Individual behavioral type captured by a Bayesian model comparison in cap making by sponge crab

Keita Harada[1] **and Katsushi Kagaya**[1, 2]

[1]**Seto Marine Biological Laboratory, Field Science, Education and Reseach Center, Kyoto University, Shirahama-cho, Nishimuro-gun, Wakayama 649-2211 Japan**
[2]**The Hakubi Center for Advanced Research, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan**

Corresponding author:
Katsushi Kagaya[1]

Email address: kagaya.katsushi.8e@kyoto-u.ac.jp

ABSTRACT

'Animal personality' is considered formed through complex interactions of an individual and its surrounding environment. How can we quantify the 'personality' of an individual? Although there is no clear consensus, quantifying intra- and inter-individual variations of behavior, or individual behavioral type, is a prerequisite in the studies of animal personality. We propose a statistical method to measure the appropriateness of our assumption of 'individual' in repeatedly measured behavioral data, from each individual from a predictive point of view. For a model case, we studied the sponge crab *Lauridromia dehaani* known to make a 'cap' from a natural sponge and carry it for camouflage. Because a cap is most likely to be rebuilt and replaced repeatedly, we hypothesized that each individual crab would grow a unique behavioral type and it would be observed under an experimentally controlled environmental condition. To test the hypothesis, we conducted behavioral experiments and newly took an approach of Bayesian model comparison to examine whether the crab has the individual behavioral type in the cap making behavior. The behavioral choices were given to a crab using artificial sponges of three different sizes. We did statistical modeling to implement hierarchical structure specifying the behavioral type. We modeled a choice of sponges, size of a trimmed part of a cap, size of a cavity of a cap, latency to produce a cap, as random variables in 26 models. In addition, we calculated widely applicable information criterion (WAIC) value for each model to evaluate the models from the predictive point of view. As a result, the crabs less than about 9 cm were found to make caps from the sponges. The body size well explained the behavioral variables, choice, trimmed and cavity size, but not the latency. The behavioral type was captured as a difference of WAIC of the models. Thus, we captured the behavioral type as a probabilistic distribution structure in the behavioral data. Our statistical approach is not limited to behavioral data but also applicable to physiological or morphological data when one would try to examine if some group structure would exist behind fluctuating empirical data.

INTRODUCTION

An individual is an important hierarchical structure in biology. We aim to capture intra- and inter-individual variations in behavior as a probabilistic distribution structure, because it is a prerequisite for the study of 'animal personality' (Sih et al., 2004; Niemelä and Dingemanse, 2018). Because the term 'individual difference' sometimes means only inter-individual difference, we use 'individual behavioral type' to refer to the two variations. Behavioral ecologists and evolutionary biologists have been interested in the behavioral type, because it can be a selective trait. At the evolutionary time scale, the structure of distribution is very likely to be related to a survival value of an individual. At the behavioral time scale, the behavioral type can be caused through complex and dynamic interactions of individual properties such as behavioral plasticity based on physiological processes, with surrounding dynamic environments. A typical interaction can be observed in body extending behaviors such as tool making and using (e.g., Hunt, 1996; Wang et al., 2014; Matsui and Izawa, 2017; Sonoda et al., 2012). A body extending be-

havior which is basically a behavior attaching non-living thing to a body, seems to require at least some information processing to infer a current body size, in order to achieve an adaptive extension through the complex interactions. Uncertainty in the inference, and accumulation of experiences accompanying the realization of body extensions, might result in the emergence of the behavioral type. Here we examine the hypothesis that the individual behavioral type would emerge in a body extending behavior. As an example of the body extending behavior, the cap making and carrying behavior by the sponge crab is experimentally examined and statistically modeled in this study. To capture the structure, we need repeated measurements and specific statistical modeling considering a hierarchical structure (Niemelä and Dingemanse, 2018). The hierarchical model such as a generalized linear mixed effect model (GLMM) is widely considered appropriate for the repeated data (Zuur et al., 2009; Niemelä and Dingemanse, 2018; Reinhart, 2015). However, it has been poorly examined in empirical data on how much appropriate a hierarchical model is against a non-hierarchical one such as a generalized linear model (GLM). One famous statistical measure used in GLM from the predictive point of view is Akaike Information Criterion (AIC) (Akaike, 1974; Sakamoto et al., 1986). To calculate AIC, the maximum log-likelihood needs to be calculated, but in general, the predictive by the maximum likelihood (ML) method is inappropriate for hierarchical models (Watanabe, 2005). This is because a model with hierarchical structures is statistically non-regular model and the assumptions set in the ML estimation is violated (Watanabe, 2005, 2010b, 2018). Alternatively, a Bayesian procedure to construct a predictive distribution is known to perform better than the ML method in the hierarchical models (Watanabe, 2018).

Although the basic Bayesian framework and its mathematical foundation of measuring predictability of an arbitrary pair of a statistical model and a prior distribution, has been rigorously established (Watanabe, 2010b,a, 2018), there is few applications of the framework to the behavioral data containing repeated measurements (Wakita et al., 2019). Specifically, a performance of a predictive distribution can be inferred by WAIC that is a measure of generalization error defined as the quantity of how well a specified predictive distribution is approximated with respect to an unknown true distribution that generates data (Akaike, 1998; Watanabe, 2018). Furthermore, there are almost no appropriate applications of WAIC to hierarchical models for the repeatedly measured data. For constructing a predictive distribution using a hierarchical model, we are usually interested in a new observation from a new cluster other than clusters that provided observations. Therefore, we need to marginalize the parameters assigned to each cluster to calculate WAIC in that situation (Watanabe, 2018; Millar, 2018). However, this point does not seem to be recognized well not only in biological communities but also in other real-world data analyses in general.

Therefore, we adopt the Bayesian model comparison framework using WAIC to study a specific individual behavioral type in the body extending behavior in the crab. In previous research, one field study dealt with the preference of dromiid crabs to materials and examined the association between a cap size with a body size (McLay, 1983). Additionally, it is reported that *Cryptodromia hilgendorfi* use caps made from many species of sponges, but they particularly prefer the sponge *Suberites carnosus*, and the crabs make sponge caps twice as large as the carapace area. In previous experimental research, the preference for material size and the suitability between body size and cap size is scarcely investigated. It is reported that *Dromia personata* mainly used sponges and ascidians (Bedini et al., 2003), while they could also make caps with paper (Dembowska, 1926). Dembowska (1926) reported non-breaking space is used qualitatively and the cap size made by *Dromia personata* (reported as *D. vulgaris*) with paper is as large as the size of those that the crabs originally carried. Because these studies sampled a body size and a camouflage size once for an individual, it is unclear whether there is an individual behavioral type. In addition, it is unknown whether the behavioral type conditional on body size exists in the cap making behavior. Thus, although the crabs in the family Dromiidae have been known to make a cap (Guinot and Wicksten, 2015), the behavior of the *Lauridromia dehaani* has not been examined so far.

Taking into account the lack of the experimental data of cap making in the crab *Dromia personata* and the limitations of the statistical approach hitherto, we set the four goals to study the individual behavioral type in the body extending behavior: (1) to perform behavioral experiments sampling behavioral data repeatedly, (2) to formulate an individual behavioral type in statistical models, and simultaneously to construct other models assuming no such a behavioral type, (3) to measure predictive performance of those models by WAIC including hierarchical models assuming an individual behavioral type and compare them with non-hierarchical models assuming no such a type, (4) to infer an association of the behavioral data with body size by conditioning the behavioral variables by the body size.

MATERIALS & METHODS

The whole data and codes are available from the supplementary material.

Animal collection

From December 2015 to April 2017, 40 individuals (21 males, 19 females) of *Lauridromia dehaani* (Brachyura: Dromiidae) (Fig. 1A) were obtained with a gill net at the Sakai fishing port, Minabe town, Wakayama, Japan (33° 44 'N, 135° 20 'E). We conducted behavioral experiments of cap making on 38 individuals (20 males, 18 females) and video recorded for 2 individuals (4.30 cm and 7.19 cm of the carapace widths for each) in a tank filled with filtered natural seawater (about 3.4 % of the salinity) at Shirahama Aquarium, Seto Marine Biological Laboratory, Kyoto University (33° 41 'N, 135° 20 'E), from December 2015 to June 2017. For the behavioral experiments, we successfully sampled repeatedly from 8 individuals (3 or 5 times for each). Thus, in the other 30 individuals, we sampled one observation from one individual. Note that the samples sizes of the behavioral acts for each individual are different, but our method is applicable. Before the experiments, all individuals were maintained in the tanks (19.5–23.8 $°C$, light on: 8–17, light off: 17–8) of the aquarium for more than two days for acclimation. We measured their carapace width (cm) (Fig. 1B) for a proxy for the body size, and divided them into three levels depending on whether they lacked any of the fourth and fifth pereiopods: (0) none of the fourth and fifth pereiopods were absent, (1) one of them was absent, (2) both of the fourth and fifth of each side were absent.

Experimental setup and procedure

We prepared white melamine foam manufactured worldwide (most notably by BASF of Germany) and often used in a general household (Gekiochi-kun, LEC, Inc.) of three classes of size (S: 20 mm × 30 mm × 40 mm; M: 30 mm × 60 mm × 85 mm; L: 30 mm × 140 mm × 150 mm). We used this sponge because it is easy to sink.

First, to confirm that the cap making behavior by the crab *Lauridromia dehaani* is similar to the behavior in the reports (Dembowska, 1926; McLay, 1983), we video-recorded the behavior from the two crabs. They were used only for the recording in the aquarium (310 mm × 180 mm × 240 mm, W × L × H). We started the recording from 9 to 10 in the morning, stopped 2 hours after the crab stopped cap making. We used red light through a polyvinyl chloride board and excluded another light source by enclosing the aquarium. We recorded 5 examples for each crab.

Second, we performed the experiment of size choice, trimming, and excavating behavior. S size sponge was smaller than all crabs, whereas L size was larger than all crabs. Each sponge was put pseudo-randomly to either side and the back center of the cage (700 mm × 470 mm × 190 mm, W × L × H, Fig. 1C), which floated in the tank. Then, crabs were introduced to the front center of a cage floating in a tank, thereby the distance between each sponge and the crab was equal. We started a trial from 9 to 10 in the morning and checked whether a crab carried any sponge once a day. We counted the days when the crab carried a sponge. The latency as days to produce a cap is modeled as a random variable. Note that the crab was assumed to make a cap at night, because it is considered nocturnal (McLay, 1983). If it did, we collected a sponge, otherwise, the crab and three sponges remained in the cage. When the crab did not carry any sponge for five days, we stopped the trial. We desiccated all sponges that the crabs processed and measured whole area (cm^2), and area of the concave part (cm^2) excavated by a crab from the pictures taken 46 cm above the sponges. The trimmed area and concave area are modeled as random variables. First, we performed one trial for one individual ($N_{animal} = 30$), but five trials for one individual after February 2017 ($N_{animal} = 8$) to examine the behavioral type after we hypothesized that the type would be formed in the behavior.

Statistical modeling

In order to quantify the behavioral type in the behavior, we constructed 26 statistical models (Table. 1) for the four different aspects: (1) choice of sponge size (6 models), (2) amount of sponge trimmed by cutting (8 models), (3) size of cavity (6 models), and (4) latency to produce a cap (6 models). In each case, we build the statistical models specifying individual behavioral types as hierarchical structures with parameters and performed MCMC samplings from the posterior distribution. Also, we conditioned the variables with the carapace width, levels of leg absence, and gender. We specified the models in the probabilistic programming language Stan (Stan Development Team, 2018). We used non-informative uniform priors

for some parameters unless otherwise explicitly described. The performed samplings from the posterior distributions using No-U-Turn Sampler (NUTS) implemented as a Hamiltonian Monte Carlo (HMC) sampler in Stan. Sampling convergence was visually diagnosed by trace plots and quantitatively via the Gelman-Rubin convergence statistic, R_{hat} (Gelman et al., 1992). All sampled draws were judged to be converged when $R_{hat} < 1.10$, were used to construct predictive distributions with WAIC on each model. All computations were performed in the R statistical environment (R Core Team, 2018), and the Stan codes for each model were compiled and executed through the R package *rstan* (Stan Development Team, 2018).

We compared the predictive performances of the models using WAIC (Watanabe, 2018, 2010b). It should be emphasized that WAIC of a hierarchical model can be defined in several ways depending on how a predictive distribution is defined. In our case, as we would like to construct a new distribution regarding a new act of a new individual, we have to marginalize the intermediate parameters assigned to each individual in the statistical model (Watanabe, 2018). This is because we are interested in the prediction of a new behavioral act when we get a new individual and get a new behavioral act instead of the prediction of a new behavioral act from the individuals sampled in this study. By performing this procedure, we can equally compare a hierarchical model with a non-hierarchical model, because the focus of the prediction in a non-hierarchical model is on a new behavioral act of a new individual.

Here we briefly describe the basic procedure based on Watanabe (2018). Let $X^n = (X_1, \ldots, X_n)$ be an i.i.d. sample from the true distribution and $p(x|w)$ a distribution with w assigned to each individual. Furthermore, w is assumed to be taken from $\varphi(w|w_0)$ to form a hierarchical structure. Then, our statistical model is built like $p_{model}(x|w_0)$ by marginalizing w out:

$$p_{model}(x|w_0) = \int p(x|w)\varphi(w|w_0)\, dw. \tag{1}$$

WAIC is a measure for the degree of accuracy of an approximation of a predictive distribution to the true distribution generating data. For our hierarchical model, the predictive distribution is defined as $\mathbb{E}_{w_0}[p_{model}(x|w_0)]$. Then, the marginal-level WAIC for a hierarchical model is defined as:

$$WAIC_h = -\frac{1}{N}\sum_{i=1}^{N}\log\mathbb{E}_{w_0}[p_{model}(X_i|w_0)] + \frac{1}{N}\sum_{i=1}^{N}\log\mathbb{V}_{w_0}[p_{model}(X_i|w_0)] \tag{2}$$

where $\mathbb{E}_{w_0}[\,]$ and $\mathbb{V}_{w_0}[\,]$ are the average and variance operator of the posterior distribution of w_0. w_0 is estimated so that practically the MCMC sample is used, thus numerical integration is required. In this study, the computation is implemented in the 'function' block in the Stan codes using a Simpson's rule and the *log_sum_exp* function provided in Stan (see supplementary material).

On the other hand, WAIC for a non-hierarchical model is defined for a statistical model $p_{model}(x|w)$:

$$WAIC_{non-h} = -\frac{1}{N}\sum_{i=1}^{N}\log\mathbb{E}_{w}[p_{model}(X_i|w)] + \frac{1}{N}\sum_{i=1}^{N}\log\mathbb{V}_{w}[p_{model}(X_i|w)] \tag{3}$$

where $\mathbb{E}_{w}[\,]$ and $\mathbb{V}_{w}[\,]$ are the average and variance operator of the posterior distribution of w. Note that the often used conditional-level WAIC is described in the Discussion.

Choice of material size (model 1_1)

To provide an overview of specified models, we here describe only the best-performing models in terms of WAIC. The other models are summarized in Table 1.

We formulate a tendency toward a choice as $\mu[n,m]$ ($m = 1, 2, 3$ for M, L, skip, respectively):

$$\mu[n,1] = 0, \tag{4}$$

$$\mu[n,2] = a_{choice_L}[ID[n]] + b_{choice_L} \cdot CarapaceWidth[n] + c_{choice_L} \cdot LegLack[n], \tag{5}$$

$$\mu[n,3] = d_{choice_0} + e_{choice_0} \cdot CarapaceWidth[n] + f_{choice_0} \cdot LegLack[n], \qquad n = 1, \ldots, N_{act} \tag{6}$$

where $N_{act} = 68$ is the total number of behavioral acts, and *ID* represents animal identity (from 1 to $N_{animal} = 38$). μ is linked to the linear predictor in terms of the carapace width, *CarapaceWidth* and the level of absence of leg, *LegLack*. The choice for M size is fixed to zero. a_{choice_L} is for each individual, thus it is hierarchized. d_{choice_0} is not hierachized. The distribution of a_{choice_L} is defined as the normal distribution with the mean $a_{choice_{L0}}$ and standard deviation $a_{choice_{Ls}}$:

$$a_{choice_L}[k] \sim Normal(a_{choice_{L0}}, a_{choice_{Ls}}), \qquad k = 1, \ldots, N_{animal}. \tag{7}$$

The actual choice *Choice* is defined as the categorical distribution with the softmax function:

$$Choice[n] \sim Categorical(softmax(\mu[n,])), \qquad n = 1, \ldots, N_{act}. \tag{8}$$

Thus, in this case, a statistical model $p_{model}(x|w_0)$ is set using the parameters:

$$w_0 = (a_{choice_{L0}}, a_{choice_{Ls}}, b_{choice_L}, c_{choice_L}, d_{choice_0}, e_{choice_0}, f_{choice_0}). \tag{9}$$

Note that $w = a_{choice_L}$ is marginalized out when we build the predictive distribution, so that it is not included in w_0. The choice *Choice*[n] is modeled as a random variable X_n. *CarapaceWidth*[n] and *LegLack*[n] are the conditioning variables.

Trimming (model 2_1)

The probability of a decision whether an animal cut off the sponge is written as ϕ_{cut} linked to the linear predictor with the carapace width *CarapaceWidth* and the selected sponge size *Choice*[n]:

$$\phi_{cut}[n] = InverseLogit(a_{cut}[ID[n]] + b_{cut} \cdot CarapaceWidth[n] + c_{cut} \cdot Choice[n]) \tag{10}$$

$$n = 1, \ldots, N_{act}. \tag{11}$$

a_{cut} is assigned for each individual. $N_{act} = 51$ and *ID* is from 1 to $N_{animal} = 30$. The distribution of a_{cut} is defined as the normal distribution with the mean a_{cut_0} and standard deviation a_{cut_s}:

$$a_{cut}[k] \sim Normal(a_{cut_0}, a_{cut_s}), \qquad k = 1, \ldots, N_{animal}. \tag{12}$$

The prior distribution of a_{cut_s} is defined as the half t distribution:

$$a_{cut_s} \sim Student\_t^+(4,0,10). \tag{13}$$

The mean area of a sponge trimmed by the crab λ is linked to the linear predictor with the log link function:

$$log(\lambda_{cut}[n]) = d_{cut}[ID[n]] + e_{cut} \cdot CarapaceWidth[n] + f_{cut} \cdot Choice[n], \qquad n = 1, \ldots, N_{act}. \tag{14}$$

207 d_{cut} is assigned for each individual. The distribution of d_{cut} is defined as the normal distribution with
208 the mean d_{cut_0} and the standard deviation d_{cut_s}:

$$d_{cut}[k] \sim Normal(d_{cut_0}, d_{cut_s}), \qquad\qquad k = 1, \ldots, N_{animal}. \qquad (15)$$

209 The prior distribution of d_{cut_s} is defined as the half t distribution:

$$d_{cut_s} \sim Student\_t^{+}(4, 0, 10). \qquad (16)$$

210 Altogether, the area of trimmed sponge is modeled as the variable *Trimmed*. The distribution of it is
211 defined as the zero-inflated Poisson distribution (ZIP) with the parameters ϕ_{cut} and λ_{cut}:

$$Trimmed[n] \sim ZIP(\phi_{cut}[n], \lambda_{cut}[n]), \qquad\qquad n = 1, \ldots, N_{act}. \qquad (17)$$

212 When a crab skips trimming behavior, *Trimmed* is set to zero even if the sponge size is smaller than
213 the defined sizes of M or L due to measurement errors. Note that *Trimmed* is rounded to integer. We
214 consider that the rounding process has no significant impact on the data distribution.

215 *Cap cavity making (model 3_1)*

216 To examine how the cap cavity size *CavitySize* is explained by the carapace width *CarapaceWidth*,
217 the gamma distribution is chosen to represent non-negative value of the cavity size. The mean of the
218 distribution is specified by $lambda_{cavity}$ with shape and rate parameters:

$$CavitySize[n] \sim Gamma(shape, \frac{shape}{\lambda_{cavity}}), \qquad (18)$$

$$log(\lambda_{cavity}) = a_{cavity}[ID[n]] + b_{cavity} \times CarapaceWidth[n], \qquad n = 1, \ldots, N_{act}. \qquad (19)$$

219 where the rate parameter was given as the shape over the log-linked linear predictor and a_{cavity} is the
220 intercept for each individual. $N_{animal} = 30, and N_{act} = 51$. The a_{cavity} is taken from the normal distribution
221 with the mean a_{cavity_0} and the standard deviation a_{cavity_s}:

$$a_{cavity}[k] \sim Normal(a_{cavity_0}, a_{cavity_s}), \qquad\qquad k = 1, \ldots, N_{animal}. \qquad (20)$$

222 *latency (model 4_1)*

223 We assumes that the latency to produce a cap, *Days*, fits the ZIP distribution which is similar to the
224 *Trimmed* case:

$$\phi_{day}[n] = InverseLogit(a_{day}), \qquad (21)$$

$$log(\lambda_{day}[n]) = b_{day}[ID[n]], \qquad (22)$$

$$b_{day}[k] \sim Normal(b_{day_0}, b_{day_s}), \qquad\qquad k = 1, ..., N_{animal}, \qquad (23)$$

$$Days[n] \sim ZIP(\phi_{day}[n], \lambda_{day}[n]), \qquad\qquad n = 1, ..., N_{act}. \qquad (24)$$

225 where $N_{animal} = 32, N_{act} = 56$. Note that b_{day} is into this model to construct a hierarchical structure.

6/12

RESULTS

We measured and modeled as random variables the four variables: the choice of sponge size, trimmed size, cavity size, and latency for completing making sponge. Furthermore, we evaluated the model predictability by WAIC (see Materials and Methods).

Cap making using an artificial sponge

The behavior of the two crabs was video recorded to confirm the behavioral sequence of cap making using an artificial sponge. They grasped either side of the sponge by the second and third pereiopods, and trimmed small pieces of the sponge by the chelae (Fig. 2A upper left, upper right, Supplementary movie 1). They visited to the two sides of the sponge. To make a cavity the crabs rotated their body backward and grasped it by the fourth and fifth pereiopods. By repeating these behaviors, the crabs made a groove to cut off a portion of a sponge. On average, it took about 50 minutes to cut the portion, and the crabs started excavating as soon as they finished the trimming behavior in 9 of 10 trials except for one trial which took 19 minutes.

Next, the crabs made cavities by tearing off small pieces of a sponge (Fig. 2A bottom, Supplementary movie 2). It took 11 minutes on average to excavate the cavity. Then, the crabs rotated their body backward in order to catch it with the fourth and fifth pereiopods while they kept the portion grasped by second and third pereiopods. Finally, the crabs released the second and third pereiopods from the cap and began to carry it (Fig. 2B, C). In the behavior, it often happened that they rotated their body forward, dorso-ventrally to make the cavity larger. It is rare to move laterally. They repeated excavating up to eleven times per night and it took up to four hours. When the crabs rotated their bodies, the direction of rotation was maintained along with the sponge. While the crabs cut the sponge, they actively moved around the sponge. In contrast, they persistently stayed under the sponge during excavating.

Sponge choice

None of the 38 animals choose the S size sponge, and 7 animals skipped the cap making behavior (Fig. 3A). Therefore, we defined the choice as a random variable taking the three values: M, L, or skipping. The hierarchical model assuming behavioral types 1_1 (Fig. 3A, B) outperformed the non-hierarchical one in terms of WAIC (0.87 to -2.13, Fig. 3A-D, Table 1). The posterior probability of the behavioral choices was more widely variable on the model 1_1 than in model 1_6 depending on the individual difference specified as a_{choice_L} (Fig. 3B). The probability of choice sampled from the posterior distribution is visualized in white lines (Fig. 3A,C). Note that the variability of the choice probability in the white curved lines is smaller than the model 1_1 even if the number of lines are the same. Although the body size of the animal indicated with the white arrowhead (Fig. 3A) is small, it preferably selected the size L. This indicates a large inter-individual difference. In either case of hierarchical or non-hierarchical model, the behavioral choice of the sponges was better explained by the carapace width (Fig. 3A,C; Table 1). The estimated information gained by the model 1_1 against model 1_3 is 1.35 (Table 1). This suggests that larger crabs tend to choose L size sponge rather than M size. However, the crabs larger than about 9 cm carapace width did not choose the sponges.

Trimming

After a choice of M or L size sponge, the crabs decided whether to trim an extra part of the sponge (Fig. 4A-C). Here, we modeled the size of an area in a sponge that was trimmed ($N_{animal} = 30$). The trimmed sponge showed the three patterns (Fig. 4C). They cut off (1) all four corners of a sponge, (2) one corner of it elliptically, or (3) two corners of it linearly. The crabs trimmed the white area (Fig. 4C) and started excavating a cavity (Fig. 4C). The twenty three crabs skipped the trimming behavior in 33 trials.

After trimming or skipping, they started excavating. For a behavioral act of trimming, a non-zero data point indicating a trimmed size of the sponge was recorded (Fig. 4A,B). The size decreased with the increase of the carapace width. If a crab skipped trimming, a data point was recorded at zero (Fig. 4A), meaning no trimming. When crabs chose M size sponges, almost all crabs except for one individual decided not to trim the sponge. Meanwhile, they less trimmed the sponges in relation to the increase of their body sizes when they chose the L size sponges.

The WAIC of the hierarchical model 2_1 was -2.08 and that of comparable non-hierarchical model 2_6 was 7.40 (Fig. 4D, Table 1), indicating that the hierarchical model is far better than the non-hierarchical one.

Cavity size

Six crabs just cut the sponge and did not excavate the sponge. We modeled the size of a cavity in a cap ($N_{animal} = 30$) as a random variable taken from the gamma distribution with the log link function (Fig. 5). The size increased with the carapace width, and the model considering individual behavioral types performed best (Table 1). WAIC of the hierarchical model 3_1 is smaller than that of the comparable non-hierarchical model 3_2 (4.45 and 4.54, respectively; Fig. 5A,B, Table 1). The individual with the arrowhead (Fig. 5A) made relatively large cap cavities, indicating an large inter-individual difference. As expected, larger crabs made larger cavities. The difference of WAIC is about 0.1 (Fig. 5B). The improvement of the predictability is relatively small against that of sponge choice, suggesting that the individual behavioral type would be less in the determination of cavity size.

Latency

We modeled the latency for cap making (from the choice of sponge to carrying) by 32 crabs as a random variable taken from the zero-inflated distribution (Fig. 6). No obvious relation was found between the carapace width and the latency, and a number of crabs had carried the cap by the next day. However, the hierarchical model 4_1 outperformed the non-hierarchical model 4_2 (WAIC values, 1.10 and 1.28 respectively).

DISCUSSION

First, we proposed a statistical method applicable to capture an individual as a probabilistic distribution structure in the repeatedly measured data. The difficulty measuring appropriateness of a statistical model was overcome by a Bayesian approach and WAIC. Second, we applied the framework to the a real problem whether individual behavioral types can be captured in a body extending behavior, cap making and carrying in the sponge crab. Thus, we demonstrated that the behavioral type existed in the behavioral variables: the choice of sponge size, the trimming size, and the cavity size, but not in the latency to produce a cap. Simultaneously, we examined the conditioned models on body size, and confirmed that behavioral types remained even when conditioned on the body size differences among individuals.

Statistical modeling from the predictive point of view

Another statistical measure R^2 to quantify the intra- and inter- 'class' ('group', 'cluster' are also used interchangeably) variation has been proposed (Nakagawa and Schielzeth, 2010) and used for empirical data (e.g. Mitchell et al., 2016). The R^2 is a measure of 'variance explained' (Nakagawa and Schielzeth, 2010). The variances are measured by fitting models and taking the ratio of variance parameters. It measures the decrease of a variance of a hierarchical model against a null model. Thus, a focus of the method is on parameters defined in models.

Because of the immense complexity in the biological phenomena, it is practically effective to explore statistical models as descriptive models from the predictive point of view which focuses a distribution of a new sample conditioned on the obtained samples (Akaike, 1985). We took the view because we considered it is a straightforward and effective way to model the variability, instead of the focus only on the obtained samples such as R^2. Akaike (1985) established the foundation of the framework and the information criterion AIC is widely used in many scientific fields. However, it is inappropriate for mixed models and hierarchical models. The reason is that an essential step of calculating AIC, the maximum likelihood estimation is inappropriate for those models (Watanabe, 2005). For the hierarchical models, an important assumption does not hold. Specifically, asymptotic normality of the maximum likelihood estimator does not hold, because the "mapping from parameters to probability distributions is not one-to-one and the set of parameters whose Kullback informations are equal to zero has singularities" (Watanabe, 2005). In the situation, AIC does not correspond to the generalization error that we want to minimize when we consider appropriateness of a model from the predictive point of view. However, recently, the mathematical basis of the measure WAIC that is widely applicable is established even in the situation (Watanabe, 2010b, 2018).

Although WAIC is beginning to be used for evaluating models with empirical data, we should be careful to compute the value on a hierarchical model. Watanabe (2018) introduces two different definitions of WAIC depending on two different predictions. The often-used definition of WAIC for the hierarchical model is the first case in the book:

$$p_{posterior}(w_k|(x_k)^l) \propto \varphi(w_k|w_0) \prod_{l=1}^{N_{act}} p_{model}(x_{lk}|w_k) \tag{25}$$

$$WAIC_k = -\frac{1}{N_{act}} \sum_{l=1}^{N_{act}} \log \mathbb{E}_{w_k}[p_{model}(x_{lk}|w_k)] + \frac{1}{N_{act}} \sum_{l=1}^{N_{act}} \log \mathbb{V}_{w_k}[p_{model}(x_{lk}|w_k)] \tag{26}$$

$$WAIC_{conditioned} = \sum_{k=1}^{N_{animal}} WAIC_k \tag{27}$$

$$k = 1,\ldots,N_{animal};\ l = 1,\ldots,N_{act} \tag{28}$$

where $(x_k)^l = (x_k^1,\ldots,x_k^{N_{act}})$ represents all given data for an individual. Note that the number of behavioral acts of the k-th animal is the same (balanced) for simplicity here (The number is unbalanced in our data). It should be noted that the statistical model $p_{model}(x_{lk}|w_k)$ is conditioned upon w_k assigned to each individual. In other words, this model focuses on a new act of the already obtained individuals. Instead, to compare the models, our focus is usually on a new act of another individual. In many cases, regarding 'animal personality', we are not usually interested in how our model explain the sampled animals itself. Instead, we are interested in a distribution of a focused species. This is the reason why we did not use this conditioned-level WAIC. However, it is used with a software without the consideration on this point. For example, Mitchell et al. (2016) uses conditioned-level WAIC to compare a hierarchical model with a non-hierarchical model. Furthermore, using the measure R^2 for the evaluation of the model in terms of 'variance explained' is focused on how we can minimize the variation in the sample obtained, not focused on the prediction of the true distribution (Nakagawa and Schielzeth, 2010). We took the predictive point of view and explored a better model not the 'correct' model, because it would be a natural assumption that a distribution we build never be correct in any empirical modeling.

Making cost and size choice: why the crab skipped carrying sponge?

The crabs in our experiments that did not carry caps were larger than those that carried caps. One possibility for the reason would be that when they grow up to some extent, the predators might avoid the crab and the relative energetic cost to make caps might increase. We speculate that this might be a reason why the large crabs did not make nor carry the caps.

Another possibility is that the sponges used in this experiment were smaller than those of the necessary size for the crabs. Dembowska (Dembowska, 1926) reported that the proportion of a cap to a size of *D. personata* tended to decrease with the increase of the size of the crabs, and considered that this was because there were few sponges that fit the large crabs. Similarly, the large crabs that skipped cap making and carrying, would carry a cap if a sponge size would be larger than the L size sponge. In addition, no individuals carried the S sponge, because it was too small for all of the crabs to carry. It is likely that a younger and smaller crab than that used in this experiment would carry the S sponge.

Difference of individual behavioral type among behavioral aspects

We focused a body extending behavior by the sponge crab, because to make the living or non-living materials suitable to the animal body, the individual appears to rely on plastic properties in the behavior. For another crustacean example of body extending, the hermit crabs are well known to prefer specific shells (Bertness, 1980; Hazlett, 1981; Wilber, 1990). Although hermit crabs cannot modify the shells by themselves, they are suggested to recognize and learn the shape of extended shells and the surrounding terrain (Sonoda et al., 2012). Therefore, the hermit crabs also might have behavioral types.

McLay (1983) showed in the *Cryptodromia hilgendorfi* the relationship of the body and cap size using a log link function and Gaussian distribution. As shown in the crab, we conditioned the variables on the carapace width of the crab *Lauridromia dehaani*. To consider 'animal personality', it is important to appropriately condition out the variables assumed to have much information about the behavioral variable. The body size is presumed to be an influential variable for cap making behavior.

Therefore, we conditioned all behavioral variables upon body size and found that the predictability improved by adding an assumption of 'individual' in all behavioral aspects. The improvement was larger in sponge choice rather than in cavity size. Because the cavity size was determined by repeated excavation and body rotation, the crab might have used the carapace as a 'measure'. However, in the

371 choice task, the information processing to measure an appropriate size would be less dependent on the
372 measure. We speculate that this makes a room for emerging of individual behavioral types dependent on
373 behavioral plasticity unique to an individual.

FIGURE LEGENDS

Figure 1. Experimental animal and setup. (A) *Lauridromia dehaani*; p—propodus of fifth pereiopod; d—dactylus of fifth pereiopod; c—chela (1st pereiopod); 2p—second pereiopod; 3p—third pereiopod; 4p—fourth pereiopod; 5p—fifth pereiopod. (B) Carapace width we measured. (C) Experimental cage floating in an aquarium tank with three different sizes of sponges. The drawings are by the authors.

Figure 2. Cap making and carrying behavior. (A) Cap making behavior. (B–C) Carrying behavior of a crab. The drawing is by the authors.

Figure 3. Sponge choice. (A) Predictive distribution on the hierarchical model 1_1 with data points of the behavioral choices, which are M or L size choices or skipping the behavior. The points connected by dotted lines represent data from the same individual. The white curved lines are ten samples from the posterior distribution in decreasing order from the highest density of a parameter representing a probability of a choice. (B) Structure of the model 1_1 in a graphical model. a_{choice_L} is a parameter assigned to each individual. The variables in the black and white ellipses represent observed data and parameters to be estimated, respectively. (C) Predictive distribution of a choice on the non-hierarchical model 1_6. (D) Structure of the model 1_6 in a graphical diagram.

Figure 4. Trimming. (A) Upper plot: the predictive distribution on the hierarchical model 2_1. The white dotted lines connect the data points from the same individual. Lower plot: the predictive distribution visualized by re-scaling the color density of the expanded area in the upper plot except for the zero in the y-axis. (B) Predictive distribution on the non-hierarchical model 2_6. (C) Outline of the trimming process from a choice of a sponge (animals larger than about 9 cm skipped the whole behavior), trimming behavior (part of animals skipped this behavior), to cavity making. The drawing is by the authors.

Figure 5. Excavated cavity in a cap. (A) Predictive distribution of a cavity size on the model 3_1. The white points connected by dotted lines are from the same individual. (B) Predictive distribution on the model 3_1. The drawing is by the authors.

Figure 6. Latency to produce a cap. (A) Outline of cap making until carrying. (B) Predictive distribution of the latency on the model 4_1. Points from the same individual are connected by dotted lines. (B) Predictive distribution on the model 4_2. The drawing is by the authors.

response variable	model	hierarchical structure	conditioning variables	link function	distribution	WAIC (nat)	dWAIC (nat)	plot
Choice	1_1	intercept_L	CW_L Leg_L CW_NO Leg_NO	softmax	categorical	-2.13	0.00	Fig.3A
Choice	1_2	intercept_L	CW_L CW_NO	softmax	categorical	-1.87	0.26	–
Choice	1_3	intercept_L	–	softmax	categorical	-0.88	1.25	–
Choice	1_4	intercept_L	Leg_L Leg_NO	softmax	categorical	-0.78	1.35	–
Choice	1_5	–	CW_L CW_NO	softmax	categorical	0.85	2.99	Fig.3C
Choice	1_6	–	CW_L Leg_L CW_NO Leg_NO	softmax	categorical	0.87	3.01	–
Trimmed size	2_1	intercept_1 intercept_2	CW Choice	logit log	ZIP	-2.08	0.00	Fig.4A
Trimmed size	2_2	intercept_2	Choice	logit log	ZIP	0.81	2.89	–
Trimmed size	2_3	intercept_2	CW Choice	logit log	ZIP	0.86	2.95	–
Trimmed size	2_4	intercept_2	–	logit log	ZIP	1.23	3.32	–
Trimmed size	2_5	intercept_2	CW	logit log	ZIP	1.37	3.46	–
Trimmed size	2_6	–	CW Choice	logit log	ZIP	7.40	9.48	Fig.4B
Trimmed size	2_7	–	CW	logit log	ZIP	10.05	12.13	–
Trimmed size	2_8	–	–	logit log	ZIP	12.55	14.63	–
Cap cavity size	3_1	intercept	CW	log	gamma	4.45	0.00	Fig.5A
Cap cavity size	3_2	–	CW	log	gamma	4.54	0.08	Fig.5B
Cap cavity size	3_3	–	CW Gender	log	gamma	4.69	0.24	–
Cap cavity size	3_4	intercept	–	log	gamma	4.71	0.26	–
Cap cavity size	3_5	–	CW	identity	normal	4.75	0.30	–
Cap cavity size	3_6	intercept cw	CW	log	gamma	6.18	1.73	–
Latency for making	4_1	intercept_2	CW	logit log	ZIP	1.10	0.00	Fig.6A
Latency for making	4_2	intercept_2	–	logit log	ZIP	1.28	0.18	–
Latency for making	4_3	–	–	logit log	ZIP	1.28	0.19	Fig.6B
Latency for making	4_4	–	Choice	logit log	ZIP	1.30	0.20	–
Latency for making	4_5	–	CW	logit log	ZIP	1.38	0.28	–
Latency for making	4_6	–	CW Choice	logit log	ZIP	1.72	0.62	–

Table 1. Summary of model structures and the predictive performances in WAIC. Abbreviations, intercept_L: intercept in the linear predictor (LP) for the choice of L; intercept_1: intercept in the LP for the decision of trimming; intercept_2: intercept in the LP for the mean of the trimmed size of the sponge; cw: slope in the LP for the carapace width; CW: carapace width; Leg: degree of the leg lack; _L and _NO: parameters for L sponge and skipping, respectively; Choice: choice whether to cut the sponge; Gender: gender of the animal; intercept_2: intercept in the LP for the mean of the days to carrying; Choice: choice of sponge size; ZIP: Zero-inflated Poisson distribution; WAIC: value of Widely-Applicable Information Criterion per a sample; dWAIC: the difference of the WAIC of the model against the best-performed model.

ACKNOWLEDGMENTS

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer.

Akaike, H. (1985). Prediction and entropy. In *Selected Papers of Hirotugu Akaike*, pages 387–410. Springer.

Akaike, H. (1998). Likelihood and the bayes procedure. In *Selected Papers of Hirotugu Akaike*, pages 309–332. Springer.

Bedini, R., Canali, M. G., and Bedini, A. (2003). Use of camouflaging materials in some brachyuran crabs of the mediterranean infralittoral zone. *Cahiers de Biologie Marine*, 44(4):375–383.

Bertness, M. D. (1980). Shell preference and utilization patterns in littoral hermit crabs of the bay of panama. *Journal of Experimental Marine Biology and Ecology*, 48(1):1–16.

Dembowska, W. S. (1926). Study on the habits of the crab *Dromia vulgaris* ME. *The Biological Bulletin*, 50(2):163–178.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Guinot, D. and Wicksten, M. K. (2015). Camouflage: carrying behaviour, decoration behaviour, and other modalities of concealment in brachyura. In *Treatise on Zoology-Anatomy, Taxonomy, Biology. The Crustacea, Volume 9 Part C (2 vols)*, pages 583–638. Brill.

Hazlett, B. A. (1981). The behavioral ecology of hermit crabs. *Annual Review of Ecology and Systematics*, 12(1):1–22.

Hunt, G. R. (1996). Manufacture and use of hook-tools by new caledonian crows. *Nature*, 379(6562):249.

Matsui, H. and Izawa, E.-I. (2017). Flexible motor adjustment of pecking with an artificially extended bill in crows but not in pigeons. *Royal Society Open Science*, 4(2):160796.

McLay, C. L. (1983). Dispersal and use of sponges and ascidians as camouflage by *Cryptodromia hilgendorfi* (Brachyura: Dromiacea). *Marine Biology*, 76(1):17–32.

Millar, R. B. (2018). Conditional vs marginal estimation of the predictive loss of hierarchical models using WAIC and cross-validation. *Statistics and Computing*, 28(2):375–385.

Mitchell, D. J., Fanson, B. G., Beckmann, C., and Biro, P. A. (2016). Towards powerful experimental and statistical approaches to study intraindividual variability in labile traits. *Royal Society open science*, 3(10):160352.

Nakagawa, S. and Schielzeth, H. (2010). Repeatability for gaussian and non-gaussian data: a practical guide for biologists. *Biological Reviews*, 85(4):935–956.

Niemelä, P. T. and Dingemanse, N. J. (2018). On the usage of single measurements in behavioural ecology research on individual differences. *Animal behaviour*, 145:99–105.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reinhart, A. (2015). *Statistics done wrong: The woefully complete guide*. No starch press.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 81.

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

Sonoda, K., Asakura, A., Minoura, M., Elwood, R. W., and Gunji, Y.-P. (2012). Hermit crabs perceive the extent of their virtual bodies. *Biology letters*, 8(4):495–497.

Stan Development Team (2018). *Stan Modeling Language Users Guide and Reference Manual Version 2.18.0.*

Wakita, D., Kagaya, K., and Aonuma, H. (2019). Generalized locomotion of brittle stars with an arbitrary number of arms. *bioRxiv*, page 616383.

Wang, L., Brodbeck, L., and Iida, F. (2014). Mechanics and energetics in tool manufacture and use: a synthetic approach. *Journal of the Royal Society Interface*, 11(100):20140827.

Watanabe, S. (2005). Algebraic geometry of singular learning machines and symmetry of generalization and training errors. *Neurocomputing*, 67:198–213.

Watanabe, S. (2010a). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Watanabe, S. (2010b). Equations of states in singular statistical estimation. *Neural Networks*, 23(1):20–34.

Watanabe, S. (2018). *Mathematical theory of Bayesian statistics*. Florida, USA: CRC Press.

Wilber, T. (1990). Influence of size, species and damage on shell selection by the hermit crab *Pagurus longicarpus*. *Marine Biology*, 104(1):31–39.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York, USA: Springer.
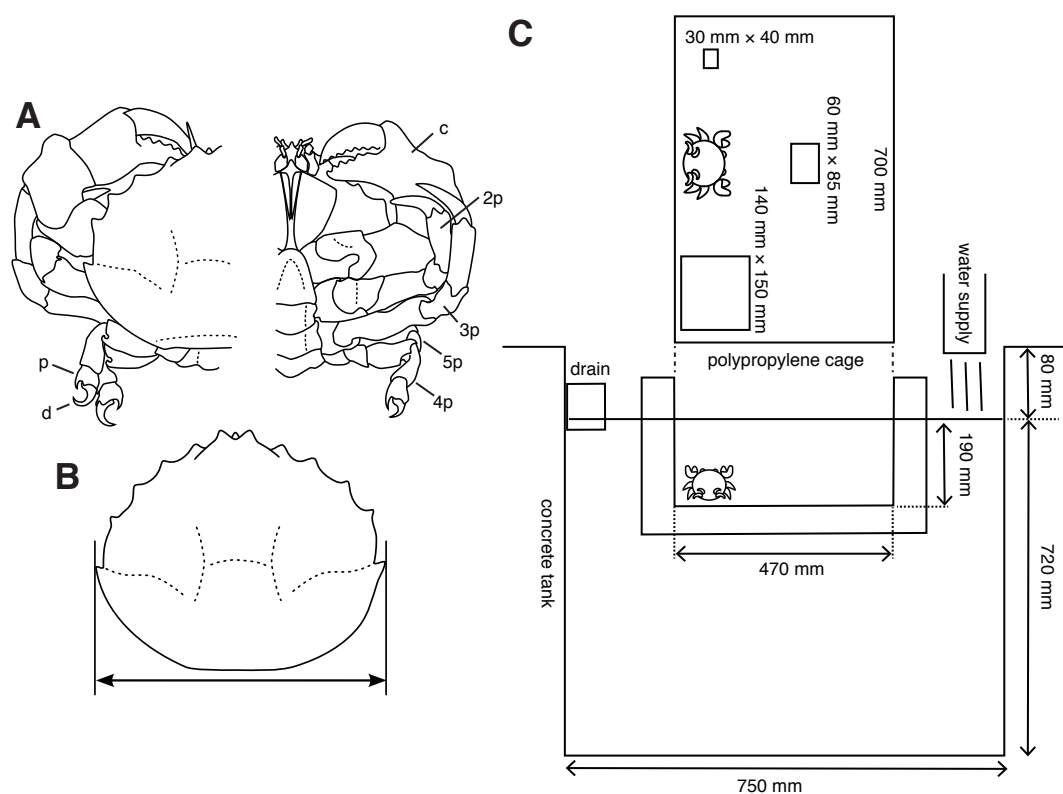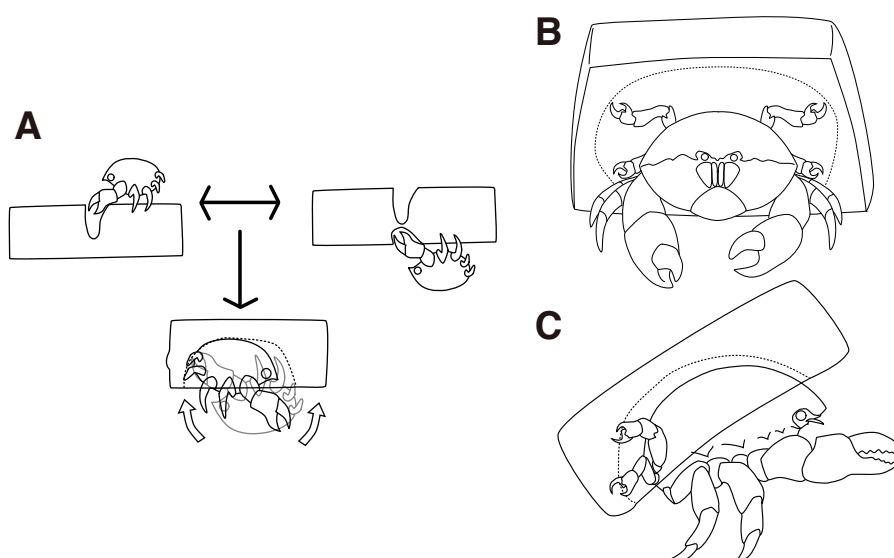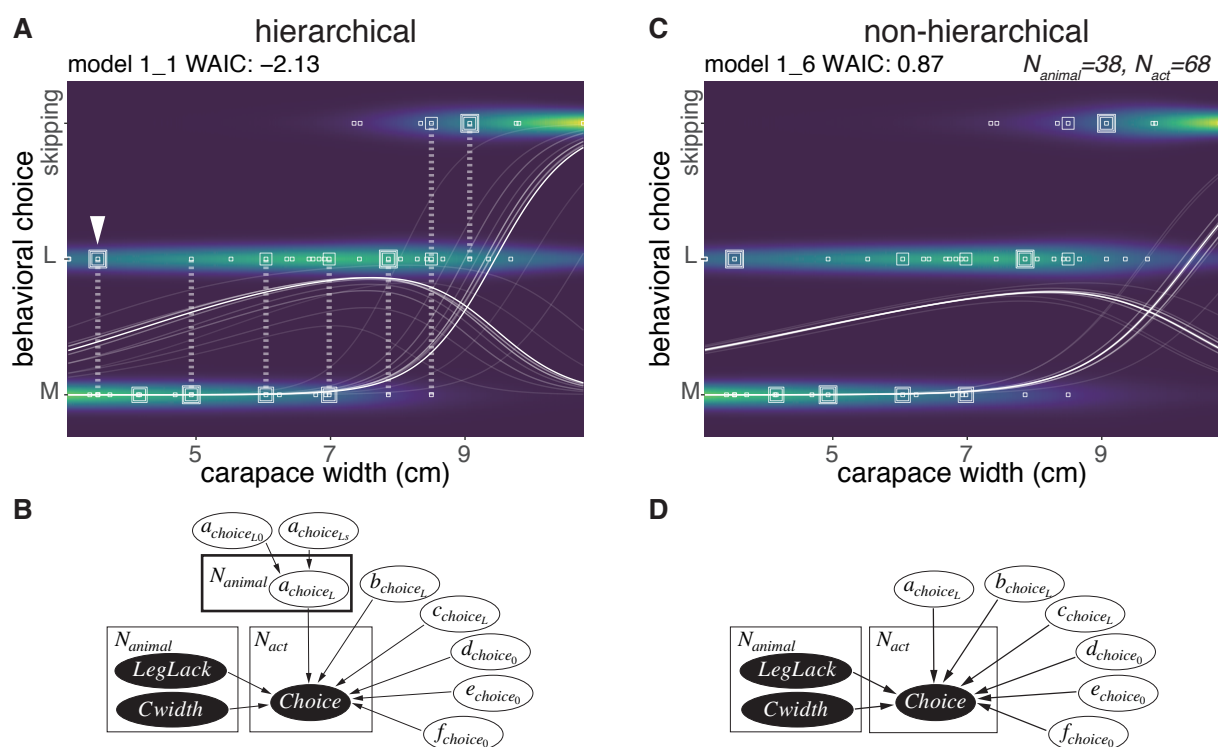
Figure 1
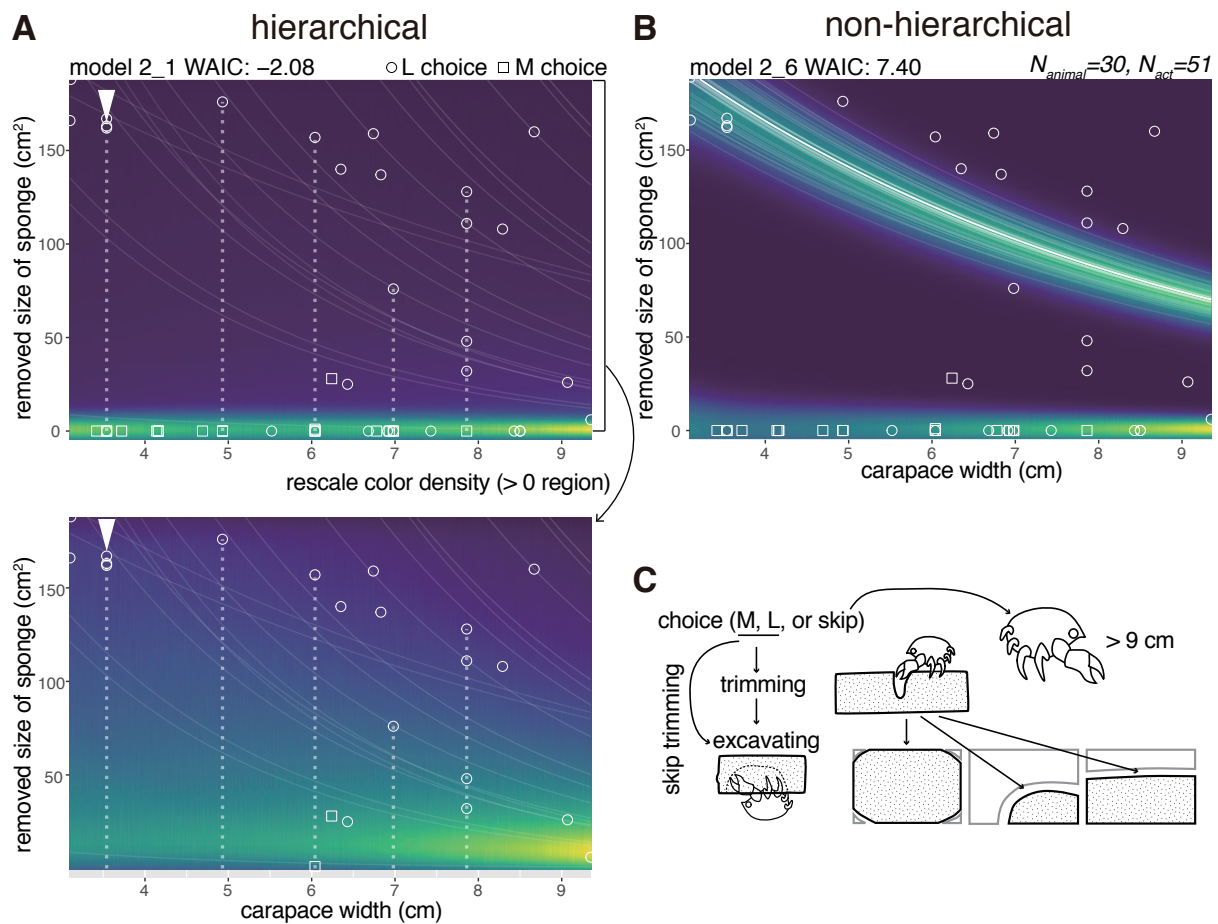
Figure 2

Figure 3

A hierarchical
model 1_1 WAIC: −2.13

C non-hierarchical
model 1_6 WAIC: 0.87 $N_{animal}=38, N_{act}=68$

B

D

Figure 4

Figure 5

Figure 6