How Red Is a Ladybeetle? Examining People's Notions of Biological Variability

Pablo León-Villagrá[1] **Olympia Mathiaparanam**[2] **Karl Rosengren**[2] **Daphna Buchsbaum**[1]
[1]Brown University, USA [2]University of Rochester, USA

Abstract

People often display essentialist biases, which can lead them to underestimate within-species variability. This bias is especially pronounced when traits are described as advantageous for survival. However, it is unclear whether this bias is limited to the specified trait or encompasses complex trait interactions. We used Markov Chain Monte Carlo with People ($MCMC_p$) to analyze people's representations of biological variability, using ladybeetles as a model species. Participants either received contextual information about the benefits of ladybeetle color for survival, or survival-irrelevant information. Overall, participants held consistent beliefs about ladybeetle features, but those with survival-relevant context produced lighter and larger ladybeetles; this difference was consistent with survey responses. However, we found no significant interaction between $MCMC_p$ variability and essentialism scores, given our context manipulation. We discuss potential explanations for these results and highlight advantages of $MCMC_p$ for assessing biological variability, particularly when studying the development of essentialist biases.

Keywords: Biological reasoning; biological variability; categories; psychological essentialism; MCMCp

Introduction

People are biased to assume that groups, such as biological species, are more uniform in their features than group members are in reality(S. A. Gelman & Rhodes, 2012). This essentialist bias poses challenges to people's ability to represent and understand variability among group members (biological variability, the diversity among members of the same biological species) and impacts how people reason about biological concepts that rely on an understanding of variability, including evolution (Emmons & Kelemen, 2015; Shtulman & Schulz, 2008). Here, we investigate the connection between adults' essentialist biases and their reasoning about the extent and dimensions of biological variability within species. Additionally, we explore the influence of context (i.e., information about a trait's function) on this relationship.

Psychological essentialism is a cognitive bias that leads people to assume a shared, unchangeable essence underlying group membership (S. A. Gelman, 2003). As a result, people perceive biological groups, such as animal species, as more homogeneous than they are in reality (S. A. Gelman & Rhodes, 2012). Consequently, essentialist reasoning tends to diminish people's acceptance of naturally occurring variability within species. However, Emmons and Kelemen (2015) demonstrated that essentialist biases are influenced by learning and development. When students learn about a trait's advantage for species survival, they endorse less within-species variability. This is likely because students reason that traits serve to fulfill a species's survival needs, and so they expect all members to have these beneficial traits.

The effect of knowledge on essentialist biases is important since learning about biological variability plays a crucial role in understanding biology, serving as a foundation for more advanced topics (AAAS, 2011). Recognized as a threshold concept, mastering biological variability is central for learning more advanced biological processes like natural selection (Walck-Shannon, Batzli, Pultorak, & Boehmer, 2019).

While previous research suggests that students' essentialist biases are more flexible than previously thought (e.g., being influenced by factors like species familiarity, Emmons & Kelemen, 2015), it remains unclear whether these biases need to be primed in order for them to emerge in student reasoning, and if they are limited to specific primed traits or extend to interactions between multiple related traits. Extending knowledge from particular traits to others (e.g., inferring a correlation between the ladybeetle's body color and its number of spots) could be rational, given that contiguous genes often covary, resulting in covarying phenotypes (Morgan, 1911). We explore this question, hypothesizing that essentialist biases will be primed by contextual information that emphasizes the advantage of a specific trait for species survival. This bias can go beyond the single trait highlighted, leading to a reduction in the accepted variability not only for that specific trait but also for other related traits.

Building on Emmons and Kelemen (2015) we investigate whether notions of variability encompass complex interactions among traits and their impact on a species' survival. For instance, when students learn that a ladybeetle's survival relies on signaling toxicity through a distinctive red body color to evade predation, this may lead students to believe that all ladybeetles must share the exact same color. We examine if students, influenced by this bias, produce these expectations for other features related to signaling toxicity (e.g., number of spots) or other traits broadly related to survival (e.g., size).

To test these complex trait interactions, we use a method inspired by statistical methods, Markov Chain Monte Carlo with People ($MCMC_p$). This experimental paradigm, introduced by Sanborn, Griffiths, and Shiffrin (2010), infers how representative individual category members are to a person (their probability of belonging to a particular category).

MCMC$_p$ is based on the statistical method Markov-Chain Monte Carlo (MCMC; for a more technical description of MCMC, see, e.g., MacKay, 2003). In MCMC, and thus MCMC$_p$, a distribution can be approximated by constructing a chain of samples from the distribution from a simple repeated forced-choice (for methodological arguments for the psychological validity of MCMC$_p$, see Sanborn et al., 2010). In each choice, the previous state is compared against a novel, proposed state. If the proposed state is accepted, it becomes the new current state; if it is rejected, the previous state is carried forward. In MCMC, the proposed state is accepted or rejected according to its likelihood. In MCMC$_p$, this decision is made by the participant. Since the decisions in MCMC and MCMC$_p$ are biased towards more likely, or better, states, over time, MCMC and MCMC$_p$ move closer to representative states for the category. Moreover, once MCMC and MCMC$_p$ are close to these representative states, the algorithm will continue to explore, producing states proportional to their likelihood, or representativeness, of the category.

As a method to elicit students' beliefs of biological variability, MCMC$_p$ has multiple potential advantages. First, MCMC$_p$ produces a larger number of judgments, is simpler, and is more flexible than previous probabilistic elicitation methods (for example, Menendez, Rosengren, & Alibali, 2022). Furthermore, previous studies were limited in the range of individual category members for which they could obtain membership ratings since all to-be-tested materials had to be specified by the experimenters, and each item had to be rated by each participant (often multiple times to ensure consistent judgments). Instead, MCMC$_p$ adaptively explores a *space* of potential materials. This flexibility also allows us to explore more naturalistic (or complex) materials.

Motivation

The present study examined students' reasoning about the biological variability of ladybeetles through the MCMC$_p$ method. Ladybeetles offer a convenient model as they display readily observable within-species variability. Additionally, while many are familiar with ladybeetles, more specific details about their biological properties (e.g., metamorphosing) are not widely known. Building on previous findings (Emmons & Kelemen, 2015), we predicted that presenting traits as advantageous for a species' survival would reduce the variability that students are willing to accept both for the specified trait as well as for other related traits. Moreover, we assessed the student's knowledge of biology and the strength of their pre-existing essentialist biases in a survey that included established essentialism measures.

We predicted that individuals with a stronger essentialist bias (indicated by lower survey scores) would correspondingly demonstrate less trait variability in the species examples they generate with MCMC$_p$. In contrast, we predicted that participants with higher biological knowledge would produce more trait variability.

Experiment

Participants

Participants ($N = 58$) were recruited from the participant pool of a private North American university for course credit and were randomly assigned to one of two experiment conditions (42 female, 13 male, and 3 non-binary, $M_{age} = 19.93$, $SD_{age} = 1.04$, 20 identified as Asian, 20 as White, 5 as Hispanic or Latino, 3 as Black, and 10 identified as multi-racial). Participants took approximately 20 minutes to complete each block (Median survey = 19.11 minutes, median MCMC$_p$ = 21.95 minutes).

According to our preregistration[1], participants were excluded if they did not complete the total number of MCMC$_p$ trials ($n = 9$), selected one or more lures in the attention-checks ($n = 1$), or exhibited a single-side bias in the MCMC$_p$ trials ($n = 6$) or a consistent alternating pattern ($n = 4$)[2]. The total number of exclusions was 17 since participants were often excluded on multiple criteria. We dropped the last extra participant to accomplish the preregistered group size of 20 participants in each group.

Materials

MCMC$_p$ items: The experiment was conducted using a custom web application on a tablet. The stimuli were variations of a stylized, colored drawing of a ladybeetle. These drawings were programmatically altered by scaling the width (*width*), height (*height*), or changing spot patterns (*spots*), or body color (*hue*, *chroma*, *lightness*), see Figure 1.

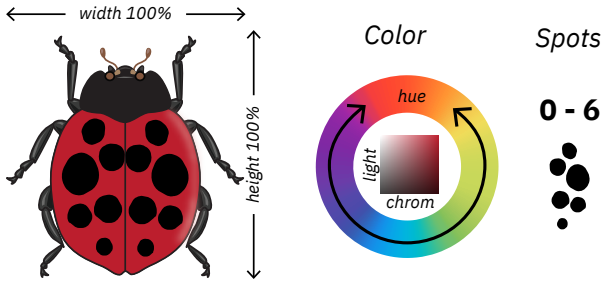

Figure 1: Stimuli were ladybeetle drawings that could be varied by manipulating six parameters. *Width* and *height* parameters scaled the beetle, changing its size and body shape. The *spots* parameter altered the spot pattern, displaying a mirrored number of spots on each side of the beetle's body. The beetle's body color was manipulated by changing *hue*, *chroma*, and *lightness*.

We chose to represent color in CIELCh space, given that it is close to perceptual uniformity. Parameters spanned 40 %

[1] https://osf.io/acht4

[2] Determined by generating bootstrapped distributions from the population and comparing each participants' frequency of one-sided or alternating choices against this distribution. We exclude participants that produced values with $p < .01$.

to 100 % for *width* and *height*, 0 to 6 spots, 0° to 360° *hue*, 0 to 180 *chroma*, and 0 % to 100 % *lightness*.

To allow participants to assess the width and height of the ladybeetle, a reference image of a pencil was shown in all MCMC_p trials; see Figure 2.

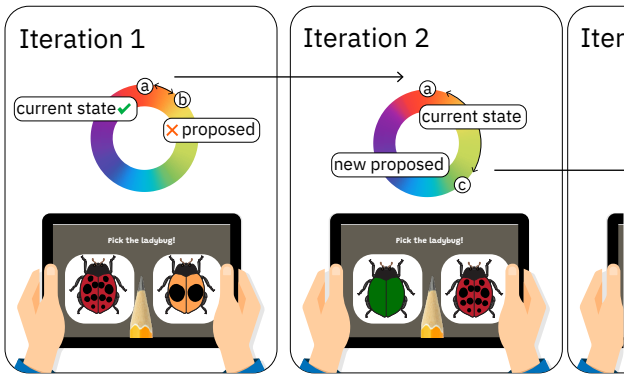

Figure 2: Illustration of the MCMC_p procedure. In each iteration, participants selected between two ladybeetle drawings. These drawings correspond to the previous choice and a proposed update (a, b; a, c). Once participants selected an option (a), it was carried forward to the next iteration and a new proposed update was generated (c).

Survey items: The survey assessed participants' biology knowledge and familiarity with ladybeetles. We also designed a set of 6 items to gauge participants' essentialist reasoning. We created four questions inspired by Emmons and Kelemen (2015), which we will refer to as $\text{Species}_\text{var}$ items. These items presented participants with one homogenous and one heterogenous set of four organisms. Participants were tasked with selecting the set they believed best represented the species shown (ladybeetle, blue weevil, monarch butterfly, blue morpho butterfly). We also included two items similar to the essentialism questions in Shtulman and Schulz (2008), which we call $\text{Evolution}_\text{var}$. Each item introduced a species (ladybeetle, grasshopper) and a crucial trait for its survival (e.g., a green body color aiding a grasshopper's camouflage). Subsequently, we presented an environmental pressure (e.g., drought leading to resilient, yellow grass dominating the once greener grasshopper habitat) or the absence of such pressure and examined participants' reasoning about how the species might change.

Procedure

The experiment consisted of two blocks: a MCMC_p block (completed on an Amazon Fire 7 tablet with a 7" display) and a subsequent survey (completed on a provided 13" 2020 MacBook Pro). For a short demonstration of the MCMC_p experiment, see `http://tinyurl.com/mcmc-p`. Before completing the experimental blocks, participants read a cover story about ladybeetles. The content of this story varied based on the conditions to which participants were randomly assigned. In the color condition, participants learned about how ladybeetles utilize body color to signal their toxicity to predators, thereby avoiding predation. In the control condition, participants learned about another beneficial feature of ladybeetles (i.e., antennae) that was not varied in the MCMC_p activity.

MCMC_p block: Participants first completed 600 iterations of the MCMC_p procedure. In each iteration, participants had to select the better of two options—two ladybeetles that were presented equidistant from the center of the screen. After each choice, the beetles were hidden, and the procedure continued after a short pause of one second. Following previous MCMC_p experiments, we split the total of 600 iterations into three chains of 200 and interleaved the presentation of each chain (e.g., Sanborn et al., 2010). Therefore, participants were presented with a sequence of choices: $\textit{chain}_1$, $\textit{chain}_2$, $\textit{chain}_3$, $\textit{chain}_1, \ldots$. Within each chain, on each trial, the two choice options always corresponded to the participants' previous choice in that chain, and the other option was a proposed update, with presentation sides randomized. Interleaving multiple chains has the advantage of making it less obvious to participants that their previous choices carried forward, as well as allowing for a more robust assessment of the statistical properties of the resulting MCMC_p data. Each of these chains was initialized at random for each participant, with parameters sampled uniformly from the parameter ranges (e.g., any possible color, shape, or number of spots). Afterward, the participants' choices were carried forward, and new options were generated by adding Gaussian distributed noise to each parameter apart from *spots*. This Gaussian was centered at zero with a variance σ^2, equal to 7 % of the parameter range, as in Sanborn et al. (2010). Since *spots* was a discrete variable, we sampled updates from [-2, 1, 1, 2], with probabilities of $1/10$, $4/10$, $4/10$, and $1/10$.

To increase the speed at which participants moved away from the random starting points, the first 70 iterations in each chain used two additional proposal functions that allowed for larger updates. One proposal function generated updates by sampling ladybeetle parameters uniformly in the parameter range. The second proposal function generated color parameters uniformly and updated the remaining parameters using the Gaussian proposals. In each step, one of the three proposal functions was sampled with equal probability, and the proposed ladybeetle was generated. These iterations are commonly regarded as *burn-in* and were discarded from the main analysis.

As in previous MCMC_p experiments, proposals outside the admissible parameter ranges were automatically rejected and not shown to participants. In these cases, we recorded the current state and the out-of-bounds proposal and counted it as an automatic rejection. Then, a new proposal was generated until the parameters were within the admissible range. These automatic rejections did not count towards the total of 600 trials that each participant completed.

Finally, four attention trials were used to check that participants were attentive. In these trials, a stereotypical ladybeetle

(from pilot experiments) and a color image of a paper kite butterfly were presented following the same procedure and presentation as the experiment trials. These trials were presented at random points during the experiment and did not contribute to the total number of trials. To help maintain participant motivation, we displayed a progress bar under the two ladybeetle options throughout the experiment.

Survey

Following the completion of the MCMC_p block, participants completed an online Qualtrics survey. The survey covered basic demographics, biology knowledge, and familiarity with ladybeetles. Participants were then asked to briefly describe their strategy for selecting ladybeetles in the preceding MCMC_p activity. Participants then completed our essentialist reasoning questions. The survey structure ensured that participants always completed the $\text{Species}_\text{var}$ items first, followed by the $\text{Evolution}_\text{var}$ items. Within each set of essentialism items, the question order was randomized.

Results

Before examining the distributions produced by MCMC_p, it is important to assess if the procedure has successfully moved from the random starting points toward the participants' intended category, as well as how effective MCMC_p was at exploring the range of plausible category members. Following standard MCMC procedures in applied statistics, we report these as MCMC_p convergence statistics.

Convergence Statistics

One measure of the efficacy of MCMC_p is the number of times a proposed update is accepted. Low acceptance values can suggest that the method was ineffective at exploring the distribution since the current state is rarely updated. In contrast, high acceptance rates can indicate that proposals are too wide, potentially reducing the ability to explore small local variations of the category.

In our experiment, proposals could be accepted by the participant or automatically (if the parameters were out of bounds). Since participant acceptances highlight the efficacy of an experimental method, and overall acceptances characterize the effectiveness of the MCMC algorithm, we report both. Overall acceptances ($M_{acc} = 10\%$, $SD_{acc} = 6\%$) were lower than the 20 % to 40 % suggested in the statistical MCMC literature (A. Gelman, Gilks, & Roberts, 1997), but comparable to previous MCMC_p experiments. Acceptance rates excluding automatic rejections were higher ($M_{acc} = 15\%$, $SD_{acc} = 9\%$). Overall acceptances and participant acceptances did not differ significantly across conditions ($U_{all} = 180.0, p = .59$, $U_{participant} = 175.0, p = .51$)[3].

A second important measure to determine the effectiveness of MCMC and MCMC_p is $\hat{R}$ (Brooks & Gelman, 1998). Intuitively, $\hat{R}$ captures how consistent individual MCMC chains evolve over time, compared to how consistent MCMC chains evolve across MCMC chains. Ideally, this ratio is very similar, resulting in values of $\hat{R}$ close to 1.0, while considerably larger values can indicate that the individual chains evolve very differently. Unsurprisingly, human $\hat{R}$ values were higher than the theoretical ideal of MCMC runs ($M = 1.8, SD = .89$). However, while the average $\hat{R}$ was high, most participants had significantly lower values (Median $\hat{R} = 1.58, min = 1.02, max = 8.28$). Furthermore, these values were comparable to previous MCMC_p experiments (León-Villagrá, Otsubo, Lucas, & Buchsbaum, 2020), and $\hat{R}$ did not differ significantly across conditions ($U = 11971.5, p = .32$).

Finally, since MCMC_p is an iterative procedure, each new state depends on the previous state, and thus the resulting samples are not independent. Therefore, the number of samples obtained by running n iterations of MCMC_p is often considerably smaller. To assess the number of independent samples that MCMC_p produced, we report the effective sample size, ESS (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2021). Across conditions and ladybeetle parameters, we obtained ESS comparable to previous MCMC_p experiments ($M = 12.62, SD = 17.54$), and ESS did not differ across conditions ($U = 11947.5, p = .29$).

Since the first MCMC_p trials are biased toward the random starting points (and we used these trials to adopt more explorative proposal functions; see Procedure), we dropped the first 70 trials in each chain, similar to Sanborn et al. (2010). This fixed burn-in period was established in pilot experiments, as 70 trials were usually sufficient to move the MCMC_p chains away from the initial starting point and towards the region of values corresponding to the participants' ladybeetle category. All subsequent analyses of the MCMC_p results were based on the trials after burn-in.

Students' Notions of Biological Variability

To evaluate participants' perceptions of biological variability, we examined their performance on essentialist reasoning questions. For the $\text{Species}_\text{var}$ items, endorsements of the homogeneous (no variability) set were coded as -1, while endorsements of the heterogeneous (variability) set were coded as $+1$. For the $\text{Evolution}_\text{var}$ items, we applied their coding scheme. Responses indicating essentialist reasoning or suggesting no species adaptation to the new environmental conditions were coded as -1. Responses exemplifying a differentiated survival style of reasoning were coded as $+1$. Two coders independently coded the $\text{Evolution}_\text{var}$ item responses. Their percentage of agreement across 116 responses was 93.1%. Disagreements were resolved through discussion. For an example of the materials and the coding scheme, see `https://osf.io/m8jrq/`.

Participants demonstrated diverse biological variability reasoning in their responses to the two essentialist item sets. For the $\text{Species}_\text{var}$ items, we identified participants who consistently endorsed the heterogeneous option ($n = 14$), as well as those who consistently favored the homogeneous option across all four questions ($n = 3$). The majority of partici-

[3] We report nonparametric two-sided Mann-Whitney tests across this analysis. Two-sided t-tests produced very similar results.

pants tended to endorse the heterogeneous option on 3 out of 4 questions ($M = 2.00, SD = 2.17$).

Similarly, for $\text{Evolution}_{\text{var}}$ items, we observed a range of responses, with some participants displaying highly essentialist reasoning for both items and others aligning more with a variability perspective ($M = 0.37, SD = 1.44$). Across both tasks, participants fell into one of three distinct categories: those classified as essentialist (scoring -1 on both items; 15%), those endorsing variability strongly (scoring $+1$ on both items; 33%), and those in between (scoring either 0 on both items or a combination of -1 and $+1$; 25%).

Unsurprisingly, the $\text{Evolution}_{\text{var}}$ items correlated significantly with participants' self-reported biology knowledge ($r(38) = .46, p < .001$) and the number of biology courses taken ($r(38) = .55, p < .001$). However, the $\text{Species}_{\text{var}}$ scores did not correlate significantly with biology knowledge, courses, or the $\text{Evolution}_{\text{var}}$ score.

Ladybeetle Exemplars

Since MCMC_{p} convergence statistics indicated satisfactory efficiency and sample sizes, we examined collections of MCMC_{p} acceptances, or posterior distributions, see Figure 3. Because the shape and size of the ladybeetle were important perceptually, we discuss *width* and *height* parameters as $size = width \times height$ and $proportion = \frac{height}{width}$ in all subsequent sections.

Across conditions, we found very consistent parameter distributions that were in line with our expectations about characteristic ladybeetle features. Both conditions produced unimodal hue distributions centered on red, with average spot numbers, small-to-average sizes, and roughly proportional shapes. To describe the resulting distributions further, we determined modes and mean absolute deviations (*AD*) from the mode by fitting kernel-density estimations[4].

This analysis corroborated the description of the posterior distributions. Both conditions produced ladybeetles with similar spot numbers ($\text{Mo}_{control} = 4$, $AD_{control} = 3.34$, $\text{Mo}_{color} = 4$, $AD_{control} = 3.55$), hues ($\text{Mo}_{control} = 36.86$, $AD_{control} = 60.15$, $\text{Mo}_{color} = 40.08$, $AD_{control} = 72.22$), and body proportions ($\text{Mo}_{control} = 1.15$, $AD_{control} = 0.21$, $\text{Mo}_{color} = 0.98$, $AD_{control} = 0.3$). However, ladybeetles in the color condition were somewhat larger ($\text{Mo}_{control} = 0.29$, $AD_{control} = 0.67$, $\text{Mo}_{color} = 0.54$, $AD_{control} = 0.46$), less saturated ($\text{Mo}_{control} = 163.44$, $AD_{control} = 113.46$, $\text{Mo}_{color} = 147.29$, $AD_{control} = 118.31$), and brighter ($\text{Mo}_{control} = 15.39$, $AD_{control} = 38.94$, $\text{Mo}_{color} = 35.49$, $AD_{control} = 43.39$). Importantly, counter to our prediction, AD_{hue} was not lower in the color condition.

Consistent with these results, only participants in the color condition reported that the most important feature for deciding which ladybeetle to select in the MCMC_{p} task was brightness ($n = 3$), while all other features were comparable (spots most important feature: $n_{color} = 13, n_{control} = 12$, the color most important: $n_{color} = 15, n_{control} = 19$, size $n_{color} = 1, n_{control} = 2$, shape $n_{color} = 2, n_{control} = 4$)[5].

MCMC_{p} and Essentialism Scores

To test the relationship between established essentialism scores and the posterior distributions produced by MCMC_{p}, we calculated participant-level modes and *AD*, following the same kernel-density approach we employed for the overall distributions. Irrespective of the experimental condition, essentialism scores did not correlate with *AD*, as hypothesized. Only the $\text{Evolution}_{\text{var}}$ scores adapted from Shtulman and Schulz (2008) significantly correlated with AD_{hue}, albeit in the opposite direction that we hypothesized ($r(38) = -0.36, p = .02$).

Similarly, our experimental manipulation did not result in a decrease of AD_{hue} as hypothesized, nor did it decrease $AD_{lightness}$, all $p > .13$.[6] This lack of difference in conditions was also apparent in the scores: While participants in the color condition scored slightly lower on essentialist bias in the $\text{Evolution}_{\text{var}}$ survey items, this difference was not significant ($M_{color} = .55$, $SD_{color} = 1.32$, $M_{control} = .2$, $SD_{control} = 1.58$, $U = 225, p = .49$). Moreover, participants across conditions scored similarly on all other measures, $\text{Species}_{\text{var}}$ $M_{control} = 2$, $SD_{control} = 1.84$, $M_{color} = 2$, $SD_{color} = 2.51$. Across both conditions, participants judged their biological knowledge similarly above average ($M_{control} = .85$, $SD_{control} = 1.46$, $M_{color} = .85$, $SD_{color} = 1.23$, from -3 (far below average) to 3 (far above average)), and reported that they had attended similar numbers of biology courses ($M_{control} = 2$, $SD_{control} = 2.22$, $M_{color} = 2.6$, $SD_{color} = 2.5$).

Discussion

In this study, we utilized MCMC_{p} to investigate how adults perceive biological variability, focusing primarily on understanding the relationship between the variability measured by MCMC_{p} and an individual's essentialist reasoning patterns.

Using MCMC_{p}, we found that across conditions, participants had very consistent beliefs about ladybeetle features. Both conditions agreed on the particular hue, number of spots, and overall proportion of ladybeetles.

However, overall, participants in the color condition produced ladybeetles with increased lightness and larger size, suggesting that other features than hue might have been inferred to convey an adaptive advantage. Moreover, we found no indication that participants who received information about the adaptive advantage of ladybeetle color produced less varied hue distributions compared to our control condition, and neither did established essentialism scores differ across conditions.

[4] Using von Mieses kernels for *hue*, and Gaussian kernels for all other parameters.

[5] Participants provided these reasons in an open response form and could provide several "most important" features. Participants provided an average of two responses, and the number of responses did not differ across conditions ($M_{color} = 1.85, SD_{color} = .99, M_{control} = 2.0, SD_{control} = .86, U = 226.5, p = .45$

[6] Testing the model hue_{AD} $condition \times Species_{\text{var}} \times Evolution_{\text{var}} \times bioknowledge$.

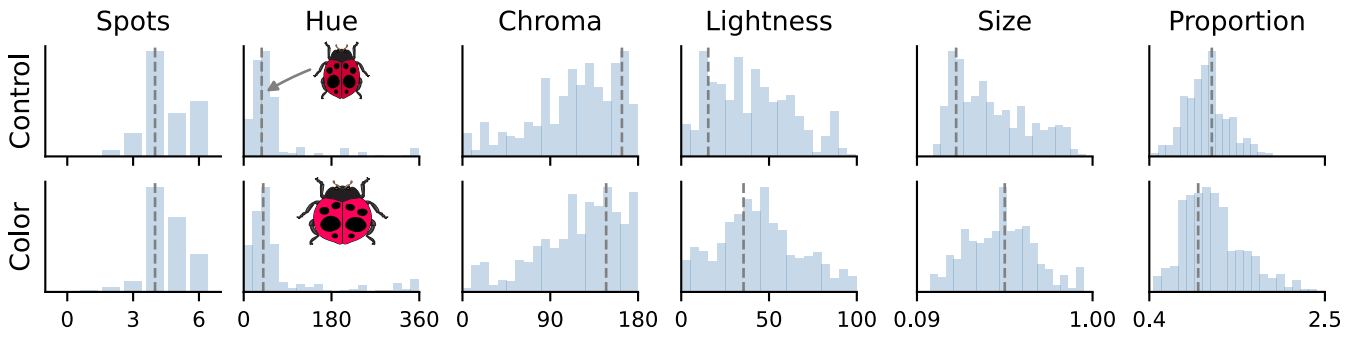

Figure 3: Posterior distributions over ladybeetle parameters resulting from the $MCMC_p$ procedure. Across both conditions, parameter distributions were remarkably similar, with similar hue distributions, spot numbers, and proportions. Overall, participants in the color condition produced slightly higher lightness distributions (modes as dashed lines) and larger ladybeetles. The resulting representative ladybeetles (see small figures) exhibited a visible increase in lightness and scale.

These findings are inconsistent with prior literature, which has reported significant relationships between essentialist beliefs, context about a trait's adaptive benefit, and biological variability reasoning (Emmons & Kelemen, 2015; Shtulman & Schulz, 2008). These prior studies typically indicate that essentialist beliefs, along with context, tend to prime essentialist biases and, as a result, minimize accepted biological variability. It is important to highlight that most previous literature has examined these factors in younger school-age children (Emmons & Kelemen, 2015; S. A. Gelman, Ware, Manczak, & Graham, 2013), although the literature illustrates these trends are also present in adults (Shtulman & Schulz, 2008).

There are several possible explanations for our findings. In line with previous experimental designs, our instruction did not specify which color feature (e.g., hue, chroma, lightness) was implied to signal toxicity and aid in ladybeetle survival. Despite participants from both conditions agreeing that ladybeetles were typically red (with very similar hue distributions and modes), unexpectedly, some participants inferred that *lighter* red colors (rather than redder colors) indicated toxicity. If a subset of participants in the color condition inferred a different adaptive mechanism (lightness rather than hue), this could have weakened any potential essentialist bias on hue even further. In addition, we chose ladybeetles as a model species due to their use in prior work and readily noticeable within-species variability. However, previous studies have suggested that species and trait familiarity may influence notions of biological variability (Emmons & Kelemen, 2015). Adults in both conditions were already familiar with ladybeetle coloring and may have also been aware of the adaptive benefit of ladybeetle coloring before participating in our study, given their high biology knowledge and the number of biology university courses they attended. Thus, it is possible that providing this information did not change their context, reducing the effectiveness of our experimental manipulation. To address this potential confound, in future studies, we aim to test less familiar animal traits and screen participants' knowledge before entering the experiment.

While these issues limit the interpretation of our results, they illustrate the potential of adopting $MCMC_p$ to study notions of biological variability. Previous studies were constrained by small sets of simple materials, limiting participants' flexibility to express their beliefs. In contrast, adopting $MCMC_p$ revealed some indications that participants differed in how they interpreted the provided information. Specifically, our $MCMC_p$ method successfully disentangled variations among participants in their understanding of which color feature (lightness vs hue) conferred a survival advantage for ladybeetles. If future studies corroborate these findings, it suggests that knowledge about a species' traits and naive theories of evolutionary survival may lead to limited acceptance of biological variability. Importantly, our study suggests that the manifestation of this essentialist bias is more complex than previously thought.

Our future studies aim to explore this idea in more detail by examining how people use information about the adaptability of traits for less familiar animals, or when given more extensive contextual information. It is plausible that adults may exhibit stronger essentialist biases when presented with novel information and have less prior knowledge about the species. Similarly, we are expanding our experiments to include preschool and primary school children. These experiments could offer insight into how essentialist biases develop over time and how these biases relate to the knowledge acquired about biological categories, as well as theories of species' survival and natural selection.

Acknowledgments

We thank the three anonymous reviewers for their helpful feedback and suggestions. We thank Angelica Carmona, Cecelia Betts, and Samantha Bauer for helping conduct the experiments. PLV was supported by a 2023 Data Science Institute @Brown grant. DB was supported by a 2023 Brown University OVPR Seed grant. This work is based on material supported by the National Science Foundation under Grant No. 2400595.

References

AAAS. (2011). *Vision and change in undergraduate biology education: A call to action* (Tech. Rep.). Washington, DC.: American Association for the Advancement of Science.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology*, *139*, 148–160. doi: 10.1016/j.jecp.2015.05.011

Gelman, A., Gilks, W. R., & Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110–120.

Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford Academic.

Gelman, S. A., & Rhodes, M. (2012). "Two Thousand Years of Stasis": How Psychological Essentialism Impedes Evolutionary Understanding. In *Evolution Challenges: Integration Research and Practice in Teaching and Learning about Evolution* (pp. 3–21). Cambridge: Oxford University Press.

Gelman, S. A., Ware, E. A., Manczak, E. M., & Graham, S. A. (2013). Children's sensitivity to the knowledge expressed in pedagogical and nonpedagogical contexts. *Developmental Psychology*, *49*(3), 491–504. doi: 10.1037/a0027901

León-Villagrá, P., Otsubo, K., Lucas, C., & Buchsbaum, D. (2020). Uncovering category representations with linked MCMC with people. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 42).

MacKay, D. J. (2003). *Information theory, Inference and Learning Algorithms*. Cambridge university press.

Menendez, D., Rosengren, K. S., & Alibali, M. W. (2022). Detailed bugs or bugging details? The influence of perceptual richness across elementary school years. *Journal of Experimental Child Psychology*, *213*, 105269. doi: 10.1016/j.jecp.2021.105269

Morgan, T. H. (1911). Random segregation versus coupling in mendelian inheritance. *Science*, *34*(873), 384–384.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology*, *60*(2), 63–106. doi: 10.1016/j.cogpsych.2009.07.001

Shtulman, A., & Schulz, L. (2008). The Relation Between Essentialist Beliefs and Evolutionary Reasoning. *Cognitive Science*, *32*(6), 1049–1062. doi: 10.1080/03640210801897864

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718.

Walck-Shannon, E., Batzli, J., Pultorak, J., & Boehmer, H. (2019). Biological Variation as a Threshold Concept: Can We Measure Threshold Crossing? *CBE—Life Sciences Education*, *18*(3), ar36. doi: 10.1187/cbe.18-12-0241