



Proactive common waxbills make fewer mistakes in a cognitive assay, the detour-reaching task

Ana Cristina R. Gomes¹ · Soraia Guerra¹ · Paulo A. Silva¹ · Cristiana I. Marques¹ · Sandra Trigo¹ · Neeltje J. Boogert² · Gonçalo C. Cardoso^{1,3}

Received: 31 December 2019 / Revised: 17 January 2020 / Accepted: 20 January 2020 / Published online: 20 February 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Reactive individuals explore and make decisions more slowly than proactive individuals, paying more attention to external cues. Reactive personality types are therefore predicted to make fewer mistakes in cognitive tasks, especially in tests of inhibitory control and reversal learning that require individuals to adjust behavioral routines. However, empirical evidence for this is mixed, and work relating cognition to a broader set of phenotypic traits has been called for. Here, using the common waxbill (*Estrilda astrild*), we relate performance in a cognitive assay, the detour-reaching task, to differences in personality, while controlling for putative confounding factors (including sex, breath rate, body size, and mass). Contrary to theoretical predictions, we found that proactive individuals, rather than reactive, made fewer mistakes in the detour-reaching task. We used body size and mass as proxies for condition, but neither predicted personality type or cognitive performance, thus providing no support for the hypothesis that condition dependence mediates the association between proactive personality and detour-reaching performance. This work adds to a growing number of studies contributing conflicting findings on how animal personality relates to cognitive performance, which, together, appear to require novel theory and more nuanced predictions.

Significance statement

Theory predicts that reactive personality types (i.e., slower-acting and more attentive individuals) should perform better on cognitive tasks that involve inhibiting a response to an irrelevant stimulus to reach food (inhibitory control). Tests of this hypothesis have generated contradictory results and studies that account for confounding factors are needed. We quantified performance in an inhibitory control assay among common waxbills (*Estrilda astrild*), while controlling for the largest set of confounding factors to date. Contrary to predictions, waxbills with a reactive personality made more mistakes in the cognitive assay. These and other recent findings suggest that a more nuanced theory on the relation between cognition and personality is necessary.

Keywords Animal cognition · Detour reaching · Animal personality · Mirror test · Breath rate · Body condition

Communicated by N. Clayton

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00265-020-2809-2>) contains supplementary material, which is available to authorized users.

✉ Ana Cristina R. Gomes
ana.gomes@cibio.up.pt

- ¹ CIBIO/InBIO—Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
- ² Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, UK
- ³ Behavioural Ecology Group, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark

Introduction

Cognition refers to the abilities of acquisition, processing, storage, and use of information, and can be important for the development and performance of various behaviors, from foraging and nest building to mate choice and social competence (Shettleworth 2010). Cognition includes learning, which is the change in an individual's state due to new experiences or information that causes the adjustment of behavioral responses (e.g., Dukas and Ratcliffe 2009). Individual differences in cognition and, in particular, learning, have been observed in virtually all species studied to date (e.g., insects: Dukas 2008; fishes: Lucon-Xiccato and Bisazza 2017; reptiles: Chung et al. 2017; birds and mammals reviewed in

Reader 2003; Boogert et al. 2011b; Thornton and Lukas 2012; Niemelä et al. 2013) and may be linked to broader differences in behavior. For example, individuals may be more or less likely to encounter new environmental cues depending on their personality traits, such as boldness, neophilia, activity, or exploration, and this, in turn, may facilitate or constrain learning opportunities (Griffin et al. 2015).

Personality refers to repeatable behavioral differences that are stable across time and contexts among individuals of the same species (e.g., Dall et al. 2004; Stamps and Groothuis 2010). Various behaviors can differ consistently among individuals, and when several of these behaviors (e.g., boldness, aggressiveness, sociability, or activity) are correlated within individuals, then together they form a broader “behavioral syndrome” (sensu Sih et al. 2004; Sih and Bell 2008; Carere and Locurto 2011; Sih and Del Giudice 2012; Garamszegi et al. 2013). Personality differences often occur along a slow-fast axis, from reactive to proactive types (e.g., Wilson et al. 1993, 1994; Sih et al. 2004; Sih and Del Giudice 2012). Proactive individuals tend to be faster explorers and more active, neophilic, and bolder, while reactive individuals tend to be slower explorers and less active, more neophobic, and shy (see Koolhaas et al. 1999; Sih et al. 2004; Carere et al. 2010).

Early in the twentieth century, Pavlov described a link between cognition (described by him as “different types of nervous systems”) and personality (described as “individuals’ fixed behavioral phenotypes”; Pavlov 1906, 1941; reviewed in Locurto 2007). He noted, for example, that “Excitable type” dogs are more active and aggressive, but slower to change a previously learned response, while “Inhibited type” dogs rest more, are more attentive, and faster in changing learned responses (Pavlov 1941). In recent years, an increasing number of studies have started to investigate relations between cognition and personality in taxonomically diverse species (reviewed in Locurto 2007; Carere and Locurto 2011; Sih and Del Giudice 2012; Griffin et al. 2015; Dougherty and Guillette 2018).

In a foundational paper, Sih and Del Giudice (2012) proposed a conceptual framework for the relation between cognition and personality, later coined the “Cognitive Style Hypothesis” (CSH, Chung et al. 2017). In line with the early findings by Pavlov, the CSH predicts that, because proactive individuals act and sample the environment faster, they should pay less attention to external cues and thus be less flexible in changing routines. The opposite is predicted for reactive individuals, who act more slowly, and would thus pay more attention to external cues, being more flexible in changing behavioral routines. Reactive individuals are predicted to perform better (i.e., make fewer errors), particularly in reversal learning and inhibitory control tests, which require, respectively, adjusting a previously learned behavior, or inhibiting a response to an irrelevant stimulus (e.g., Sih and Del Giudice 2012).

Empirical studies provide mixed support for the CSH (conflicting results are reviewed in Table 1 of Mazza et al. 2018). Several studies have reported findings that support predictions from the CSH (e.g., Verbeek et al. 1994; Guillette et al. 2009; Carere et al. 2010; Ruiz-Gomez et al. 2011; Cockrem 2013; Guenther et al. 2014; reviewed in Groothuis and Carere 2005; Cockrem 2007; Coppens et al. 2010; Carere and Locurto 2011; Sih and Del Giudice 2012; Griffin et al. 2015; Mazza et al. 2018). However, the recent meta-analysis by Dougherty and Guillette (2018) found that, while significant associations between personality types and cognitive performance exist, the direction of those associations is inconsistent across species, or even within the same species (e.g., between the sexes; Dougherty and Guillette 2018). Faced with this puzzling state of the art, Dougherty and Guillette (2018) called for empirical work that tests for associations between cognition and several aspects of personality, rather than focusing on a single personality trait, and that also takes into account confounding factors such as sex or physiological differences. This could provide a better understanding of the phenotypic correlates of cognition and help formulate more nuanced theory and predictions than the CSH currently affords (Dougherty and Guillette 2018).

Here, we follow Dougherty and Guillette’s (2018) call for measuring a diverse range of phenotypic traits, in addition to the personality traits of interest, to assess whether these drive or confound the relation between cognition and personality type. Our goals are to test for a relation between cognition and different aspects of personality, and whether such a relation is robust to, or confounded by, additional phenotypic differences among individuals. Our model species is the common waxbill (*Estrilda astrild*), a gregarious finch species for which it has been shown that individual differences in personality are stable through time and are consistent across social and exploration contexts (Carvalho et al. 2013; Funghi et al. 2015).

As cognitive assay, we chose one of the most widely used inhibitory control assays for animals, the detour-reaching task, in which individuals need to detour around a transparent barrier to obtain food or a desired object (Boogert et al. 2011a). Previous work in other species has reported within-species correlations between performance in the detour-reaching task and phenotypic traits (bird song repertoire size, body condition; Boogert et al. 2011a; Shaw 2017), and a cross-species comparison showed that performance on the detour-reaching task correlates positively with brain size (MacLean et al. 2014; but see Healy and Rowe 2007; Chittka and Niven 2009). As personality assays, we used the mirror test, which in the common waxbill indicates stable differences in behavior along a proactive-reactive axis, and the tonic immobility test, which is related with fear or stress towards a predator (Carvalho et al. 2013; Funghi et al. 2015). Finally, as phenotypic traits that might confound the relation between cognitive performance and personality, we considered body size, size-

corrected mass, and breath rates, all of which have been previously proposed or found to correlate with either cognitive performance or personality types. For example, larger or heavier individuals tend to be more proactive (Niemelä and Dingemanse 2018) or may show better cognitive performance (Shaw 2017; van Horik et al. 2017). Also, basal metabolic rates have been reported to differ among personality types (reviewed in Mathot et al. 2019), and here we use breath rates in isolation, thus avoiding acute stress, as a proxy for individual differences in metabolic rates. Controlling for all these phenotypes in a single model species, this work is the most comprehensive attempt, to date, for assessing whether and how phenotypic traits mediate a potential relation between cognition and personality type.

Methods

We captured a total of 60 common waxbills (28 males and 32 females), in two different years, using mist nets in agricultural areas in the northwest of Portugal: 14 males and 16 females were captured in fields near Póvoa de Varzim in October 2016, and 14 males and 16 females were captured in fields around Lagoa da Apúlia in September 2017. All captures took place in the morning. Immediately after capture, we fitted each individual with a unique numbered metal leg band and weighed it with a Pesola spring scale to the nearest 0.1 g. We did not use birds that were unambiguously juveniles (i.e., absence of a red bill), nor breeding females (presence of brood patch or signs of an egg, when examined through palpation of the abdomen). Such birds were released from the nets. We transported the birds in metal cages (36.5 cm × 21.5 cm × 35.5 cm), with perches, food, and water, and reached the aviary facilities within a maximum of 4 h after capture. The aviary facilities were in a room (ca. 8 m²) where two of the walls were netted, and cages were therefore exposed to natural daylight and temperature, but sheltered from the rain, wind, and direct sunlight. Natural sunlight was supplemented by full-spectrum lights with a photoperiod similar to the natural light–dark cycle (lights switched on 30 to 45 min before sunrise, and switched off 30 to 45 min after sunset). In each year, birds were kept in 6 mixed-sex groups of 5 individuals (3 males + 2 females per cage, or vice versa, except for one cage with one male and four females), because waxbills are highly gregarious and form mixed-sex flocks. Cages (88.5 cm × 30 cm × 40 cm) had 4 perches and were open along the front, with bars. Birds had ad libitum water, mixed grit with crushed oyster shell, and seeds (Versele-Laga Prestige: Tropical Finches Australian Waxbills in 2016, and Tropical Finch in 2017) in enough feeders (2 feeders per cage, each 23 cm × 5.5 cm × 7.5 cm) for all individuals to feed simultaneously. Feeders were opaque (covered with black adhesive paper) to prevent birds from habituating to feed from a

transparent object, as this could confound their performance on the detour-reaching task (van Horik et al. 2018). Vitamins (Avigold Advance, Aviform) diluted in water were provided twice a week. We observed the birds daily to ensure they adapted to the cages and behaved normally. All birds remained in these cages for ca. 8 weeks, during which time we conducted behavioral assays. Birds were 5 or 6 days in their cages before the start of the mirror and tonic immobility assays described below (Fig. 1).

Genetic sexing

A small blood sample (< 100 µL) was collected, with a small puncture of the brachial vein, for 56 out of our 60 birds. Blood was immediately stored at –20 °C until DNA extraction. Birds were sexed following Griffiths et al. (1998), using a set of primers (P2 and P8) to amplify the CHD-W and CHD-Z genes. The PCR protocol consisted of 5 µL of MyTaq (Bioline, UK), 0.5 µL of each primer, and 2 µL of DNA template for a total volume of 10 µL. The conditions were as follows: 15 min at 95 °C, 35 cycles of 30s at 95 °C, 30s at 55 °C, and 80s at 72 °C, followed by 10 min at 72 °C. PCR products were separated on a ABI 3130 xl Genetic Analyzer (Applied Biosystems, USA) and results visualized in software GeneMapper V5.0 (Applied Biosystems, USA). Individuals showing double (ZW) and single (ZZ) bands were identified as females and males, respectively. Bands were 38 bp apart. From the 56 individuals genetically sexed, 3 had been wrongly sexed phenotypically (ca. 5%, a similar percentage of error as reported in Cardoso et al. 2014). For the 4 individuals without blood sample, we kept their phenotypic sexing.

Detour-reaching task

We conducted detour-reaching assays using protocols adapted from Boogert et al. (2011a), which have been used in several species, including in the zebra finch (*Taeniopygia guttata*; MacLean et al. 2014), another estrildid finch like the waxbill. First, for habituation to the test apparatus, birds were exposed in their home cages (i.e., containing 4 other individuals) to two opaque cylinders (in the middle of each side of the cage) containing a small spoonful of seeds identical to their diets in the home cages, with the open cylinder ends oriented to the front and back of the cage. Birds could feed freely from the cylinders by reaching the inside of the cylinders with their head. Figure 2 and Movie S1 illustrate these cylinders, made

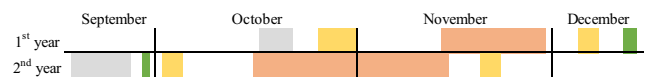


Fig. 1 Timeline of procedures in the 2 years of experiments. Gray indicates periods for capture plus weighting wild birds, green for morphometrics, yellow for the mirror and tonic immobility tests (first and second rounds), and orange for the detour-reaching task

from acrylic sheeting (0.5-mm thickness), wrapped in opaque black adhesive paper, and glued to a wooden base (4×3 cm and 0.4 cm high). Cylinders were 4 cm in length and 3.5 cm in diameter, so that only the head of one individual could fit inside the cylinder at a time. These cylinders were presented daily in each group cage, from ca. 9.30 a.m. to ca. 2.00 p.m. The habituation phase was used to minimize effects of neophobia on an individuals' performance. The habituation phase started as soon as birds were captured and brought to the aviary, and continued throughout the entire period of experiments, except during the actual behavioral assays. At the beginning of the habituation period, the seeds inside the cylinder were not being eaten in every group cage, but after ca. 5 days, the seeds inside both cylinders were already being eaten. We started the detour-reaching assays after yet another week of the habituation to the cylinder (i.e., 12 habituation days in total).

The detour-reaching assays consisted of a training phase and a test phase. For each of these two phases, each individual was isolated in the right side of its home cage by dividing the cage with an opaque white plastic partition (Fig. 2a). Food and water were removed for 60 min before the assay, to increase birds' motivation to engage with the task. The training and test phases consisted of successive trials where the bird was presented with an opaque cylinder (training phase) or a transparent cylinder (test phase) containing in its center seeds of yellow and red panicum, and white millet (3 seeds in total, 1 of each type; all preferred seeds from their diets in the home cages; ACRG and ST pers. obs.). We observed all trials standing quietly behind a dark blue curtain, permanently placed at a distance of 153 cm in front of the cages, hanging on the net wall of the aviary. We (ACRG and ST in 2016, and PAS in 2017) watched the trials through a small opening (7 cm \times 30 cm) in the curtain. Following Boogert et al. (2011a) and MacLean et al. (2014), each trial ended when the bird fed from the cylinder, whether or not the bird also pecked the cylinder earlier. If during the 10 min of a trial, the bird did not eat from the cylinder, the trial ended and, in this work, conservatively was counted as "invalid," whether or not the bird pecked at the cylinder, because this could be due to neophobia or insufficient motivation rather than inability to be successful in the

task. There was a 1-min interval between consecutive trials, and at the beginning of each trial, we removed the cylinder from the previous trial and replaced it with a different cylinder containing new seeds. For each valid trial, we recorded whether birds reached inside the cylinder immediately to eat seeds without pecking the cylinder wall first ("correct" trial) or if the first interaction with the cylinder was pecking the sides in an attempt to reach the seeds, before eating the seeds inside the cylinder ("incorrect" trial; Movie S1). We also recorded the time, in seconds, of the first interaction of the bird with the opaque cylinder (either pecking or feeding), to calculate the latency to the first attempted feeding (see below).

The training phase consisted of a series of trials using the same opaque cylinders that had been used in the habituation phase placed on the cage floor as before, with the open ends of the single presented cylinder oriented to the front and back of the cage (Fig. 2b). The purpose of the training phase was to confirm that each individual had learned to reach seeds inside the opaque cylinder to feed. The training phase ended, and the bird continued to the test phase, when three out of four consecutive trials were correct. If an individual did not feed after 60 min of trials, we stopped the assay. Of the 60 birds, 55 successfully completed the training phase, with only 2 of those pecking the side of the opaque cylinder, only in one trial. The mean number of trials in the training phase was 3.51, meaning that generally very few trials were required to confirm that individuals knew how to feed from the opaque cylinder.

The test phase took place immediately after the training phase and differed from the training phase only in that now the cylinder was transparent rather than opaque. The cylinders were identical to those described above, but not wrapped in opaque adhesive (Fig. 2c). The use of a transparent cylinder tests for detour reaching, because birds may attempt to reach the seeds inside the cylinder directly through the transparent wall (the new stimulus individuals need to inhibit responding to), rather than detouring to reach seeds by one of the cylinder openings, as they were trained to do with the opaque cylinder. The test phase lasted 15 trials, unless the bird did not feed from the cylinder in 60 min of successive trials, in which case we stopped the assay. Of the 60

Fig. 2 Detour-reaching task apparatus. **a** Focal individual isolated in the right side of its home cage, divided with an opaque partition; cage mates remain in the left side. **b** Opaque cylinder in training phase. **c** Transparent cylinder in test phase



birds, 48 passed the training and test phase. Of the 12 birds that did not pass, we re-tested 8 once more at a later date, after all other birds were tested, and only 2 passed the training and test phases; we did not re-test the other 4 birds for logistic reasons and because only 2 of the 8 re-tested birds were able to pass the training and test phases. Thus, a total of 50 birds completed the detour-reaching task, and the remaining 10 individuals were not included in further analyses of cognitive performance. Training and test phases were conducted from 10.00 a.m. to 4.30 p.m., on November 14 to 29, 2016, and on October 16 to November 14, 2017 (Fig. 1).

For each bird, we computed the proportion of correct trials in the test phase (number of correct trials divided by the total number of valid trials), as the measure of performance in the detour-reaching task. Detour-reaching scores ranged from 0.53 to 1.00 (0.80 ± 0.15 , mean \pm standard deviation (SD)). In order to control for possible differences in hunger or willingness to interact with the apparatus, we measured the cumulative amount of time (in seconds) that each bird took to interact for the first time with the opaque cylinder in the training phase; if the first interaction did not happen during the first trial, then the total time from the previous trials was summed. Hereafter, we refer to this as “latency to feed in training phase”, and it ranged from 6 to 2130 s (mean \pm SD = 307.98 ± 510 s). We used the time to first feed, rather than the number of trials, because there was much more variation in the time than the number of trials to first feeding.

Finally, we measured how performance changed along the test phase, i.e., whether it increased or decreased across trials, calculating the mean order of correct trials and dividing by the total number of valid trials. For example, if only 8 trials were valid and, of those 8, only the last 3 were correct (trials 6 to 8; mean = 7), then the mean order of correct trials is $7/8 = 0.875$. A mean order above 0.5 indicates increasing performance, because correct trials are more common late than early in the test phase; likewise, mean orders below 0.5 indicate decreasing performance. The purpose of quantifying the mean order of correct trials is to measure plastic improvement when given the detour-reaching task with the transparent cylinder, which the birds had never encountered before. We can use this measure to test whether proactive and reactive individuals differ in the extent to which they plastically improve their performance during the test phase (mean order of correct trials above 0.5) as compared with not improving it (mean order of correct trials at ca. 0.5). Mean order of correct trials ranged from 0.38 to 0.67 (mean \pm SD = 0.53 ± 0.05).

Mirror and tonic immobility tests

We conducted mirror tests and tonic immobility tests, both of which have previously been shown to be repeatable in common waxbills (Carvalho et al. 2013; Funghi et al. 2015). Mirror tests

have been used to assay aggressiveness in several species (reviewed in Cattelan et al. 2017), but waxbills do not react aggressively towards their mirror image (Carvalho et al. 2013; Funghi et al. 2015), perhaps because they are highly gregarious and sociable. Thus, in waxbills, this test pertains to non-aggressive responses towards a social stimulus (mirror image). More active behavior during the mirror test (vocalizing and moving more) and less attentiveness to the mirror image have been shown to correlate with more exploratory behavior in open-field tests (Carvalho et al. 2013), indicating a behavioral syndrome. Therefore, the mirror test can be used as an assay of personality along a proactive-reactive axis in waxbills, where more proactive individuals explore more in open-field tests and are more active but less attentive towards a mirror image, and the reverse for more reactive individuals (Carvalho et al. 2013). The tonic immobility test consists of turning a waxbill upside down, mimicking what may happen in encounters with predators or aggressive interactions, which can trigger a state of paralysis (tonic immobility). The duration of this state of paralysis is commonly used as an index of fear (e.g., Gallup 1979; Nakayama and Miyatake 2010; Edelaar et al. 2012; Pusch et al. 2018). Previous work with waxbills showed that the duration of tonic immobility is not correlated with behavior in the open-field test, the mirror test, or levels of aggression when caught by a human (Carvalho et al. 2013; Funghi et al. 2015). Therefore, we interpret individual differences in the duration of tonic immobility as a behavioral trait unrelated to the waxbill proactive-reactive behavioral syndrome, as assayed with the mirror test described above.

We conducted these tests twice on each bird, in 2 rounds separated by ca. 6 weeks (Fig. 1), to assess repeatability of each behavior. Tests were performed between 9.30 a.m. and 1.00 p.m., and the tonic immobility test always immediately followed the mirror test. Tests were performed by ACRG in 2016 on October 26–31 and repeated on December 5–7 for birds captured in 2016, and by ACRG and SG in 2017 on October 2–4 and repeated on November 20–22 for birds captured in 2017 (Fig. 1). All tests were video-recorded (Canon LEGRIA HF M306) and behaviors then quantified using Observer XT 11 (Noldus Information Technology bv; version 11.0.630), by ACRG and SG, in 2016 and 2017, respectively.

Similarly to previous work on common waxbills (Carvalho et al. 2013; Funghi et al. 2015), we conducted the mirror test in a small bird cage (24.5 cm \times 17 cm \times 15 cm) with three perches equally distanced (Fig. 3a, b), which we placed 73.4 cm above the floor in the center of an empty room (4 m²) illuminated with full spectrum lamps. A mirror (17 cm \times 15 cm), initially covered with removable cardboard, was located at one end of the cage, and a fixed cardboard of the same size was positioned at the opposite end of the cage. This assay lasted for 10 min: in the first 5 min, the mirror was covered (Fig. 3a), while in the last 5 min, the mirror was exposed (Fig. 3b). The mirror was uncovered by pulling the cardboard with a nylon string from outside the room. Each bird was taken by hand from its home cage and, in less than

1 min, was transported to the mirror test room inside a paper bag and released into the small cage. The mirror test began within 1 min of the bird being released into the small cage. Following Carvalho et al. (2013), we quantified 7 behaviors, both before and after exposing the mirror: the position relative to the mirror (time-weighted mean distance categories to the mirror; from 1: nearest to the mirror, to 5: most distant from the mirror), the proportion of time facing the mirror, the number of vocalizations, the activity inside the cage (sum of movements in the cage), the time spent in fast movements (flying or hopping continuously, in seconds), resting, and grooming. Detailed descriptions of these measurements are provided by Carvalho et al. (2013). Resting and grooming were not analyzed further because the former was very rare (<0.01 s performed by only 6 individuals), and the latter was never observed. We then tested which behaviors increased from the periods before to after uncovering the mirror (Carvalho et al. 2013; Funghi et al. 2015). Several behaviors increased significantly from before to after uncovering the mirror: time facing the mirror (means before and after uncovering mirror = 3.86 s and 4.86 s; Wilcoxon paired-sample tests: $V = 2771$, $N = 120$, 60 birds times 2 rounds, $P = 0.02$), time in fast movement (means = 1.59 s and 2.16 s; $V < 0.01$, $N = 120$, $P < 0.01$), and activity (means = 34.79 and 46.34; $V = 2476$, $N = 120$, $P = 0.01$). The number of vocalizations also increased from before to after uncovering the mirror, though not significantly so (means = 2.46 and 4.47; $V = 367$, $N = 120$, $P = 0.20$). Since in previous work with larger samples the number of vocalizations did increase significantly after exposing the mirror (Carvalho et al. 2013), we also included it in our analyses. Position relative to the mirror did not change significantly ($V = 2990$, $N = 120$, $P = 0.21$), similar to previous work (Carvalho et al. 2013; Funghi et al. 2015) and was not included in further analyses. We summarized these selected four behaviors (time facing the mirror, fast movements, activity, and vocalizations) with a principal component analysis (PCA), using only data from the period with the mirror uncovered. We log-transformed the number of vocalizations and activity ($\log(x + 1)$, in both cases), and also the time spent performing fast movements ($\log(1000 * x + 1)$) to approach normality. We did not transform the time facing the mirror as this was unfeasible because of its near-platykurtic distribution. The PCA on these four behaviors returned one principal component (PC) with an eigenvalue > 1

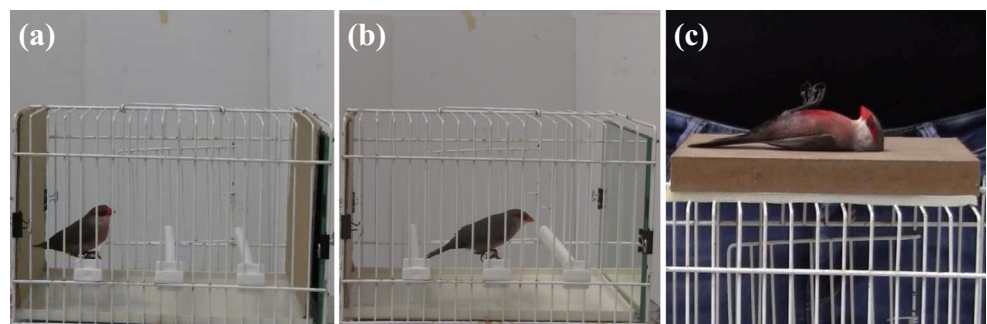
(eigenvalue = 2.49), explaining 62% of the variation. Activity, fast movements, and number of vocalizations had high positive loadings on this PC (0.85, 0.84, and 0.69, respectively), while time facing the mirror had a high negative loading (-0.77). Thus, high values on this PC indicate more active responses to the mirror (meaning more proactive individuals), and low values indicate less active responses, but more attentive towards the mirror image (meaning more reactive individuals), similarly to previous work using this behavioral test in waxbills (Carvalho et al. 2013; Funghi et al. 2015). Mirror test values ranged from -1.86 to 3.71 (mean \pm SD = 0.00 ± 1.45).

For each individual, we conducted the tonic immobility test immediately following the mirror test. To measure individual variation in tonic immobility, we used a wooden platform (5 cm \times 5.1 cm \times 1.7 cm) on top of the mirror test cage (91 cm above the floor), then removed the bird from the cage and, within 1 min, gently placed the bird on its back on the wooden platform, with its side facing the researcher (Fig. 3c). Then, the researcher, always at approximately 25 cm from the platform and always wearing dark clothes, pulled back her arms and waited for a maximum of 60 s for the bird to overturn and fly away. Afterwards, through the video recordings, we quantified the exact time, in seconds, that each bird stayed in tonic immobility, from 0 (did not stay in tonic immobility state) to 60 s (maximum duration of the test). This behavioral test was previously shown to be repeatable within individuals, but unrelated to the behavior in the open-field or the mirror tests (Carvalho et al. 2013; Funghi et al. 2015). Tonic immobility values ranged from 0 to 60 s (mean \pm SD = 24.62 ± 20.97 s).

Breath rates

Individual differences in breath rate can be difficult to quantify in small birds independently of stress responses. Breath rates are often measured by keeping a bird in the hand while feeling and counting breast movements (e.g., Carere et al. 2001; Fucikova et al. 2009; Kluehn et al. 2014). These measurements are strongly influenced by the birds' acute stress response to being handled, and Calder (1968) advised that natural breath rates should be quantified in standardized conditions, with the birds free to move normally. Here, we measured each

Fig. 3 a, b Mirror test. a With the mirror covered. b With the mirror uncovered. c Tonic immobility test



individual's breath rate from videos in standardized conditions in a cage, which, although not completely avoiding stress, avoids the acute stress of having the bird in the hand, and we optimized an algorithm to buffer measurements from observational error.

We quantified each individual's breath rates from the videos of the first 5 min of the mirror test, the period with the mirror covered (Fig. 3a). A single observer (CIM) watched the videos at half speed on a large screen with the software Observer XT 11, registering every noticeable breath movement as detected by continuously observing the belly movement. There were instances where respiratory movements were not observable, for example, because the bird was hopping, flying, or facing away, thus generating missing values. In rare instances, the keyboard key could also be pressed twice in a row, thus introducing some noise to the data in the form of duplicate breaths. The resulting dataset is a series of keystroke times, which we cleaned with an algorithm to correct for missing and duplicate breaths (algorithm explanation in supplementary material methods, R code in supplementary material code, and worked example in supplementary file S1 and supplementary Fig. S1). With the cleaned data, we computed mean breath rate across the entire 5 min of each video, its standard deviation, and the difference between the mean breath rate in the second half (seconds 151 to 300) and the first half of the videos (0 to 150 s; hereafter, change in breath rate). All three variables were approximately normally distributed, as evaluated by QQ-plots and histograms.

To evaluate repeatability of these three measurements of breath rate and breath rate variation, we randomly selected 20 videos, ensuring that multiple videos from the same individual were not selected, and repeated the entire process of scoring the videos and cleaning the data, blind to the identity of the individual birds. We calculated repeatability across the two measurements of the same videos with the "rptGaussian" function of the R (version 3.4.3; R Core Team 2017) package "rptR" (version 0.9.21; Stoffel et al. 2017) using video ID as random effect, and we checked with histograms and QQ-plots that the residuals of the model were approximately normal. We found high repeatability for mean breath rate ($R = 0.95$, $N = 20$, $P < 0.01$) and change in breath rate ($R = 0.93$, $N = 20$, $P < 0.01$). Repeatability for the standard deviation in breath rate was significant but only moderate in value ($R = 0.60$, $N = 20$, $P < 0.01$), and we therefore did not analyze standard deviation in breath rate further. The two remaining measures, mean breath rate and change in breath rate, were not correlated across the 60 individuals (Pearson correlation of mean values per individual: $r_p = 0.02$, $N = 60$, $P = 0.91$). Mean breath rate ranged from 2.43 to 4.64 Hz (mean \pm SD = 3.25 ± 0.46 Hz), and the change in breath rate ranged from -0.80 to 0.86 Hz (mean \pm SD = -0.236 ± 0.27 Hz).

Body size and size-corrected mass

We used body size and size-corrected mass as proxies for condition. We quantified body size as the first PC from a PCA on two skeletal size measurements: tarsus length (from the last complete scale before toes diverge to the notch at the intertarsal joint) and head plus bill length (length from the furthest point at the back of the skull to the tip of the bill). We used those skeletal measurements because they are more stable and reliably measured than mass or feather measurements. Measurements were made on December 12 and 13, 2016, and on September 29, 2017, for birds captured in 2016 and 2017, respectively (Fig. 1), with a digital caliper to the nearest 0.1 mm. PC1 explained 64% of the variance (eigenvalue = 1.29) and had high loadings (0.80) for both morphological measurements.

We used the weight of each bird at the time of capture and the size PC1, above, to compute size-corrected mass as the residual mass on size, which is advisable to render this measurement independent of body size (Jakob et al. 1996). Capture weight and body size PC1 were positively correlated ($r_p = 0.37$, $N = 60$, $P < 0.01$). We computed residuals (hereafter, size-corrected mass) from a linear model of body mass on body size PC1, with the "lm" function in the R package "stats." Size-corrected mass was highly correlated with capture weight ($r_p = 0.93$, $N = 60$, $P < 0.01$), and both measures were approximately normal, as observed in QQ-plots and histograms. Body size ranged from -2.64 to 1.97 (mean \pm SD = 0.00 ± 1.14), and size-corrected mass ranged from -0.74 to 0.76 g (mean \pm SD = 0.00 ± 0.38 g).

Statistical analyses

First, we assessed repeatability across the two rounds of tests per individual, separately for the mirror test, tonic immobility test, measurements of breath rate, and change in breath rate. We calculated repeatability with the "rptGaussian" function of the R package "rptR," as before, using round number as an independent factor and individual ID as random effect. Residuals of most models were approximately normal, as indicated by their histograms and QQ-plots. The exception was the model for the tonic immobility, which clearly violated the assumption of normality, and data cannot be transformed due to a floor and ceiling effect (values have minimum and maximum absolute values of 0 and 60 s) with peaks at 0 and 60. Therefore, we instead calculated consistency for tonic immobility using a non-parametric Spearman correlation, correlating the score of tonic immobility for the first and second rounds of each individual. In all following analyses, we used the mean value per individual across the two rounds for each of these behavioral phenotypes.

Prior to relating performance in the detour-reaching task to its predictor variables (behavior in the mirror and tonic

immobility tests, breath rate and its change, body size, size-corrected mass, and latency to feed in training phase), we checked if any of those variables were strongly correlated, which could cause multicollinearity problems in statistical models. We conducted Pearson correlation tests for all the pairwise combinations of these variables, except for those involving tonic immobility, which deviated strongly from normality and for which we used non-parametric Spearman correlations. We used Bonferroni corrections for these bivariate analyses to account for multiple testing and false positives. These analyses, and the PCA described earlier, were conducted in the R “stats” package.

Finally, to test which phenotypic traits were more strongly associated to cognitive performance, we used a model selection approach, fitting generalized linear models (GLM) and weighting them by their Akaike’s information criterion (corrected for small sample sizes; AICc, Hurvich and Tsai 1989). We ran GLM with either the performance in detour reaching (i.e., the proportion of correct trials in the detour-reaching task) or the mean order of correct trials as dependent variable. As independent variables, we used behavior in the personality assays (mirror and tonic immobility tests), and, to account for putative confounding effects, we used the remaining traits measured (breath rate and its change, body size, size-corrected mass, and latency to feed in training phase), and also sex and the year of capture. We did not include interaction terms in these models because they include many terms, and we have no a priori biological predictions for the ecological relevance of such interactions. We computed AICc for models with all possible combinations of independent variables, and, because models differing by less than 2 AICc are not considered different (Burnham and Anderson 2002), we used models within 2 AICc from the best for model averaging. Following Symonds and Moussalli (2011), we averaged those models weighted by their AICc weight values. We report AICc model-averaged standardized partial coefficients (β_{st}) for each predictor variable and the relative importance (RI) of each predictor variable (i.e., the probability of the predictor to be in the true best model, computed as the weighted sum of models in which that variable appears; Symonds and Moussalli 2011), the standard error (SE), the 95% confidence intervals of the model-averaged coefficients (95% CI), and also AICc model-averaged values of P for those coefficients (Garamszegi and Mundry 2014). Furthermore, we also report details for top models (i.e., those within 2 AICc from the best model). As a complement, we report β_{st} and the associated values of P from a single GLM, not model-averaged (hereafter, single model). For this, we made a GLM that only contains the important predictors, as indicated by their RI in the model selection, thus including only the predictors with a RI higher than

0.5. For single models, we confirmed that residuals approximated a normal distribution by inspecting QQ-plots and that variances were homogeneous. Model selection and model averaging were conducted with the “MuMIn” (version 1.40.0; Bartón 2017) R package. We used the function “dredge” for model selection, indicating that coefficient estimates should be standardized by the standard deviation (argument “beta” = “sd”). We used the function “model.avg” for model averaging, keeping, as referred above, the model with the lowest AICc and those differing from it by less than 2 AICc (argument “subset” = delta < 2). The final single GLM was conducted using the “glm” function from the R “stats” package. 95% confidence intervals were obtained, in both cases, using the function “confint” from the R “stats” package.

Data availability

All data are available in electronic supplementary material.

Results

Repeatability across the two rounds of the tests of the same individual waxbills (separated by ca. 6 weeks; Fig. 1) was high for the mirror test ($R = 0.72$, $P < 0.01$, 95% CI = [0.58; 0.83]) and mean breath rate ($R = 0.58$, $P < 0.01$, 95% CI = [0.39; 0.73]), and individual consistency in the duration of tonic immobility was moderate but still significant ($r_s = 0.37$, $P < 0.01$). The change in breath rate during the assay was weakly repeatable ($R = 0.20$, $P = 0.06$, 95% CI = [0; 0.435]); $N = 60$ individuals in all cases.

There were no strong pairwise correlations between behavior in the mirror test, tonic immobility test, mean breath rate, change in breath rate, body size, size-corrected mass, and latency to feed in the training phase (all $|r| \leq 0.27$, $N = 60$ birds, uncorrected $P \geq 0.03$; none were significant with Bonferroni correction for multiple comparisons; Table 1). Therefore, multicollinearity does not appear to be an issue for these data. Only one of those bivariate correlations was nominally significant, and this was the association between large body size and low mean breath rate ($r_p = -0.27$, uncorrected $P = 0.03$; Table 1).

The model selection approach to assess which variables were associated with detour-reaching performance retained only 4 predictors: behavior in the mirror test and capture year, both with high relative importance (RI; 0.88 and 0.81, respectively), and tonic immobility and size-corrected mass, both with lower RI (0.38 and 0.12, respectively; Table 2, see also Table S1 for the top models, i.e., those within 2 AICc from the best model). Both model-averaged results and the single GLM using predictors with

Table 1 Correlations between all measured predictors

	Mirror test		Tonic immobility test [†]		Mean breath rate		Change in breath rate		Body size		Size-corrected mass	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>
Tonic immobility test [†]	0.04	0.77										
Mean breath rate	0.15	0.27	-0.07	0.60								
Change in breath rate	-0.04	0.74	-0.17	0.18	0.02	0.91						
Body size	-0.06	0.63	0.04	0.76	-0.27	0.03*	0.04	0.74				
Size-corrected mass	-0.04	0.76	0.05	0.71	0.03	0.83	-0.02	0.87	< 0.01	1.00		
Latency to feed in training phase	-0.07	0.64	-0.04	0.78	-0.14	0.35	-0.20	0.16	0.05	0.75	-0.19	0.19

[†]Correlations involving this trait were non-parametric (Spearman correlations), due to its strong deviation from normality

*Significant values

Significance level with correction for multiple tests (Bonferroni correction) = 0.002

RI ≥ 0.5 showed a significant association with behavior in the mirror test (model-averaged $\beta_{st} = 0.29$, SE = 0.14, $P = 0.04$, 95% CI = [0.02; 0.57], and single $\beta_{st} = 0.29$, SE = 0.14, $P = 0.03$, 95% CI = [0.03; 0.56]; Table 2), with more active individuals performing better on the detour-reaching task (Fig. 4), and a non-significant association with year (model-averaged $\beta_{st} = 0.27$, SE = 0.15, $P = 0.07$, 95% CI = [-0.02; 0.55], and single $\beta_{st} = 0.48$, SE = 0.27, $P = 0.08$, 95% CI = [-0.05; > 0.99]; Table 2). Both of the variables with lower RI did not show a significant association with detour-reaching performance (tonic immobility: model-averaged $\beta_{st} = 0.21$, SE = 0.14, $P = 0.15$, 95% CI = [-0.07; 0.49]; size-corrected mass: model-averaged $\beta_{st} = -0.10$, SE = 0.14, $P = 0.46$, 95% CI = [-0.38; 0.17]).

For the mean order of correct trials (on average 0.53, across 50 birds with detour-reaching data), model selection retained 5 predictors: tonic immobility had high RI (0.79), and change in breathing rate, mirror test, and mean breath rate had lower RI (0.16, 0.16, and 0.13, respectively; Table 3, see also top models in Table S2). But none of these predictors were significantly related to mean order of correct trials (all $|\beta_{st}| \leq 0.25$, $P \geq 0.08$, and all 95% CI included zero; Table 3).

Table 2 Model average standardized coefficients (β_{st}), standard errors (SE), *P* values (*P*), and 95% confidence intervals of the model-averaged coefficients (95% CI) of the relation between performance in detour-

	Model averaging			Single model	
	β_{st} (SE; <i>P</i>)	95% CI	Relative importance	β_{st} (SE; <i>P</i>)	95% CI
Mirror test	0.29 (0.14; 0.04)*	(0.02; 0.57)	0.88	0.29 (0.14; 0.03)*	(0.03; 0.56)
Capture year	0.27 (0.14; 0.07)	(-0.02; 0.55)	0.81	0.48 (0.27; 0.08)	(-0.05; > 0.99)
Tonic immobility	0.21 (0.14; 0.15)	(-0.07; 0.49)	0.38		
Size-corrected mass	-0.10 (0.14; 0.46)	(-0.38; 0.17)	0.12		

*Significant values

Discussion

Cognition and proactive-reactive personalities

As in many other species (MacLean et al. 2014), the large majority of common waxbills completed our cognitive assay, the detour-reaching task (50 out of 60 birds), with performance ranging from ca. 50 to 100% correct trials. To relate this cognitive performance to personality differences, we used two behavioral assays: the mirror test assesses differences along a proactive-reactive axis of personality in waxbills, while the tonic immobility test assesses unrelated but consistent behavioral differences in waxbills (Carvalho et al. 2013; Funghi et al. 2015). We found that individual differences in the duration of tonic immobility were not related to performance in the detour-reaching task. The time that animals spend in tonic immobility is commonly interpreted as an index of fear or stress (e.g., Gallup 1979; Nakayama and Miyatake 2010; Edelaar et al. 2012; Pusch et al. 2018). Accepting this interpretation, our results suggest that differences in stress levels among individual waxbills do not explain their performance in the detour-reaching task. We also found that more proactive individuals made fewer mistakes in this cognitive

reaching task and all predictors, and relative importance (RI) of predictors. Also reported are results from a single GLM using all predictors with RI > 0.50

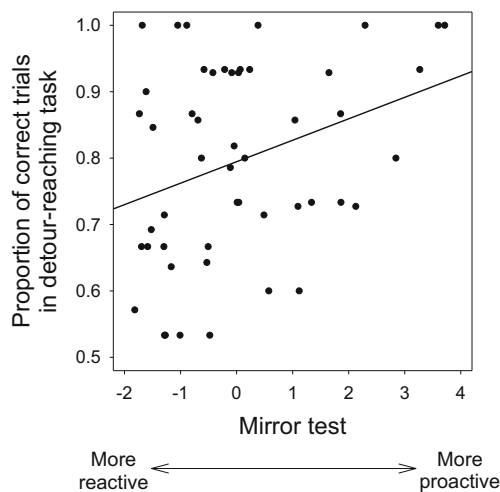


Fig. 4 Scatterplot of performance in the detour-reaching task (proportion of correct trials) as a function of scores in the mirror test. Represented are the non-standardized data from the proportion of correct trials in detour-reaching task and the mirror test, with the regression line from the bivariate relation between the two variables

task, and this is in apparent contradiction to the cognitive style hypothesis (CSH). The CSH predicts that reactive individuals (not proactive individuals) should perform better in tasks that require adjusting previously established behavioral routines, or inhibiting response to irrelevant stimuli in order to pursue a specific goal, such as the inhibitory control required by the detour-reaching task (reviewed in Hauser 1999).

Personality traits correlate with differences in cognitive performance in several species (reviewed in Locurto 2007; Carere and Locurto 2011; Sih and Del Giudice 2012; Griffin et al. 2015; Dougherty and Guillette 2018), but sometimes in the direction predicted by the CSH and, nearly as often, in the opposite direction (reviewed in Dougherty and Guillette 2018). We assayed waxbill personality using mirror tests, which in this species are repeatable over time (Carvalho et al. 2013; Funghi et al. 2015), even a year apart or in different seasons (Guerra et al. submitted). Waxbill responses to the mirror image indicate differences along a proactive-reactive behavioral syndrome, because individuals' mirror test responses correlate with their exploratory behavior in open-field tests (more exploratory individuals are less attentive to

the mirror image and vocalize or move more; less exploratory individuals are more attentive to the mirror image and move or vocalize less; Carvalho et al. 2013). Furthermore, differences along this proactive-reactive axis correlate with climatic stability across waxbill populations, suggesting that this is an ecologically-relevant aspect of personality (Carvalho et al. 2013). Thus, finding better detour-reaching performance in proactive waxbills appears contrary to the CSH. The accumulation of results similar to ours in other species (e.g., Titulaer et al. 2012 only for males; Trompf and Brown 2014 only for females; Chung et al. 2017) suggests that the CSH needs refinement (Dougherty and Guillette 2018).

It is generally accepted that the detour-reaching task measures inhibitory control (e.g., Vlamings et al. 2010; Boogert et al. 2011a; MacLean et al. 2014; Guillette et al. 2015; Shaw et al. 2015; Shaw 2017), defined as the ability to pursue cognitively represented goals while inhibiting responses to other stimuli (Hauser 1999; Guillette et al. 2015), and inhibitory control has in turn been argued to indicate more general problem-solving abilities (e.g., Diamond 1990; Dempster 1992; Hauser 1999; Carlson and Moses 2001; Hare et al. 2009; Guenther and Brust 2017; Medina-García et al. 2017). In support of this, and using brain size as a proxy for general cognitive abilities (e.g., Deaner et al. 2007), performance in the detour-reaching task correlates with relative brain size across various species (MacLean et al. 2014). However, several studies also found that inhibitory control does not correlate with performance in other cognitive tasks (Boogert et al. 2011a; Guillette et al. 2015; Shaw et al. 2015). It is therefore debatable whether or not performance on the detour-reaching task used here can indicate more general cognitive abilities (reviewed in Boogert et al. 2018; Völter et al. 2018). It is thus possible that future work on waxbills using different cognitive tasks may find associations with personality more in line with predictions of the CSH.

The CSH also predicts that proactive individuals should have more stereotyped and less flexible behavioral routines, and perhaps stereotyped routines are given an advantage in the detour-reaching task if, when tested with the transparent cylinder, stereotyped individuals simply keep performing the same routine that they learned earlier with

Table 3 Relation between mean order of correct trials in the detour-reaching task and all predictors. Statistics as in Table 2

	Model averaging			Single model	
	β_{st} (SE; <i>P</i>)	95% CI	Relative importance	β_{st} (SE; <i>P</i>)	95% CI
Tonic immobility	0.24 (0.14; 0.10)	(−0.05; 0.53)	0.58	0.22 (0.14; 0.13)	(−0.06; 0.50)
Change in breath rate	0.17 (0.14; 0.24)	(−0.12; 0.46)	0.31		
Body size	−0.15 (0.14; 0.31)	(−0.44; 0.14)	0.27		
Sex	−0.12 (0.14; 0.43)	(−0.40; 0.17)	0.08		
Latency to feed in training phase	−0.09 (0.15; 0.54)	(−0.39; 0.20)	0.07		

the opaque cylinder, unaware of the fact that the cylinder is no longer opaque. This alternative interpretation runs counter the generally accepted interpretation that the detour-reaching task measures inhibitory control. If the alternative interpretation were true, we would expect that proactive individuals behaved in a consistent manner throughout the test trials, while reactive individuals, having noticed the change from an opaque to a transparent cylinder, should have made mistakes mostly in the first few test trials and then gradually adjust their behavior over consecutive trials. To address this prediction, we computed the mean order of correct trials along the test phase and tested if reactive individuals had a higher mean order of correct trials than proactive individuals. However, we did not find evidence for this alternative interpretation of the detour-reaching task, as the mean order of correct trials was not related to personality type, thus supporting the generally accepted interpretation of the detour-reaching task as accessing inhibitory control.

Putative confounding factors

We evaluated associations between detour-reaching performance and personality while taking into account putative confounding factors (sex, size and size-corrected mass, breath rates, latency to feed in training phase, year, and site of capture), as advised by Dougherty and Guillette (2018). This is the most comprehensive attempt so far at controlling confounding factors. However, none of these traits were related to performance in the detour-reaching task. For example, sex does not appear to be a confounding factor in waxbills because, unlike in some other species where the sexes differ in aspects of learning or cognition (e.g., Range et al. 2006; Titulaer et al. 2012; Brust et al. 2013; Carazo et al. 2014; Mamuneas et al. 2014; Etheredge et al. 2018; Mazza et al. 2018), male and female waxbills did not differ in cognitive performance, and previous work reported that they do not differ in the personality traits we studied either (Carvalho et al. 2013). Another example of a trait often related to cognition is sociability (e.g., Lalot et al. 2017; Ashton et al. 2018; reviewed in Thornton and Clutton-Brock 2011; Sih and Del Giudice 2012; but see Trompf and Brown 2014; Chung et al. 2017). Our battery of phenotypes encompasses a proxy for social dominance (body size predicts dominance in common waxbills; Funghi et al. 2015), and this too was not related to cognitive performance.

One trait that we suggest could be important to control for when testing associations between personality type and cognitive performance, but that has not been accounted for in past studies, is individual condition. In several species, heavier or larger individuals are more proactive (meta-analysis in Niemelä and Dingemanse 2018). This suggests

a relation between condition and personality, to the extent that mass can be taken as indicating better current condition and that size can be taken as indicating better developmental conditions during the time when most skeletal growth took place. Also, a few studies found that larger individuals show better cognitive performance (Shaw 2017; van Horik et al. 2017; but see contrary examples in Cole et al. 2011; Thornton and Samson 2012). A mechanism by which condition could affect both personality and cognition is through its effect on brain development. The brain is particularly sensitive to developmental conditions (reviewed in Pravosudov 2009; Boogert et al. 2018), and aspects of brain development could mediate a correlation between cognition and personality. Brain lateralization, for example, has been associated with better cognitive multi-tasking (e.g., Rogers 2000; Rogers et al. 2004; Dadda and Bisazza 2006; Vallortigara and Versace 2017) and, in some species, bolder individuals are more strongly lateralized (Reddon and Hurd 2009; Goursot et al. 2018) or have higher neuron density in certain brain regions (Wiese et al. 2018). Therefore, if cognitive performance and proactive personality are both dependent on current condition or conditions during development, then differences among individuals in current or in developmental condition could drive positive correlations between these traits, even if, as posited by the CSH, all else being equal proactive behavior reduces performance in cognitive tasks involving inhibitory control. This hypothesis follows a classic reasoning in life-history theory whereby, even if two traits are predicted to trade-off with each other due to resource limitation, positive correlations between those traits can arise when individuals differ in the amount of resources they can access (van Noordwijk and de Jong 1986). For example, despite the prediction that proactive individuals should live less due to their more risk-prone behavior (Réale et al. 2010; Dammhahn et al. 2018), proactive individuals in the wild are found to live longer on average than reactive individuals (reviewed in Moiron et al. 2020), thus suggesting that proactive individuals are in some aspect of higher quality than reactive individuals.

To test if our finding of a positive correlation between proactive personality and detour-reaching performance, contrary to predicted by the CSH, could be mediated by individual differences in condition, we used size-corrected mass as a proxy for present condition and body size as a proxy for condition during development. But we did not find evidence for either of these traits being related to detour-reaching performance or to proactive personality. Condition and, especially, condition during development, are difficult to infer based on current phenotype (Barnett et al. 2015). Therefore, we suggest that effects of condition on the relation between cognition and personality should be further investigated, for example, by experimental manipulation of condition during development.

No relations between personality and breath rate or body size

We quantified breath rates from video recordings in standardized conditions, with the birds free to move normally. Although breath rates are easily disturbed, this approach allowed measuring mean breath rate and its change during the video recordings in a highly repeatable manner when testing the same individual ca. 6 weeks apart. This approach provides considerably more detail and standardization than typically achieved with birds in the hand (Kluen et al. 2014; Trnka et al. 2018), where measurements of breath rate are strongly affected by the birds' acute stress response (Calder 1968). Our approach should be particularly useful when, as here, researchers want to use breath rates as an indication of individual differences in metabolic rate (rather than individual differences in stress response), since basal metabolic rates have been reported to differ among personality types (Mathot et al. 2019).

We found, from the pairwise correlations, that larger body size was related to slower breath rates in waxbills, as has also been found across species (Calder 1968), and that differences along the waxbill proactive-reactive personality axis were not predicted by body size, size-corrected mass, or breath rates. In contrast, studies in other species found that personality types associated with fear or stress had higher breath rates (e.g., Carere et al. 2001; Carere and van Oers 2004; Fucikova et al. 2009). This would be expected if faster breath rates are indicative of fear (Carere and van Oers 2004) or stress (Fucikova et al. 2009) and, as per the pace of life hypothesis (Réale et al. 2010; Mathot and Frankenhuis 2018), more fearful individuals are less bold and less aggressive (e.g., Verbeek et al. 1996; Malmkvist and Hansen 2002; Carere and van Oers 2004). Similar to our results, however, some previous studies also did not find associations between breath rates or other proxies of stress and animal personality (David et al. 2012; Kluen et al. 2014).

Finally, we found that the proactive-reactive personality axis in waxbills (as evaluated by mirror and exploration tests) is unrelated to the tonic immobility test, which is often interpreted as indicative of fear. This confirms previous results (Carvalho et al. 2013; Funghi et al. 2015) showing that tonic immobility in waxbills quantifies a different aspect of personality, unrelated to their proactive-reactive behavioral syndrome. Since spending a longer time in tonic immobility is commonly taken as indicating fear or stress (e.g., Gallup 1979; Nakayama and Miyatake 2010; Edelaar et al. 2012; Pusch et al. 2018), these results suggest that the reactive or proactive personalities in waxbills may not be related to differences in fear or stress physiology.

Conclusion

The cognitive style hypothesis (CSH) predicts that reactive personality types should perform better in cognitive tasks that require adjusting previously learned behavior, or in tasks in which individuals need to inhibit responding to irrelevant stimuli, to pursue a specific goal. However, we found the opposite pattern of proactive waxbills performing better on an assay of inhibitory control, the detour-reaching task. While it may be possible to reconcile this result with the CSH, for example, if the correlation between cognition and proactive personality appears as a secondary consequence of both traits being strongly influenced by a third variable, we found no evidence for this. Together with other recent studies providing conflicting results, our findings suggest that the CSH might need further formal theoretical development and refinement.

Acknowledgments We thank Ricardo J. Lopes for guidance on molecular sexing and two anonymous reviewers for their helpful comments.

Funding information This work was funded by grant PTDC/BIA-EVF/4852/2014 and fellowships DL57/2016/CP1440/CT0011 to GCC and SFRH/BD/129002/2017 to ACRG, from Fundação para a Ciência e a Tecnologia. NJB is funded by a Royal Society Dorothy Hodgkin Research Fellowship.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Work on waxbills was done under permits 690/2016/CAPT and 57/2017/CAPT from the Instituto da Conservação da Natureza e das Florestas and approval from CIBIO/InBIO ethics committee (ORBEA reference 2017-01); therefore, all applicable international, national, and/or institutional guidelines for the care and use of animals were followed, and all procedures were in accordance with the ethical standards of the institution at which the studies were conducted.

References

- Ashton BJ, Ridley AR, Edwards EK, Thomson A (2018) Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature* 554:364–367
- Barnett CA, Suzuki TN, Sakaluk SK, Thompson CF (2015) Mass-based condition measures and their relationship with fitness: in what condition is condition? *J Zool* 296:1–5
- Bartón K (2017) MuMIn: multi-model inference. R package version 1.40.0. <https://cran.r-project.org/web/packages/MuMIn/index.html>
- Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S (2011a) Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Anim Behav* 81: 1209–1216
- Boogert NJ, Fawcett TW, Lefebvre L (2011b) Mate choice for cognitive traits: a review of the evidence in nonhuman vertebrates. *Behav Ecol* 22:447–459

- Boogert NJ, Madden JR, Morand-Ferron J, Thornton A (2018) Measuring and understanding individual differences in cognition. *Philos Trans R Soc Lond B Biol Sci* 373:20170280
- Brust V, Wuerz Y, Krüger O (2013) Behavioural flexibility and personality in zebra finches. *Ethology* 119:559–569
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference. Springer Science & Business Media, New York
- Calder WA (1968) Respiratory and heart rates of birds at rest. *Condor* 70:358–365
- Carazo P, Noble DW, Chandrasoma D, Whiting MJ (2014) Sex and boldness explain individual differences in spatial learning in a lizard. *Proc R Soc B* 281:20133275
- Cardoso GC, Leitão AV, Funghi C, Batalha HR, Lopes RJ, Mota PG (2014) Similar preferences for ornamentation in opposite- and same-sex choice experiments. *J Evol Biol* 27:2798–2806
- Carere C, Locurto C (2011) Interaction between animal personality and animal cognition. *Curr Zool* 57:491–498
- Carere C, van Oers K (2004) Shy and bold great tits (*Parus major*): body temperature and breath rate in response to handling stress. *Physiol Behav* 82:905–912
- Carere C, Welink D, Drent PJ, Koolhaas JM, Groothuis TGG (2001) Effect of social defeat in a territorial bird (*Parus major*) selected for different coping styles. *Physiol Behav* 73:427–433
- Carere C, Caramaschi D, Fawcett TW (2010) Covariation between personalities and individual differences in coping with stress: converging evidence and hypotheses. *Curr Zool* 56:728–740
- Carlson SM, Moses LJ (2001) Individual differences in inhibitory control and children's theory of mind. *Child Dev* 72:1032–1053
- Carvalho CF, Leitao AV, Funghi C, Batalha HR, Reis S, Mota PG, Lopes RJ, Cardoso GC (2013) Personality traits are related to ecology across a biological invasion. *Behav Ecol* 24:1081–1091
- Cattelan S, Lucon-Xiccato T, Pilastro A, Griggio M (2017) Is the mirror test a valid measure of fish sociability? *Anim Behav* 127:109–116
- Chittka L, Niven J (2009) Are bigger brains better? *Curr Biol* 19:R995–R1008
- Chung M, Goulet CT, Michelangeli M, Melki-Wegner B, Wong BBM, Chapple DG (2017) Does personality influence learning? A case study in an invasive lizard. *Oecologia* 185:641–651
- Cockrem JF (2007) Stress, corticosterone responses and avian personalities. *J Ornithol* 148:169–178
- Cockrem JF (2013) Corticosterone responses and personality in birds: individual variation and the ability to cope with environmental changes due to climate change. *Gen Comp Endocr* 190:156–163
- Cole EF, Cram DL, Quinn JL (2011) Individual variation in spontaneous problem-solving performance among wild great tits. *Anim Behav* 81:491–498
- Coppens CM, de Boer SF, Koolhaas JM (2010) Coping styles and behavioural flexibility: towards underlying mechanisms. *Philos Trans R Soc Lond B Biol Sci* 365:4021–4028
- Dadda M, Bisazza A (2006) Does brain asymmetry allow efficient performance of simultaneous tasks? *Anim Behav* 72:523–529
- Dall SRX, Houston AI, McNamara JM (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol Lett* 7:734–739
- Dammhahn M, Dingemanse NJ, Niemelä PT, Réale D (2018) Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history. *Behav Ecol Sociobiol* 72:62
- David M, Auclair Y, Dechaume-Moncharmont FX, Cezilly F (2012) Handling stress does not reflect personality in female zebra finches (*Taeniopygia guttata*). *J Comp Psychol* 126:10–14
- Deaner RO, Isler K, Burkart J, van Schaik C (2007) Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain Behav Evol* 70:115–124
- Dempster FN (1992) The rise and fall of the inhibitory mechanism: toward a unified theory of cognitive development and aging. *Dev Rev* 12:45–75
- Diamond A (1990) Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Ann N Y Acad Sci* 608:637–676
- Dougherty LR, Guillette LM (2018) Linking personality and cognition: a meta-analysis. *Philos Trans R Soc Lond B Biol Sci* 373:20170282
- Dukas R (2008) Evolutionary biology of insect learning. *Annu Rev Entomol* 53:145–160
- Dukas R, Ratcliffe JM (2009) Cognitive ecology II. University of Chicago Press, Chicago
- Edelaar P, Serrano D, Carrete M, Blas J, Potti J, Tella JL (2012) Tonic immobility is a measure of boldness toward predators: an application of Bayesian structural equation modeling. *Behav Ecol* 23:619–626
- Etheredge RI, Avenas C, Armstrong MJ, Cummings ME (2018) Sex-specific cognitive-behavioural profiles emerging from individual variation in numerosity discrimination in *Gambusia affinis*. *Anim Cogn* 21:37–53
- Fucikova E, Drent PJ, Smits N, van Oers K (2009) Handling stress as a measurement of personality in great tit nestlings (*Parus major*). *Ethology* 115:366–374
- Funghi C, Leitão AV, Ferreira AC, Mota PG, Cardoso GC (2015) Social dominance in a gregarious bird is related to body size but not to standard personality assays. *Ethology* 121:84–93
- Gallup GG (1979) Tonic immobility as a measure of fear in domestic fowl. *Anim Behav* 27:316–317
- Garamszegi LZ, Mundry R (2014) Multimodel-inference in comparative analyses. In: Garamszegi LZ (ed) Modern phylogenetic comparative methods and their application in evolutionary biology. Springer-Verlag, Berlin, pp 305–331
- Garamszegi LZ, Markó G, Herczeg G (2013) A meta-analysis of correlated behaviors with implications for behavioral syndromes: relationships between particular behavioral traits. *Behav Ecol* 24:1068–1080
- Goursot C, Düpjan S, Kanitz E, Tuchscherer A, Puppe B, Leliveld LMC, Pruitt J (2018) Assessing animal individuality: links between personality and laterality in pigs. *Curr Zool* 65:541–551
- Griffin AS, Guillette LM, Healy SD (2015) Cognition and personality: an analysis of an emerging field. *Trends Ecol Evol* 30:207–214
- Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex most birds. *Mol Ecol* 7:1071–1075
- Groothuis TG, Carere C (2005) Avian personalities: characterization and epigenesis. *Neurosci Biobehav Rev* 29:137–150
- Guenther A, Brust V (2017) Individual consistency in multiple cognitive performance: behavioural versus cognitive syndromes. *Anim Behav* 130:119–131
- Guenther A, Brust V, Dersen M, Trillmich F (2014) Learning and personality types are related in cavies (*Cavia aperea*). *J Comp Psychol* 128:74–81
- Guillette LM, Reddon AR, Hurd PL, Sturdy CB (2009) Exploration of a novel space is associated with individual differences in learning speed in black-capped chickadees, *Poecile atricapillus*. *Behav Process* 82:265–270
- Guillette LM, Hahn AH, Hoeschele M, Przyluski AM, Sturdy CB (2015) Individual differences in learning speed, performance accuracy and exploratory behaviour in black-capped chickadees. *Anim Cogn* 18:165–178
- Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:646–648
- Hauser MD (1999) Perseveration, inhibition and the prefrontal cortex: a new look. *Curr Opin Neurobiol* 9:214–222
- Healy SD, Rowe C (2007) A critique of comparative studies of brain size. *Proc R Soc Lond B* 274:453–464
- Hurvich C, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76:297–293

- Jakob EM, Marshall SD, Uetz GW (1996) Estimating fitness: a comparison of body condition indices. *Oikos* 77:61–67
- Kluen E, Siitari H, Brommer JE (2014) Testing for between individual correlations of personality and physiological traits in a wild bird. *Behav Ecol Sociobiol* 68:205–213
- Koolhaas JM, Korte SM, de Boer SF, van der Vegt BJ, van Reenen CG, Hopster H, de Jong IC, Ruis MAW, Blokhuis HJ (1999) Coping styles in animals: current status in behavior and stress-physiology. *Neurosci Biobehav Rev* 23:925–935
- Lalot M, Ung D, Péron F, d’Ettorre P, Bovet D (2017) You know what? I’m happy. Cognitive bias is not related to personality but is induced by pair-housing in canaries (*Serinus canaria*). *Behav Process* 134:70–77
- Locurto C (2007) Individual differences and animal personality. *Comp Cogn Behav Rev* 2:67–78
- Lucon-Xiccato T, Bisazza A (2017) Individual differences in cognition among teleost fishes. *Behav Process* 141:184–195
- MacLean EL, Hare B, Nunn CL et al (2014) The evolution of self-control. *Proc Natl Acad Sci U S A* 111:E2140–E2148
- Malmkvist J, Hansen SW (2002) Generalization of fear in farm mink, *Mustela vison*, genetically selected for behaviour towards humans. *Anim Behav* 64:487–501
- Mamuneas D, Spence AJ, Manica A, King AJ (2014) Bolder stickleback fish make faster decisions, but they are not less accurate. *Behav Ecol* 26:91–96
- Mathot KJ, Frankenhuys WE (2018) Models of pace-of-life syndromes (POLS): a systematic review. *Behav Ecol Sociobiol* 72:41
- Mathot KJ, Dingemanse NJ, Nakagawa S (2019) The covariance between metabolic rate and behaviour varies across behaviours and thermal types: meta-analytic insights. *Biol Rev* 94:1056–1074
- Mazza V, Eccard JA, Zaccaroni M, Jacob J, Dammhahn M (2018) The fast and the flexible: cognitive style drives individual variation in cognition in a small mammal. *Anim Behav* 137:119–132
- Medina-García A, Jawor JM, Wright TF, Pruitt J (2017) Cognition, personality, and stress in budgerigars, *Melopsittacus undulatus*. *Behav Ecol* 28:1504–1516
- Moiron M, Laskowski KL, Niemelä PT (2020) Individual differences in behaviour explain variation in survival: a meta-analysis. *Ecol Lett* 23:399–408
- Nakayama S, Miyatake T (2010) Genetic trade-off between abilities to avoid attack and to mate: a cost of tonic immobility. *Biol Lett* 6:18–20
- Niemelä PT, Dingemanse NJ (2018) Meta-analysis reveals weak associations between intrinsic state and personality. *Proc R Soc B* 285:20172823
- Niemelä PT, Vainikka A, Forsman JT, Loukola OJ, Kortet R (2013) How does variation in the environment and individual cognition explain the existence of consistent behavioral differences? *Ecol Evol* 3:457–464
- Pavlov IP (1906) The scientific investigation of the psychical faculties or processes in the higher animals. *Science* 24:613–619
- Pavlov IP (1941) Lectures on conditioned reflexes, vol. 2. Conditioned reflexes and psychiatry. Lawrence & Wishart, London
- Pravosudov VV (2009) Development of spatial memory and the hippocampus under nutritional stress: adaptive priorities or developmental constraints in brain development? In: Dukas R, Ratcliffe JM (eds) *Cognitive ecology II*. The University of Chicago Press, Chicago, pp 88–110
- Pusch EA, Bentz AB, Becker DJ, Navara KJ (2018) Behavioral phenotype predicts physiological responses to chronic stress in proactive and reactive birds. *Gen Comp Endocrinol* 255:71–77
- R Core Team (2017) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna. <http://www.R-project.org>
- Range F, Bugnyar T, Schlogl C, Kotrschal K (2006) Individual and sex differences in learning abilities of ravens. *Behav Process* 73:100–106
- Reader S (2003) Innovation and social learning: individual variation and brain evolution. *Anim Biol* 53:147–158
- Réale D, Dingemanse NJ, Kazem AJ, Wright J (2010) Evolutionary and ecological approaches to the study of personality. *Philos Trans R Soc Lond B Biol Sci* 365:3937–3946
- Reddon AR, Hurd PL (2009) Individual differences in cerebral lateralization are associated with shy–bold variation in the convict cichlid. *Anim Behav* 77:189–193
- Rogers LJ (2000) Evolution of hemispheric specialization: advantages and disadvantages. *Brain Lang* 73:236–253
- Rogers LJ, Zucca P, Vallortigara G (2004) Advantages of having a lateralized brain. *Proc R Soc Lond B* 271:S420–S422
- Ruiz-Gomez ML, Huntingford FA, Overli O, Thornqvist PO, Hoglund E (2011) Response to environmental change in rainbow trout selected for divergent stress coping styles. *Physiol Behav* 102:317–322
- Shaw RC (2017) Testing cognition in the wild: factors affecting performance and individual consistency in two measures of avian cognition. *Behav Process* 134:31–36
- Shaw RC, Boogert NJ, Clayton NS, Burns KC (2015) Wild psychometrics: evidence for ‘general’ cognitive performance in wild New Zealand robins, *Petroica longipes*. *Anim Behav* 109:101–111
- Shettleworth SJ (2010) Cognition, evolution, and behavior. Oxford University Press, Oxford
- Sih A, Bell AM (2008) Insights for behavioral ecology from behavioral syndromes. *Adv Study Behav* 38:227–281
- Sih A, Del Giudice M (2012) Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philos Trans R Soc Lond B Biol Sci* 367:2762–2772
- Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol Evol* 19:372–378
- Stamps J, Groothuis TG (2010) The development of animal personality: relevance, concepts and perspectives. *Biol Rev* 85:301–325
- Stoffel MA, Nakagawa S, Schielzeth H, Goslee S (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* 8:1639–1644
- Symonds MRE, Moussalli A (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion. *Behav Ecol Sociobiol* 65:13–21
- Thornton A, Clutton-Brock T (2011) Social learning and the development of individual and group behaviour in mammal societies. *Philos Trans R Soc Lond B Biol Sci* 366:978–987
- Thornton A, Lukas D (2012) Individual variation in cognitive performance: developmental and evolutionary perspectives. *Philos Trans R Soc Lond B Biol Sci* 367:2773–2783
- Thornton A, Samson J (2012) Innovative problem solving in wild meerkats. *Anim Behav* 83:1459–1468
- Titulaer M, van Oers K, Naguib M (2012) Personality affects learning performance in difficult tasks in a sex-dependent way. *Anim Behav* 83:723–730
- Trnka A, Samas P, Grim T (2018) Stability of a behavioural syndrome vs. plasticity in individual behaviours over the breeding cycle: ultimate and proximate explanations. *Behav Process* 153:100–106
- Trompf L, Brown C (2014) Personality affects learning and trade-offs between private and social information in guppies, *Poecilia reticulata*. *Anim Behav* 88:99–106
- Vallortigara G, Versace E (2017) Laterality at the neural, cognitive, and behavioral levels. In: Call J (ed) *APA handbook of comparative psychology*, vol. 1: Basic concepts, methods, neural substrate, and behavior. American Psychological Association, Washington DC, pp 557–577

- van Horik JO, Langley EJ, Whiteside MA, Madden JR (2017) Differential participation in cognitive tests is driven by personality, sex, body condition and experience. *Behav Process* 134:22–30
- van Horik JO, Langley EJG, Whiteside MA, Laker PR, Beardsworth CE, Madden JR (2018) Do detour tasks provide accurate assays of inhibitory control?. *Proceedings of the Royal Society B: Biological Sciences* 285(1875):20180150
- van Noordwijk AJ, de Jong G (1986) Acquisition and allocation of resources: their influence on variation in life history tactics. *Am Nat* 128:137–142
- Verbeek MEM, Drent PJ, Wiepkema PR (1994) Consistent individual differences in early exploratory behaviour of male great tits. *Anim Behav* 48:1113–1121
- Verbeek MEM, Boon A, Drent PJ (1996) Exploration, aggressive behaviour and dominance in pair-wise confrontations of juvenile male great tits. *Behaviour* 133:945–963
- Vlamings PH, Hare B, Call J (2010) Reaching around barriers: the performance of the great apes and 3-5-year-old children. *Anim Cogn* 13:273–285
- Völter CJ, Tinklenberg B, Call J, Seed AM (2018) Comparative psychometrics: establishing what differs is central to understanding what evolves. *Philos Trans R Soc Lond B Biol Sci* 373:20170283
- Wiese AS, Needham EK, Noer CL, Balsby TJS, Dabelsteen T, Pakkenberg B (2018) The number of neurons in specific amygdala regions is associated with boldness in mink: a study in animal personality. *Brain Struct Funct* 223:1989–1998
- Wilson DS, Coleman K, Clark AB, Biederman L (1993) Shy-bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*): an ecological study of a psychological trait. *J Comp Psychol* 107:250–260
- Wilson DS, Clark AB, Coleman K, Dearstyne T (1994) Shyness and boldness in humans and other animals. *Trends Ecol Evol* 9:442–446

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.