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral Syndromes: An Integrative Overview. *The Quarterly Review of Biology*, *79*(3), 241–277. https://doi.org/10.1086/422893

Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *367*(1603), 2762–2772. https://doi.org/10.1098/rstb.2012.0216

Skelhorn, J., & Rowe, C. (2015). Cognition and the evolution of camouflage. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 283, Issue 1825). Royal Society of London. https://doi.org/10.1098/rspb.2015.2890

Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: a meta-analysis. *Behavioral Ecology*, *19*(2), 448–455.

Smith, C., Philips, A., & Reichard, M. (2015). Cognitive ability is heritable and predicts the success of an alternative mating tactic. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1809). https://doi.org/10.1098/rspb.2015.1046

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35. https://doi.org/10.1017/S0016672300014634

Sneddon, L. U. (2003). The bold and the shy: Individual differences in rainbow trout. *Journal of Fish Biology*, *62*(4), 971–975. https://doi.org/10.1046/j.1095-8649.2003.00084.x

Soltis, J., Thomsen, R., & Takenaka, O. (2001). The interaction of male and female reproductive strategies and paternity in wild Japanese macaques, Macaca fuscata. *Animal Behaviour*, *62*(3), 485–494.

https://doi.org/10.1006/anbe.2001.1774

Sonesson, A. K., Ødegård, J., & Rönnegård, L. (2013). Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (Salmo salar). *Genetics Selection Evolution*, *45*(1). https://doi.org/10.1186/1297-9686-45-41

Sonnenberg, B. R., Branch, C. L., Pitera, A. M., Bridge, E., & Pravosudov, V. V. (2019). Natural Selection and Spatial Cognition in Wild Food-Caching Mountain Chickadees. *Current Biology*, *29*(4), 670-676.e3. https://doi.org/10.1016/j.cub.2019.01.006

Sorato, E., Zidar, J., Garnham, L., Wilson, A. J., & Løvlie, H. (2018). Heritabilities and co-variation among cognitive traits in red junglefowl. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756). https://doi.org/10.1098/rstb.2017.0285

Sorensen, D., & Waagepetersen, R. (2003). Normal linear models with genetically structured residual variance heterogeneity: A case study. *Genetical Research*, *82*(3), 207–222. https://doi.org/10.1017/S0016672303006426

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: Individual differences in intraindividual variability (IIV). *Animal Behaviour*, *83*(6), 1325–1334. https://doi.org/10.1016/j.anbehav.2012.02.017

Stan Development Team. (2018). *RStan: the R interface to Stan. R package version 2.17.3.* Http://Mc-Stan.Org/. https://doi.org/10.3168/jds.S0022-

0302(63)89186-9

Stearns, S. C. (1992). *The Evolution of Life Histories*. Oxford University Press.

Sternberg, R. J. (2002). Handbook of Intelligence. *American Journal of Psychiatry*, *159*(2), 332-a-333. https://doi.org/10.1176/appi.ajp.159.2.332-a

Sterne, J. A. C., & Smith, G. D. (2001). Sifting the evidence—what's wrong with significance tests? *Physical Therapy*, *81*(8), 1464–1469. https://doi.org/10.1093/ptj/81.8.1464

Stram, D. O., & Lee, J. W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, *50*(4), 1171. https://doi.org/10.2307/2533455

Tatliyer, A., Cervantes, I., Formoso-Rafferty, N., & Gutiérrez, J. P. (2019). The Statistical Scale Effect as a Source of Positive Genetic Correlation Between Mean and Variability: A Simulation Study. *G3: Genes|Genomes|Genetics*, *9*(9), 3001–3008. https://doi.org/10.1534/g3.119.400497

Thornton, A., & Boogert, N. J. (2019). The Nature and Nurturing of Animal Minds. In *Genes and Behaviour* (pp. 181–201). https://doi.org/10.1002/9781119313663.ch9

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology*, *25*(6), 1299–1301. https://doi.org/10.1093/beheco/aru095

Thornton, A., & Lukas, D. (2012a). Individual variation in cognitive performance:

Developmental and evolutionary perspectives. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 367, Issue 1603, pp. 2773–2783). The Royal Society. https://doi.org/10.1098/rstb.2012.0214

Thornton, A., & Lukas, D. (2012b). Individual variation in cognitive performance: Developmental and evolutionary perspectives. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 367, Issue 1603, pp. 2773–2783). Royal Society. https://doi.org/10.1098/rstb.2012.0214

Thornton, A., & Wilson, A. J. (2015). In search of the Darwinian Holy Trinity in cognitive evolution: A comment on Croston et al. In *Behavioral Ecology* (Vol. 26, Issue 6, pp. 1460–1461). https://doi.org/10.1093/beheco/arv119

Tolman, E. C. (1924). The Inheritance of Maze-Learning Ability in Rats. *Journal of Comparative Psychology*, *4*(1), 1–18. https://doi.org/10.1037/h0071979

Trompf, L., & Brown, C. (2014). Personality affects learning and trade-offs between private and social information in guppies, Poecilia reticulata. *Animal Behaviour*, *88*, 99–106.

Tryon, R. C. (1940). Genetic differences in mazelearning ability in rats. *Yearbook of the National Society for the Study of Education*, *Vol 39(I)*(1), 111–119. https://psycnet.apa.org/record/1940-02917-001

van Horik, J. O., Beardsworth, C. E., Laker, P. R., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2019). Unpredictable environments enhance

inhibitory control in pheasants. *Animal Cognition*, *22*(6), 1105–1114. https://doi.org/10.1007/s10071-019-01302-0

van Horik, J. O., & Emery, N. J. (2018). Serial reversal learning and cognitive flexibility in two species of Neotropical parrots (Diopsittaca nobilis and Pionites melanocephala). *Behavioural Processes*, *157*, 664–672. https://doi.org/10.1016/j.beproc.2018.04.002

van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875). https://doi.org/10.1098/rspb.2018.0150

van Horik, J. O., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2017). Differential participation in cognitive tests is driven by personality, sex, body condition and experience. In *Behavioural Processes* (Vol. 134, pp. 22–30). https://doi.org/10.1016/j.beproc.2016.07.001

van Horik, J. O., & Madden, J. R. (2016). A problem with problem solving: Motivational traits, but not cognition, predict success on novel operant foraging tasks. *Animal Behaviour*, *114*, 189–198. https://doi.org/10.1016/j.anbehav.2016.02.006

Van Oers, K., De Jong, G., Van Noordwijk, A. J., Kempenaers, B., & Drent, P. J. (2005). Contribution of genetics to the study of animal personalities: A review of case studies. *Behaviour*, *142*(9), 1185–1206. https://doi.org/10.1163/156853905774539364

Vardi, R., Goulet, C. T., Matthews, G., Berger-Tal, O., Wong, B. B. M., & Chapple, D. G. (2020). Spatial learning in captive and wild-born lizards: heritability and environmental effects. *Behavioral Ecology and Sociobiology*, *74*(2). https://doi.org/10.1007/s00265-020-2805-6

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505–522. https://doi.org/10.1111/j.1558-5646.1985.tb00391.x

Visscher, P. M. (2006). A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics*, *9*(4), 490–495. https://doi.org/10.1375/183242706778024928

Wallace, K. J., & Hofmann, H. A. (2021). Equal performance but distinct behaviors: sex differences in a novel object recognition task and spatial maze in a highly social cichlid fish. *Animal Cognition*, *1*, 3. https://doi.org/10.1007/s10071-021-01498-0

Walsh, B., & Blows, M. W. (2009). Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 41–59. https://doi.org/10.1146/annurev.ecolsys.110308.120232

Wang, M. Y., Brennan, C. H., Lachlan, R. F., & Chittka, L. (2015). Speed-accuracy trade-offs and individually consistent decision making by individuals and dyads of zebrafish in a colour discrimination task. *Animal Behaviour*, *103*, 277–283. https://doi.org/10.1016/j.anbehav.2015.01.022

Warren, E. W., & Callaghan, S. (1976). The response of male guppies (Poecilia reticutata, Peters) to repeated exposure to an open field. *Behavioral Biology*, *18*(4), 499–513. https://doi.org/10.1016/S0091-6773(76)92524-4

Wasserman, E. A., & Zentall, T. R. (2006). Comparative Cognition: Experimental Explorations of Animal Intelligence. In *Comparative Cognition: Experimental Explorations of Animal Intelligence*. https://doi.org/10.1093/acprof:oso/9780195377804.001.0001

Westneat, D. F., Schofield, M., & Wright, J. (2013). Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. In *Behavioral Ecology* (Vol. 24, Issue 3, pp. 598–604). https://doi.org/10.1093/beheco/ars207

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Wey, T. W., Réale, D., & Kelly, C. D. (2019). Developmental and genetic effects on behavioral and life-history traits in a field cricket. *Ecology and Evolution*, *9*(6), 3434–3445. https://doi.org/10.1002/ece3.4975

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring Impulsivity and Examining Its Relationship to Delinquency. *Journal of Abnormal Psychology*, *103*(2), 192–205. https://doi.org/10.1037/0021-843X.103.2.192

White, S. J., Houslay, T. M., & Wilson, A. J. (2019). Evolutionary genetics of

personality in the Trinidadian guppy II: sexual dimorphism and genotype-by-sex interactions. *Heredity*, *122*(1), 15–28. https://doi.org/10.1038/s41437-018-0083-0

White, S. J., Kells, T., & Wilson, A. J. (2016). Metabolism, personality and pace of life in the Trinidadian guppy, Poecilia reticulata. *Behaviour*, *153*(13–14), 1517–1543. https://doi.org/10.1163/1568539X-00003375

White, S. J., Pascall, D., & Wilson, A. J. (2020). Towards a comparative approach to the structure of animal personality variation. *Behavioral Ecology*, *31*(2), 340–351. https://doi.org/10.1093/beheco/arz198

White, S. J., & Wilson, A. J. (2019). Evolutionary genetics of personality in the Trinidadian guppy I: maternal and additive genetic effects across ontogeny. *Heredity*, *122*(1), 1–14. https://doi.org/10.1038/s41437-018-0082-1

White, S. L., Wagner, T., Gowan, C., & Braithwaite, V. A. (2017). Can personality predict individual differences in brook trout spatial learning ability? *Behavioural Processes*, *141*, 220–228. https://doi.org/10.1016/j.beproc.2016.08.009

Wilkinson, A., Kuenstner, K., Mueller, J., & Huber, L. (2010). Social learning in a non-social reptile (Geochelone carbonaria). *Biology Letters*, *6*(5), 614–616. https://doi.org/10.1098/rsbl.2010.0092

Williams, G. C. (1967). Ethological Isolating Mechanisms in Four Sympatric Species of Poeciliid Fishes . N. R. Liley. *The Quarterly Review of Biology*, *42*(4), 579–579. https://doi.org/10.1086/405618

Wilson, A. D. M., Godin, J. G. J., & Ward, A. J. W. (2010). Boldness and reproductive fitness correlates in the eastern mosquitofish, gambusia holbrooki. *Ethology*, *116*(1), 96–104. https://doi.org/10.1111/j.1439-0310.2009.01719.x

Wilson, A. J. (2018). How should we interpret estimates of individual repeatability? *Evolution Letters*, *2*(1), 4–8. https://doi.org/10.1002/evl3.40

Wilson, A. J., Gelin, U., Perron, M. C., & Réale, D. (2009). Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1656), 533–541. https://doi.org/10.1098/rspb.2008.1193

Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Coltman, D. W., Mifsud, D. V, Clutton-Brock, T. H., & Kruuk, L. E. B. (2006). Environmental coupling of selection and heritability limits evolution. *PLoS Biology*, *4*(7), 1270–1275. https://doi.org/10.1371/journal.pbio.0040216

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. . B., & Nussey, D. H. (2010). An ecologists guide to the animal model. In *J Anim Ecol* (Vol. 79, Issue 1, pp. 13–26). http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2009.01639.x/full

Wolf, M., Van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, *447*(7144), 581–584.

Wolf, M., & Weissing, F. J. (2012). Animal personalities: Consequences for

ecology and evolution. In *Trends in Ecology and Evolution* (Vol. 27, Issue 8, pp. 452–461). https://doi.org/10.1016/j.tree.2012.05.001

Wong, R. Y., French, J., & Russ, J. B. (2019). Differences in stress reactivity between zebrafish with alternative stress coping styles. *Royal Society Open Science*, *6*(5), 181797. https://doi.org/10.1098/rsos.181797

Wong, S. C., Dykstra, M., Campbell, J. M., & Earley, R. L. (2008). Measuring water-borne cortisol in convict cichlids (Amatitlania nigrofasciata): Is the procedure a stressor? *Behaviour*, *145*(10), 1283–1305. https://doi.org/10.1163/156853908785765863

Wourms, J. P. (1981). Viviparity: The maternal-fetal relationship in fishes. *Integrative and Comparative Biology*, *21*(2), 473–515. https://doi.org/10.1093/icb/21.2.473

Zidar, J., Balogh, A., Favati, A., Jensen, P., Leimar, O., Sorato, E., & Løvlie, H. (2018). The relationship between learning speed and personality is age- and task-dependent in red junglefowl. *Behavioral Ecology and Sociobiology*, *72*(10), 168. https://doi.org/10.1007/s00265-018-2579-2

Zucca, P., Antonelli, F., & Vallortigara, G. (2005). Detour behaviour in three species of birds: Quails (Coturnix sp.), herring gulls (Larus cachinnans) and canaries (Serinus canaria). *Animal Cognition*, *8*(2), 122–128. https://doi.org/10.1007/s10071-004-0243-x

8 Appendix 1

Electronic Supplementary Material: S1

9 Appendix 2

Supplementary Information: Chapter 2

Table S2.1: Fixed effect estimates from the full random intercept and random slope model for maze time in both maze A and B

Table S2.2: Fixed effect estimates from the full random intercept and random slope model for relative area.

Table S2.3: Character state representation of among-individual correlation structure between trial specific *maze time* (both mazes) and *relative area*.

Figure 2.1: Characterising individual variation in learning performance

Table S2.1: Fixed effect estimates (with standard errors in parentheses) from the full random intercept and random slope model for maze time in both maze A and B

Maze	Fixed effect	Effect size (SE)	DF	F	P
A	Intercept	0.140 (0.117)	1, 114.6	1.369	0.075
	Trial	-0.043 (0.014)	1, 59.8	10.140	0.003
	Maze position (top)	0.105 (0.061)	1, 600.3	2.987	0.085
	Order	-0.025 (0.015)	1, 635.1	3.189	0.096

B	Intercept	0.172 (0.140)	1, 107.7	2.552	0.111
	Trial	0.014 (0.014)	1, 53.6	1.193	0.301
	Maze position (top)	0.066 (0.070)	1, 540.4	0.892	0.345
	Order	-0.030 (0.020)	1, 565.8	2.760	0.116

Table S2.2: Fixed effect estimates (with standard errors in parentheses) from the full random intercept and random slope model for relative area from the open field trials (OFT).

Model	Fixed effect	Effect size (SE)	DF	F	P
OFT	Intercept	0.229 (0.451)	1, 159	0.003	0.613
	Trial	-0.262 (0.066)	1, 116.8	15.710	< 0.001
	Time	0.000 (0.000)	1, 150.3	0.003	0.953

Table S2.3: Character state representation of among-individual correlation structure between trial specific *maze time* (both mazes) and *relative area*.

Table 1 in the main text presents the estimated among individual (**ID**) covariance matrix of reaction norm (RN) intercepts and slopes for *Maze time*$_A$, *Maze time*$_B$ and *relative area* (intercept only). Assuming the assumption of linear reaction norms hold true this can be transformed to the corresponding 'character state' (CS) among-individual covariance matrix of trial specific maze times and *relative area* (designated $\mathbf{ID_{CS}}$).

For a single trait (e.g. *Maze time*$_A$), $\mathbf{ID_{CS}} = \mathbf{Q} \cdot \mathbf{ID_{RN}} \cdot \mathbf{Q^T}$ (following e.g. equation 5.8 in Roff et al., 2014), where $\mathbf{ID_{RN}}$ is the 2x2 covariance matrix of reaction norm (RN) intercepts and slopes, $\mathbf{Q^T}$ is the transpose of matrix **Q**, and **Q** itself contains the values of the covariate (trial number) at which we wish to evaluate $\mathbf{ID_{CS}}$. Where we want $\mathbf{ID_{CS}}$ to be an 11x11 matrix containing the among-individual variance in maze time at each trial number (1-11) on the diagonal, and the covariance between each pair of trial numbers in the off diagonal elements

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \\ 1 & 10 \\ 1 & 11 \end{bmatrix}$$

Following this, but expanded to the multivariate case, we transformed the estimated covariance matrix (**ID**) formulated under the trivariate model described in the main text (i.e., with individual effects on Maze A and B modelled as first order random regressions of trial number) to the corresponding character state matrix. This was then rescaled to yield point estimates of the among-individual correlation between trial specific performance within- and across-mazes, and between these performances and stress responsiveness. For simplicity we do not similarly attempt to transform estimates of uncertainty, but note that this table is a mathematical consequence (and transformation) of the statistical estimated presented in Table 1 of the main text (i.e. this is same set of results presented a different way).

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	RA	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
A1																							
A2	0.993																						
A3	0.969	0.991																					
A4	0.925	0.963	0.990																				
A5	0.860	0.914	0.959	0.989																			
A6	0.778	0.846	0.909	0.958	0.990																		
A7	0.684	0.765	0.843	0.910	0.960	0.991																	
A8	0.586	0.678	0.768	0.85	0.917	0.965	0.992																
A9	0.491	0.590	0.690	0.785	0.867	0.929	0.971	0.994															
A10	0.402	0.507	0.615	0.719	0.812	0.888	0.943	0.977	0.995														
A11	0.321	0.430	0.544	0.656	0.759	0.845	0.910	0.955	0.983	0.996													
RA	0.286	0.293	0.295	0.292	0.284	0.269	0.249	0.228	0.205	0.183	0.162												
B1	0.686	0.725	0.757	0.777	0.782	0.770	0.744	0.707	0.665	0.621	0.577	0.024											
B2	0.691	0.733	0.769	0.792	0.800	0.791	0.767	0.733	0.691	0.648	0.605	0.062	0.996										
B3	0.689	0.735	0.774	0.801	0.812	0.806	0.784	0.752	0.712	0.670	0.627	0.101	0.981	0.995									

B4	0.681	0.729	0.771	0.801	0.815	0.812	0.794	0.764	0.726	0.686	0.645	0.140	0.957	0.980	0.995								
B5	0.666	0.716	0.760	0.793	0.810	0.811	0.795	0.768	0.733	0.694	0.655	0.178	0.922	0.955	0.979	0.995							
B6	0.644	0.696	0.742	0.777	0.797	0.801	0.789	0.764	0.732	0.696	0.659	0.214	0.879	0.920	0.954	0.980	0.995						
B7	0.617	0.670	0.717	0.754	0.777	0.784	0.775	0.754	0.724	0.691	0.656	0.246	0.829	0.878	0.921	0.956	0.981	0.995					
B8	0.586	0.639	0.687	0.726	0.752	0.761	0.755	0.737	0.711	0.680	0.648	0.274	0.775	0.831	0.882	0.925	0.959	0.982	0.996				
B9	0.552	0.606	0.655	0.695	0.722	0.734	0.731	0.716	0.693	0.665	0.636	0.299	0.718	0.781	0.838	0.889	0.931	0.963	0.984	0.996			
B10	0.518	0.571	0.620	0.661	0.690	0.704	0.704	0.692	0.672	0.647	0.620	0.319	0.661	0.729	0.793	0.851	0.900	0.939	0.968	0.987	0.997		
B11	0.484	0.536	0.585	0.627	0.657	0.673	0.675	0.666	0.649	0.627	0.602	0.336	0.605	0.678	0.747	0.810	0.866	0.911	0.947	0.972	0.989	0.997	

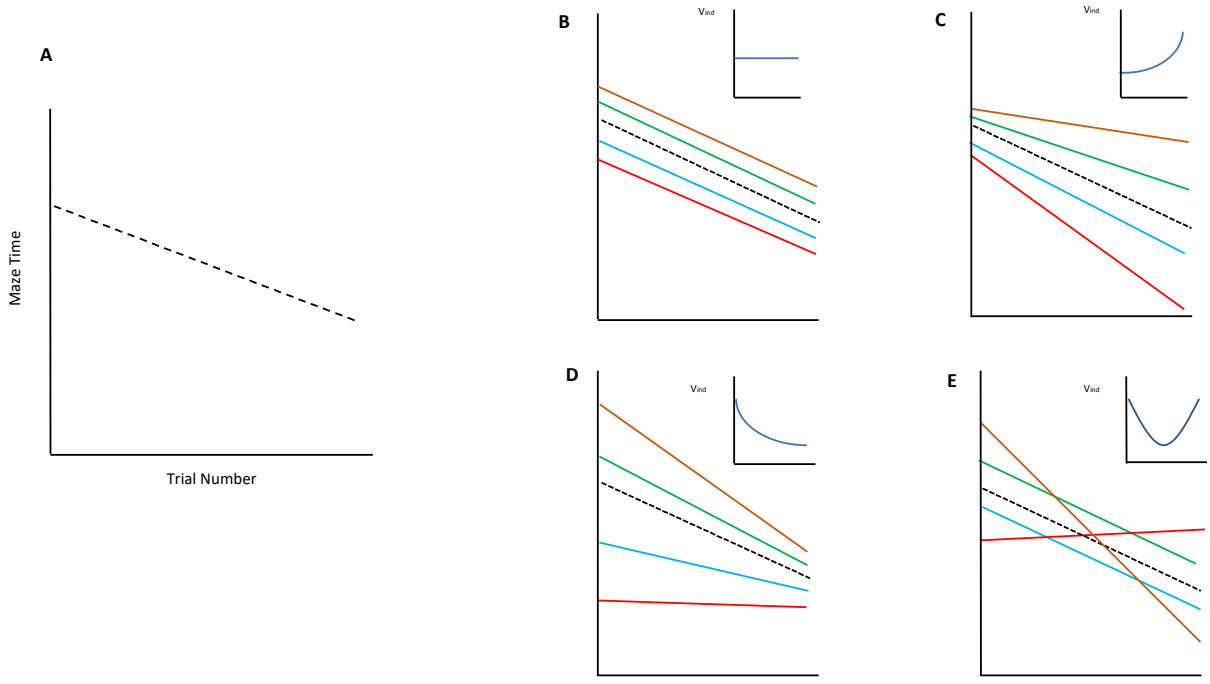


Figure. S2.1: Characterising individual variation in learning performance. Main panel (A) shows an average (black dashed line) decrease in maze time with trial number from 1 to 11 consistent with learning. Inset panels show how individual trajectories may vary around this because of differences in reaction norm intercepts (B) and or slopes (C-E). Where slopes vary (C-E), a corollary of this is that the among-individual variance (V_{ind}) in maze time will change across trials. This could potentially increase (C) or decrease (D) monotonically, or there could be an intermediate trial number at which variance is minimised (E) or maximised (not shown). Where reaction norms tend to cross a lot within the range of trial numbers explored (E), this will result in low (and potentially negative) among-individual correlations between early and late trials.

10 Appendix 3

Supplementary Information: Chapter 3

Table S3.1: Fixed effects estimates for the full univariate models for each trait

Table S3.2: Principle components loadings (eigen decomposition) of ID matrix

Figure S3.1: Plots of raw data of all traits across each set of trials

Table S3.1: Fixed effects estimates for the full univariate models Fixed effect estimates (with standard errors in parentheses) from full univariate models for each trait. Intercept represents estimate at Trial number = 1, sex = Female, Stack = A, Colour = Blue and Reward side = Left.

Table S3.1a – *emergence*

Trait		Fixed effect	Effect size (SE)	DF	F	P
emergence	Mean	Intercept	0.555 (0.213)	1, 81.9	0.002	0.966
		Trial number (2)	-0.082 (0.136)	8, 555.7	9.240	<0.001
		Trial number (3)	0.425 (0.138)			
		Trial number (4)	0.580 (0.138)			
		Trial number (5)	0.475 (0.118)			
		Trial number (6)	0.523 (0.125)			

Trial number (7)	0.572 (0.121)			
Trial number (8)	0.603 (0.126)			
Trial number (9)	0.363 (0.125)			
Sex (Male)	-0.324 (0.168)	1, 82.4	3.705	0.057
Stack (B)	0.101 (0.168)	1, 82.6	2.328	0.550
Colour (Green)	-0.030 (0.152)	1, 81.9	0.034	0.846
Reward side (Right)	-0.113 (0.070)	1, 561.8	2.585	0.108

Table S3.1b – AL_{time}

Trait		Fixed effect	Effect size (SE)	DF	F	P
AL_{time}	Mean	Intercept	0.151 (0.227)	1, 58.5	0.146	0.704
		Trial number (2)	-0.224 (0.149)	17, 801.8	2.246	0.002
		Trial number (3)	-0.250 (0.147)			
		Trial number (4)	-0.043 (0.154)			
		Trial number (5)	-0.298 (0.150)			
		Trial number (6)	-0.313 (0.174)			
		Trial number (7)	-0.265 (0.170)			
		Trial number (8)	-0.417 (0.158)			
		Trial number (9)	-0.403 (0.149)			
		Trial number (10)	-0.377 (0.152)			
		Trial number (11)	-0.567 (0.154)			
		Trial number (12)	-0.269 (0.176)			
		Trial number (13)	-0.414 (0.158)			
		Trial number (14)	-0.527 (0.155)			
		Trial number (15)	-0.587 (0.178)			
		Trial number (16)	-0.512 (0.154)			
		Trial number (17)	-0.688 (0.182)			

Trial number (18)	-0.553 (0.227)			
Sex (Male)	0.315 (0.181)	1, 59.5	2.381	0.087
Stack (B)	0.430 (0.181)	1, 59.2	3.606	0.021
Colour (Green)	-0.362 (0.146)	1, 87.8	5.894	0.015
Reward side (Right)	-0.095 (0.065)	1, 803.4	2.110	0.147

Table S3.1c – AL_{speed}

Trait		Fixed effect	Effect size (SE)	DF	F	P
AL_{speed}	Mean	Intercept	-0.657 (0.216)	1, 81.7	0.002	0.970
		Trial number (2)	0.131 (0.134)	8, 555.0	7.380	<0.001
		Trial number (3)	0.423 (0.136)			
		Trial number (4)	0.659 (0.137)			
		Trial number (5)	0.499 (0.117)			
		Trial number (6)	0.648 (0.124)			
		Trial number (7)	0.631 (0.120)			
		Trial number (8)	0.508 (0.124)			
		Trial number (9)	0.395 (0.216)			
		Sex (Male)	0.197 (0.172)	1, 82.3	1.347	0.261
		Stack (B)	-0.102 (0.172)	1, 82.4	1.277	0.554
		Colour (Green)	0.223 (0.155)	1, 81.7	2.011	0.156
		Reward side (Right)	0.148 (0.070)	1, 560.9	4.552	0.033

Table S3.1d – $AL_{accuracy}$

Trait		Fixed effect	Effect size (SE)	DF	F	P
$AL_{accuracy}$	Mean	Intercept	-0.645 (0.177)	1, 79.3	0.002	0.963
		Trial number (2)	-0.137 (0.140)	8, 558.0	4.602	<0.001
		Trial number (3)	0.203 (0.142)			
		Trial number (4)	0.330 (0.142)			
		Trial number (5)	0.361 (0.122)			
		Trial number (6)	0.374 (0.124)			
		Trial number (7)	0.395 (0.124)			
		Trial number (8)	0.308 (0.129)			
		Trial number (9)	0.230 (0.177)			
		Sex (Male)	0.103 (0.127)	1, 79.3	1.049	0.418
		Stack (B)	-0.205 (0.127)	1, 79.5	4.229	0.112
		Colour (Green)	0.907 (0.115)	1. 79.3	61.770	<0.001
		Reward side (Right)	0.030 (0.072)	1, 567.4	0.161	0.689

Table S3.1e – MC_{time}

Trait		Fixed effect	Effect size (SE)	DF	F	P
MC_{time}	Mean	Intercept	0.242 (0.244)	1, 48.1	0.002	0.967
		Trial number (2)	0.028 (0.174)	17, 692.6	1.174	<0.001
		Trial number (3)	-0.099 (0.174)			
		Trial number (4)	0.143 (0.149)			
		Trial number (5)	-0.288 (0.171)			
		Trial number (6)	0.178 (0.154)			
		Trial number (7)	-0.037 (0.171)			
		Trial number (8)	0.386 (0.174)			
		Trial number (9)	0.104 (0.172)			
		Trial number (10)	-0.332 (0.147)			
		Trial number (11)	-0.516 (0.229)			
		Trial number (12)	-0.468 (0.150)			
		Trial number (13)	-0.309 (0.151)			
		Trial number (14)	-0.272 (0.152)			
		Trial number (15)	-0.452 (0.148)			
		Trial number (16)	-0.166 (0.149)			
		Trial number (17)	-0.233 (0.152)			

Trial number (18)	-0.502 (0.146)			
Sex (Male)	-0.063 (0.192)	1, 48.3	-0.974	0.742
Stack (B)	0.262 (0.193)	1, 48.3	3.742	0.180
Colour (Green)	-0.560 (0.158)	1, 48.1	12.61	0.001
Reward side (Right)	0.279 (0.072)	1, 694.8	15.07	<0.001

Table S3.1f – RL_{time}

Trait		Fixed effect	Effect size (SE)	DF	F	P
RL_{time}	Mean	Intercept	-0.341 (0.307)	1, 47.7	0.005	0.946
		Trial number (2)	0.196 (0.197)	17, 581.2	1.565	0.060
		Trial number (3)	0.272 (0.214)			
		Trial number (4)	0.328 (0.195)			
		Trial number (5)	-0.012 (0.193)			
		Trial number (6)	0.332 (0.190)			
		Trial number (7)	0.224 (0.197)			
		Trial number (8)	0.360 (0.196)			
		Trial number (9)	0.344 (0.213)			
		Trial number (10)	1.132 (0.594)			
		Trial number (11)	0.343 (0.234)			
		Trial number (12)	0.449 (0.218)			
		Trial number (13)	0.272 (0.205)			
		Trial number (14)	0.487 (0.254)			
		Trial number (15)	0.494 (0.206)			
		Trial number (16)	0.616 (0.200)			
		Trial number (17)	0.503 (0.206)			

Trial number (18)	0.310 (0.205)			
Sex (Male)	-0.104 (0.219)	1, 47.7	0.169	0.638
Stack (B)	-0.153 (0.227)	1, 47.9	0.380	0.505
Colour (Green)	0.348 (0.182)	1, 47.4	3.529	0.066
Reward side (Right)	0.043 (0.078)	1, 578.4	0.308	0.579

Table S3.1g – $RL_{accuracy}$

Trait		Fixed effect	Effect size (SE)	DF	F	P
$RL_{accuracy}$	Mean	Intercept	-0.402 (0.263)	1, 46.3	0.110	0.742
		Trial number (2)	0.206 (0.211)	17, 584.2	1.085	0.271
		Trial number (3)	0.104 (0.230)			
		Trial number (4)	0.350 (0.209)			
		Trial number (5)	0.280 (0.206)			
		Trial number (6)	0.290 (0.203)			
		Trial number (7)	0.279 (0.211)			
		Trial number (8)	0.185 (0.210)			
		Trial number (9)	0.315 (0.228)			
		Trial number (10)	0.058 (0.636)			
		Trial number (11)	0.001 (0.260)			
		Trial number (12)	0.327 (0.233)			
		Trial number (13)	0.359 (0.219)			
		Trial number (14)	0.225 (0.272)			
		Trial number (15)	0.548 (0.220)			
		Trial number (16)	0.514 (0.214)			
		Trial number (17)	0.610 (0.220)			

Trial number (18)	0.437 (0.219)			
Sex (Male)	-0.008 (0.166)	1, 46.4	0.006	0.963
Stack (B)	-0.303 (0.173)	1, 48.1	5.464	0.086
Colour (Green)	0.631 (0.138)	1, 46.1	20.860	<0.001
Reward side (Right)	0.025 (0.084)	1, 579.9	0.089	0.764

Table S3.2: – Principle components analyses (eigen decomposition) of ID matrix

Table S3.2a: Phenotypic variance (%) explained by six principle components of ID (on a correlation scale), with 95% confidence intervals from 5000 bootstrap replicates

PC1	PC2	PC3	PC4	PC5	PC6
0.571 [0.427, 0.698]	0.254 [0.126, 0.309]	0.121 [0.074, 0.156]	0.037 [0.047, 0.099]	0.017 [0.008, 0.055]	0.001 [0, 0]

Table S3.2b: Loadings of six cognitive traits onto six principle components of ID (on a correlation scale)

	PC1	PC2	PC3	PC4	PC5	PC6
AL_{speed}	-0.435	0.431	-0.022	-0.217	0.758	-0.047
$AL_{accuracy}$	-0.384	0.543	0.081	-0.198	-0.555	0.451
AL_{time}	-0.513	0.056	0.254	0.423	-0.240	-0.658
MC_{time}	-0.396	-0.471	0.343	0.373	0.193	0.571
RL_{time}	-0.367	-0.523	0.022	-0.737	-0.135	-0.174
$RL_{accuracy}$	-0.329	-0.139	-0.900	0.231	-0.065	0.070

Figure S3.1: Plots of observed means for all performance time and decision accuracy measures across sets and trials

Figure S3.1a: *emergence*

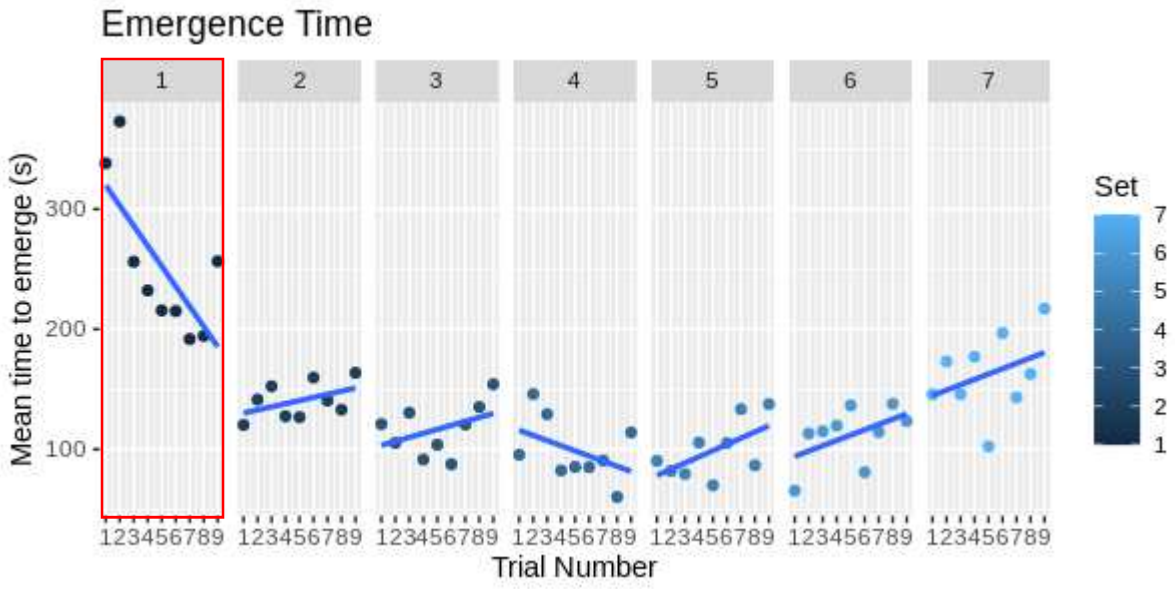


Figure S3.1b: *AL*$_{time}$

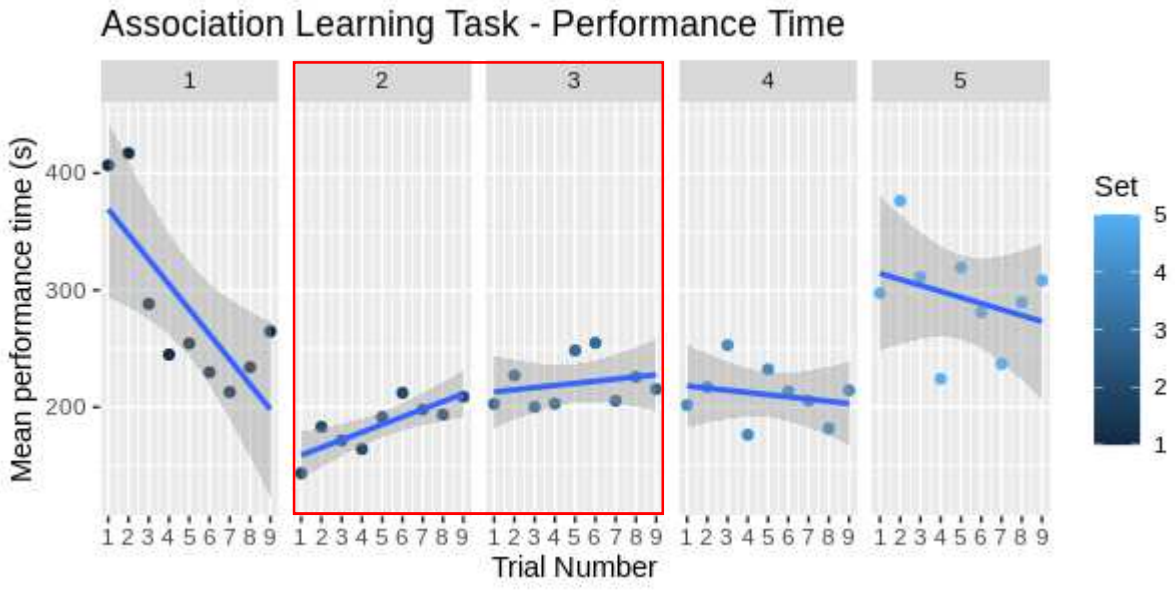


Figure S3.1c: *AL*$_{speed}$

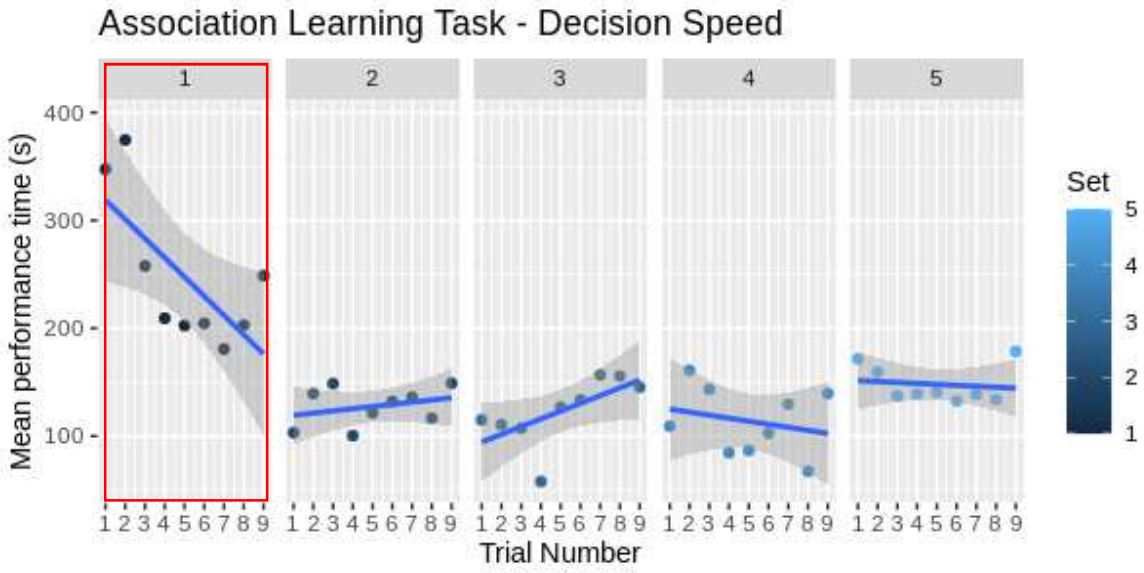


Figure S3.1d: *AL*$_{accuracy}$

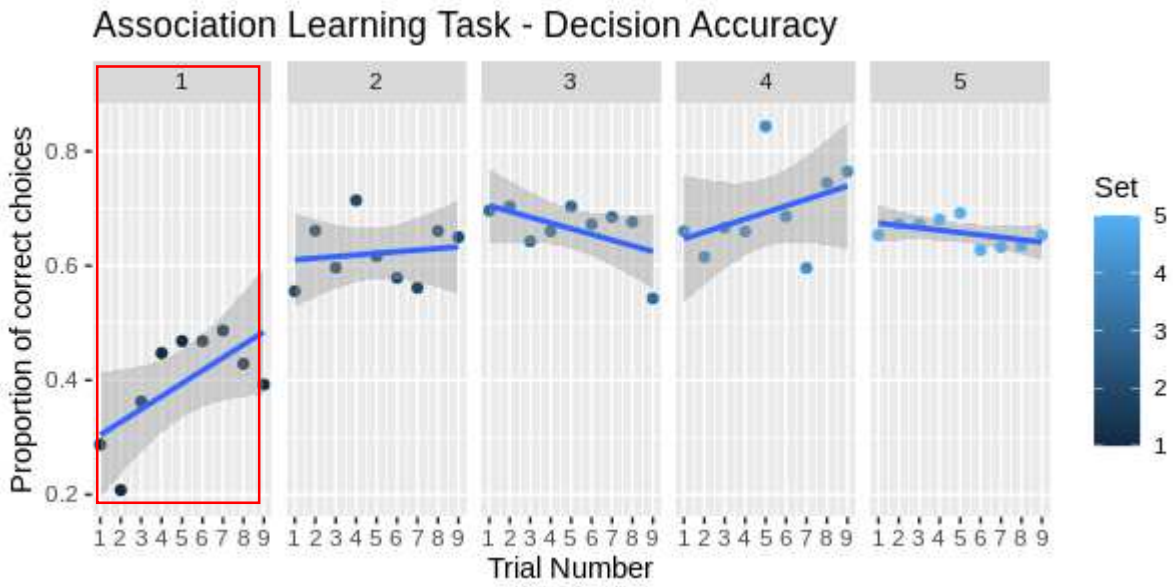


Figure S3.1e: *MC*$_{time}$

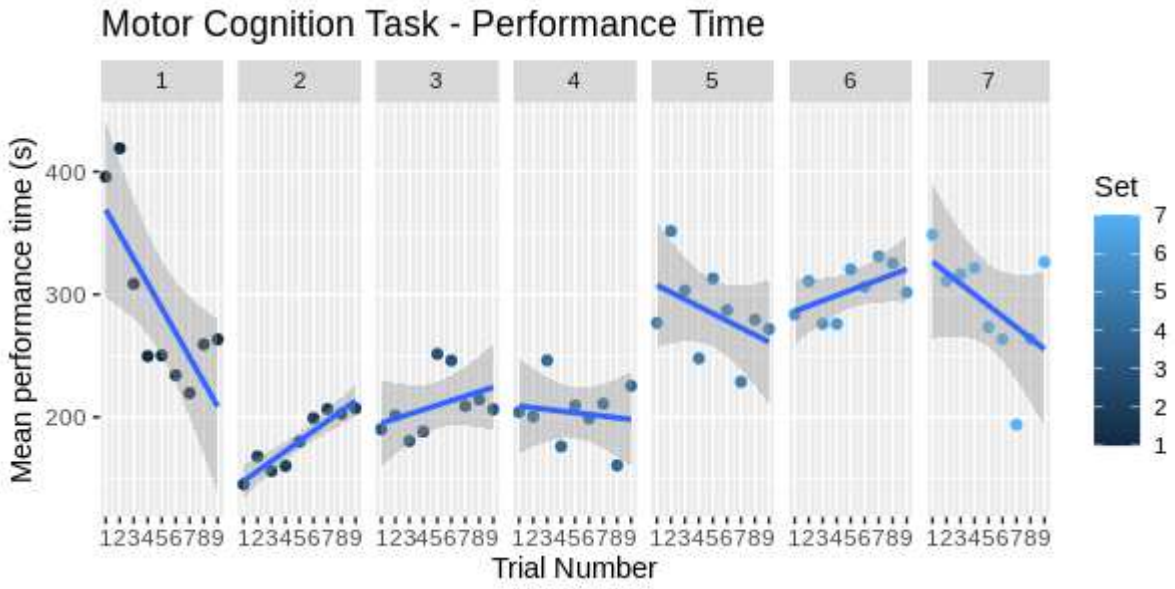


Figure S33.1f: *RL*$_{time}$