Exploring the Impact of Early Life Stress on Learning in Dynamic Environments

First Year Project

Pablo Cáceres

10/18/2018

Abstract

Learning in dynamic environments is a critical skill to successful adaptation across contexts. Different context demands the implementation of different learning strategies to maximize reward-value. Previous literature has shown that growing up in chaotic and stressful environments may impact the learning mechanism that makes such flexible adaptation possible. In this project, we investigate the relationship between lifetime stress and feedback integration speed in a probabilistic learning task. Sixty 11-16 years-old children (36 females) participated in a probabilistic learning task. Our results show that children from higher levels of lifetime stress tend to integrate feedback faster relative to children from lower levels of lifetime stress. This suggest that children from high levels of lifetime stress tend to be more sensitive to feedback relative to children from low levels of lifetime stress, which can make adjustment to probabilistic settings more difficult.

Introduction

A critical aspect of successful adaptation to a dynamic environment is the ability to learn the associations between actions and outcomes. According to normative theories of optimal decision-making, in times of uncertainty, when associations among events become volatile, a relative fast updating of action-outcome values is key to adjusting to rapidly changing circumstances. Conversely, in times of stability, a slower rate of updating is adaptive, since associations among events are unlikely to change (Behrens, Hunt, Woolrich, & Rushworth, 2008; Behrens, Woolrich, Walton, & Rushworth, 2007; Nassar, Wilson, Heasly, & Gold, 2010). In this project we explore whether chronic exposure to stressful and chaotic environments may impact how children approach learning in a probabilistic learning setting. Such insight may help to understand some of the behavioral problems associated with chronic stress exposure in children.

Early life stress and learning

There is a growing literature showing that stress during the prenatal period, childhood and adolescence, plays a key role in shaping the neural circuitry and behavioral mechanisms that support learning skills (Hanson et al., 2017; Harms, Shannon Bowen, Hanson, & Pollak, 2018, p. 2; Hollon, Burgeno, & Phillips, 2015; B. S. McEwen, Nasca, & Gray, 2016) Nonetheless, despite recent progress, more precise characterizations of the mechanisms underlying this relationship are still needed to better understand the impact of lifetime stress on patterns of behavioral adjustment. Importantly, alterations in learning and reward processing as a consequence of stress may put children at risk of maladaptive patterns of behavior, academic difficulties, and both physical and mental illness (Fareri & Tottenham, 2016; C. A. McEwen & McEwen, 2017; Peters, McEwen, & Friston, 2017). Thus, identifying the impact of stress and chaos on learning and their neural underpinnings is a key challenge to understanding the relationship between adversity and both maladaptive and positive adjustment to adverse circumstances (B. S. McEwen, Gray, & Nasca, 2015)

In the stress literature, there is a diverse group of contexts characterized as ‘stressful’, such as poverty, domestic violence or institutionalization. Various accounts about stress function have proposed that what makes such diverse environments stressful is the high prevalence of *uncertainty* and *threat*, which are the fundamental ingredients to trigger the stress response physiology (Peters et al., 2017). Thus, children that are exposed to persistent uncertainty and threat face the challenge of adjusting their behaviors and 'internal working models' to such environments. In this context, one possibility is that children who have experienced unstable and stressful lives acquire a tendency to rapidly incorporate and adjust to feedback from the environment. Such relative fast speed of feedback integration may help to adapt to constantly changing circumstances. Nevertheless, chronic exposure to unstable environments may generate maladaptive patterns of behavior by impairing their ability to flexibly adapt learning strategies in response to changes in the environment. One way to approach this question is by assessing the speed of feedback integration by fitting a reinforcement learning model to human behavior and using that measure to compare the speed at which children from different levels of lifetime stress incorporate feedback to evaluate probabilistic contingencies. Additionally, we can compare the speed of feedback integration implemented by human participants with the 'behavior' of a Bayesian ideal learning algorithm that serves as benchmark of optimal behavior (i.e., the feedback integration speed that maximize reward-value). In the next section we explain our modelling strategy.

A Bayesian approach to learning in a dynamic environment

Adaptive decision-making in a dynamic environment is a challenging task because the information about action-outcomes associations is unknown or incomplete. To overcome this problem, individuals often rely on learning through trial and error to gain knowledge about such associations. In the learning theory literature, reinforcement learning models have arisen as a productive formal description of such process (Sutton, Barto, & Bach, 1998). In this realm, a widely used reinforcement learning algorithm is the 'Delta Rule' (Rescorla & Wagner, 1972), which operates according to the following equation:

$$B_{t+1} = B_t + \alpha_t x \, \delta_t \tag{1}$$

where B_{t+1} represents a belief about contingencies in the environment at the time *t+1* (i.e., the next trial); B_t represent beliefs at time *t* (i.e., the current trial); α_t represent the learning rate (i.e., the rate of integration of new information to update previous beliefs); and δ_t represent the reward prediction error, (i.e., the value difference between the expected and actual outcomes).

Accordingly, in this model, new beliefs are updated based on previous beliefs and some proportion of the prediction error. The rate at which the new information replaces old information is controlled by the learning rate. At one extreme, a learning rate = 1 indicates that the new information completely replaces previous beliefs, while in the other extreme, a learning rate = 0 indicates that new information is completely disregarded, and beliefs remain unchanged. Current research on perception and learning has stressed the importance of adopting an 'optimal learning rate' in accordance to the higher order statistical properties of the environment (i.e., the degree of volatility in action-outcomes contingencies over time), proposing that Bayesian decision theory, applied to reinforcement learning, may help to predicts and explain how such adjustment occurs (Behrens et al., 2008, 2007; Nassar et al., 2010).

Bayesian decision theory is a method for optimal decision-making under uncertainty informed by Bayesian inference (Peterson, Abbey, & Eckstein, 2009; Trenti, Barraza, & Eckstein, 2010).To illustrate this concept, imagine needing to determine your travel time to work by bus. Under this framework, you would approach the inference problem starting with an assumption (a 'belief') about the probability of a set of 'possible travel times', which is known as the 'prior probability distribution'. Even if the bus schedule is fixed, random contingencies can alter the bus travel time (e.g., traffic accidents, the weather, etc.). To get an estimate less affected by random variations, you can register your daily time travel and compute the average, which is your 'best guess' available (i.e., the most likely travel time, assuming a gaussian distribution). Next, you can sequentially update your travel time probability distribution after each trip by adding some proportion of each new piece of information; this additional information allows you to have the best possible estimate available to inform your decisions. The result of such updating process results in what is known as the 'posterior probability distribution', which is your new 'best guess' about your travel time. Under this framework, the rate at which agents integrate new information to prior estimations (i.e., the learning rate parameter in equation [1]) should be adjusted in accordance to the levels of uncertainty in the estimate of the action's value.

When agents face more volatile environments (i.e., high uncertainty about action's values) a faster rate of integration of new information is desirable to quickly adjust prior expectations to rapidly changing actions-outcomes contingencies; in more stable environments the opposite is true, as the history of actions-outcomes associations is more predictive of the current state of the environment. In our previous example, a stable environment may be a city during the summer, when traffic is less affected by random incidents, whereas a volatile environment may be the same city during the rest of the year, when the increment of the traffic makes random incidents and delays more frequent (i.e., the travel time become more unpredictable). Previous research has confirmed this proposal, showing that adults adjust their learning rates based on the degree of volatility of the environment in a similar fashion that an ‘ideal Bayesian learner’ predicts, which is a computational algorithm that represent a benchmark for optimal decision-making under uncertainty (Behrens et al., 2008, 2007; Nassar et al., 2010).These results have been interpreted as evidence that adults adjust their prior beliefs about volatility/stability in the environment in a flexible and near to optimal manner. However, the possibility that long-term exposure to adverse life-circumstances may alter the formation of prior beliefs about volatility/stability and impair the flexible updating of such estimations has not yet been examined. Exploring this possibility may be a productive avenue to mechanistically understand how lifetime stress impacts behavioral adaptation.

Hypotheses

Our primary hypothesis is that children with higher levels of lifetime stress exposure will tend to integrate feedback faster relative to children with lower levels of lifetime stress (i.e., a relatively higher learning rate). We expect this relationship hold in both the stable and volatile phases of the task. We will test this hypothesis measuring the ‘learning rate’ parameter of a simple ‘Delta Rule’ learning model applied to subjects’ decisions and testing for statistical differences. If the value of the learning rate parameter is significantly higher for children from high stress backgrounds overall and in both contexts (i.e., positive correlation between learning rate and lifetime stress), this will provide evidence supporting our hypothesis. Additionally, we will also estimate the same model for a computational algorithm (an 'ideal Bayesian learner') as a benchmark of optimal behavior. If our hypothesis is confirmed, these data will provide support to the idea that children from high stress backgrounds tend to be more reactive to feedback because of a history of instability and stressful life life-circumstances. **Fig. 1** displays a graphical representation of the hypothesis[1]

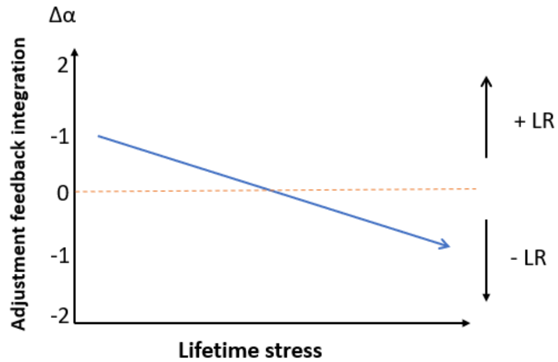
Fig. 1

Our second hypothesis proposes that children with high levels of lifetime stress will be less flexible adjusting the speed of feedback integration relative to children from lower levels of lifetime stress (this is, when changing from a stable to a volatile environment, an vice versa). We will test this hypothesis measuring the change in the learning rate parameter (i.e., subtracting the learning rate from the volatile phase to the learning rate of the stable phase) of a simple 'Delta Rule' learning model applied to subjects' decisions, and testing for statistical differences. If the change on the learning rate parameter is significantly lower (i.e., negative correlation between *change* in the learning rate and lifetime stress) for children from high stress backgrounds, this will provide evidence supporting our hypothesis. If our hypothesis is confirmed, these data will provide support to the idea that children from high stress backgrounds are less flexible in adjusting their feedback integration speed in response to changes in the environment because of a history of instability and stressful life life-circumstances. **Fig 2** displays a graphical representation of the hypothesis[2]

---

[1] Arbitrary values for illustrative purposes
[2] Arbitrary values for illustrative purposes

Fig. 2



Method

Sample. 60 children (36 females) ages 11-16 were recruited from the Madison metropolitan area. Our recruiting strategy was oriented to incorporate children representing the whole spectrum of lifetime stress. This sample size is based on previous similar studies in this field, suggesting moderately high effect sizes for the effects of early life stress on cognitive task performance (Hanson et al., 2017; Harms et al., 2017).

Procedure. We used a probabilistic learning task that has been effective with adults (Behrens et al., 2008, 2007). In this task, participants are asked to repeatedly make predictions by choosing between green and blue boxes. In each trial, just one outcome may be correct (i.e., blue box or green box), in the same manner that when flipping a coin just heads or tails can be correct. The participants are explicitly instructed that the probability of each color being correct depends only on the recent outcome history, and not by the reward displayed at the center of each box, which is randomly selected from 0 to 100. The first part of the task consists of 120 trials, with a constant probability of 75% of a blue outcome, which defines the stable environment phase. The second part consists of 170 trials where reward probabilities shift to 80% blue and 80% green every 30 or 40 trials, which defines the volatile environment phase. A red bar in the bottom of the screen shows the total amount of points earned during the task, which growth in proportion to the points displayed at the center of each box, only if the prediction was correct. Two lines at the bottom right corner of the screen represent a small prize (grey line) and a big prize (gold line), which can be reached as more points are accumulated during the task. The temporal sequence of the task consists of two stages: first, the two options are presented with the associated reward at the center of each box. Here the participant makes the selection, with a grey square highlighting the chosen box; second, a new screen displays the correct response at the center, with the respective increment in the red bar if the prediction were correct. In the example below, the participant chose blue, which means that the participant obtained the amount of points at the center of the card.

Fig. 3

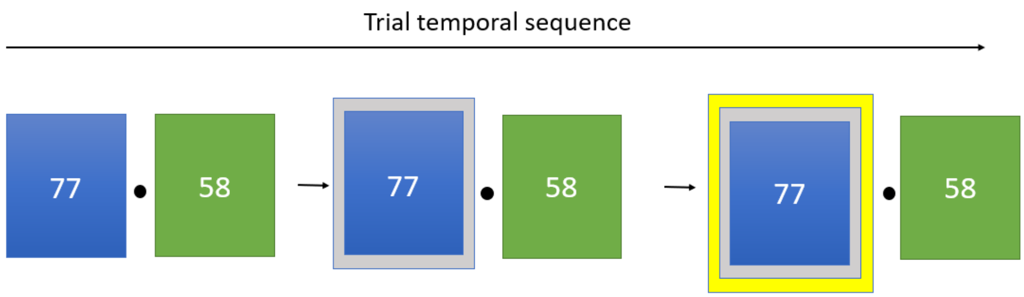


Figure 3. Binary choice trial sequence. Participants were presented with two options, blue and green, and instructed to select one of the options using the keyboard. On this example, the blue option is selected, and it is subsequently highlighted with grey. Immediately after selection, participants received feedback in the form of the yellow highlighting as depicted in the figure. Correct choices lead to adding the amount of points indicated at the center of the card to their total score. Higher scores were associated with bigger rewards at the end of the task.

Measures

Stress exposure. We assessed stress exposure using the Youth Life Stress Interview (Rudolph et al., 2000). Children and their parents completed this interview independently. The interviews were applied by trained interviewers using a semi-structured questionnaire. A group of three to seven raters used the information from the interviews to provide a consensual rating on a 10-point scale reflecting cumulative life stress.

*Feedback integration speed***.** To analyze learning and decision-making dynamics, we fitted a reinforcement learning model previously used to describe adult (Behrens et al., 2008, 2007) and children behavior (Manning, Kilner, Neil, Karaminis, & Pellicano, 2017) in probabilistic learning settings. The estimation procedure has been described by Behrens and colleagues (2007) and it can be found in the Supplementary Information Appendix.

Working memory. To rule out the possibility that our results are confounded by general cognitive impairments, we measured numeric Memory span capacity using the digit-span test from the Wechsler Intelligence Scale for Children (WISC-III).

Table 1

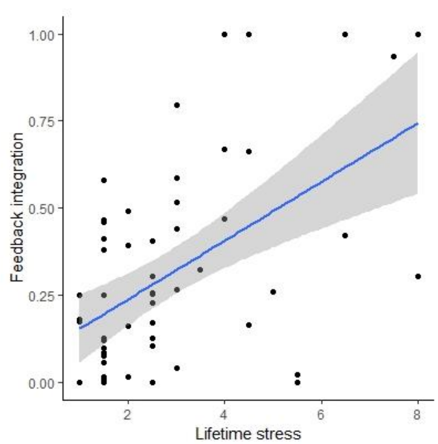
	M (sd)	Range
Lifetime stress	2.89 (1.82)	1-8
Feedback integration speed	0.31(0.29)	0.1-0.99
Memory span	15.93(3.85)	8-28

Results

Is lifetime stress associated with feedback integration speed?

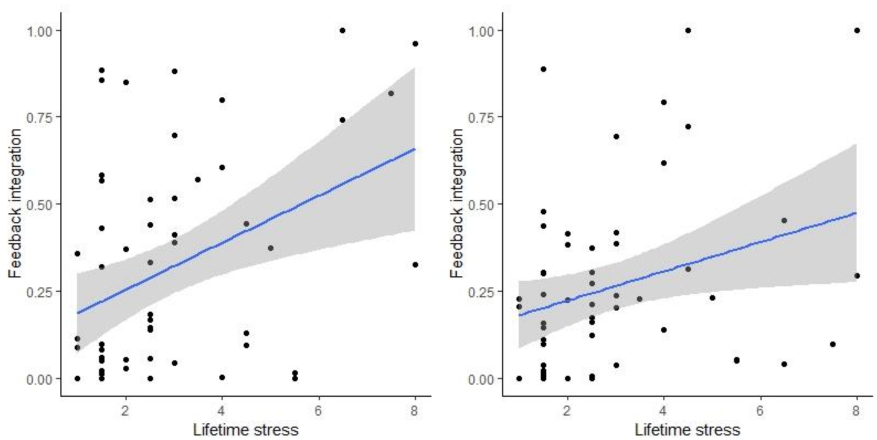
To examine the relationship between lifetime stress and the speed of feedback integration, we first ran a Pearson's correlation test. Our analysis revealed a positive and significant correlation between children's lifetime stress exposure and the speed of their integration of feedback ($r(54)$ = .52, 95% *CI* [0.30, 0.69], $t(54)$ = 4.51, p < .001). This result indicates that children with higher levels of stress exposure integrate feedback and update their value estimation of cues faster than children who experience normative lifetime stress (see *Fig. 4*).

Fig. 4 Association feedback integration speed and lifetime stress



We were particularly interested in assessing if the relationship between feedback integration speed and lifetime stress differs depending on environmental statistics. To test this, we ran two separate Pearson's correlation tests, one for the stable phase (120 trials) of the task and one for the volatile phase (170 trials). Our analysis revealed a positive and significant correlation between children's lifetime stress exposure and the speed of feedback integration for both the stable phase ($r(54) = .39$, *95% CI* [0.14, 0.59], $t(54) = 3.113$, $p = .003$), and the volatile phase ($r(54) = .30$, *95% CI* [0.04, 0.52], $t(54)= 3.28$, $p = .026$; see Fig. 5).

Fig 5. Association feedback integration speed and lifetime stress (stable: left; volatile: right)

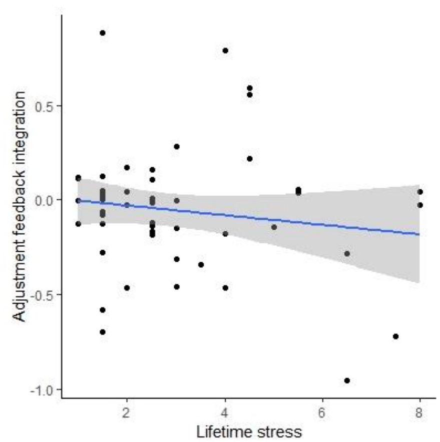


To test the possibility that this association is cofounded by general cognitive skills or task-order, we fitted a regression model adding WISC Memory Span and task-order (1= volatile trials first; 0= stable trials first) as covariates. Our analysis showed that lifetime stress remains positively and significantly associated with speed of feedback integration after controlling for such covariates ($\beta = 0.08$, $t(53) = 4.54$, $p < .001$, partial $\eta^2 = .28$). Neither Memory Span ($p = 0.96$) nor task-order (p = 0.133) were significantly associated with speed of feedback integration. We repeated this analysis for stable and volatile phases of our task separately. The results for the stable phase revealed that feedback integration speed is positively and significantly correlated with both lifetime stress ($\beta = 0.08$, $t(53) = 3.64$, $p < .001$, partial $\eta^2 = .28$) and task-order ($\beta = 0.2$, $t(53) = 2.7$, $p = .009$, partial $\eta^2 = .\ 0.13$), but not with Memory span ($p = 0.51$). Finally, the results for the volatile phase showed that feedback integration speed is positively and significantly correlated with lifetime stress ($\beta = 0.04$, $t(53) = 2.21$, $p = .032$, partial $\eta^2 = 0.087$) but not with Memory span ($p = 0.90$) or task-order ($p = 0.32$). Overall, our results confirm that lifetime stress is positively associated with feedback integration speed regardless of task-order and Memory span, as expected by our initial hypothesis.

Is lifetime stress associated with adjustment feedback integration (i.e., learning rate adjustment)?

We also investigated the possibility of lifetime stress being associated with the adjustment of the speed of feedback integration. Adjustment of feedback integration was defined as the magnitude of the change in the speed of feedback integration when transitioning from the first to the second block of our task (i.e., speed integration block 1 – speed of integration block 2 = speed adjustment). As shown in Figure 6, our analysis did not reveal an association among lifetime stress and speed of feedback integration (*p* = .294). We repeated this analysis splitting our sample between participants in the in the Stable first condition and Volatile first condition. Neither Stable first (*p* = .098) or Volatile first (*p* = .99) revealed significant associations.

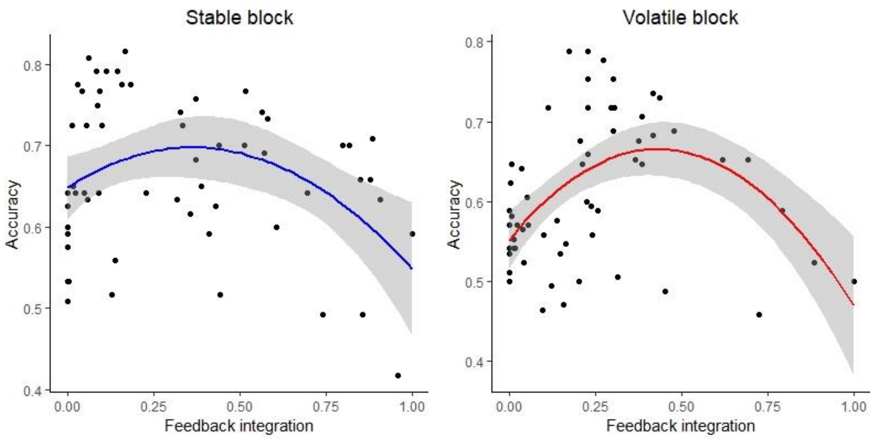
Fig. 6: Adjustment speed of feedback integration



Is feedback integration speed associated with accuracy in a u-shape inverted manner?

Theories of optimal decision-making propose that to maximize reward-value, the speed of feedback integration should be neither too fast nor too slow. In other words, accuracy in our task should be maximized by implementing a moderate speed of feedback integration, with slightly faster updating in the Volatile phase relative to the Stable phase. From this framework, we predicted that the speed of feedback integration would display an inverted u-shape relationship with accuracy. To test this idea, we first fitted a polynomial regression model incorporating speed of feedback integration in the Stable phase as a quadratic term to correlate that with accuracy in the Stable phase of our task. We added Memory span and task-order as covariate to rule out the possibility of general cognitive skills or task-order acting as cofounder. Our results showed that, as predicted, accuracy in the Stable phase is significantly correlated with speed of feedback integration in a quadratic fashion ($\beta = 0.29$, $t(56) = 2.17$, $p = .035$, *partial* $\eta 2 = .075$; β-*squared* $= -0.38$, $t(56) = -2.63$, $p = .011$, *partial* $\eta 2 = .11$). Neither Memory span ($p = 0.93$) nor task-order ($p = 0.54$) showed a significant correlation. Specifically, this means that integrating feedback too slow or too fast is negatively associated with performance in the task. Finally, we fitted the same polynomial regression model for the Volatile phase of our task. Again, our results showed a significant correlation between accuracy in the Volatile phase and speed of feedback integration in a quadratic fashion ($\beta = 0.52$, $t(56) = 11.18$, $p < .001$, *partial* $\eta 2 = .25$; β-*squared* $= -.61$, $t(56) = -4.47$, $p < .001$, *partial* $\eta 2 = .27$ Neither Memory span ($p = 0.98$) nor task-order ($p = 0.80$) showed a significant correlation. Is important to consider that polynomial regressions have a high-rate of false positive results ((Simonsohn, 2017)), reason why these preliminary results must be interpreted carefully.

Figure 7: Accuracy and feedback integration association by condition



Discussion

In this study, we evaluated the relationship between lifetime stress and children's ability to incorporate feedback to adjust their behavior. The goal of this experiment was to better understand whether a history of chaotic and stressful life circumstances impacts learning. Our more specific goal was to begin to explore a potential mechanism linking early life adversity with patterns of behavioral adaptation. As expected, children with higher levels of cumulative lifetime stress incorporated feedback faster relative to children who had experienced more normative levels of lifetime stress. We next examined this same relationship by breaking down our results into stable and volatile phases of our task. The results for both the stable and volatile phases revealed the same pattern of results: Children with higher levels of lifetime stress integrated feedback faster relative to children from lower levels of lifetime stress. These results remained significant after accounting for task-order effects and Memory span. Is important to mention that the task-order effect was specific for the stable phase of our task. Specifically, this means that when participant experienced the volatile block of the task first, they tended to use a high speed of feedback integration for the remainder of the trials across both conditions. This suggests that experiencing a volatile environment first might prime behavior towards fast feedback integration. Future research might test this effect further by implementing a between-subjects design to better examine the impact of experiencing volatile task conditions in subsequent behavior.

Overall, the pattern of results obtained in this experiment is consistent with the idea that exposure to high levels of chaos and stress might bias learning of environmental contingencies into the direction of fast feedback integration. According to Bayesian decision-making theory, the speed of feedback integration should be adjusted to the higher-order statistics of the environment to maximize reward-value. In other words, to maximize reward-value, agents need to adhere to what we might call a "Goldilocks Effect." Agents do not want to learn neither too slowly or too quickly, but just at the right speed given the predictability of environmental contingencies. Therefore, incorporating feedback too quickly may be part of the causal chain in behavioral problems observed in stress-exposed children. In the context of our study, children with higher levels of lifetime stress tended to incorporate feedback faster than necessary in both the stable and volatile phase of the task, which might have negatively impacted their ability to learn and maximize reward-value. This last point is illustrated by the fact that children who incorporated feedback closer to what is expected according to a Bayesian ideal learner algorithm (i.e., ~0.1 for stable phase; ~0.2 in volatile phase), achieved higher levels of accuracy in the task. This last fact raises the possibility that children from high levels of lifetime stress may actually perform well in a task with higher levels of volatility than the used here, and might be consistent with anecdotal evidence that children from stressful family backgrounds tend to choose environments and careers with high levels of stress.

The literature on behavioral development of children who suffered significant levels of early life adversity, has documented how such experiences are linked with impulse control problems (Lovallo, 2013), emotion dysregulation (Kim & Cicchetti, 2009), and difficulties on learning in probabilistic reversal learning settings (Hanson et al., 2017; Harms et al., 2018). Our results add the possibility of a general learning mechanism potentially mediating the link between early adversity and behavioral difficulties. In general, learning to predict which cues in the environment are more likely to lead to rewards and punishments is difficult when the quality of the feedback signal is ambiguous and inconsistent. This is the kind of environment that many children confront when caregivers, peers, and other sources of rewards and punishments, display inconsistent patterns of behavior. Such situations become more problematic in dynamic environments, when learning strategies must be flexibly adjusted to changing circumstances. Even though the results for our secondary hypothesis regarding flexibility on adjusting the speed of feedback integration did not emerge as significant, the fact that children from higher stress backgrounds consistently displayed a fast speed of feedback integration is indicative of inflexible behavior.

One problem with the measure of feedback integration adjustment used in our study is ceiling effects: once you implement a high speed of feedback integration, going higher becomes more difficult because there is not much room left to growth. The same problem is present with low speed feedback integration and moving to lower levels. Another potential issue is that the volatile phase of our task may not have been volatile enough, meaning that even when it was more volatile than the stable phase, it might have not been perceived as such. In future research we plan to address this issue by incorporating higher levels of volatility to facilitate the distinction between task phases.

The nature of the notion of volatility itself is another challenge to consider. In the context of our task volatility is operationalized as recurrent changes in the probability of two options. In other words, what we need to capture is the inconsistency in reward and punishment contingencies in the daily environment. There are multiple proxies of such phenomena, like parenting behaviors, house chaos, and lifetime stress, but none of them is solely focused on reward and punishments stability. Additionally, there are multiple sources of rewards and punishments in children lives, like siblings, peers, extended family, teachers, and others, that should be considered in an ideal measure. Incorporate multiple measurements at the same time, is not necessarily a solution, given the high level of collinearity among these variables.

Finally, the correlational nature of our study does not allow to make strong claims regarding causality. Nonetheless, we believe that it is a significant contribution to the understanding of the domain-general mechanism that govern learning on dynamic environments in the context of child development.

Conclusion

This study provides evidence suggesting that early life stress alters important processes of learning and decision-making in children. Specifically, we showed that higher levels of lifetime stress are associated with a faster feedback integration in a dynamic probabilistic learning task. Future research should address how early life tress is associated with varying levels of volatility in action-outcome contingencies, including extreme levels of volatility. An additional key challenge is to develop and use better measures of environmental instability in rewards and punishments contingencies over the life course. Our results contribute with important knowledge oriented to understand the relationship between early life adversity and patters of adaptative and maladaptive behavior in dynamic environments.

References

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Fareri, D. S., & Tottenham, N. (2016). Effects of early life stress on amygdala and striatal development. *Developmental Cognitive Neuroscience*, *19*, 233–247. https://doi.org/10.1016/j.dcn.2016.04.005

Hanson, J. L., van den Bos, W., Roeber, B. J., Rudolph, K. D., Davidson, R. J., & Pollak, S. D. (2017). Early adversity and learning: implications for typical and atypical behavioral development. *Journal of Child Psychology and Psychiatry*, *58*(7), 770–778. https://doi.org/10.1111/jcpp.12694

Harms, M. B., Shannon Bowen, K. E., Hanson, J. L., & Pollak, S. D. (2018). Instrumental learning and cognitive flexibility processes are impaired in children exposed to early life stress. *Developmental Science*, *21*(4), e12596. https://doi.org/10.1111/desc.12596

Hollon, N. G., Burgeno, L. M., & Phillips, P. E. M. (2015). Stress effects on the neural substrates of motivated behavior. *Nature Neuroscience*, *18*(10), 1405–1412. https://doi.org/10.1038/nn.4114

Kim, J., & Cicchetti, D. (2009). Longitudinal pathways linking child maltreatment, emotion regulation, peer relations, and psychopathology: Pathways linking maltreatment, emotion regulation, and psychopathology. *Journal of Child Psychology and Psychiatry*, *51*(6), 706–716. https://doi.org/10.1111/j.1469-7610.2009.02202.x

Lovallo, W. R. (2013). Early life adversity reduces stress reactivity and enhances impulsive behavior: Implications for health behaviors. *International Journal of Psychophysiology*, *90*(1), 8–16. https://doi.org/10.1016/j.ijpsycho.2012.10.006

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2017). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science*, *20*(5), e12435. https://doi.org/10.1111/desc.12435

McEwen, B. S., Gray, J. D., & Nasca, C. (2015). Recognizing resilience: Learning from the effects of stress on the brain. *Neurobiology of Stress*, *1*, 1–11. https://doi.org/10.1016/j.ynstr.2014.09.001

McEwen, B. S., Nasca, C., & Gray, J. D. (2016). Stress Effects on Neuronal Structure: Hippocampus, Amygdala and Prefrontal Cortex. *Neuropsychopharmacology*, *41*(1), 3–23. https://doi.org/10.1038/npp.2015.171

McEwen, C. A., & McEwen, B. S. (2017). Social Structure, Adversity, Toxic Stress, and Intergenerational Poverty: An Early Childhood Model. *Annual Review of Sociology*, *43*(1), 445–472. https://doi.org/10.1146/annurev-soc-060116-053252

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, *156*, 164–188. https://doi.org/10.1016/j.pneurobio.2017.05.004

Peterson, M. F., Abbey, C. K., & Eckstein, M. P. (2009). The surprisingly high human efficiency at learning to recognize faces. *Vision Research*, *49*(3), 301–314.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Rudolph, K. D., Hammen, C., Burge, D., Lindberg, N., Herzberg, D., & Daley, S. E. (2000). Toward an interpersonal life-stress model of depression: The developmental context of stress generation. *Development and Psychopathology*, *12*(2), 215–234. https://doi.org/10.1017/S0954579400002066

Simonsohn, U. (2017). Two-lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *SSRN, Accessed on May 1, 2018 At*.

Sutton, R. S., Barto, A. G., & Bach, F. (1998). *Reinforcement learning: An introduction*. MIT press.

Trenti, E. J., Barraza, J. F., & Eckstein, M. P. (2010). Learning motion: Human vs. optimal Bayesian learner. *Vision Research*, *50*(4), 460–472.

Supplementary Information Appendix

*Mathematical model of learning***.** Behrens and colleagues (2007) describe – in verbatim- the estimation procedure in the following words:

The predictor is in the form of a simple delta-learning rule. This rule has a single free parameter, the learning rate. The delta-learning rule estimates outcome probabilities using the following equation

$$\hat{r}_{i+1} = \hat{r}_i + \alpha\varepsilon_i \tag{1}$$

where $\hat{r}_{i+1}$ is the predicted outcome probability for the (i+1)th trial, $\hat{r}_i$ is the predicted outcome probability for the ith trial, ε_i is the prediction error at the ith trial, and a is the learning rate. By choosing different values for a, the model can make different approximations of the subject's outcome probability estimates

The selector model explains subject decisions on the basis of these estimates. Here, decisions are determined by both the estimated reward likelihood, $\hat{r}_{i+1}$, and by the reward magnitude on each option. Optimal action selection would involve computing the estimated Pascalian value (outcome size X outcome probability) of each option as follows:

$$g_{blue\ i+1} = \hat{r}_{i+1} f_{blue\ i+1} \tag{1.2}$$

$$g_{green\ i+1} = (1 - \hat{r}_{i+1}) f_{green\ i+1} \tag{1.2}$$

where $f_{green,i}$ and $f_{blue,i}$ are the known reward sizes of each color. The optimal response is then the color with the highest predicted profit. However, we do not make the assumption that human subjects weigh reward likelihood with reward magnitude in this optimal Pascalian fashion. Instead, we include a free parameter that allowed subjects to increase the weight of either reward likelihood or reward magnitude when valuing an outcome (respectively representing risk-averse and risk-prone behavior). Subjects are taken to value each option according to the following equations:

$$g_{blue\ i+1} = F(\hat{r}_{i+1}, y) f_{blue\ i+1} \tag{1.3}$$

$$g_{green\ i+1} = F(1 - \hat{r}_{i+1}, y) f_{green\ i+1} \tag{1.4}$$

where function $F(r, y)$is a simple linear transform within the bounds of 0 and 1:

$$F(r, y) = \max[\min[(y(r - 0.5) + 0.5), 1], 0] \tag{1.5}$$

and $y = 1$, $y < 1$ and $y > 1$ imply optimal, risk-prone and risk-averse behavior, respectively. Subjects were then assumed to generate actions stochastically, according to a sigmoidal probability distribution (for example):

$$P(C = Green) = \frac{1}{1 + \exp(-\beta(g_{green} - g_{blue}))} \tag{1.6}$$

We fit this model using Bayesian estimation techniques (using direct numerical integration) to compute the expected value of the marginal posterior distribution on a for each subject in each task phase (p.1219-1220).