

Original Article

Tonic immobility is a measure of boldness toward predators: an application of Bayesian structural equation modeling

Pim Edelaar,^a David Serrano,^a Martina Carrete,^b Julio Blas,^a Jaime Potti,^a and José Luis Tella^a

^aDepartment of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Av. Américo Vespucio s/n, 41092, Spain and ^bDepartment of Physical, Chemical and Natural Systems, University Pablo de Olavide, Ctra. de Utrera, km. 1 41013, Spain

Animals attacked by predators often enter a state of tonic immobility (TI) in which individuals appear to simulate death. Despite the fact that TI is often used as a proxy of fear in domesticated animals, quantitative data on individual variation is very scarce for wild vertebrates. As a consequence, we lack ecological interpretations for the variability in TI that may exist in wild populations. Here, we tested whether there are consistent differences among individuals in 2 components of TI within wild populations of 2 avian species, the Yellow-crowned bishop (*Euplectes afer*) and the Tree sparrow (*Passer montanus*). We next tested whether this variation reflects variation in boldness toward predators (measured as the response to 2 predator models) or is simply related to variation in general activity/restlessness (measured as baseline activity) in the bishop. We analyzed our data by means of Bayesian structural equation modeling (SEM), which has several general advantages and, moreover, allowed us to analyze censored (truncated) data. We found good support for relatively high repeatability within individuals of both components of TI. Measures of TI appeared to be uncorrelated with baseline activity. In contrast, our results suggest that individual variation in TI in a wild vertebrate can be interpreted in a context of boldness toward predators, making TI a meaningful and practical behavioral trait for studies involving personality and antipredation behavior in wild populations. In addition, we show that the Bayesian structural equation modeling approach to analyze censored data had greater statistical power than other approaches. Hence, this rarely implemented technique deserves to be more widely used. *Key words:* activity, animal personality, antipredation behavior, Bayesian structural equation modeling, boldness, fear, repeatability, tonic immobility. [*Behav Ecol* 23:619–626 (2012)]

INTRODUCTION

Tonic immobility (TI) is a behavioral state characterized by lack of movements and an apparent lifeless position (Gallup and Rager 1996; Miyatake et al. 2009). Alternative labels for this behavior include death feigning, thanatosis, immobility reflex, contact defense immobility, righting time, catatonia, playing possum, playing dead, and even animal hypnosis. TI has been recorded in a great variety of taxa, such as insects, decapods, spiders, fish, amphibians, reptiles, birds, and mammals (Gallup and Rager 1996). A number of hypotheses have been forwarded to explain TI (Ruxton 2006; Miyatake et al. 2009), some of which are specific to certain taxa such as insects that take a rigid position in order to prevent being swallowed whole by predators that do not chew their prey (Ruxton 2006). In most vertebrates, TI seems to be best explained by the hypothesis of “loss of predator’s interest.” Under that hypothesis, TI is a last resort to prevent death by predation, after freezing (remaining still in a normal posture to avoid detection), escape, or fighting has failed. In support of this functional explanation for TI, Sargeant

and Eberhardt (1975) showed that dozens of ducks attacked and grabbed by foxes all went into TI. Subsequently, they recorded frequently that foxes dropped or cached the living ducks while in TI, after which many ducks managed to escape. Similar support comes from Thompson et al. (1981). They exposed quail to cats and recorded that quail previously induced to TI had lower rates of attack than quail that were not in TI, and that spontaneous TI reduced the time of handling by the predator. Presumably, the lack of movement makes predators believe they successfully killed their prey after which they stop attacking it, such that TI increases prey survival (Thompson et al. 1981).

Parallel studies have interpreted and used TI as a general measure of fear, and the evidence for such an interpretation is particularly good for chickens (Gallup 1974, 1977; Boissy 1995; Forkman et al. 2007). In addition, TI has been shown to be heritable (Gallup 1974; Forkman et al. 2007; Ohno and Miyatake 2007; Nakayama et al. 2010) and correlate with a number of other behavioral and physiological variables (Gallup 1974; Boissy 1995; Jones 1996; Forkman et al. 2007; Miyatake et al. 2008; Zulkifli et al. 2009). As such, it has been interpreted as a component of animal personality, where individuals that are easily induced into TI and that stay longer in that state are seen as reactive individuals, while those that are not easily induced and stay shorter in TI are seen as proactive individuals (Boissy 1995; Jones 1996; Erhard et al. 1999; Cockrem 2007). However, the generality of much of this has yet to be established. For example, TI has been predominantly investigated

Address correspondence to Pim Edelaar, Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Avenida Américo Vespucio s/n, 41092 Sevilla, Spain. E-mail: edelaar@ebd.csic.es.

Received 28 November 2011; revised 28 December 2011; accepted 28 December 2011.

in domesticated chicken and quail and, as far as we know, there are no estimates of repeatability of components of TI for any wild vertebrate. There are clear limitations in extrapolating experimental paradigms between domesticated and wild species, since differences of ecological characteristics and motivations have been identified (Forkman et al. 2007). Here, we aim to determine if individual variation in TI is repeatable in wild vertebrates using 2 avian species as study models. In addition, we aim to test whether TI and the perceived risk of predation are correlated at the *individual* level, which would support its use as a measure of boldness toward predators and as a component of animal personality (Reale et al. 2007). To our knowledge, such a correlation at the individual level has never been studied in wild vertebrates. Typically, the link between TI and fear is investigated by comparing TI among groups exposed to different experimental treatments that are related to fear (e.g., aversive conditioning or administration of tranquilizers—see Gallup 1977; Forkman et al. 2007). Instead, we ask the question whether the individual variation in behavior during exposure to a model of a predator correlates with individual variation in TI measures. However, since the response to a predator model is often expressed as movement rate, and inherently more restless individuals might also stay shorter in TI, this could also explain any encountered correlation (Forkman et al. 2007; Miyatake et al. 2008; Nakayama et al. 2010). We therefore also tested this alternative hypothesis and measured baseline activity prior to the exposure to a predator and determined to what extent baseline activity and response to a predator model are correlated with TI.

TI is commonly induced in vertebrates by manually restraining or covering the individual for a short time and exerting light pressure on its body, which supposedly mimics the grip of a predator. TI is then scored using several component variables, such as the number of inductions that are necessary to attain TI, the time to first head movement, the rate of head movements between first head movement and the end of TI, and the total time of TI (Gallup and Rager 1996). Often these measures are subsequently combined into a single one to facilitate (univariate) statistical analysis. However, there is little insight on how these components should be combined (normally simple additive rules are used), if they are each repeatable, and if they measure the same underlying characteristics of the individual. We therefore tested for correlation among 2 components of TI (“number of inductions to attain TI” and “duration of TI”), and analyzed their separate relationships with the response to a predator and baseline activity in order to determine if the patterns in the data were replicated for the 2 TI measures.

The concept of consistent individual behavioral differences across contexts or types of behaviors (i.e., personalities) is now firmly implemented in behavioral ecology and related fields (Reale et al. 2007, 2010). There are several statistical methodologies available for the uncovering of personalities in a given system, such as examining bivariate correlations or principal components of the measured behaviors. One method that is particularly powerful and versatile for the study of personalities is structural equation modeling (SEM; Grace 2006). The application of SEM has a long and rich history in studies of social science, psychology, and human behavior (Arhonditsis et al. 2006; Grace 2006). In contrast, it has been hardly applied in behavioral studies of wild animals, despite its many advantages (Dochtermann and Jenkins 2007; Dingemanse, Dochtermann, et al. 2010). In this study, we apply the statistical framework of SEM to our investigation of the ecological relevance and role of TI in personality. More specifically, we applied a Bayesian version of SEM because this allows for the analysis of censored (truncated) data, in our case due to a fixed-length protocol to measure the duration of TI.

MATERIALS AND METHODS

Study species, capture and housing

We studied the individual consistency of 2 components of TI and the correlation among these in 2 wild avian species. The Yellow-crowned bishop (*Euplectes afer*) is native to sub-Saharan Africa but invasive and locally common in Spain, while the Tree sparrow (*Passer montanus*) is a common native species. Birds were caught with mist nets in the surroundings of Doñana National Park (SE Spain) in January 2010 during the nonbreeding season and transported within a few hours after capture to the lab, using transport cages with food available ($n = 25$ for Yellow-crowned bishop, $n = 39$ for Tree sparrow; all mature birds of unknown sex). Birds were kept in captivity under permit SGYB/FOA/AFR from the Consejería de Medio Ambiente, Junta de Andalucía. They were initially housed in a communal outdoor aviary of 4 m³ (each species separately), and after 2 days of acclimatization were transferred to identical individual cages (35 × 35 × 40 cm) within the same room where they stayed for the duration of the experiments under natural light and temperature regimes (no significant effects of cage position on the behavioral measures: not shown). These cages were fitted with a feeding station (ad libitum standard tropical finch seed mixture), a drinking station and 2 perches. The spatial configuration of these items was identical in each cage. Tree sparrows were released 5 weeks later at the catching site, well before the next breeding season. Yellow-crowned bishops remained in captivity for follow-up studies, including captive breeding, and because releasing non-native species is not allowed in Spain.

Tonic immobility

After 2 weeks of acclimatization to captivity, we measured components of TI in the following way. Birds were taken from their cage and, as part of another study, at 0 min and 30 min body temperature and breathing rate were scored; in between, they were kept in a cloth bag. After this, each individual was tested for TI in a separate room in order to avoid disturbance by sound or movement. TI was measured in a small wire cage, whose walls were lined with thin black rubber sheeting to prevent visibility side ways. TI was induced by placing the individual on its back, fully covering the bird with one hand, and exerting light pressure to the breast area (this supposedly mimics immobilization by a predator). After 15 s, the hand was slowly removed and the door of the cage carefully closed. If a bird righted itself within 5 s, the bird was recaptured and the procedure was repeated, up to a maximum of 5 times (“number of inductions to attain TI”). If a bird stayed on its back >5 s, we measured the time until the bird righted itself (“duration of TI”). The session was terminated if a bird was still in TI after 10 m. All birds recovered activity instantaneously without any apparent lasting effects. During all the experiments, the same person (P.E.) stared the bird in the eye from a distance of about 60 cm as this promotes TI (Gallup 1977; Jones 1980), presumably because the stare resembles that from a true (vertebrate) predator. As is standard practice in TI assessment in higher vertebrates (Forkman et al. 2007), we used a human observer instead of some other live predator to induce TI because it increases standardization and, this way, we could observe directly how the birds behaved during TI and assess without error when TI was terminated. Observer facial, eye, and other movements were minimized, since movements or sounds can terminate TI (Gallup 1977). TI was assessed twice, with 7 days in-between each assessment. This allowed us to test for the individual consistency of each of the TI components. Tests were performed between 9:30 and 16:00 hours (i.e., several hours after sunrise and several hours before sunset).

Response to predators

Eighteen weeks later, the response to a predator model was determined for the Yellow-crowned bishops. We used a taxidermic owl (Little owl *Athene noctua*) and a life-like sculpted and painted model of a falcon (Hobby *Falco subbuteo*). Both are common avian predators in the study area, which frequently take small passerines the size of our bishops. The models were fixed on top of an extendable pole, which could be adjusted such that it was exactly in front of each cage, at a distance of 30 cm. To avoid unnecessary stress and habituation to predator exposure, we covered the front of all cages with thin wooden panels 2 days prior to the exposure (day light still came in from the top). Next, for a focal individual, we removed the covering panel and recorded baseline activity as the number of movements (hops or flights within or between perches) for 1 min (“baseline activity”). This was done from behind a permanent screen at 70 cm from the cage, observing the birds through a hole covered with black meshing. All birds appeared to feel comfortable and undisturbed when watched from behind this screen; many would sit quietly or preen for periods of time, apparently unaware of our presence. After recording baseline activity, we then placed the predator model in front of the cage, and again recorded the number of movements for 1 min (“activity with predator”). After this, the panel was replaced and the same sequence was repeated for the next individual. When exposed to the predator models, individuals mostly greatly increased their rates of movement, clearly moved away from the front of the cages and appeared to be looking for escape possibilities at the backside of the cage, often frantically so or were periodically hiding underneath their drinking stations. Even if they actually did not recognize our models as predators, we interpret this behavior to mean that they indeed were afraid of our predator models (see also Gallup 1977; Feenders and Bateson 2011). We did not detect any short- or long-term change in the normal appearance and behavior of the study subjects after our experiments were finished. The 2 different models were tested with 6 days in-between (first falcon, then owl), and these 2 predator exposures were treated as independent replicate exposures to a predator. Tests were performed between 9:30 and 14:30 hours.

Bayesian SEM of censored data

We analyzed the individual consistency of the components of TI, and their relationships with the response to predators and baseline activity, by means of SEM. SEM is a combination of regression, factor analysis, and path analysis (Grace 2006), thereby SEM allows one to model and test a degree of complexity that cannot be achieved using traditional approaches. In general, SEM can handle variables that are dependent variable with respect to one variable but that are predictor variable with respect to another (as in a chain of events), can analyze all data in 1 model which reduces type I and II error, can be used to model underlying (unmeasured but hypothesized) latent variables, can incorporate and even suggest directional (supposedly causal) effects, and can include direct and indirect effects (mediated through correlations with other variables), which also helps to reveal hidden covariance patterns (e.g. a negative correlation between 2 variables that appeared to be positive because of positive correlations with other confounding variables). On the other hand, SEM performs just as well for simple models such as multiple regression and gives similar results. In fact, even for simple statistical models, SEM might be preferred because it does not assume that predictor variables are measured without error, which is an assumption (of doubtful validity) in traditional generalized linear models (GLMs).

Here, we preferred SEM over traditional GLM approaches because it is not a priori clear which variables are dependent

(responses) and which are independent (predictors), and stochasticity/measurement error must be assumed in all measured variables. Because we obtained a number of censored data, we applied a Bayesian approach to SEM. Censored data is truncated data that has been limited in their values because of constraints to the temporal or spatial range of the study. Censored data are frequent in both experimental and observational studies, but their inclusion in traditional GLM or SEM is problematic. Omitting all cases with censored values leads to a bias in the data set and the results since the most extreme individuals are deleted. In addition, statistical power is lost. Especially in behavioral studies of animals, where sample sizes are often below 25 individuals (Taborsky 2010), this is highly undesirable. Alternatively, simply using the minimum or maximum values for the censored values underestimates the variance among individuals, also leading to biased estimates, inflated repeatability estimates, and violation of distributional (e.g., normality) assumptions of the data. Some methodology has been developed that allows for the proper use of censored data, such as data imputation or survival analysis (see Quinn and Keough 2002). The option chosen here is a Bayesian analysis that implements an estimation of the posterior probability distribution of each censored data point as part of the fitting procedure, using the information contained in other correlated variables. Hence, likely values for censored data were estimated based on the fitted model (and its credibility), and variability in the censored data was iteratively incorporated during the model fitting (Arbuckle). A second advantage is that Bayesian estimation in general performs better with smaller samples and deviations from normality, and avoids the large-sample assumptions of maximum likelihood used in many other statistical estimation techniques (Grace 2006). Finally, the Bayesian philosophy avoids problems related to traditional null hypothesis significance testing, such as the testing of nonsensical null hypotheses, the use of a fixed, arbitrary, and dichotomous significance criterion, unequal type I and type II errors, and the acceptance of unrejected null hypotheses (reviewed in Quinn and Keough 2002; McCarthy 2007).

For our SEM analyses, we used AMOS 19.0 (Arbuckle). While other, including free, programs may well be able to do the same kind of analyses, we provide some background as how to use this menu-driven program for Bayesian SEM of censored data (see also Arbuckle and Grace 2006). We first have to recode the censored data in our data base (e.g., Excel) to indicate which data is censored, and in which direction. For example, those individuals that remained in TI until the end of the observation period of 10 min received a score of “>10.00” for duration of TI, but censored data of the type “<X” (the unknown value is smaller than X) or “X < < Y” (the unknown values lies between X and Y) can also be analyzed. Next, the data sheet is imported after checking the option “Allow non-numeric data.” To analyze a certain model in AMOS one has to draw this model, like ours in Figure 1, which is assisted by a graphical interface. From the list of variables of the imported data file, one drags the variables of interest into the corresponding boxes of the drawn model. Generally, in any SEM application, one has to identify (fix) certain parameters as known if the model needs to fit more information than there is available (i.e., when it is “unidentified”). The final steps are the fitting of the model by clicking the corresponding icon, either traditionally or by the Bayesian approach, and interpretation of the output. For our Bayesian implementation of SEM, we used the default settings, including an uninformative flat prior ranging from -10^{34} to 10^{34} , but prevented the estimation of negative variances (the “admissibility check” option in AMOS) in order to improve model convergence. The Markov Chain Monte Carlo sampling for the Bayesian estimation was continued until subsequent runs were sufficiently uncorrelated, which is

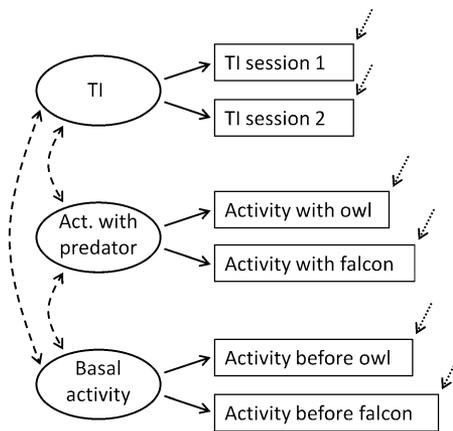


Figure 1

Graphical representation of the SEM fitted to estimate the relationships between different components of TI, activity in the presence of a predator, and basal activity. Variables in rectangles are measured (“indicator”) variables; variables in ovals are underlying (“latent”) variables. The labels “TI session 1” and “TI session 2” are generic labels and are replaced in the different models by the first and second measurement of duration of TI or the number of inductions to attain TI, respectively (i.e., yielding 2 different models with the same structure). Dashed double-headed arrows represent covariances among latent variables, continuous arrows represent causal regression weights, and dotted single-headed arrows represent (residual) error variances, which are all estimated from the observed data.

indicated by a “convergence criterion” icon in AMOS (this indicates whether the value of the Gelman–Carlin–Stern–Rubin convergence statistic is less than the conservative default value of 1.002), but always with a minimum of 50,000 sampling events even if convergence of posterior summaries was reached sooner. Prior to analysis, basal activity and activity with predator were square-root transformed in order to improve the normality of the data (Quinn and Keough 2002), since this improves SEM. Measures of activity as well as TI can depend on the time of day, but in our case effects of time of day on all measures were statistically and biologically negligible in both species (evaluated visually and by statistical testing: all $P > 0.05$) and hence not corrected for.

Model structure

When building any kind of model with several variables, a number of different models are possible. However, based on biological considerations and expectations, here, we are only interested in the correlations between repeated measures of the same kind of behavior and the correlations between the different behaviors. For this reason, we only ran a single basic model and present its resulting parameter estimates, without comparison with more simple or more complex models that perhaps fit the data better, in order to avoid inflation of type 1 error (Forstmeier and Schielzeth 2011). Hence, the model structure was composed of 2 layers (Figure 1): the ability to construct such synthetic model layers is one of the advantages of SEM. We first modeled latent (unmeasured) variables that are expressed in the actual measurements in each of the replicate sessions (the “measurement model” of SEM: Grace 2006), therefore modeled graphically in AMOS by means of a directional arrow from latent to measured variable. For example, a value for the latent variable “activity with predator” for each individual is estimated taking into account the regression weights of the measured activities during exposure to the owl or to the falcon on that latent variable. In a traditional GLM, one would probably just use the mean of these

2 sessions as an independent predictor. However, modeling a latent variable has two advantages. First, it allows the model to focus more on the information where the replicated measures coincide. Thus, if the data of one session had more random measurement error, its regression weight would be lowered relative to the other session, partly shielding the final results from this noninformative measurement error (Grace 2006). Second, it is impossible to calculate a mean for censored data, whereas the estimation of a latent variable uses the information that a given individual has only 1 or 2 censored values. Hence, the 2 sessions of basal activity, of exposure to a predator and of a particular measurement of TI, were each treated as replicated measures, yielding the 6 independent observed variables in Figure 1. The measurement model provided the Bayesian credible intervals for the correlations between these replicate measures. These correlations are a measure of consistency repeatability (intraclass correlation coefficient) since they do not take into account any differences in mean values among different sessions (Nakagawa and Schielzeth 2010). The second layer of our model (Figure 1) was composed of the relationship among the latent variables (the “structural model” of SEM: Grace 2006). In this case, the covariances among the components of TI, baseline activity and the response to predators, and the posterior distributions of these covariances were estimated. In AMOS, the relationships among these latent variables are graphically modeled by double-headed arrows because, in this case, we cannot assume that a certain variable only has a causal, unidirectional relationship with another variable.

Presentation of results

Bayesian analyses normally do not test null hypotheses but present the relative support for a given hypothesis or parameter estimate in the form of posterior distributions. We therefore present a few of these for visual interpretation by the reader. However, we also summarize all relevant posterior distributions by the median and a Bayesian Credible Interval. The median is a measure of central tendency that is more informative than the mean when the posterior distribution is skewed (Quinn and Keough 2002), as was sometimes the case here. The credible interval has similarities with a frequentist confidence interval, but gives the probability that the true value of the parameter will be within that interval (McCarthy 2007). We chose a 95% credible interval for presentation because this allows direct comparison with a two-sided null hypothesis test using the traditional (though arbitrary) significance criterion α fixed at 0.05, for those readers that prefer this way of assessing the data. In cases where the output of AMOS only produced (unstandardized) covariances, we calculated a correlation using the estimated median variances and covariance (correlation = covariance/square root [variance 1 \times variance 2]).

RESULTS

There was good support (Table 1) for the presence of repeatable variation in duration of TI (see also Figure 2A) and number of inductions to attain TI for both species, although in general, repeatability appears higher in Yellow-crowned bishops than in Tree sparrows. In addition, repeatable variation in basal activity and activity in the presence of a predator model were also well supported in Yellow-crowned bishops. The 95% credible intervals did not include zero, indicating that these repeatabilities can be seen as significant under conventional null hypothesis testing.

The median correlation between duration of TI and number of inductions was negative as expected but very low in both

Table 1
Repeatability of TI components, basal activity, and activity in presence of a predator model (expressed as the correlation between 2 repeated measures) in Yellow-crowned bishops and Tree sparrows

Trait	Median	95% credible interval
Yellow-crowned bishop		
Duration of TI	0.42	0.11/0.73
Number of inductions TI	0.65	0.20/0.87
Basal activity	0.90	0.77/0.96
Response to predator	0.63	0.34/0.81
Tree sparrows		
Duration of TI	0.25	0.03/0.59
Number of inductions TI	0.32	0.05/0.62

Estimated by Bayesian SEM of censored data; posterior distributions of correlations are summarized by median and 95% credible intervals.

sessions for both species, and the 95% credible intervals widely overlapped with zero (Table 2), suggesting that these 2 components of TI could be measuring independent aspects of individual variability.

In the Yellow-crowned bishop, the correlation between activity in the presence of a predator and basal activity appeared very low (median: -0.08 , credible interval: $-0.65/0.43$), despite one being measured straight after the other. In addition, there was little support that either of the 2 components of TI covaried much with basal activity: 95% credible intervals of covariances widely overlapped with zero, and estimated median correlations were mostly low and in the wrong direction (Table 3, Figure 2B). In contrast, activity in the presence of a predator consistently showed a higher median correlation with either component of TI, and the covariance between duration in TI and activity in the presence of a predator can be seen as significant under conventional null hypothesis testing (Table 3, Figure 2C). Individuals that showed higher activity in the presence of a predator model had a shorter duration of TI and needed more inductions to attain TI. When analyzing

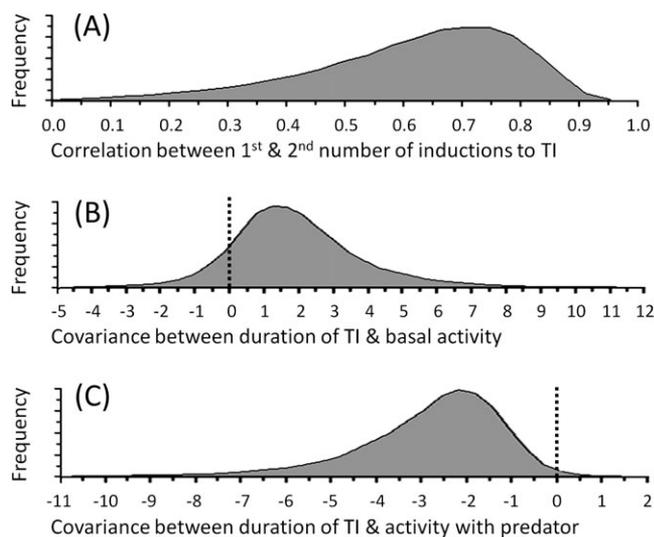


Figure 2
 Bayesian posterior distributions of various estimated correlations and covariances of the SEM of Figure 1 for the Yellow-crowned bishop. The density under the curve gives the probability that the estimated parameter has that value. Dotted lines indicate the parameter value of zero where relevant.

Table 2
Correlations (median and 95% credible interval) between duration of TI and number of inductions to attain TI in Yellow-crowned bishop and Tree sparrow

Species	Session	Median	95% credible interval
Yellow-crowned bishop			
	1	-0.10	$-0.47/0.31$
	2	-0.12	$-0.57/0.31$
Tree sparrow			
	1	-0.07	$-0.35/0.21$
	2	-0.07	$-0.34/0.19$

Estimated by Bayesian SEM of censored data; posterior distributions of correlations are summarized by median and 95% credible intervals.

a structurally identical model as before (Figure 1) but simply using a value of 10 min for birds that remained in TI at the end of the session (i.e., ignoring the uncertainty in censored values), duration of TI was no longer significantly correlated with activity in the presence of a predator under conventional null hypothesis testing. This was true both when conducting a Bayesian SEM (credible interval: $-6.66/0.14$) or when conducting a standard maximum-likelihood SEM ($P = 0.057$).

The 2 species seemed to differ in how many inductions they needed to attain TI. Tree sparrows needed on average only 1.69 and 1.47 (mean: 1.58) inductions in session 1 and 2, respectively, whereas Yellow-crowned bishops needed 1.92 and 2.13 (mean: 2.03) inductions; the credible intervals did not overlap among species in session 2. The species also seemed to differ in duration of TI. Tree sparrows stayed on average 3.54 and 3.82 (mean: 3.68) min in session 1 and 2, respectively, whereas Yellow-crowned bishops stayed only 2.47 and 2.67 (mean: 2.57) min. In that sense, the number of inductions to attain TI and the duration of TI does appear negatively correlated across species (or populations), but this cannot be evaluated meaningfully with only 2 independent data points (the 2 species).

DISCUSSION

Variation in TI

We obtained good support that both components of TI in Yellow-crowned bishops and Tree sparrows were repeatable. This means that there are consistent differences among individuals in these measures over the time period of our study. To what extent environmental or genetic effects underlie this temporal consistency is unknown, but heritability estimates are currently attempted using breeding experiments. Repeatability is sometimes seen as providing the upper limit to the heritability of a trait (Lessells

Table 3
Covariances (median and 95% credible interval) and correlations between components of TI and basal activity or activity in the presence of a predator model in Yellow-crowned bishops

Traits	Median	95% credible interval	Correlation
Duration of TI—basal activity	1.71	$-1.51/6.48$	0.34
Duration of TI—activity with predator	-2.50	$-6.97/-0.41$	-0.71
Number of inductions—basal activity	0.26	$-1.53/2.28$	0.10
Number of inductions—activity with predator	0.43	$-0.74/2.10$	0.23

Estimated by Bayesian SEM of censored data; posterior distributions of correlations are summarized by median and 95% credible intervals.

and Boag 1987), but this is not necessarily the case, for example, if the traits are measured with error (Sæther et al. 2007). In our case, random measurement error is likely present. Hence, we should interpret the reported correlations as minimum estimates of the true correlations, since random measurement error will reduce correlation, even when estimated with SEM. Some of the observed repeatabilities (Table 1) are quite high for behavioral traits: Bell et al. (2009) found a mean repeatability of about 0.3 for activity, 0.5 for antipredator behavior, and 0.4 across all sorts of behaviors for birds.

TI has been studied frequently as a measure of fear, and several lines of evidence support this interpretation, especially for domesticated chickens (see INTRODUCTION). However, quantitative studies of TI (including repeatability) are scarce for wild vertebrates and virtually lacking for wild birds. Lack of tests of repeatability of TI may limit the interpretation of any results. For example, Rubolini et al. (2005) found no effect of administration of corticosterone in the egg stage on TI in gull chicks, but they measured TI in 2-day-old hatchlings. Previous studies on gulls (Montevecchi 1978) and chickens (Forkman et al. 2007) show that TI is not or hardly developed by that age, so their measures of TI may well have mostly been random noise, and it is possible that they missed interesting effects of corticosterone on TI in older individuals (see Dingemanse, Edelaar, et al. 2010). Here we show, as far as we know, for the first time that TI is repeatable in wild vertebrates. This suggests that levels of fear of the same stimulus (which includes here the standardized prehandling and measurement of temperature and breathing rate prior to measuring TI) vary consistently among wild individuals. A similar conclusion was drawn by Carrete and Tella (2010) based on flight initiation distance (the distance when an individual flies away when approached), which is also interpreted as a measure of fear and antipredation behavior. Importantly, the repeatability of TI is replicated in 2 species, suggesting this might be generally true. In addition, the level of fear seems to differ among the 2 species we tested, and such consistent interspecific (or interpopulation) differences in fear can have important consequences, for example, for adaptation to habitat changes (Carrete and Tella 2011). Potentially, species and populations are positioned along a gradient of alternative antipredator strategies of hiding or fleeing as a response to local predator pressures. Moreover, our results also show that both the number of inductions to attain TI and the duration of TI are independently repeatable. Both measures have been used often (Gallup and Rager 1996), and our study validates their use in terms of representing consistent individual variability. On the other hand, number of inductions and TI duration are hardly correlated. This would mean that individuals can differ independently in these 2 traits, and we should take this into account, as outlined below.

Relationship between tonic immobility and risk of predation

Our results validate the earlier interpretation that TI is related to a response toward a predator, and show that this is even true at the individual level. We found a fair amount of support for a correlation between the response to a predator model and the duration of TI. This suggests that an individual's response to the experimental exposure to a predator model is linked to the individual's response to the conditions during TI sessions. Presumably, in both cases, the response is toward the perceived risk of predation. In contrast to the observed genetic correlation between activity and TI in a flour beetle (Miyatake et al. 2008), we found only weak support that basal activity is linked to TI and the effect actually had an opposite direction for duration of TI, so we discard the link between activity and TI components as important in our study species. Thus, indi-

vidual variation in TI seems to be related to individual responses to the risk of predation under more natural conditions and could be interpreted in a context of boldness toward predators. Such variability in boldness may well be permanent, because we assessed TI and the response to predator models with an 18-week difference. Note that if individuals that stayed shorter in TI are interpreted as being bolder, then bolder individuals moved more in the presence of a predator in our experiments (perhaps representing graded alternative antipredator strategies: flee vs. hide). This agrees with results from Feenders and Bateson (2011). They also found that both wild-caught and hand-reared starlings moved more when a human entered the room, but this was less so for the wild-caught birds, which presumably were more afraid of humans. We found virtually no support for a link between baseline activity and activity in the presence of a predator, which means that boldness and activity were uncorrelated in the Yellow-crowned bishops.

TI is often measured by various indicator variables, and then these measures are frequently combined into a single one to facilitate (univariate) statistical analysis (following Jones and Mills 1983). Combining indicator variables might be convenient if they have only weak signals and their error partly cancels out, resulting in a stronger signal. However, our results show that the different components of TI we used (number of inductions and duration of TI) are only weakly correlated and seem to have different relationships with the other variables in our model. In agreement with earlier studies (Forkman et al. 2007) and despite its lower repeatability, the duration of TI in general showed stronger patterns than the number of inductions to attain TI. These findings argue against the combination of these and additional components of TI into a single measure, or at least in the arbitrary way it is usually done (e.g., by weighting each component equally). It might be better to use SEM to derive one or more latent variables that explain the observed measurements (measurement model), and test for the partial effects of these latent variables and the remaining effects of the measured indicator variables on the other variables of interest (structural model).

Advantages of Bayesian SEM

In the MATERIALS AND METHODS, we discussed the advantages of using SEM when testing and comparing specific yet sometimes complex hypotheses (see also Dingemanse, Dochtermann, et al. 2010). Here, we discuss in more depth the specific advantages of applying the Bayesian approach to SEM for our study.

The preset limits for the assessment of number of inductions to attain TI (maximum 5 times) and duration of TI (maximum 10 min) resulted in many censored values. By estimating their posterior distribution, the Bayesian approach to SEM allowed us to incorporate the uncertainty of these censored data into the analyses. In short, likely values for censored data were estimated based on the fitted model (and its credibility), and variability in the censored data was iteratively incorporated during the model fitting (Arbuckle). As such, the informative part of censored data is put to good use, but at the same time the uncertainty in censored data is propagated into the final results of the model. This approach is theoretically more defensible and preferable than other approaches omitting censored data or not taking the effect of censoring into account because those approaches assume that extreme individuals do not exist in the measured population, leading to biases in the parameter estimates. In addition, we found that our methodological approach resulted in numeric estimates that were statistically significant under conventional null hypothesis testing, but which were not significant if ignoring the effects of censoring. Altogether, these results suggest that the Bayesian approach increased

statistical power. Generally, the advantages of the Bayesian approach over the classical methods are the ability (although not done here) to incorporate prior knowledge about the parameters (McCarthy 2007) and the fact that the modeling process does not rely on asymptotic theory (Arhonditsis et al. 2006). The latter issue is particularly important when the sample size is small (commonly experienced in studies of animal behavior: Taborsky 2010), and thus the classical estimation methods (maximum likelihood, generalized and weighted least squares) are not robust. Monte Carlo Markov chain samples are taken from the posterior distribution, and as a result the procedure works for all sample sizes and various sources of non-normality (Arhonditsis et al. 2006). Moreover, Bayesian analyses can detect multimodality in parameter estimates. In our case, some of the posterior distributions were not normal (Figure 2), which means that the mean and median are not the same, an insight that is not obvious from standard maximum-likelihood estimation. Finally, the Bayesian philosophy avoids problems related to traditional null hypothesis significance testing as outlined in the MATERIALS AND METHODS. For these reasons, we encourage others to consider this type of approach. We have been unable to find any application of a Bayesian approach to SEM in studies of animal behavior, despite its many advantages, especially when analyzing censored or completely missing data. It is our hope that this paper will help to promote the wider application of Bayesian SEM.

CONCLUSIONS

We established that several aspects of TI have quite good repeatability in 2 wild avian species. This individually repeatable variation in TI is correlated with individually repeatable variation in the response to predators, but not with repeatable variation in baseline activity. This makes TI a candidate behavioral trait for studies on personalities (especially boldness) or for other studies involving antipredation behavior. There also seems to be a consistent difference in TI among similar-sized species, and it would be exciting to know whether this difference is similarly correlated to species or population differences in responses to predators. The statistical technique of SEM allowed for the simultaneous assessment of repeatability of each trait and correlations among traits. The Bayesian approach allowed us to meaningfully model censored data, and this appeared to provide more statistical power than approaches that did not take the censoring into account.

FUNDING

This work was supported by Excellence Projects P07RNM02918 and P08RNM4014 from Junta de Andalucía (Spain), by Fundación Repsol, and by JAE-DOC and Ramon y Cajal contracts from the Spanish Ministry of Science to P.E., M.C., and J.B.

We thank J. Ayala, A. Jurado, and M. Vázquez for their bird-netting help, A. Jurado for his help with care of captive birds, N. Dingemans and 2 anonymous reviewers for advice and comments that improved the paper, and N. Dochtermann for instruction and advice on use of SEM.

REFERENCES

Arbuckle JL. IBM SPSS Amos 19 user's guide. Chicago (IL): SPSS. Available from: <http://www.amosdevelopment.com/download/amos.pdf>.
 Arhonditsis GB, Stow CA, Steinberg LJ, Kenney MA, Lathrop RC, McBride SJ, Reckhow KH. 2006. Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecol Modell*. 192:385–409.
 Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behavior: a meta-analysis. *Anim Behav*. 77:771–783.

Boissy A. 1995. Fear and fearfulness in animals. *Q Rev Biol*. 70:165–191.
 Carrete M, Tella JL. 2010. Individual consistency in flight initiation distance in burrowing owls: a new hypothesis on disturbance-induced habitat selection. *Biol Lett*. 6:167–170.
 Carrete M, Tella JL. 2011. Inter-individual variability in fear of humans and relative brain size are related to contemporary urban invasion in birds. *PLoS One*. 4:e18859.
 Cockrem JF. 2007. Stress, corticosterone and avian personalities. *J Ornithol*. 148:S169–S178.
 Dingemans NJ, Dochtermann N, Wright J. 2010. A method for exploring the structure of behavioral syndromes to allow formal comparisons within and among data sets. *Anim Behav*. 79:439–450.
 Dingemans NJ, Edelaar P, Kempenaers B. 2010. Why is there variation in baseline glucocorticoid levels? *Trends Ecol Evol*. 25:261–262.
 Dochtermann NA, Jenkins SH. 2007. Behavioural syndromes in Merriam's kangaroo rats (*Dipodomys merriami*): a test of competing hypotheses. *Proc R Soc B Biol Sci*. 274:2343–2349.
 Erhard HW, Mendl M, Christiansen SB. 1999. Individual differences in tonic immobility may reflect behavioral strategies. *Appl Anim Behav Sci*. 64:31–46.
 Feenders G, Bateson M. 2011. Hand-rearing reduced fear of humans in European starlings, *Sturnus vulgaris*. *PLoS One*. 6:e17466.
 Forkman B, Boissy A, Meunier-Salaün M-C, Canali E, Jones RB. 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiol Behav*. 92:340–374.
 Forstmeier W, Schielzeth H. 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav Ecol Sociobiol*. 65:47–55.
 Gallup GG. 1974. Animal hypnosis: factual status of a fictional concept. *Psychol Bull*. 81:836–853.
 Gallup GG. 1977. Tonic immobility: the role of fear and predation. *Psychol Rec*. 1:1–41.
 Gallup GG, Rager DR. 1996. Tonic immobility as a model of extreme states of behavioral inhibition. In: Sanberg PR, Ossenkopp K-P, Kavaliers M, editors. Motor activity and movement disorders: research issues and applications. New Jersey (NY): Humana Press. p. 57–80.
 Grace JB. 2006. Structural equation modeling and natural systems. Cambridge (UK): Cambridge University Press.
 Jones RB. 1980. Reactions of male domestic chicks to two-dimensional eye-like shapes. *Anim Behav*. 28:212–218.
 Jones RB. 1996. Fear and adaptability in poultry: insights and implications and imperatives. *Worlds Poult Sci J*. 52:131–174.
 Jones RB, Mills AD. 1983. Estimation of fear in two lines of the domestic chick: correlations between various methods. *Behav Process*. 8:243–253.
 Lessells CM, Boag PT. 1987. Unrepeatable repeatabilities—a common mistake. *Auk*. 104:116–121.
 McCarthy MA. 2007. Bayesian methods for ecology. Cambridge (UK): Cambridge University Press.
 Miyatake T, Nakayama S, Nishi Y, Nakajima S. 2009. Tonic immobilized selfish prey can survive by sacrificing others. *Proc R Soc B Biol Sci*. 276:2763–2767.
 Miyatake T, Tabuchi K, Sasaki K, Okada K, Katayama K, Moriya S. 2008. Pleiotropic antipredator strategies, fleeing and feigning death, correlated with dopamine levels in *Tribolium castaneum*. *Anim Behav*. 75:113–121.
 Montevecchi W. 1978. Tonic immobility responses of Herring gull chicks. *Condor*. 80:248–249.
 Nakayama S, Nishi Y, Miyatake T. 2010. Genetic correlation between behavioral traits in relation to death-feigning behavior. *Popul Ecol*. 52:329–335.
 Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev*. 85:935–956.
 Ohno T, Miyatake T. 2007. Drop of fly? Negative correlation between death-feigning intensity and flying ability as alternative anti-predator strategies. *Proc R Soc B Biol Sci*. 274:555–560.
 Quinn GP, Keough MJ. 2002. Experimental design and data analysis for biologists. Cambridge (UK): Cambridge University Press.
 Reale D, Dingemans NJ, Kazem AJN, Wright J. 2010. Evolutionary and ecological approaches to the study of personality. *Philos Trans R Soc Lond B Biol Sci*. 365:3937–3946.

- Reale D, Reader SM, Sol D, McDougall PT, Dingemans NJ. 2007. Integrating animal temperament within ecology and evolution. *Biol Rev.* 82:291–318.
- Rubolini D, Romano M, Boncoraglio G, Ferrari RP, Martinelli R, Galeotti P, Fasola M, Saino N. 2005. Effect of elevated egg corticosterone levels on behavior, growth, and immunity of yellow-legged gull (*Larus michahellis*) chicks. *Horm Behav.* 47:592–605.
- Ruxton G. 2006. Grasshoppers don't play possum. *Nature.* 440:880.
- Sargeant AB, Eberhardt LE. 1975. Death feigning by ducks in response to predation by red foxes (*Vulpes fulva*). *Am Midl Nat.* 94:108–119.
- Sæther SA, Fiske P, Kålås JA, Kuresoo A, Luigujõe E, Piirtney SB, Sahlman T, Höglund J. 2007. Inferring local adaptation from Q_{ST} - F_{ST} comparisons: neutral genetic and quantitative trait variation in European populations of great snipe. *J Evol Biol.* 20:1563–1576.
- Taborsky M. 2010. Sample size in the study of behavior. *Ethology.* 116:185–202.
- Thompson RKR, Foltin RW, Boylan RJ, Sweet A, Graves CA, Lowitz CE. 1981. Tonic immobility in Japanese quail can reduce the probability of sustained attack by cats. *Anim Learn Behav.* 9:145–149.
- Zulkifli I, Al-Aqil A, Omar AR, Sazili AQ, Rajion MA. 2009. Crating and heat stress influence blood parameters and heat shock protein 70 expression in broiler chickens showing short or long tonic immobility reactions. *Poult Sci.* 88:471–476.