Integrating Information in Stress Research

Valeria Bonapersona

UMC Utrecht Brain Center

To my brother

Doctoral thesis, Utrecht University, Utrecht, The Netherlands

Author: Valeria Bonapersona
Cover: Valeria Bonapersona & Cecilia De Paola
Lay-out: Cecilia De Paola
Printed by: Proefschrift All In One (AIO)
ISBN: 978-90-393-7494-8

Integrating Information in Stress Research

Informatie Integratie in Stress Onderzoek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 30 augustus 2022 des middags te 2.15 uur

door

Valeria Bonapersona

geboren op 13 september 1992
te Varese, Italië

Promotors:
Prof. dr. M. Joëls
Prof. dr H.J.A. Hoijtink

Co-promotor:
Dr. R. A. Sarabdjitsingh

Beoordelingscommissie:
Prof. dr. R.M. Dijkhuizen (voorzitter)
Prof. dr. A.G.J. van de Schoot
Prof. dr. P.J. Lucassen
Prof. dr. S.J.T. Branje
Prof. dr. M. MacLeod

Contents

Appendices

Chapter 1

Introduction

1

Charlie is counting backwards from 1022 in steps of 13 for his final math exam. He is clenching his sweaty hands to keep the focus. Jessica is checking the position of her fingers on the athletics track, while her heart has already started racing. Cornelius has been going through a divorce. Although it has been several weeks since last time he cried, he has bags under his eyes and looks soullessly at the television. Cornelius's children (David and Emma) are "doing fine". Emma is excelling at school (besides in her gymnastic class due to her sudden weight loss), and David is spending a lot of time with new friends at the park. He has been drinking, but Cornelius says that's to be expected for kids his age.

We can all give real-life names to our fictional characters. We can also recognize the sweaty hands, increased heart rate, problems with sleeping, and changes in (various forms of) habits. These experiences are relatable, and also interconnected. What do sweaty hands and sleeping problems have in common? In biology, we give both the faults and merits to a unique biological system: the stress system. Be it a situation of acute stress (the counting or start of an athletic race in the examples above) or chronic stress (the impact of Cornelius' divorce on him and his children).

In the classical view[1], the stress system is responsible for maintaining homeostasis, the equilibrium every organism thrives towards to. This equilibrium is dynamic, and it is continuously adjusted by our bodies to promote adaptation, a processed referred to as allostasis[2,3]. By mediating our ability to adapt to the environment, the stress system is the first line of response to a potentially threatening situation (real or imaginary, psychological or physiological), that is subjectively experienced as stress. It is often operationalized as the activity of the sympathetic nervous system and the hypothalamic-pituitary-adrenal (HPA) axis[4] (Box 1). The sympathetic nervous system prepares the body for the fight-or-flight response[5], while the HPA axis regulates the neuro-endocrine response, whose activity is reflected (among others) by the blood concentration of corticosteroid hormones[1]. Both systems act directly and indirectly on the brain, where they not only affect their own activity (e.g. through negative feedback) and coordinate the physical response to the situation at hand, but also modulate memory, how we think and make decisions (cognitive and executive functioning), as well as our emotions (emotional control)[1]. Effective functioning of the stress response

Box 1 – Sympathetic nervous system and HPA axis

As part of the autonomic nervous system, the sympathetic nervous system plays an essential role in regulating (unconscious) body homeostasis[5], such as cardiovascular fitness, blood glucose levels and metabolism[8]. Upon a potentially threatening situation, the sympathetic nervous system is rapidly activated, causing the release of adrenaline into the circulation by action of the adrenal glands and (indirectly) the release of noradrenaline in synapses. The (potentially threatening) situation at hand is perceived via various brain areas, and next processed and contextualized. The brain integrates information about the situation and signals the hypothalamus to release corticotropin-releasing hormone (CRH), followed by the release of ACTH by the pituitary, and ultimately secretion of glucocorticoids (mainly cortisol in humans and corticosterone in rodents) by the adrenal glands[15]. The hormones released by the sympathetic nervous system and the hypothalamus-pituitary-adrenal (HPA) axis directly regulate many bodily responses associated with a stressful event, with effects differing per organ partially depending on specific receptors[16,17]. At first glance, it may seem that the stress system is extremely redundant[18], since it involves neurotransmitters, peptides, hormones, etc. Another possible explanation is that – rather than being redundant – the multiplicity of these mediators underlies the remarkable ability of our brain to dynamically adapt to a changing environment[18]. These dynamic changes are therefore considered to be an adaptive response to stress, which aid an organism's quick response in acute situations, as well as memory storage for future use.

In an experimental setting, various manipulations can be used to induce acute stress. In rodents, we used a single inescapable foot-shock (*Chapter 2*). Among other advantages[19], we selected this stressor because it offers an experimental advantage: it is very brief, thereby providing a clear starting time, essential for the time-dependent effects investigated in this thesis. Furthermore, since this model has been frequently used in the literature, it offered extensive prior knowledge which we could use to validate our findings. In humans, our understanding of the acute stress response is generally derived from experimental (i.e. laboratory based) acute stress studies. These use different (versions of) paradigms to induce acute stress, such as the Trier Social Stress Test[20], the Cold Pressor Test[21], the Maastricht Acute Stress Test[22], and virtual reality (VR)-based tests[23]. All of these experimental paradigms model two aspects of the acute stress response: physical and/or social. All paradigms are able to reliably increase the plasma and salivary cortisol concentrations of participants, and are therefore often used to induce an acute stress response.

implies that the stress system is rapidly activated and efficiently terminated when needed[1]. When the stress response is inadequate, excessive, prolonged or frequent, the cost of reinstating the current homeostasis may be too high[3]. This cost is generally referred to as allostatic load[3]. Allostatic load is the wear-and-tear of the body due to the over- or in- activity of those systems whose function is to maintain the equilibrium. In other words, allostatic load is the price to pay for continuous adaptation[3], with long-term consequences on brain[6], body[7,8] and behavior[9] that differ from person to person[10].

The consequences may be especially long-lasting if they occur during sensitive periods of development, such as early in life – events that are collectively described by the umbrella term "early life adversity" (Box 2). In the historical definition by Selye, the stress response was referred to as a syndrome, a predictable and generalized response of the body that would function as a "general alarm signal"[11]. Although there is no consensus on a definition of stress[12], it is generally agreed that the stress response is: 1) multimodal, since it acts upon numerous effector systems, 2) multifaceted, since it acts with spatial and temporal specificity, and 3) malleable, since its functioning is dependent on genetic predisposition, as well as early- and later- life events. In that respect, chronic stress may lead to different responses than a single, acute stress situation, as mentioned at the start.

How can we study such a complex system? Historically, biologist have approached the study of the stress system by isolating a specific feature, and performing specific experiments[13]. This traditional approach of taking things apart has been very successful. Nowadays, we can be extremely precise about which subset of cells (e.g. CRH+ cells) in which brain area (e.g. paraventricular nucleus, PVN) is responsible for a specific behavior (e.g. amplify the acute stress reaction via co-release of vasopressin[14]). However, the components of the stress system are intrinsically related to one another, and it's becoming exceedingly clear that this property cannot be forgotten. The functioning, robustness and adaptation of a system is not merely the sum of its subparts[13], but it also depends on their relationship and interrelations. For example, in the case of stress, the production of glucocorticoids and its feedback mechanism need to be equally balanced for the successful management of the stress response. Understanding the stress response as a system, rather than a collection of organs and hormones, brings us closer to comprehend how it actually works in nature, in health and for future improvement of disease. Yet, as a stress research community, we are just starting to address stress as a system. While moving from collections to systems is intuitive to understand, it is intricate to implement. The next challenge in stress research is therefore the one of information integration (Box 3).

Box 2 – Early life adversity

Early life adversity (ELA) is a broad term to describe negative environmental conditions early in life that impact normal brain development[24]. During early life, the brain is still developing and it is influenced by life experiences, which prepare the child (or pup, in rodents) for later (adult) life[25]. While genetics provides the "clay" for development, early life sculps it. If the experiences during this period are "adverse", e.g. related to abuse and neglect[26], they may have long-lasting consequences on mental health, lasting well into adulthood[27]. The first large-scale study on this topic was conducted in 1998, the ACE (Adverse Childhood Experience) study. This study evaluated the relationship between various forms of deprivation, abuse and neglect on later life outcomes, including mental health. The results were staggering: exposure to one or more ACEs accounted for 54% of the population attributable risk for depression28, 67% for suicide attempts[28] and 64% for illicit drug use[29]. To put these percentages into context, the population attributable risk of death of lung cancer due to cigarette smoking is 52.2% for males and 11.8% for females[30]. Since the ACE report appeared, numerous epidemiological studies have consistently identified ELA as a main risk factor for poorer (mental) health later in life (for example,[31–33]). ELA has even been described as a pleiotropy[34], meaning that its consequences are multiple and apparently uncorrelated. Decades of research have highlighted elements that can help navigate through this heterogeneity. Among these, the type, timing, and recurrence of the adverse experiences have received most attention[35].

Currently, there are two overall competing frameworks to explain the effects of ELA: the cumulative-risk and the multi-dimensional models. In the cumulative risk model, multiple stressors are added to predict health outcomes[36], similarly to the ACE study. Conversely, the multi-dimensional model argues that the effects of ELA are not necessarily additive, since different ELA types can be associated to different ELA outcomes[37]. Specifically, the multi-dimensional approach categorizes the ELA experiences into dimensions, such as deprivation, threat/harshness and unpredictability[38]. This model assumes that these dimensions have distinct biological underpinnings, which can be then used to directly categorize and investigate mechanisms of ELA[37]. More recently, others have proposed a "research domain criteria" framework, suggesting that new studies should map phenotypes to neural circuits, rather than linking genes to multifaceted clinical syndromes[34]. Furthermore, new methodological approaches have been proposed to simultaneously research multiple dimensions of ELA at a population level, by using public health records such as hospital admissions[39]. However, the authors note that a portion of children is exposed simultaneously to repeated adverse

experiences across social/health/family dimensions. For this sub-population, which carries the highest disease burden, it is still not possible to disentangle the various dimensions of ELA, not even with datasets of hundreds of thousands of participants[39]. Despite the disagreements about which theoretical framework to use, the bottleneck of human ELA research is clear: mechanistically investigating what ELA causes to the human brain is limited by the availability of data.

Many have therefore turned to rodents to study the underlying biological mechanisms of (aspects of) ELA. In *Chapter 9*, we provide a rationale for the use of ELA models in rodents, as well as their limitations. The ELA rodent models investigated in this thesis focus on a specific aspect of postnatal ELA, i.e. altered maternal care. Rodents and humans are altricial species: they are born under-developed and require the care of a primary caregiver upon birth. In rodents, there are three main paradigms to model altered maternal care. The first paradigm is based on the absence of (a relationship with) the mother for certain periods during the day. Mother and pups are separated for 1-8 hours per day over multiple days (maternal separation) during the first postnatal weeks or for a single prolonged period of 24h (maternal deprivation). Pups can also be individually separated, a variant of maternal separation referred to in this thesis as "isolation". The second paradigm is based on a disruption of maternal care, rather than its complete absence. In this model, the dam is housed with limited nesting and bedding material[40]. The behavior of the mother becomes more fragmented and unpredictable: the quality rather than the quantity of maternal care is altered, which is assumed to be more comparable to human ELA than the separation models. Lastly, the third paradigm is based on the natural variation of maternal care[41,42], as a proxy of quality of life. In this model, maternal care provided by the dam is analyzed in terms of pup-directed behaviors, such as licking, grooming and arch-backed nursing. Dams are then categorized as giving low or high amounts of care, referring to those mothers 1 standard deviation below and above respectively the average amount of maternal care. Pups receiving low amounts of care have neuroendocrine and behavioral phenotypes resembling those of pups with a history of maternal separation and deprivation[41].

Although all of these models have been applied at different times during the rodent brain development, in this thesis we focus on altered maternal care during the first two postnatal weeks. The choice of timing is linked to the development of the stress system. Early postnatally, rodents are hyporesponsive to stress[43]. During this period, the adrenals are less responsive to ACTH (see **Box 1**), as a consequence the HPA axis is relatively insensitive to mild insults[44]. This hyposensitivity is maintained by maternal care. Disruption in maternal care

therefore lifts the control system on the HPA axis, which – by becoming overactive for this period – will impact brain development and behavior[45]. The presence of a stress hyporesponsive period has been observed also in humans[46] and other mammals[47]; furthermore, the importance of parental care in mediating the HPA axis has been shown in human as well as nonhuman primates[48]. In all, although rodent studies can only model specific aspects of the human experience of ELA, they hold good construct and face validity[48,49], and can be used to increase our understanding of the mechanisms behind ELA and brain development.

Box 3 – Integration of information

Integration of information is a general term to describe the merging of information from multiple sources. An example of structure that integrates information is the human brain: the brain receives information from its sensory systems, which are then integrated to make decisions[50]. The job of a scientist is similar to that of the brain: a scientist gathers data from experiments (i.e. his/her 'sensory systems'), which he/she interprets to understand how biology works (i.e. his/her 'decisions'). Often however, the results of the experiments to integrate are not in agreement one another. A scientist will therefore need to understand where the discrepancy comes from. He/she will generally evaluate three categories of plausible reasons: 1) biological (e.g. different sexes explored), 2) methodological (e.g. use of reagent), and 3) statistical (e.g. linked to outlying or missing values). For example, Kanatsou and colleagues (2017,[51]) reported in their publication that early life adversity (ELA) both increases and decreases memory in mice depending of the behavioral test. A biological explanation could be that the behavioral tests used measure two different types of memory (stressful vs non-stressful). A methodological explanation could be (e.g.) that a female researcher performed the first and a male researcher performed the second experiment. A statistical explanation could be that memory does not increase, and that the difference in experiments is linked to sampling variation. Here, the experiments were performed in the same animals. If instead the experiments would have been performed in two different laboratories, additional confounders should be considered, e.g. the strain used or any other element of the (often heterogeneous) experimental design. Dissecting where the discrepancy comes from is therefore not straightforward, especially when the experiments are conducted from multiple laboratories, hence with heterogeneous methodologies. While obtaining

discordant findings is part of the self-correcting nature of science[52], a 2005 inflammatory paper suggested that the discrepancies in published study results are especially linked to (lack of) methodological rigor[53], including small sample sizes[54,55], p- hacking[56] and HARKing[57] (reviewed in [58]). Indeed, effect sizes are often inflated[59] and scientific results' replication fails between 30 and 70% of times, depending on research field[60–62]. This is generally referred to as "reproducibility crisis" when the same data and code are used to re-generate the results, or as "replicability crisis" when new data is collected to reach the same conclusion[63]. The proposed solutions to address these issues regard i) changes in the incentive structure of science, ii) increased transparency of methods and data, and iii) a reform in academic publishing[64]. While these reforms are already taking place and will be essential for the future, scientists (will) still face the challenge of interpreting information from multiple sources. Robust information integration methodologies are therefore an essential part of a scientist's toolkit to make choices and decisions[65]. Although integration of information is a general term, how we integrate information depends specifically on the type of data to integrate.

In this thesis, we focus on integrating past data, mostly already generated and analyzed, with the intent to give it a new life. We collected and re-analyzed data from other laboratories, either as summary statistics (e.g. from literature) or as raw data (e.g. through consortia). The main methodology at our disposal is evidence synthesis, which can be used when multiple studies investigate – in principle – the same construct[66]. Evidence synthesis can be supported by different statistical methodologies, such as meta-analysis[66] (Chapter 3-9) or Bayesian updating (Chapter 10)[67]. Meta-analysis is a statistical process to quantitatively compare and summarize separate studies[65]. Meta-analysis does not pool data to achieve a larger sample size; it adapts systematic methods to account for differences in e.g. sample size, variability (heterogeneity) in study approach, and dependency of observations[68]. Bayesian updating is a specific technique of Bayesian evidence synthesis, i.e. the Bayesian approach to meta-analyses[69]. Bayesian updating describes the process of sequentially updating one's knowledge based on the data. For example, the knowledge a scientist has before experiment A is prior knowledge. After experiment A, it is posterior knowledge. This posterior knowledge, however, becomes prior knowledge before experiment B is conducted, etc. While both meta-analyses and Bayesian evidence synthesis can be used to achieve integration of information, each confers specific possibilities. For example, Bayesian evidence synthesis 1) can integrate evidence from multiple sources (e.g. different types of experiments, or experiment vs expert knowledge), 2) describes uncertainty as conditional to the currently

available evidence, and 3) can be used to test informative hypotheses. However, due to the limited availability of software and the challenges in determining prior beliefs[69], frequentist meta-analysis often remains the methodology of choice. Despite the differences in approach available, here we used evidence synthesis not to "summarize" literature in a number (e.g. an effect size), but rather to apply a systematic methodology[70] (typical of the meta-analyses) that could serve an hypothetical-deductive approach to research[71] (typical of the Bayesian framework). The systematic methodology required assessment of the methodological quality of the included studies, of their risk of bias, of their assumptions, of their analyses (e.g. missing values). It accounts for study-specific effects, and it assesses heterogeneity. Evidence synthesis rationalizes and structures our ability to make conclusions. The hypothetical-deductive approach conceptualizes experiments as a cumulation of knowledge, and views data and data analysis with respect of consistency and replicability of the findings. We therefore used evidence synthesis as our tool to distinguish real effects from potential biases; promising directions from problems in reproducibility. Our goal was to make the best conclusions – possible at this time, with the data currently available – on the effects of stress on brain and behavior.

Aim and approach

The studies described in this thesis address the goal of information integration in stress research. Specifically, the **objectives** of this thesis are a first step towards:

A. integrating information related to the healthy acute stress response (in rodents and humans);
B. integrating information related to chronic stress experienced early in life (in rodents);
C. developing methodologies for information integration.

Information integration is here viewed not only in a biological sense, e.g. which brain areas at which time-point are important for stress, but more broadly in the way in which we (collectively) perform science. As a single brain area cannot explain a behavior, and a single hormone cannot explain the HPA axis, this thesis is based on the view that different laboratories should act together for knowledge integration to be achieved. The approach proposed in this thesis reflects these values:

I. we performed experiments in mice investigating the whole-brain, rather than pre-specified brain areas;
II. we founded the RELACS consortium and actively participated in the stress-NL consortium to collect information at the individual participant level of multiple laboratories, for humans' as well as rodents' data;
III. we extensively reviewed literature with systematic approaches, thereby categorizing and summarizing decades of previous research;
IV. we provide all data and all codes freely available online, following in full spirit an Open way to Science.

Outline of the thesis

This thesis is divided into three parts, each addressing an aim (A-C) described above.

In **Part A** (*Chapter 2 and 3*), we investigate the acute stress response with two separate studies focusing on the rodent brain (*Chapter 2*) and on salivary cortisol concentration in humans (*Chapter 3*). Although these studies have been conducted respectively in mice and humans, they share as a common feature the investigation of an acutely stressful situation as a dynamic process that occurs in time. In *Chapter 2*, we used whole-brain immunohistochemistry to investigate in mice the activation of the entire brain after foot-shock, with single cell resolution and over time. By using a whole-brain approach, we challenge the view that the effects of acute stress are limited to the historical stress-sensitive brain areas (e.g. the limbic system), but rather impact the whole brain. We propose this as a translational technique: it enables the investigation of functional networks like in human fMRI studies, while maintaining the advantages of rodent experiments, i.e. single cell resolution, availability of tissue, and ease of manipulation. The challenge of this chapter was analytical: it required the

development of a pre-processing and analytical pipeline, which is now available as R package. In *Chapter 3*, we created a database of human stress studies conducted in the Netherlands, the stress-NL database. Of note, while we developed the database, the stress-NL consortium already existed when we became involved. This data-sharing initiative was created to accurately portray the multivariable essence of the acute stress response. It is an accurate inventory of (neuro)biological, physiological and behavioral data from laboratory-based human studies that used acute stress paradigms. We provide example analyses on salivary cortisol concentrations. This chapter showcases the potential of combining and reusing existing data for meta-analytical, proof-of-concept and exploratory analyses.

Part B of this thesis (*Chapter 4 to 9*) focuses on another aspect of stress: continuous, repeated and/or severe exposure during childhood, potentially causing changes in how the brain develops and functions. Specifically, Part B aims to provide a general overview of the effects of early life adversity on adult behavior and neurobiological changes in rodents. Rodent models have been extensively used to mechanistically investigate the long-term effects of ELA; however, methodologies are often heterogeneous and the resulting findings rather incoherent and difficult to interpret. To overcome this limitation, here we systematically synthetize the scientific knowledge on the effects of ELA, by integrating information with systematic reviews and meta-analyses of the literature. In *Chapter 4*, we aimed to establish a causal link between ELA and changes in behavior in rodents, specifically on anxiety, memory and social behavior. *Chapters 5 to 8* investigated different aspects of the neurobiology, with a focus on the effects of ELA on the monoamines' systems (*Chapter 5 and 6*), on the brain's expression of immediate early genes such as c-fos (*Chapter 7*) and on structural plasticity (*Chapter 8*), including morphology, neurogenesis and BDNF expression. Finally, in *Chapter 9* we extracted 7 principles of ELA in rodents based on the entire body of information that we gathered in the previous chapters of Part B. Specifically, *Chapter 9* can be conceived as a discussion of Part B of this thesis. This chapter experiments with a new way of integrating information: it provides a detailed yet broad overview of ELA, which is a commentary, yet supported by quantitative meta-analytical statements.

In **Part C**, we focus on methodological aspects of integration information. In *Chapter 10*, we introduce a statistical method to integrate historical control data into new experiments, to increase the statistical power and reliability of animal research. This chapter was inspired by the meta-analytic work described in Part B, and it aims to answer a societal need as well as a scientific feasibility issue. The central idea is that we can give a new life to past control data, i.e. to re-use these data to improve the reliability of future experiments. In *Chapter 11*, we provide an overview of all the software developed for this thesis. We put a specific focus on interactive visualizations to increase the utilization and value of generated data.

Finally, *Chapter 12* concludes the thesis with a summary of the main findings of the previous chapters. We discuss specifically methodological limitations of the approach and provide suggestions for future directions.

References

1. de Kloet, E. R., Joëls, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nat Rev Neurosci* 6, 463–475 (2005).
2. Sterling, P. & Eyer, J. Allostasis: A new paradigm to explain arousal pathology. in *Handbook of life stress, cognition and health* 629–649 (John Wiley & Sons, 1988).
3. McEwen, B. S. Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences* 840, 33–44 (1998).
4. Jackson, M. Evaluating the Role of Hans Selye in the Modern History of Stress. in *Stress, Shock, and Adaptation in the Twentieth Century* (eds. Cantor, D. & Ramsden, E.) (University of Rochester Press, 2014).
5. Scott-Solomon, E., Boehm, E. & Kuruvilla, R. The sympathetic nervous system in development and disease. *Nat Rev Neurosci* 22, 685–702 (2021).
6. McEwen, B. S. et al. Mechanisms of stress in the brain. *Nat Neurosci* 18, 1353–1363 (2015).
7. Chrousos, G. P. Stress and disorders of the stress system. *Nat Rev Endocrinol* 5, 374–381 (2009).
8. Goldstein, D. S. Differential responses of components of the autonomic nervous system. *Handb Clin Neurol* 117, 13–22 (2013).
9. Lupien, S. J., McEwen, B. S., Gunnar, M. R. & Heim, C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci* 10, 434–445 (2009).
10. Sapolsky, R. M. Stress and the brain: individual variability and the inverted-U. Nat Neurosci 18, 1344–1346 (2015).
11. Selye, Hans. A Syndrome Produced by Diverse Nocuous Agents. *Nature* 138, 32 (1936).
12. Harris, B. N. Stress hypothesis overload: 131 hypotheses exploring the role of stress in tradeoffs, transitions, and health. *General and Comparative Endocrinology* 288, 113355 (2020).
13. Yam, Y. B. *Introduction to the Study of Complex Systems.* (Edward Elgar Handbook of Research Methods in Complexity Science, 2020).
14. Herman, J. P. & Tasker, J. G. Paraventricular Hypothalamic Mechanisms of Chronic Stress Adaptation. *Front Endocrinol (Lausanne)* 7, 137 (2016).
15. Mcewen, B. S. Protection and Damage from Acute and Chronic Stress: Allostasis and Allostatic Overload and Relevance to the Pathophysiology of Psychiatric Disorders. *Annals of the New York Academy of Sciences* 1032, 1–7 (2004).
16. Nixon, M. et al. ABCC1 confers tissue-specific sensitivity to cortisol versus corticosterone: A rationale for safer glucocorticoid replacement therapy. *Sci Transl Med* 8, 352ra109 (2016).
17. Lopez, J. P. et al. Single-cell molecular profiling of all three components of the HPA axis reveals adrenal ABCB1 as a regulator of stress adaptation. *Sci Adv* 7, eabe4497 (2021).
18. Joëls, M. & Baram, T. Z. The neuro-symphony of stress. *Nat Rev Neurosci* 10, 459–466 (2009).
19. Bali, A. & Jaggi, A. S. Electric foot shock stress: a useful tool in neuropsychiatric studies. *Reviews in the Neurosciences* 26, 655–677 (2015).
20. von Dawans, B., Kirschbaum, C. & Heinrichs, M. The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology* 36, 514–522 (2011).
21. Silverthorn, D. U. & Michael, J. Cold stress and the cold pressor test. *Adv Physiol Educ* 37, 93–96 (2013).
22. Smeets, T. et al. Introducing the Maastricht Acute Stress Test (MAST): a quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology* 37, 1998–2008 (2012).
23. Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E. & Domes, G. Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology* 101, 186–192 (2019).
24. Kaffman, A., Herringa, R. J. & Sanchez, M. M. Editorial: Effects of Early Life Stress on Neurodevelopment and Health: Bridging the Gap Between Human Clinical Studies and Animal Models. *Frontiers in Human Neuroscience* 15, 596 (2021).
25. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nat Rev Neurosci* 17, 652–666 (2016).
26. World Health Organization. Preventing child maltreatment : a guide to taking action and generating evidence / World Health Organization and International Society for Prevention of Child Abuse and Neglect. *Guide sur la prévention de la maltraitance des enfants : intervenir et produire des données* (2006).
27. Shonkoff, J. P. et al. The Lifelong Effects of Early Childhood Adversity and Toxic Stress. *Pediatrics* 129, e232–e246 (2012).

28. Dube, S. R., Felitti, V. J., Dong, M., Giles, W. H. & Anda, R. F. The impact of adverse childhood experiences on health problems: evidence from four birth cohorts dating back to 1900. *Prev Med* 37, 268–277 (2003).
29. Dube, S. R. et al. Childhood Abuse, Neglect, and Household Dysfunction and the Risk of Illicit Drug Use: The Adverse Childhood Experiences Study. *Pediatrics* 111, 564–572 (2003).
30. Ando, M. et al. Attributable and absolute risk of lung cancer death by smoking status: findings from the Japan Collaborative Cohort Study. *Int J Cancer* 105, 249–254 (2003).
31. Teicher, M. H., Gordon, J. B. & Nemeroff, C. B. Recognizing the importance of childhood maltreatment as a critical factor in psychiatric diagnoses, treatment, research, prevention, and education. *Mol Psychiatry* 1–8 (2021) doi:10.1038/s41380-021-01367-9.
32. McLaughlin, K. A. et al. Childhood Adversities and First Onset of Psychiatric Disorders in a National Sample of US Adolescents. *Archives of General Psychiatry* 69, 1151–1160 (2012).
33. Struck, N. et al. Childhood maltreatment and adult mental disorders - the prevalence of different types of maltreatment and associations with age of onset and severity of symptoms. *Psychiatry Res* 293, 113398 (2020).
34. Montalvo-Ortiz, J. L., Gelernter, J., Hudziak, J. & Kaufman, J. RDoC and Translational Perspectives on the Genetics of Trauma-Related Psychiatric Disorders. *Am J Med Genet B Neuropsychiatr Genet* 171, 81–91 (2016).
35. Gee, D. G. Early Adversity and Development: Parsing Heterogeneity and Identifying Pathways of Risk and Resilience. *AJP* 178, 998–1013 (2021).
36. Evans, G. W., Li, D. & Whipple, S. S. Cumulative risk and child development. *Psychol Bull* 139, 1342–1396 (2013).
37. McLaughlin, K. A. & Sheridan, M. A. Beyond Cumulative Risk: A Dimensional Approach to Childhood Adversity. *Curr Dir Psychol Sci* 25, 239–245 (2016).
38. McLaughlin, K. A., Sheridan, M. A., Humphreys, K. L., Belsky, J. & Ellis, B. J. The Value of Dimensional Models of Early Experience: Thinking Clearly About Concepts and Categories. *Perspect Psychol Sci* 16, 1463–1472 (2021).
39. Rod, N. H., Bengtsson, J., Elsenburg, L. K., Taylor-Robinson, D. & Rieckmann, A. Hospitalisation patterns among children exposed to childhood adversity: a population-based cohort study of half a million children. *The Lancet Public Health* 6, e826–e835 (2021).
40. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A Novel Mouse Model for Acute and Long-Lasting Consequences of Early Life Stress. *Endocrinology* 149, 4892–4900 (2008).
41. Liu, Dong et al. Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science* 277, (1997).
42. Meaney, M. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual review of neuroscience* (2001) doi:10.1146/ANNUREV.NEURO.24.1.1161.
43. Levine, S. Infantile experience and resistance to physiological stress. *Science* 126, 405 (1957).
44. Rosenfeld, P., Suchecki, D. & Levine, S. Multifactorial regulation of the hypothalamic-pituitary-adrenal axis during development. *Neurosci Biobehav Rev* 16, 553–568 (1992).
45. Stanton, M. E., Gutierrez, Y. R. & Levine, S. Maternal deprivation potentiates pituitary-adrenal stress responses in infant rats. *Behav Neurosci* 102, 692–700 (1988).
46. Gunnar, M. R. & Donzella, B. Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology* 27, 199–220 (2002).
47. Wang, L. et al. Neuroendocrine responses to social isolation and paternal deprivation at different postnatal ages in Mandarin voles. *Dev Psychobiol* 56, 1214–1228 (2014).
48. Suchecki, D. Maternal regulation of the infant's hypothalamic-pituitary-adrenal axis stress response: Seymour 'Gig' Levine's legacy to neuroendocrinology. *Journal of Neuroendocrinology* 30, e12610 (2018).
49. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr Opin Psychol* 15, 174–181 (2017).
50. Toker, D. & Sommer, F. T. Information integration in large brain networks. *PLOS Computational Biology* 15, e1006807 (2019).
51. Kanatsou, S. et al. Overexpression of Mineralocorticoid Receptors in the Mouse Forebrain Partly Alleviates the Effects of Chronic Early Life Stress on Spatial Memory, Neurogenesis and Synaptic Function in the Dentate Gyrus. *Front Cell Neurosci* 11, 132 (2017).
52. Stroebe, W. & Strack, F. The Alleged Crisis and the Illusion of Exact Replication. *Perspect Psychol Sci* 9, 59–71 (2014).
53. Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLOS Medicine* 2, e124 (2005).
54. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14, 365–376 (2013).
55. Bonapersona, V., Hoijtink, H., RELACS Consortium, Sarabdjitsingh, R. A. & Joëls, M. Increasing the statistical power of animal experiments with historical control data. *Nat Neurosci* 24, 470–477 (2021).
56. Wagenmakers, E.-J. A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review* 14, 779–804 (2007).
57. Kerr, N. L. HARKing: Hypothesizing After the Results are Known. *Pers Soc Psychol Rev* 2, 196–217 (1998).
58. Lewandowsky, S. & Oberauer, K. Low replicability can support robust and efficient science. *Nat Commun* 11, 358 (2020).
59. Ioannidis, J. P. A. Why Most Discovered True Associations Are Inflated. *Epidemiology* 19, 640–648 (2008).
60. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*

349, aac4716 (2015).
61. Camerer, C. F. et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nat Hum Behav* 2, 637–644 (2018).
62. Errington, T. M., Denis, A., Perfito, N., Iorns, E. & Nosek, B. A. Challenges for assessing replicability in preclinical cancer biology. *eLife* 10, e67995 (2021).
63. National Academies of Sciences, E. et al. Understanding Reproducibility and Replicability. *Reproducibility and Replicability in Science* (National Academies Press (US), 2019).
64. Nelson, N. C., Ichikawa, K., Chung, J. & Malik, M. M. Mapping the discursive dimensions of the reproducibility crisis: A mixed methods analysis. *PLOS ONE* 16, e0254090 (2021).
65. Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R. & Howells, D. W. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. J Cereb *Blood Flow Metab* 34, 737–742 (2014).
66. Higgins, Julian PT *et al.* Cochrane Handbook for Systematic Reviews of Interventions. in *Cochrane Handbook for Systematic Reviews of Interventions* i–xxi (John Wiley & Sons, Ltd, 2008). doi:10.1002/9780470712184.fmatter.
67. Gelman, Andrew, Carlin, John B., Stern, Hal S., & Rubin, Donald B. *Bayesian Data Analysis.* (Chapman and Hall/CRC, 1995).
68. Egger, M. & Smith, G. D. Meta-analysis: Potentials and promise. *BMJ* 315, 1371–1374 (1997).
69. Sutton, A. J. & Abrams, K. R. Bayesian methods in meta-analysis and evidence synthesis. 27.
70. Hox, J. J. *Multilevel analysis: Techniques and applications*, 2nd ed. x, 382 (Routledge/Taylor & Francis Group, 2010).
71. Gelman, A. & Shalizi, C. R. Philosophy and the practice of Bayesian statistics. *Br J Math Stat Psychol* 66, 8–38 (2013).

PART A

CHAPTER 2

The mouse brain after foot-shock: temporal dynamics at a single cell resolution.

PNAS (2022), doi: 10.1073/pnas.2114002119

Bonapersona V[1]
Schuler H[1,2]
Damsteegt RD[1]
Adolfs Y[1]
Pasterkamp RJ[1]
van den Heuvel MP[3]
Joëls M[1,4^]
Sarabdjitsingh RA[1^]

1 Dept. Translational Neuroscience, UMC Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands.
2 Integrated Program in Neuroscience, McGill University, Montréal, QC, Canada.
3 Connectome Lab, Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam, The Netherlands.
4 University Medical Center Groningen, Groningen, The Netherlands.

^ Contributed equally

Abstract

Acute stress leads to sequential activation of functional brain networks. A biologically relevant question is exactly which (single) cells belonging to brain networks are changed in activity over time after an acute stress, across the entire brain. We developed a novel pre-processing and analytical pipeline to chart whole-brain immediate early genes' expression – as proxy for cellular activity – after a single stressful foot-shock, in 4 dimensions; that is, from functional networks up to 3D single-cell resolution, and over time. The pipeline is available as R-package. Most brain areas (96%) showed increased numbers of c-fos+ cells after foot-shock, yet hypothalamic areas stood out as being most active and prompt in their activation, followed by amygdalar, prefrontal, hippocampal and finally thalamic areas. At the cellular level, c-fos+ density clearly shifted over time across subareas, as illustrated for the basolateral amygdala. Moreover, some brain areas showed increased numbers of c-fos+ cells, while others – like the dentate gyrus – dramatically increased c-fos intensity in just a subset of cells, reminiscent of engrams; importantly, this 'strategy' changed after foot-shock in half of the brain areas. One of the strengths of our approach is that single-cell data were simultaneously examined across all of 90 brain areas and can be visualized in 3D in our interactive web-portal.

Introduction

Acute stress leads to the activation of multiple functional brain networks, as demonstrated in humans using functional magnetic resonance imaging (fMRI, for reviews[1,2]). Yet, spatial resolution beyond the level of (networks of) nuclei is currently not possible with fMRI. This severely limits our ability to answer an important biological question: which (single) cells belonging to brain networks are changed in activity over time after an acute stress, across the entire brain? And do all cells respond in the same way? This most likely is not the case. Previous rodent studies have established that even small areas such as the basolateral amygdala (BLA) have a heterogeneous cellular composition[3] and contribute to a wide array of behaviours[4], presumably linked to long-range connectivity[3]. Other studies in animals have highlighted that just a few cells within the dentate gyrus are greatly responsive to acute stress; this high degree of responsiveness was linked to demethylation of specific CpG sites[5,6]. Thus, studies confined to subparts of the brain point to heterogeneity in the cellular response to stress and emphasize the necessity of a whole-brain approach with cellular resolution.

In principle, whole-brain microscopy can be used to address these questions. This technique can provide a snapshot of transcriptional[7] cellular activity throughout the whole-brain[8] by staining for immediate early genes (IEG)[9]. The analytical challenges are not trivial. Several tools have been developed to detect active cells and to register them to an atlas (for an excellent review of open-source tools, see[10]). Most of these tools even offer built-in options for visualization, but to date no study has thoroughly explored the subsequent steps of data analysis, i.e. dealing with missing values, batch effect corrections, normalization and transformation. Yet, these steps are essential: they can influence results and the interpretation of findings[11], as previously shown in several other fields (for example,[12–14]). Before embarking on complex whole-brain analyses as well as introducing a time-dynamic, we therefore first tackled how to clean and preprocess the data.

Whole-brain microscopy has excellent spatial resolution (~5µm,[8]) yet very poor time resolution, usually confined to a single time-point. Solving this conundrum was the second novel step in our approach. For this, we used a method developed and commonly used by many labs before (for example, [8,15–18]), using the IEG c-fos (Supplementary Note 1) as a post-mortem marker of cellular activity[19], to which we added a pseudo-time. Previous studies report that c-fos mRNA can peak at different times across brain areas after swim or restraint stress[20]. This suggests that there may be multiple waves of c-fos activation throughout the brain, which could be used to map the temporal dynamics across all brain areas up to the level of single cells and from minutes to hours after the initial stimulation. Determining shifts in activity during the different phases of the stress response can be a first step to clarify the temporal dimension of the stress response at a single cell level across the whole brain, a topic that has received little attention so far[21]. This approach moves the field beyond important earlier studies that looked at

c-fos expression – some even brain-wide (for example,[16–18]) – after a variety of stimuli and demonstrated both general patterns of activation that are typical for arousal, as well as transcriptional changes that seem to be stressor-specific, yet all confined to a single time-point[22].

Overall, to understand how single cells across the brain in 3D adapt their activity at various time-points after stress, we exposed adult male mice to a single stressful foot-shock and charted cellular activity across 89 areas, using c-fos staining as a proxy of cellular activity. A pipeline for data preprocessing and analysis at different spatial resolutions was developed, allowing investigation from the macro- (functional networks) to micro- (single cell resolution) scales and with a pseudo-time scale.

Results

Overview of the pipeline and quality control

To approach our biological questions, we first optimized, combined and expanded available methodological tools. Figure 1 summarizes the main features of the pipeline including experimental procedure, image processing (Step 1), data cleaning (Step 2), data pre-processing (Step 3) and analysis.

In brief, cell detection (Step 1) was performed with Imaris's spot object (Supplementary Figure 1a), after which it was aligned to the Allen Brain Reference Atlas (ABA) with Clearmap. Precision of alignment was assessed by comparing how sample images and template images would distort landmarks which were previously manually placed (Supplementary Figure 1b). The average absolute difference was 8.39 ± 5.88 µm (mean ± SD) in the horizontal plane and 10.92 ± 12.44 µm (mean ± SD, maximal displacement = 23.36 µm) in the sagittal plane, with more laterally placed landmarks being less precise; i.e., on average an uncertainty of roughly one soma. Alignment did not differ per condition, suggesting that alignment error should not affect our results. Of note, we excluded from the analysis 6 brain areas because we deemed their size too small for a reliable quantification (Supplementary Table 1). Until this step, we adapted tools developed by others (Imaris, Elastix[23] and Clearmap[8]); alternative tools (e.g. CellFinder[24], WholeBrain[25]) could have also been used for the same purpose.

Subsequent steps in the pipeline, however, were newly developed. Thus, during data cleaning (Step 2), first unspecific binding was mitigated by removing background signals and applying a mask of 3 voxels (~75 µm) around the borders of the brain and ventricles (Movie 1), and by removing cells with abnormally high intensity ($n_{\text{cells removed}}$=12). The background and mask step accounted for ~97% of the removed cells. Second, across all samples, ~5% of the brain areas showed some form of damage; these were removed from the analyses and re-imputed. Ultimately, the number of cells removed during the quality control procedure did not differ between groups (Supplementary Figure 2).

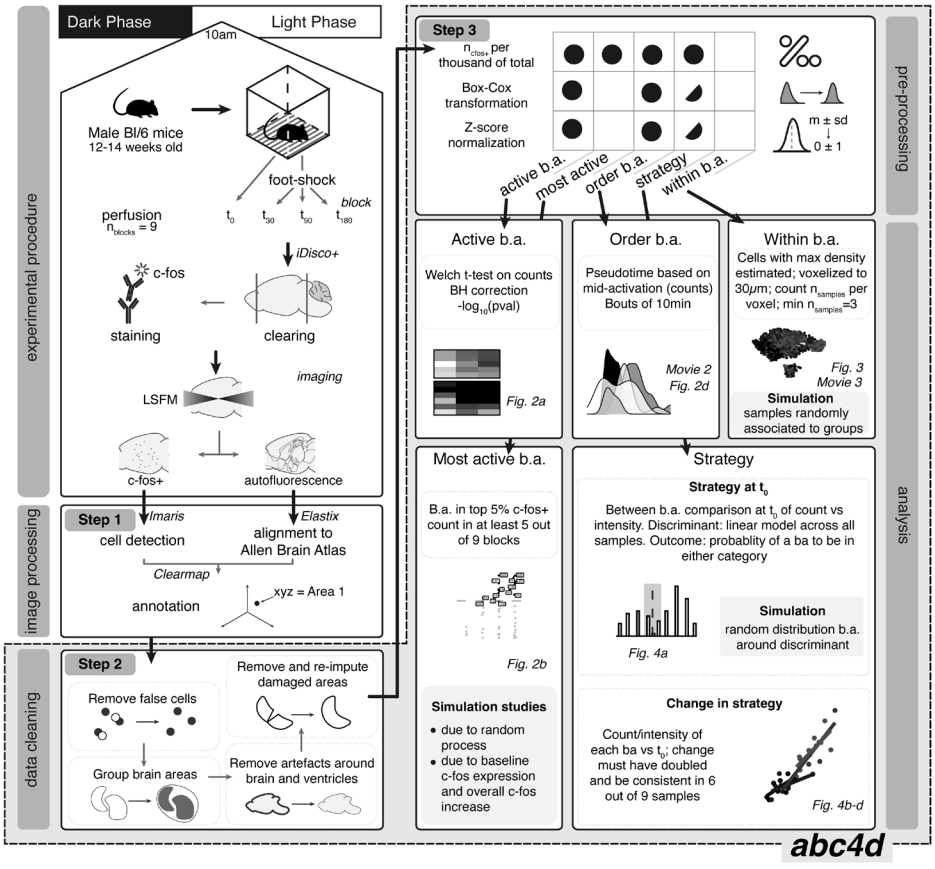


Figure 1. Schematic overview of the pipeline. Animals were perfused at different time-points ($n_{animals}$ = $n_{time\ point}$*n_{blocks} = 4*9 = 36) after foot-shock. Whole-brain samples were processed with iDisco+ protocol; c-fos+ cells imaged with light-sheet fluorescent microscopy. Cells were detected with Imaris and annotated to the Allen Brain Atlas with Clearmap. Output yielded xyz coordinates per cell. Quality control (data cleaning) consisted of removal of various artefacts and of grouping brain areas (b.a.) to the spatial resolution of interest. Data preprocessing of b.a. was performed for each of the analyses. Circle: step required; half-circle: step recommended but not required. Strategy refers to t_0 strategy categorization, as well as change of strategy over time. The frame at the bottom right summarizes the analyses conducted, and the main statistical decisions made. The Image processing step uses software developed by others (i.e. Imaris (v9.2.0, Bitplane), Elastix[23], Clearmap[8]). The steps by us developed (data cleaning (Step 2), pre-processing (Step 3) as well as analyses) are explained in detail in the Methods section and have been implemented in the *abc4d* package.

The next step in the pipeline is data pre-processing (Step 3). As summarized in the figure (Figure 1 upper left panel), data-preprocessing is specific for each analysis type. Of note, we used a block design, meaning that a 'block' (i.e., mice from the same cage, one animal for each time-point, processed simultaneously to avoid isolation stress of the last mouse in the cage) was the experimental unit of randomization and processing of samples. This type of design is essential for effective batch effect correction. The

data cleaning (Step 2) and pre-processing pipelines (Step 3) are available for similar future questions in the R package developed for the purpose, *abc4d* ("Analysis Brain Cellular activation in 4 Dimensions"), which is interoperable with several annotation/alignment tools (Step 1, e.g. Clearmap[8], CellFinder[24]).

Single cell activity is increased after foot-shock throughout the brain, but with spatial and temporal specificity

In answer to our biological question, we observed that – compared to control animals (see below) – the total number of c-fos+ cells (n_{cfos+}) across the brain was increased 30 minutes (t_{30}) after foot-shock induction and remained elevated at t_{90} and t_{180}. It returned to t^0 levels 300 minutes after foot-shock (Supplementary Figure 3). Across batches, n_{cfos+} was comparable to that of previous literature[8,15]. Of note, control animals (t_0) were placed in a foot-shock chamber but did not receive a foot-shock. As a consequence, they should be considered as a "mildly stressed" (novelty stressor) group rather than true baseline controls (for more information on all control groups, see Supplementary Information and Supplementary Figure 3).

To test the extent of c-fos+ expression throughout the brain, we performed pairwise comparisons (Welch t-test, one sided, Benjamini-Hochberg p-value correction) between each foot-shock time point (t_{30}, t_{90}, t_{180}) and t_0 (Figure 2A). 86 out of the 89 brain areas had a significant increase in c-fos+ cells in at least one of the time points. Only three brain areas were not significantly changed, i.e. medial preoptic nucleus, ventral anterior-lateral complex of the thalamus, and ventro-posterior complex of the thalamus. The time point t_{180} had the highest number of significant brain areas ($n_{sig\ brain\ areas}$=85), followed by t_{90} ($n_{sig\ brain\ areas}$=79) and t_{30} ($n_{sig\ brain\ areas}$=40). The effect sizes (g $\pm$ SD) ranged between -0.32 $\pm$ 0.23 (Midbrain raphe nuclei, t_{30} vs t_0) and 5.17 $\pm$ 0.96 (Subiculum, t_{90} vs t_0) with a mean of 1.62 $\pm$ 0.57.

Since nearly all brain areas were active in at least one time point, we aimed to identify which brain areas were more active than others. To answer this question, we identified for each block (i.e., a unique set of each time-point) the brain areas that had the highest (i.e., top 5% of the distribution) c-fos+ cell count density (per thousand of total, $n_{cfos+/tot}$). Under random circumstances, the same brain area would be in the top 5% in at least 5 out of 9 samples in about 1% of the cases, as illustrated by a simulation study (Methods). Being selected by at least 5 samples was therefore used as a criterion to define consistency of highly active brain areas. In our experimental data, the criterion was met by 8 brain areas (Figure 2 b), which belonged mostly to the hypothalamus (Figure 2 c). This number ($n_{highly\ active}$=8) was much higher than the 1% expected ($n_{randomly\ active}$=1) by sheer randomness. With a simulation study, we confirmed that this activation could also not be attributed to the spatial localization of c-fos throughout the brain, as reported by the Allen Brain Atlas (Supplementary Figure 4).

Next, we hypothesized that although in most brain areas the number of c-fos+ cells is

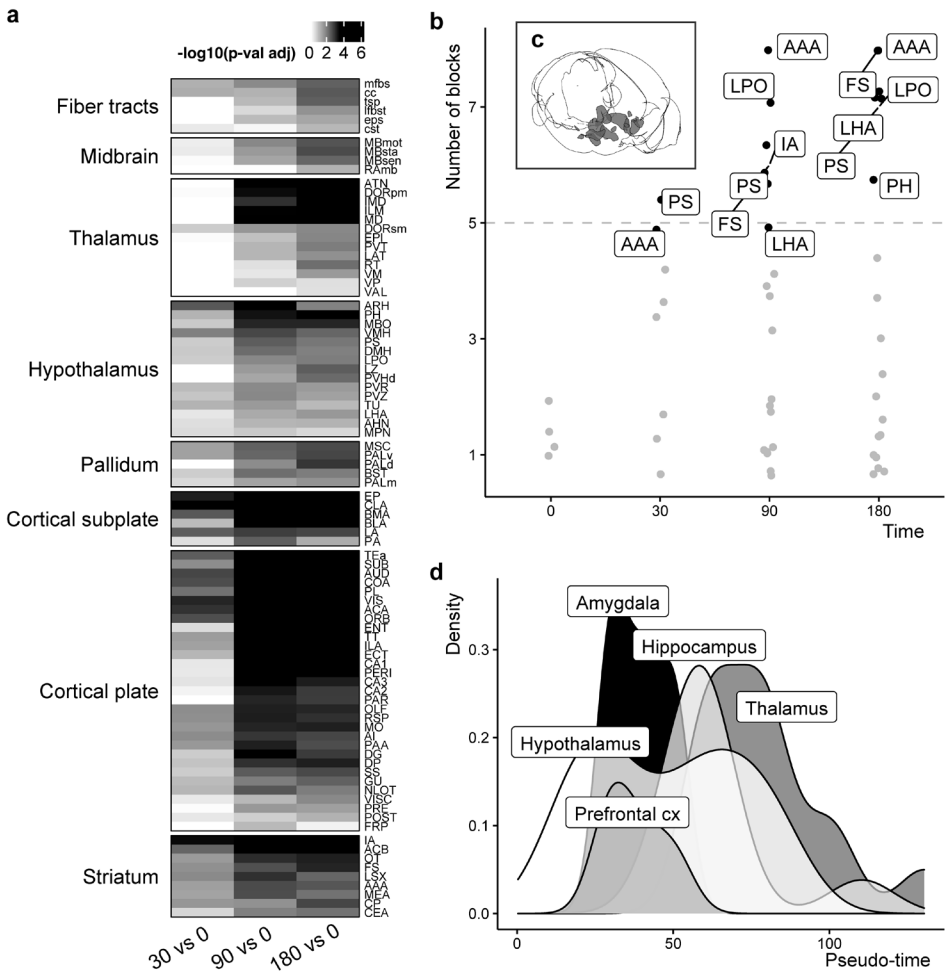


Figure 2. Most brain areas are activated by foot-shock, with spatial and temporal specificity. a) Heatmap of -$\log_{10}$ p-values derived from pairwise comparisons of each foot-shock time point (t_{30}, t_{90}, t_{180}) against t_0 for each brain area. White: p-val_{adj} >=0.05; $grey_{shades}$: p-val_{adj} < 0.05. The legend numbers correspond to the log values. b) A set of hypothalamic areas was consistently found to have the highest (top 5% of the distribution) number of c-fos+ cells (per thousand of total). The criterion for consistency was 5 out of 9 samples. Abbreviations are explained in Suppl. Table 1. c) Cartoon of the brain areas identified in b; created with *brainrender*[50] d). Functional order of brain areas' c-fos activity following foot-shock. Brain areas were ordered based on a pseudo-time depending on c-fos+ activation across the time points, and grouped based on functional categorization important for the stress response[26]. Hypothalamic areas are the first to reach the mid-point of their activation, followed by amygdalar, prefrontal, hippocampal and lastly thalamic areas. Of note, the functional order is based on the point of mid-activation of brain areas, rather than the first instance in which brain areas were activated. For an interactive visualization of the single brain areas rather than the categorization, see Movie 2.Time-dependent wave of activation within the Basolateral Amygdala

increased after foot shock, the peak of activation would not occur at the same time for every (network of) brain area(s). Brain areas are expected to be involved at different stages – and therefore at different times – of the stress response as earlier proposed, based on human fMRI studies[1]. We organized brain areas on a pseudo-time scale, based on the time-point in which a brain area (median across blocks) would reach the middle of its activation. This pseudo-time should only be interpreted relatively. We visualized the order of activation (Figure 2 d) of (networks of) the brain areas, using a functional categorization valuable to the stress response[26]. Based on this classification, hypothalamic areas were found to be activated first, followed by amygdalar and prefrontal, hippocampal, and finally thalamic areas. Movie 2 shows a visualization of all brain areas over time.

While the results so far confirm – in rodents – insights at the network level earlier obtained in humans with fMRI1, our main goal was to investigate dynamic brain activity after foot-shock with higher spatial resolution, up to the single cell level, which is a great advantage over e.g. fMRI studies. Rather than highlighting all areas, we here illustrate the findings for the BLA, an area key for the cognitive processing of a foot-shock[27] and stressful conditions in general[28]; for the remaining areas we refer to an open-source dynamic database (https:// utrecht-university.shinyapps.io/brain_after_footshock/), with which one can browse through all other regions investigated.

We hypothesized that the increase of c-fos+ cells was not uniform across the BLA; but, rather, may be restricted to different sub-parts or cells. For each sample independently, we identified the most densely activated part of the BLA, i.e., the part with the highest number (density) of c-fos+ cells relative to the rest of the BLA. All samples considered, there are obvious regional distributions across time points (Figure 3 a, Movie 3), which are not evident when samples are randomly associated to the experimental groups (Supplementary Figure 5 a).

We voxelized the xyz coordinates (voxel size: ~30µm x 30µm x 30µm) and visualized per time point which voxels have at least one cell from three different samples. As shown in Figure 3 b, at later time-points after foot-shock, i.e. t_{90} and t_{180}, the highest density of c-fos+ cells in the BLA was found to be more posterior (difference of ~96µm from anterior to posterior, 23% of BLA) than at t_0 and t_{30}.

Cells use different strategies of activation, which can change after foot-shock

A third biological question is whether all cells use a comparable activation 'strategy' at the various time-point after foot-shock. With 3D microscopy, one can count the n_{cfos+}, but also quantify the intensity of c-fos staining per cell. Among other parameters, a cell is considered c-fos+ if the intensity of c-fos is higher than the background (i.e., signal-to-noise ratio). As a consequence, one would expect the n_{cfos+} per brain area

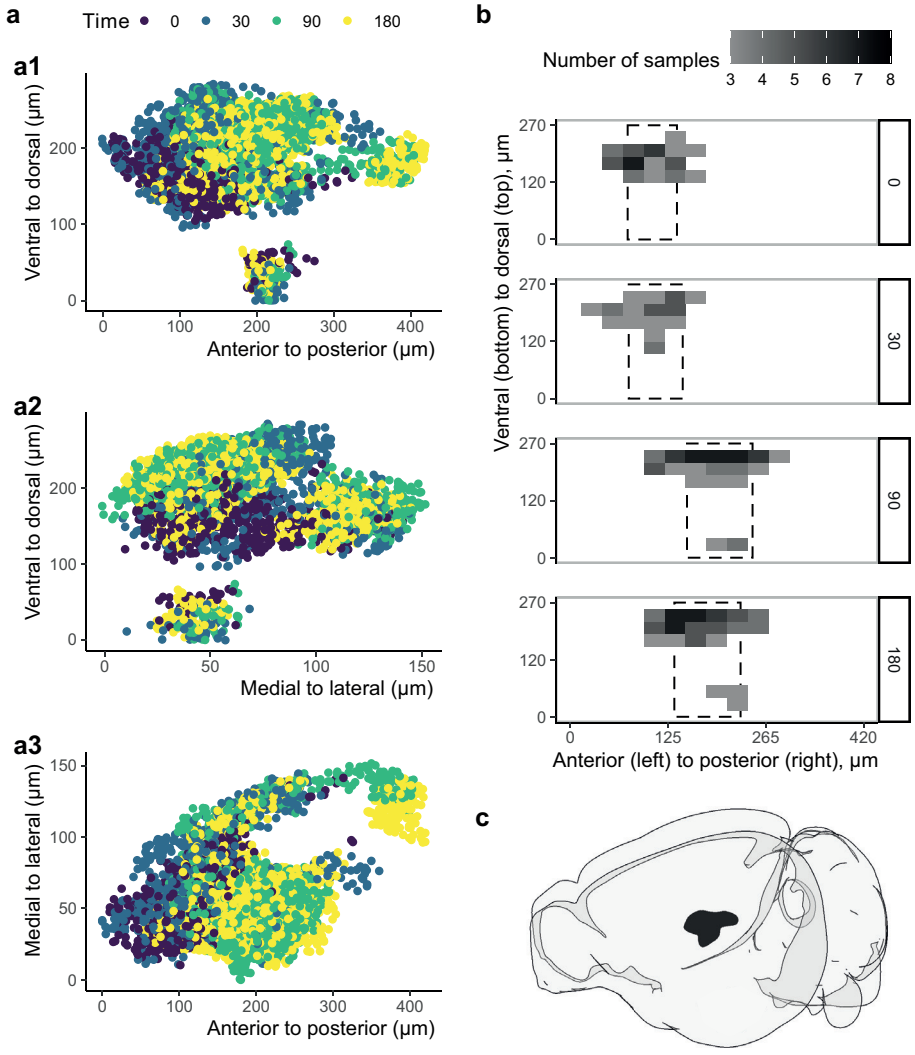


Figure 3. Changes in c-fos+ cell density within the Basolateral Amygdala. a) Cells in high density regions of the right BLA. The 3D cell coordinates are represented as a set of three 2D graphs, one for each couple of coordinates (xy, yz, xz). Each dot is a cell of a sample in a region with highest density. The colors refer to the different time points. b) The densest c-fos+ sub-part of the BLA moved from more anterior (t_0 and t_{30}) to more posterior after foot-shock (t_{90} and t_{180}). The BLA has been voxelized (voxel size 30x30x30µm), and the fill color refers to the number of samples with at least one cell in that voxel. The dashed box indicates the mean and SD per group along the posterior-anterior axis. c) Cartoon visualization of the right BLA in the same orientation of a1 and b, created with *brainrender*[50].

to strongly correlate with the average c-fos intensity. In other words, brain areas are expected to be normally distributed along the correlation line, as shown by our simulation (Supplementary Figure 5 b).

However, this was not the case (Supplementary Figure 5 c). Rather, different brain areas have a preferential strategy of activation: a few very active (i.e. low count, high intensity) cells, versus many lowly active cells (i.e. low intensity, high count). We categorized the brain areas of each t^0 sample (corresponding to a very mildly stressful condition) based on their 'strategy', i.e. their preference for increasing in count or intensity. Across samples, we then calculated the probability of each brain area to belong to either categorization (Figure 4 a). The results showed a bimodal distribution (Figure 4 a), which is clearly different from the normal distribution expected under our hypothesis (Supplementary Figure 5 b). Therefore, whether a brain area is activated by increasing n_{cfos+} or by increasing the average intensity of c-fos per cell is unlikely to be the result of a technical characteristics or of a random process.

Intensity and count are therefore expected to be related within brain areas, rather than across the whole brain. This relationship should be constant across all groups; if not, foot-shock must have induced transcriptional changes in specific subsets of cells. We therefore next examined whether the strategy of a brain area changes after foot-shock, relatively to t_0. For each time-point after foot-shock (t_{30}, t_{90}, t_{180}), we selected brain areas with a consistent (at least 6 out of 9 samples) change in either count and intensity, and calculated to what extent count and intensity were increasing compared to each other. We categorized brain areas as "changing strategy" if they at least doubled the increase in one category. 43 brain areas met these criteria (Figure 4 b). Of these, 30 increased activation by means of intensity rather than c-fos+ cell count, especially in the amygdala, hypothalamus and thalamus. Of note, the increase in intensity for the amygdalar nuclei was present only for the time points t_{90} and t_{180}. Figure 4 c displays two brain areas (BLA and Subiculum) as representative examples of activation strategy towards intensity and count, respectively. We also added a visualization of the dentate gyrus (Figure 4 c, part 3) as a validation, since this area has been described to increase intensity of IEG staining after stress in a very limited subset of cells[5,6].

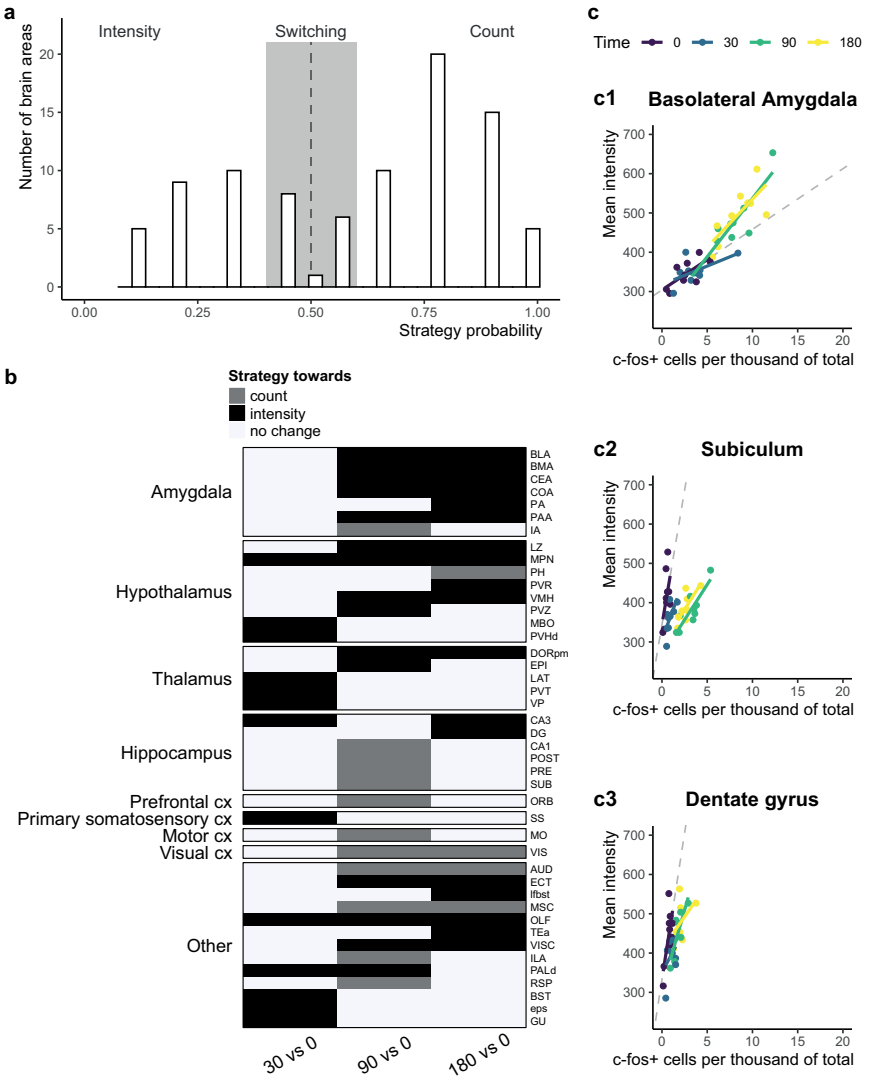


Figure 4. Brain areas have a preferential strategy of activation, which may change after foot-shock. a) Preferential strategy of brain areas based on t_0 data, i.e. the relationship between intensity and count per brain area. Histogram of the number of brain areas across the strategy probability. If a brain area would be activated by indiscriminately increasing c-fos+ cells and expression, the distribution would be Normal (Supplementary Figure 5b), with μ around 0.5. The bimodal distribution suggests that certain brain areas preferentially increase the number of c-fos+ cells (probability > 0.5, count), whereas other increase the mean c-fos expression (probability < 0.5, intensity). b) The strategy of brain areas can change after foot-shock. Binary heatmap of how strategy can change across brain areas for pairwise comparisons of time points. White corresponds to no change in strategy, black to a change towards intensity and grey to a change towards count. Of note, a change towards intensity does not necessarily mean that the brain area does not increase count; rather, it means that the increase in intensity cannot be explained by the increase in count alone. c) Representative examples of a brain area that after foot-shock changes strategy towards intensity (BLA, c1) or count (Subiculum, c2). The dentate gyrus (c3) was added for literature validation (see Discussion). Each dot represents one sample; the line represents that correlation between count and intensity per group. Dashed line represents what one would expect based on t_0 activation.

Discussion

Human fMRI studies over the past decades have shown that acute stress activates multiple functional brain networks, with hypothalamic and amygdalar areas being among the first to be activated, and areas linked to higher cognitive functions such as prefrontal cortex and hippocampus following in due course[1]. Yet, beyond the scale of (networks of) nuclei, e.g. up to single cells, very little is known about stress-induced effects at a whole brain level and over time. Are all cells among or within nuclei equally affected at various time-points after stress? Animal studies focusing on specific areas, e.g. the dentate gyrus[5,6], suggest not.

To address this question, we first had to develop a thorough and robust analytical pipeline, investigating changes in cellular activity after a highly stressful foot-shock in mice over time, by staining for the IEG c-fos and introducing a pseudo-time metric; this is summarized in Figure 1. Although we adapted tools developed by others to transform images into numeric data[8,23], we outline for the first time how to conduct the required subsequent steps of data analysis, i.e. data cleaning and preprocessing. Also, so far analyses were limited to a region-based approach, where the number of active cells is calculated for each brain area separately. Voxel-based analyses similar to MRI have only recently been developed[29]. Here, we took this one step further, and suggest analyses for i) the time dynamics and ii) the single cell level. The resolution up to the level of single cells is certainly one of the major advantages of whole-brain microscopy. In the future, c-fos+ cells could be characterized in more detail, being able to distinguish excitatory from inhibitory neurons, or neurons from glia cells, as the current findings confirm that c-fos staining is not confined to neurons only (for example,[30]). This could be achieved by multiple concurrent stainings (for example,[31,32]), or by computationally categorizing cellular morphology[33]. Re-stainings could also be an option, for example by using SWITCH[34] rather than iDisco+. Our pipeline can be applied independently of the type of clearing method or software used for alignment and annotation, and it can analyze 3D (i.e. whole-brain) or 4D (over time) experimental designs. Furthermore, it is interoperable with several other annotation/alignment tools. It is available in the newly developed R package *abc4d*, to which new improvements can be easily added in future. *Abc4d* also includes a framework of simulation studies, where the null hypotheses for different analyses can be investigated. An overview of the package is provided in the cheat-sheet (Supplementary Figure 6).

With this toolbox in place, we addressed our biological questions. Although foot-shock increased the activation of 96% of the brain areas, distinct temporal dynamics in networks of brain areas stood out. Thus, foot-shock first activated (cells in) hypothalamic areas, followed by amygdalar and prefrontal, hippocampal and lastly thalamic areas. This is largely in line with the earlier human literature using fMRI[1]. Importantly, while foot-shock is not a common stressor in humans, it nevertheless captures crucial elements of stress exposure in humans, involving physical characteristics such as discomfort and

psychological aspects like uncontrollability. This lends credibility to the common patterns seen across species. The percentage of active brain areas is higher than previously reported after a single prolonged stress[16], which suggests that investigating multiple time-points offers a more complete (dynamic) view of brain areas activated after stress. Furthermore, previous c-fos studies have identified common patterns of activation for rewarding and aversion stimuli, presumably linked to the aroused state[17,18,35]. We here identified the same "aroused pattern" of nuclei previously observed (at t_{90}), including – but not limited to - the cingulate cortex, nucleus accumbens, bed nucleus of the stria terminalis, lateral hypothalamus, periventricular hypothalamic nucleus, paraventricular thalamic). The absolute numbers of c-fos+ cells were comparable to previous literature investigating whole-brain c-fos with other paradigms[8,15]. We also provide a temporal pattern of functional activation, as visible in Figure 1.

Additionally, we demonstrated within the basolateral amygdala – a key area in the processing of a stressful foot-shock[27,28] – a clear shift after foot-shock from activation of cells in the lateral-anterior part towards a more posterior-medial subset of cells. Although we here present information about the basolateral amygdala only, data about all other areas investigated is available for closer scrutiny. To dive deeper into any area of interest, we provide an interactive interface on the data at our web portal (https://utrecht-university.shinyapps.io/brain_after_footshock/). For example, subparts of the bed nucleus of the stria terminalis (BSNT) were previously found to be similar in c-fos expression in 2D 60 minutes after a multimodal stress[18]. Our data (visible on our web portal) indeed confirms that 90 and 180 minutes after foot-shock the c-fos activation is widespread across the BSNT. However, this was not the case for earlier time points (t_0 and t_{30}), where the highest distribution of c-fos+ cells was more ventral.

Lastly, our approach allowing single cell investigation revealed that brain areas follow specific c-fos expression strategies that are skewed towards either an increase in number of c-fos+ cells or c-fos intensity per cell. Importantly, in a subset of areas the strategy changed after foot-shock. The finding that brain areas use a distinct strategy in c-fos cell activation under mildly stressful conditions such as a novel environment (in this case the shock-box at t_0) compared to exposure to a very stressful situation like an inescapable foot-shock is very novel. One can currently only speculate about the functional relevance of the two main cellular 'strategies'. Earlier studies showed that the expression of c-fos is proportional to the rate of firing of the cell[36]. If so, one could hypothesize that in certain brain areas many cells are slightly more activated after stress (i.e. express c-fos above detection threshold), whereas other brain areas may have only a few cells that are very strongly excited, which would lead to an increase in their c-fos intensity. Several previous studies (for example,[35,37,38]) suggested that c-fos expression may be different across cell populations. This could now be investigated brain-wide with new experiments. Of note, in amygdalar areas the increase in

intensity was particularly observed at 90 and 180 minutes after foot-shock, which is compatible with a gene-mediated, possibly glucocorticoid-dependent mechanism[21]. This observation that a limited number of cells gets highly activated fits extremely well with current views on engrams[39]. For instance, engram cells in the dentate gyrus were shown to be powerful in reversing behavior caused by chronic stress[40]. Moreover, using inducible IEG promotor approaches in areas of interest, others have shown that immediately after contextual fear conditioning – also involving foot-shock exposure – a consistent ~10% of (baso)lateral amygdala neurons becomes part of the engram, while participation of dentate granule cells is much lower[41]. This resembles the ~5% of BLA neurons with very high intensity staining we observed 90-180 minutes post foot-shock and the far lower number of high-intensity c-fos+ cells in the dentate. The somewhat lower percentage of c-fos+ cells in the BLA we observed may be explained by the recency of the foot-shock[41], the level of excitability prior to foot-shock[42] and/or the fact that we used c-fos rather than arc as IEG. Very high intensity of IEG staining in a small subgroup of dentate cells was also reported after swim stress, which was found to increase DNA demethylation in the dentate gyrus at specific CpG sites close to the c-fos transcriptional start site, in the gene promoter region of early growth response protein 1[5,6]. Overall, the fact that dentate cells indeed follow an 'intensity-strategy' lends credibility to our approach; the strength of our study is that we do not focus on a single area but can simultaneously examine and compare 89 regions.

There are some limitations to consider. The choice of c-fos as an activity marker is arguably appropriate in the case of acute stress exposure[19], but it is by no means the only IEG one could choose for the current approach. The cellular role of c-fos remains largely unknown[38]; therefore it is not possible to determine its exact function in our experimental set-up. Other markers of cellular activity may afford additional insights into the circuits being activated after stress. More than 100 genes have been classified as IEG[38,43], although only a subset is expressed in neurons[44]. *Arc* and *Egr1* were reported to be transcriptionally activated following acute stress in a multi-omics approach[45]. *Egr1* has a high expression maintained already by normal ongoing cellular activity[46], so this gene could potentially be used to investigate de-activation of brain areas. Lastly, IEGs are not equally expressed by all cells in all brain areas[38]. This means that the sensitivity of our method may not be equal across the whole brain, specifically it may be lower for subcortical/thalamic structures. In our analysis, we moderated this limitation with normalization/standardization steps, which showcases the importance of proper data processing. We also performed simulations using the baseline mRNA c-fos distribution of the Allen Brain Atlas to validate our findings against the non-homogeneity of c-fos across brain regions. Ideally, a combination of markers should be applied to get a completer view. Another technical limitation is linked to the current size restraints of imaging with light-sheet microscopy. In our study, we trimmed the most frontal and most caudal parts of the brains, and we excluded brain areas of small

volume (Supplementary Table 1) that could not be reliably measured. Researchers interested in small structures (especially laterally placed and without strong landmarks) may opt for other alignment methods that do not rely on autofluorescence, although to the best of our knowledge no previous study has quantified the displacement due to alignment. Although this does not impact the current methodology and the main finding of between- and within-area cellular differentiation in response to stress, some brain areas involved in the acute stress response (e.g. locus coeruleus) are missing. A solution could be to divide the brain for scanning, but to analyze the data together, after the appropriate corrections. This would also be a solution for those interested in hemisphere-specific effects (i.e. lateralization). In our experimental design, we did not randomize the direction of the brain within the microscope chamber. The right/left hemispheres were always each scanned by the same laser. Although we took care in laser calibration, a "laser-specific" effect cannot be excluded. We therefore refrained from investigating lateralization, although it is plausible to occur after acute stress[47]. A third consideration concerns the pseudo-time approach. On the one hand, it overcomes the absence of high frequency sampling (as possible with fMRI). On the other hand, it is a mere approximation of real-time processes. For example, it is likely that place cells[48] within the hippocampus are activated promptly when in a new environment. Our pseudo-time metric is based on the mid-point of activation of a brain area, rather than the instance when activation was first measured. As a consequence, it misses the temporal resolution to pick up the earliest changes. Thus, while the pseudo-time metric might be an acceptable approximation of the activation phase, the method gives little insight in the gradual turning off of brain areas, since this also depends on the half-life time of the c-fos protein. The half-life time may differ across brain areas[19], something that could be investigated with a meta-analytic approach.

Despite these limitations, the ready-to-use pipeline for 4D immunohistochemical whole-brain analysis presented in this report (and supported by a new R package) revealed that stressors like an acute foot-shock not only sequentially activate functional networks in the brain, but also specifically activate subsets of neurons, using different strategies of activation.

Materials and Methods

For an in-depth description of the methods, see Supplementary Methods. The protocol, data, scripts, acb4d R package and additional experimental information are available at https://osf.io/8muvw/. Data can also be interactively visualized at our web portal (https://vbonapersona.shinyapps.io/brain_after_footshock/). All animal procedures were approved by the Animal Ethical Committee at Utrecht University (license: AVD1150020184806), the Netherlands.

Experimental design

We used a block design (n_{block}=9), where each block had an animal for each experimental group (n_{time}=4) from the same cage. We used a total of $n_{animals}$=36. Control animals were identical to experimental, but did not receive the foot-shock.

Brains were cleared with iDisco+[8] for c-fos and imaged with a light-sheet microscopy. c-fos+ cells were detected with Imaris and aligned to the Allen Mouse Brain reference Atlas (25mm)[49] with Elastix[23] via Clearmap[8].

Samples underwent a thorough quality control. Pre-processing was required for region-based analyses (Figure 1 and Supplementary Methods).

Summary of analyses

To test activation from baseline, we used pairwise comparisons (Welch t-test, one-sided, alpha = 0.05, pval corrected with Benjamini-Hochberg (BH) procedure) for t_{30},t_{90},t_{180} against t_0 on $n_{cfos+/tot}$.

Brain areas were defined as "most active" if in at least 5 out of 9 blocks they were in the top 5% of most activated areas.

To order brain areas, we considered the time points on a "continuum" of pseudo-time, calculated per block the point of mid-activation, and grouped each brain area to the closest 10 minutes bout (binning).

To identify highest density within a brain area, we calculated how many samples (minimum 3) had at least one cell in each voxel (30µm per side). In each xyz direction, we calculated per time point the median and interquartile of the voxels' position.

To categorize the strategy, we calculated across samples the probability of a brain area to be towards count/intensity, using a linear model on t_0 (n_{cfos+} vs mean intensity of each brain area) as a criterion. Brain areas "changed strategy" if their rate of change relative to t^0 was at least doubled in either count or intensity.

We performed several simulation studies to exclude our findings were due to chance.

References

1. Hermans, E. J., Henckens, M. J. A. G., Joëls, M. & Fernández, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends Neurosci.* 37, 304–314 (2014).
2. Berretz, G., Packheiser, J., Kumsta, R., Wolf, O. T. & Ocklenburg, S. The brain under stress-A systematic review and activation likelihood estimation meta-analysis of changes in BOLD signal associated with acute stress exposure. *Neurosci. Biobehav.* Rev. 124, 89–99 (2021).
3. O'Leary, T. P. *et* al. Extensive and spatially variable within-cell-type heterogeneity across the basolateral amygdala. *eLife* 9, e59003 (2020).
4. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* 517, 284–292 (2015).
5. Chandramohan, Y., Droste, S. K., Arthur, J. S. C. & Reul, J. M. H. M. The forced swimming-induced behavioural immobility response involves histone H3 phospho-acetylation and c-Fos induction in dentate gyrus granule neurons via activation of the N-methyl-d-aspartate/extracellular signal-regulated kinase/mitogen- and stress-activated kinase signalling pathway. *Eur. J. Neurosci*. 27, 2701–2713 (2008).
6. Saunderson, E. A. *et* al. Stress-induced gene expression and behavior are controlled by DNA methylation and methyl donor availability in the dentate gyrus. *Proc. Natl. Acad. Sci.* U. S. A. 113, 4830–4835 (2016).

7. Clayton, D. F. The genomic action potential. *Neurobiol. Learn. Mem.* 74, 185–216 (2000).
8. Renier, N. *et* al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. *Cell* 165, 1789–1802 (2016).
9. Guzowski, J. F. *et* al. Mapping behaviorally relevant neural circuits with immediate-early gene expression. *Curr. Opin. Neurobiol.* 15, 599–606 (2005).
10. Tyson, A. L. & Margrie, T. W. Mesoscale microscopy and image analysis tools for understanding the brain. *Prog. Biophys. Mol. Biol.* (2021) doi:10.1016/j.pbiomolbio.2021.06.013.
11. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* 11, 702–712 (2016).
12. Goh, W. W. B., Wang, W. & Wong, L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol.* 35, 498–507 (2017).
13. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36, 421–427 (2018).
14. Tran, H. T. N. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 21, 12 (2020).
15. Kim, Y. et al. Whole-Brain Mapping of Neuronal Activity in the Learned Helplessness Model of Depression. *Front. Neural Circuits* 10, (2016).
16. Azevedo, H., Ferreira, M., Mascarello, A., Osten, P. & Guimarães, C. R. W. Brain-wide mapping of c-fos expression in the single prolonged stress model and the effects of pretreatment with ACH-000029 or prazosin. *Neurobiol. Stress* 13, 100226 (2020).
17. Cho, J.-H., Rendall, S. D. & Gray, J. M. Brain-wide maps of Fos expression during fear learning and recall. *Learn. Mem.* 24, 169–181 (2017).
18. Lin, X. et al. c-Fos mapping of brain regions activated by multi-modal and electric foot shock stress. *Neurobiol. Stress* 8, 92–102 (2018).
19. Kovács, K. J. Invited review c-Fos as a transcription factor: a stressful (re)view from a functional map. *Neurochem. Int.* 33, 287–297 (1998).
20. Cullinan, W. E., Herman, J. P., Battaglia, D. F., Akil, H. & Watson, S. J. Pattern and time course of immediate early gene expression in rat brain following acute stress. *Neuroscience* 64, 477–505 (1995).
21. Joels, M., Karst, H. & Sarabdjitsingh, R. A. Unraveling the Time Domains of Corticosteroid Hormone Influences on Brain Activity: Rapid, Slow, and Chronic Modes. *Pharmacol. Rev.* 64, 901–938 (2012).
22. Mukherjee, D. *et al.* Salient experiences are represented by unique transcriptional signatures in the mouse brain. *eLife* 7, e31220 (2018).
23. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans. Med. Imaging* 29, 196–205 (2010).
24. Tyson, A. L. *et al.* A deep learning algorithm for 3D cell detection in whole mouse brain image datasets. *PLOS Comput. Biol.* 17, e1009074 (2021).
25. Fürth, D. et al. An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* 21, 139–149 (2018).
26. Henckens, M. J. A. G. *et al.* Stress-induced alterations in large-scale functional networks of the rodent brain. *NeuroImage* 105, 312–322 (2015).
27. Silva, B. A., Gross, C. T. & Gräff, J. The neural circuits of innate fear: detection, integration, action, and memorization. *Learn. Mem.* 23, 544–555 (2016).
28. Sharp, B. M. Basolateral amygdala and stress-induced hyperexcitability affect motivated behaviors and addiction. *Transl. Psychiatry* 7, e1194–e1194 (2017).
29. Vandenberghe, M. E. et al. Voxel-Based Statistical Analysis of 3D Immunostained Tissue Imaging. *Front. Neurosci.* 12, (2018).
30. Condorelli, D. F. et al. Induction of protooncogene fos by extracellular signals in primary glial cell cultures. *J. Neurosci. Res.* 23, 234–239 (1989).
31. Brignani, S. *et al.* Remotely Produced and Axon-Derived Netrin-1 Instructs GABAergic Neuron Migration and Dopaminergic Substantia Nigra Development. *Neuron* 107, 684-702.e9 (2020).
32. Belle, M. et al. Tridimensional Visualization and Analysis of Early Human Development. *Cell* 169, 161-173.e12 (2017).
33. Tyson, A. L. & Margrie, T. W. Mesoscale microscopy for micromammals: image analysis tools for understanding the rodent brain. *ArXiv210211812* Q-Bio (2021).
34. Murray, E. et al. Simple, scalable proteomic imaging for high-dimensional profiling of intact systems. *Cell* 163, 1500–1514 (2015).
35. Johnson, Z. V., Revis, A. A., Burdick, M. A. & Rhodes, J. S. A similar pattern of neuronal Fos activation in 10 brain regions following exposure to reward- or aversion-associated contextual cues in mice. *Physiol. Behav.* 99, 412–418 (2010).
36. Fields, R. D., Eshete, F., Stevens, B. & Itoh, K. Action Potential-Dependent Regulation of Gene Expression: Temporal Specificity in Ca2+, cAMP-Responsive Element Binding Proteins, and Mitogen-Activated Protein Kinase Signaling. *J. Neurosci.* 17, 7252–7266 (1997).
37. Moench, K. M., Breach, M. R. & Wellman, C. L. Chronic stress produces enduring sex- and region-specific alterations in novel stress-induced c-Fos expression. *Neurobiol. Stress* 10, 100147 (2019).

38. Gallo, F. T., Katche, C., Morici, J. F., Medina, J. H. & Weisstaub, N. V. Immediate Early Genes, Memory and Psychiatric Disorders: Focus on c-Fos, Egr1 and *Arc. Front. Behav. Neurosci.* 12, 79 (2018).
39. Rao-Ruiz, P., Yu, J., Kushner, S. A. & Josselyn, S. A. Neuronal competition: microcircuit mechanisms define the sparsity of the engram. *Curr. Opin. Neurobiol.* 54, 163–170 (2019).
40. Ramirez, S. et al. Activating positive memory engrams suppresses depression-like behaviour. *Nature* 522, 335–339 (2015).
41. Kitamura, T. et al. Engrams and circuits crucial for systems consolidation of a memory. *Science* 356, 73–78 (2017).
42. Gouty-Colomer, L. A. et al. Arc expression identifies the lateral amygdala fear memory trace. *Mol. Psychiatry* 21, 364–375 (2016).
43. Sheng, M. & Greenberg, M. E. The regulation and function of c-fos and other immediate early genes in the nervous system. *Neuron* 4, 477–485 (1990).
44. Sheng, H. Z., Lin, P. X. & Nelson, P. G. Combinatorial expression of immediate early genes in single neurons. *Mol. Brain Res.* 30, 196–202 (1995).
45. Ziegler, L. M. von et al. Molecular roadmap of the healthy stress response in the mouse hippocampus. *bioRxiv* 2021.03.26.436970 (2021) doi:10.1101/2021.03.26.436970.
46. Worley, P. F. et al. Constitutive expression of zif268 in neocortex is regulated by synaptic activity. *Proc. Natl. Acad. Sci.* U. S. A. 88, 5106–5110 (1991).
47. Ocklenburg, S., Korte, S. M., Peterburs, J., Wolf, O. T. & Güntürkün, O. Stress and laterality – The comparative perspective. *Physiol. Behav.* 164, 321–329 (2016).
48. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175 (1971).
49. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953. e20 (2020).
50. Claudi, F. *et al. Brainrender: a python-based software for visualizing anatomically registered data.* http://biorxiv.org/lookup/doi/10.1101/2020.02.23.961748 (2020) doi:10.1101/2020.02.23.961748.

Chapter 3

The STRESS-NL database: a resource for human acute stress studies across the Netherlands.

Psychoneuroendocrinology (2022),
doi: 10.1016/j.psyneuen.2022.105735

Bonapersona V[1]
Born FJ[1,2]
Bakvis P[3]
Branje S[4]
Elzinga B[3]
Evers AE[3]
van Eysden M[1]
Fernandez G[5]
Habets PC[6]
Hartman CJ[7]
Hermans EJ[5]
Meeus W[4]
van Middendorp H[7]
Nelemans S[4]
Oei NY[8]
Oldehinkel AJ[7]
Roelofs K[5]
de Rooij SR[6]
Smeets T[9]
Tollenaar MS[3]
Joëls M[10]
Vinkers CH[6]

1 Utrecht University, Utrecht, the Netherlands.
2 Charité University, Berlin, Germany.
3 Leiden University, the Netherlands
4 Utrecht University, Utrecht, the Netherlands
5 Radboud University Medical Center, Nijmegen, the Netherlands.
6 Amsterdam UMC, Amsterdam, the Netherlands.
7 University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
8 University of Amsterdam, Amsterdam, the Netherlands
9 Tilburg University, Tilburg, the Netherlands.
10 University Medical Center Groningen, Groningen, the Netherlands.

Abstract

Stress initiates a cascade of (neuro)biological, physiological, and behavioral changes, allowing us to respond to a challenging environment. The human response to acute stress can be studied in detail in controlled settings, usually in a laboratory environment. To this end, many studies employ acute stress paradigms to probe stress-related outcomes in healthy and patient populations. Though valuable, these studies in themselves often have relatively limited sample sizes. We established a data-sharing and collaborative interdisciplinary initiative, the STRESS-NL database, which combines (neuro)biological, physiological, and behavioral data across many acute stress studies in order to accelerate our understanding of the human acute stress response in health and disease (www.stressdatabase.eu). Researchers in the stress field from 12 Dutch research groups of 6 Dutch universities created a database to achieve an accurate inventory of (neuro)biological, physiological, and behavioral data from laboratory-based human studies that used acute stress tests. Currently, the STRESS-NL database consists of information on 5529 individual participants (2281 females and 3348 males, age range 6-99 years, mean age 27.7 $\pm$ 16 years) stemming from 57 experiments described in 42 independent studies. Studies often did not use the same stress paradigm; outcomes were different and measured at different time points. All studies currently included in the database assessed cortisol levels before, during and after experimental stress, but cortisol measurement will not be a strict requirement for future study inclusion. Here, we report on the creation of the STRESS-NL database and infrastructure to illustrate the potential of accumulating and combining existing data to allow meta-analytical, proof-of-principle analyses. The STRESS-NL database creates a framework that enables human stress research to take new avenues in explorative and hypothesis-driven data analyses with high statistical power. Future steps could be to incorporate new studies beyond the borders of the Netherlands, or build similar databases for experimental stress studies in rodents. In our view, there are major scientific benefits in initiating and maintaining such international efforts.

Introduction

Stress initiates a cascade of neurochemical and physiological changes that enable an individual to rapidly deal with a stressor and recover thereafter. It is clear that our stress response is extremely complex [1,2] and our understanding of stress has its roots in a rich research history stemming from Cannon, Selye, Benard, to McEwen.[3] To adequately respond to acute or chronic stress, an integrated response at the level of emotions, behavior, physiology and (neuro)biology is vital, including temporally distinct changes in brain networks, and stress systems (i.e. the HPA-axis, sympathetic nervous systems, and immune system)[4-6]. The integrated and well-orchestrated stress response is individual-specific, depending on biological and psychological factors, previous experiences, but also the ecological context of an individual's life[1,7]. Stress initiates a cascade of neurochemical and physiological changes which enable an individual to rapidly deal with a stressor and recover thereafter[2,8].

The integrated and well-orchestrated stress response is individual-specific, depending on biological (e.g. genetic) and psychological factors, as well as previous experiences. Moreover, it depends on the context of acute stress (e.g. stress type, intensity, controllability) and the ecological context of an individual's life at large[1,7]. Thorough study of the human stress response is of high relevance not only to understand the normal stress response, but also how stress can result in the development of psychiatric and somatic disorders, including depression[9,10].

Our current understanding of the human stress response stems from a large body of scientific literature based to a great extent on experimental (laboratory-based) acute stress studies in humans, which induce acute stress in a controlled setting using different (versions of) stress-inducing paradigms. This includes the well-known and often-used Trier Social Stress Test (TSST) in individual or group form[11,12], the Cold Pressor Test (CPT[13]) including the socially-evaluated CPT[14], the Maastricht Acute Stress Test[15], but, more recently, also online stress tests[16], and virtual reality (VR)-based stress tests[17,18]. In these acute stress studies, a physical or socially evaluated challenge is monitored through by outcome measurement, with often salivary cortisol as a biomarker to investigate the HPA-axis[66]. Studies differ in timing (when are outcomes assessed following acute stress) and correlates (which predictors and outcomes are measured). With regard to timing, cortisol levels are often measured at different time points and time periods following acute stress[9]. This is relevant as the stress response has a clear dynamic pattern, with well-known time-dependent effects following stress across (neuro)biological, physiological, endocrine, and behavioral outcomes[5,19]. For example, Schlotz and colleagues showed a strong coupling of the psycho-endocrine response, once an endocrine lag due to the dynamic of the system is considered[20]. With regard to predictors and outcomes, (neuro)biological but also psychological and psychiatric assessments differ from study to study, as do assessments of psychiatric history, current and previous stress and trauma exposure. Importantly, even though

exceptions in large cohorts exist, sample sizes of acute stress studies are often limited due to their labor-intensive nature, and this is even more pressing with well-known effects of age, sex, menstrual cycle, and time of day on for example stress-induced cortisol outcomes[21-23]. However, if one wishes to combine data from acute stress studies, it can be very challenging to identify, compare, and combine relevant studies. It is therefore of utmost importance to develop metadata that allow to identify and synthesize data from multiple studies.

To make progress in our understanding of the complexity of the human stress response, collaboration and integration across the field is called for. Therefore, the STRESS-NL database consortium was founded to actively collaborate and capitalize on domain-specific expertise. In this manuscript, we describe the conception and building of the STRESS-NL database, and we present preliminary analyses to demonstrate the content and usability of this collection of acute stress studies. Although currently the database consists only of studies performed in the Netherlands, it can be equally relevant and open for stress researchers from other countries.

Materials and Methods

Study identification and selection

The main objective of the STRESS-NL database was to develop a stress database for aggregation, curation and archival of information of most of human acute stress studies in The Netherlands. For an overview of the process how the STRESS-NL database was created, see the research flow chart (Fig. 1). Principal investigators (PIs, the main initiators of the acute stress studies) were identified within the network of the STRESS-NL consortium (www.stress-nl.nl) and invited to participate in this initiative for data sharing of experimental stress studies. PIs were encouraged to share the invitation with other researchers who were potentially interested in the initiative and asked to share data of (un)published research that met prespecified inclusion criteria: i) any type of study design in human subjects (e.g., experimental, longitudinal, cohort, repeated measures); ii) investigating the effects of acute stress in humans with a behavioral intervention (e.g. psychosocial such as a variant of the Trier Social Stress Test; or physical such as the Cold Pressor Test); and iii) at least have measured cortisol concentrations after acute stress per participant. Although this third criterion was required for the current wave of data collection, it won't be required for future waves. Pharmacological intervention studies, for example related to the HPA-axis (e.g. cortisol administrations) were excluded. No exclusion criteria were specified regarding the presence or type of control condition, nor were any limits set on age, gender, diagnosis or any other population characteristics, but all these factors are systematically indicated.

Data collection and harmonization

Interested PIs were initially contacted for an informal discussion about the eligibility of their data. Eligible PIs were then requested to provide the data in two files: 1) individual anonymized participant data of a selection of variables (gender, age, contraceptive use (where applicable), clinical diagnosis (where applicable), cortisol concentrations) in the PI's preferred format (e.g. excel, SPSS); and 2) meta-information for each study on available data (e.g. questionnaires, cognitive tests, structured interviews, biological outcomes, neuroimaging, and EEG). These two files per experiment were then manually processed and added to the database.

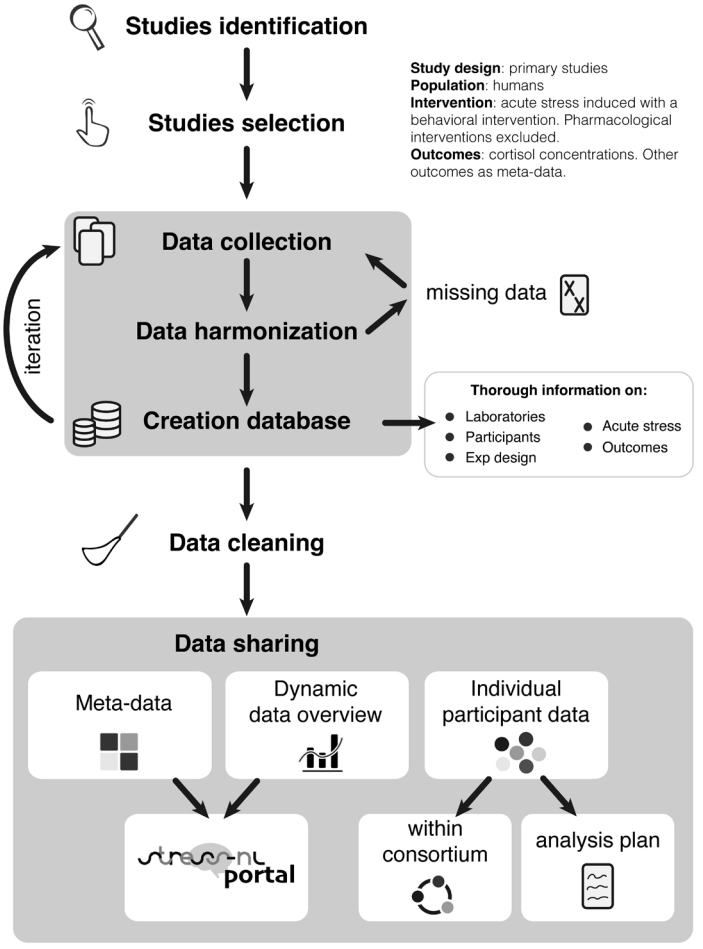


Fig. 1. Overview of the research flow chart of the STRESS-NL database. For details on the contents of the database, see (link). The database can be accessed at three levels:1) meta-data, 2) dynamic data overview (freely available via our web-portal), and 3) individual participant data, which can be accessed only by members of the consortium or via an analysis plan accepted by the consortium. Exp design = experimental design.

In order to integrate data from different studies, data harmonization procedures were set in place. First, we established a naming convention common to all different experiments. For example, we re-calculated the cortisol time points of each experiment so that the baseline value was t=0 for each experiment. Then, we identified a minimal unique set of variables that could be used to adequately categorize the experiments' meta information. For example, the type of intervention, the concentration/frequency/timing of the cortisol concentration etc. (for a complete list, see Supplementary Table S1). Lastly, we classified particularly heterogeneous variables into subgroups. For example, modifications of the TSST were categorized as "TSST variations". With these steps, we created a stress taxonomy to comprehensively categorize stress research data.

Following data harmonization, a database prototype was created. We focused on four main objectives. 1) The database had a high informational content: information about laboratories, experiments, participants, methods, and various outcomes was accurately and comprehensively represented. 2) The database had to provide intuitive and user-friendly solutions for (meta)data exploration. Variable names therefore were explicit, non-ambiguous, and aligned with customs in the stress field. 3) The database had to be scalable and flexible, with the possibility to accommodate future growth. 4) The database had to comply with the highest ethical standards, and with international, EU, and national law (including European Privacy Protection laws); and provide applications to restrict data access.

A database template was created where experimental studies could be added using an iterative process (Fig 1). Where necessary, additional information was collected from PIs or from the publications associated with the studies. If information at the individual participant level was missing for continuous variables (e.g. age), we used the group range or, if range was not available, the mean. Data were verified for completeness and consistency. In this first final form, the STRESS-NL database contained two tables, one for the experiments' meta information, and one for the anonymized individual participant data for the limited dataset centered around gender, age, and cortisol values over time.

Missing data

Despite our intent to be as comprehensive as possible, missing data were encountered for two main reasons. First, we did not perform a systematic search for acute stress studies and PIs contributed data voluntarily. The current version of the STRESS-NL database is therefore not comprehensive of all acute stress studies in the Netherlands. Second, in some studies, missing data was present. In the database, we distinguished between information that was 'not available', for example due to a discrepancy between metadata and individual data, or truly missing, for example due to a technical problem with an assay (e.g. missing cortisol values).

Results

The STRESS-NL database collects information of Dutch acute stress studies by categorizing them in the following categories: 1) information about the laboratories (section 3.1), 2) characterization of the participants (section 3.1), 3) description of the acute stress intervention (section 3.2) and of 4) the experimental design, and lastly 5) a thorough description of anonymized outcomes (section 3.3). Furthermore, the database currently contains individual participant data of cortisol concentration after acute stress, although this will not be a strict requirement for data inclusion in the future. We believe that these elements are exhaustive to describe each study; yet, more elements can be easily added in the future if deemed appropriate.

Database content: meta-study information and participants

In 2021, the STRESS-NL database consists of 12 Dutch laboratories across six different universities, with data from 57 acute stress experiments stemming from 41 independent datasets (reported in 38 published and 4 unpublished manuscripts, with some experiments included in more than one paper)[15,19,24-61]. Supplemental table S2 summarizes the general characteristics of each study included in the database.

The STRESS-NL database contains individual participant information on 5529 participants (Fig. 2a), of which 2281 are females and 3348 males. The age ranged between 6 and 99 years (females: mean [sd] = 29.4 [±17.7]; males: mean [sd] = 26.5 [±14.8], Fig. 2b). Age had a bimodal distribution, with a clear peak in the early 20's. This overrepresentation of young adults is due to the recruitment strategy of the included studies. 64% of participants were described as healthy individuals, and 16% had confirmed past or current psychiatric or neurological conditions (Fig. 2c). Information about the use of oral contraceptives is available for 61% of women and information about the menstrual cycle for 17% (Fig. 2d).

Type of stress tests

The database includes studies that induced acute stress in humans in a laboratory-setting. Several behavioral paradigms can be used to induce acute stress, which can be roughly categorized by typology (Fig. 2e). Acute stress was induced by social evaluation (i.e., TSST, SECPT, and PST and respective variations, nexp = 40; npart = 4204), emotional (i.e., aversive movies, n_{exp} = 2; n_{part} = 400) and physical (i.e., cold pressure test, n_{exp} = 3; n_{part} = 465) stressors, or a combination of the two (i.e., M-PASAT, P-SECPT, MAST and variations, n_{exp} = 10; n_{part} = 460). For a list of the available paradigms and their categorizations, see Supplemental table S2. Overall, in 83% of participants in the STRESS-NL database acute stress tests were used, and the remaining 17% of participants were exposed to a non-stressful control condition.

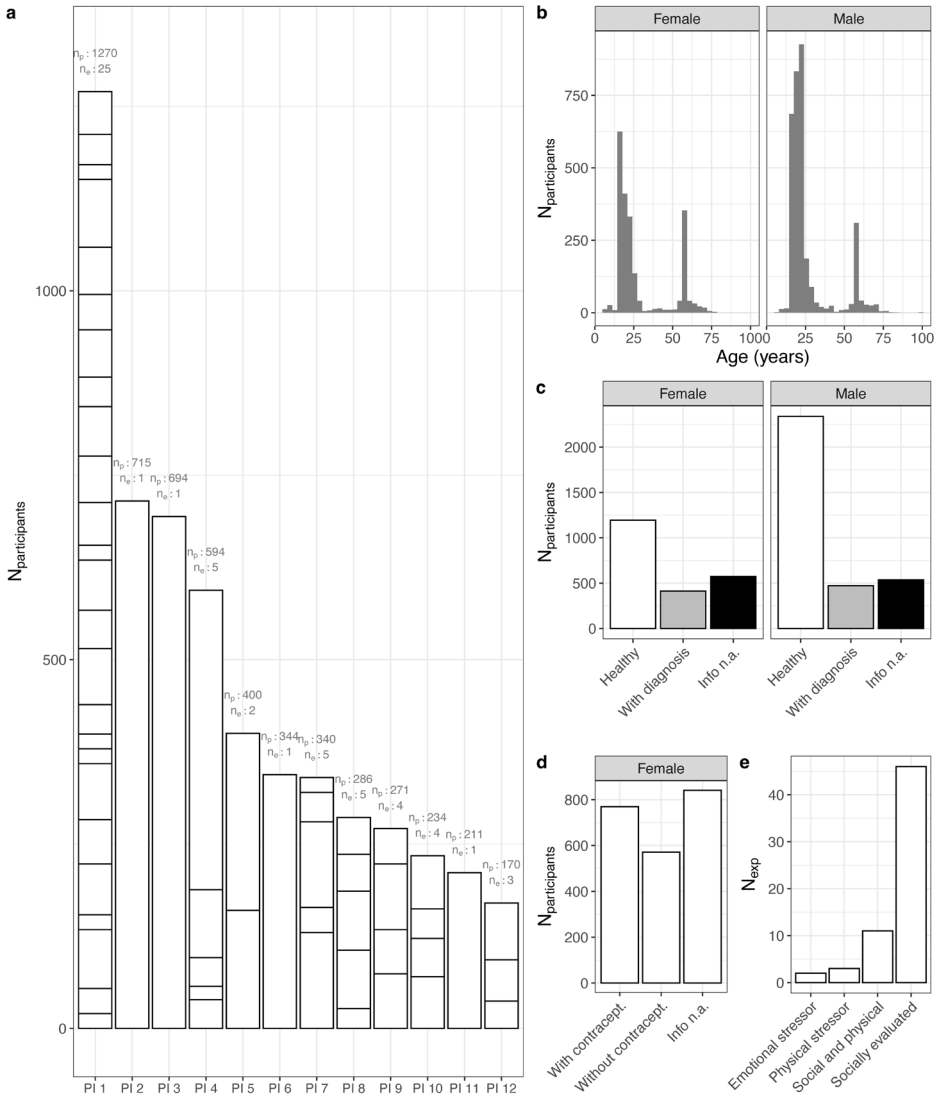


Fig. 2 Demographics, population and intervention. a) Number of participants across laboratories. Each rectangle represents a separate experiment, of height equal to the number of participants, and stacked by principle investigator (PI). n_p: number of participants; n_e = number of experiments. b) Distribution of age across males and females; b) Number of participants based on the presence/absence of a diagnosis. c) Number of female participants (not) using oral contraceptives. d) Number of experiments using different acute stress tests. For a complete list of available acute stress tests, see **Supplemental table S1**. n.a. = not available

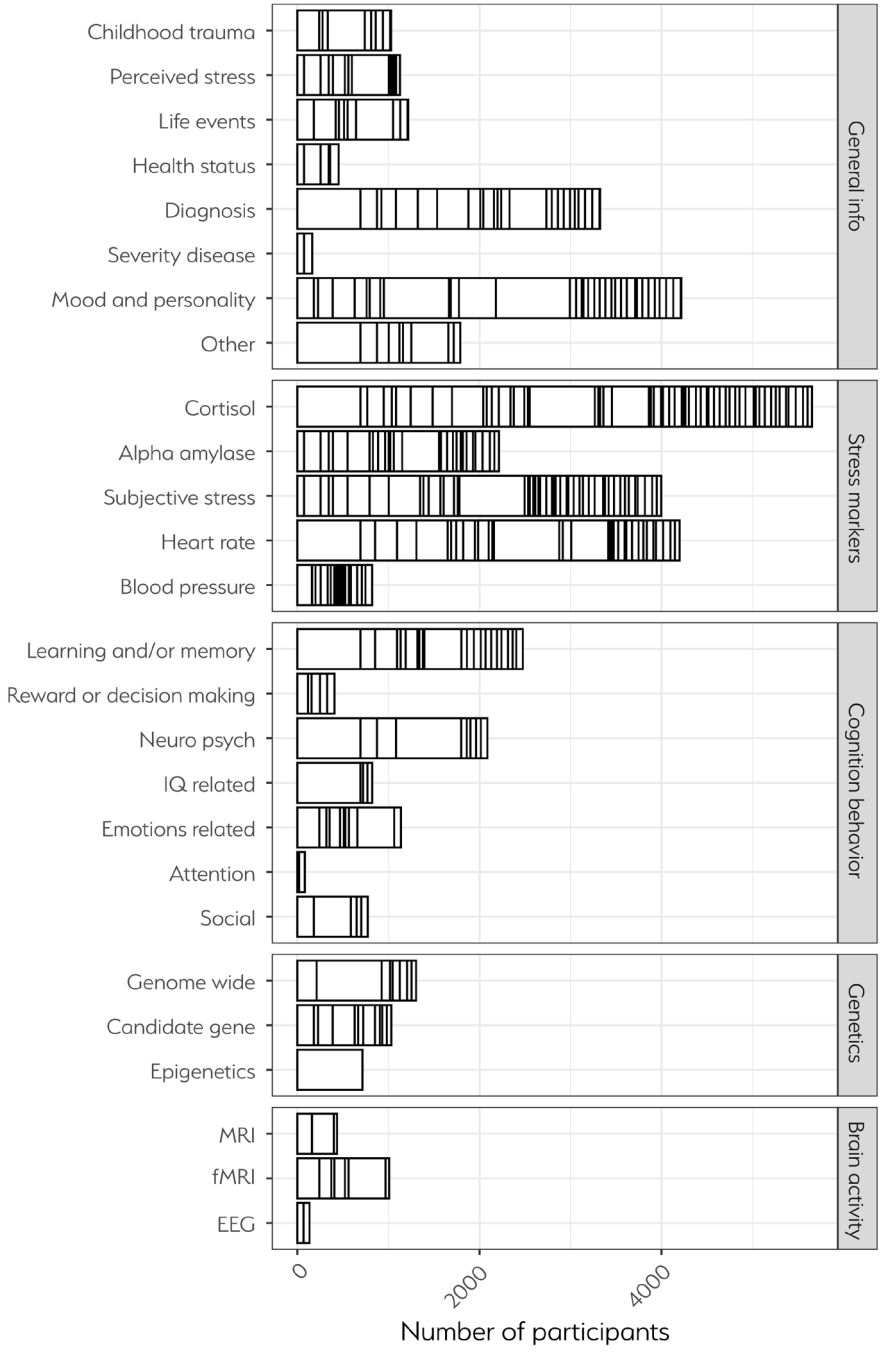


Fig. 3 Overview of participants across outcomes. Number of participants (stress and control groups) for each (grouped) outcome available. Each rectangle forming the frequency bar plots represent a unique study. Of note, all studies have provided individual participant cortisol values.

Available Stress-related Outcomes

The main (required) outcome of the STRESS-NL database is cortisol concentration, and all studies provided pseudonymized individual participant information for cortisol for all measured time points in the study. All cortisol values belonged to saliva samples. Across studies, between 2 and 11 cortisol timepoints were collected, with a mean of 5,7. Most of the studies were conducted in the afternoon (58%), with a small percentage in the morning (22,6%) or with a combination of the two (19,4%). Additionally, we collected meta information of several available secondary stress-related outcomes (Fig. 3). These can roughly be categorized in 1) *stress markers*, such as alpha amylase, blood pressure, heart rate, and subjective stress ratings; 2) *questionnaires*, related to general information, such as for childhood trauma, life events, or health status; 3) *genetic outcomes*, such as genome-wide, epigenetic or candidate gene analyses; 4) *cognition/behavioral* tests, such as related to learning and memory, IQ, reward/decision making, attention, emotion, sociality, social anxiety and neuropsychiatric; 5) *brain activity measures*, such as (f)MRI and EEG. The STRESS-NL database contains meta-information on all acquired outcomes, i.e. which tests were performed, what type and quantity of data is available (including questionnaires, subjective stress, physiology, (epi)genetics, and fMRI data). Through our online portal (www.stressdatabase.nl), all outcome information can be found to identify a population of interest. For example, a researcher may be interested in cortisol values after a TSST, but only if information on childhood maltreatment is also available, or search studies that have included fMRI outcomes following stress.

Cortisol outcomes as an example from the current STRESS-NL database

The STRESS-NL database centrally stores meta-data of all participating studies, but also limited anonymized individual participant data related to descriptives, such as sex, age, and contraceptive use, and one specific stress outcome, that is, cortisol timepoints and concentrations. In this section, we showcase analyses that can be performed using the STRESS-NL database on human cortisol levels following acute stress. In total, 18 experiments (42%) measured baseline cortisol concentrations (Fig 4a). No experiment measured cortisol concentration later than 2 hours after stress induction (except one study assessing cortisol after 24 hours[34]), with 85% of cortisol measurements taken within the first hour after stress induction. Since cortisol is dynamically and transiently expressed after acute stress, differences in measured time-points across experiments may highlight biologically relevant heterogeneity.

To illustrate the possibilities of the database, we here calculated the difference between males and females. Across the available data, we selected experiments investigating male as well as female participants, for a total of 23 studies. With the summary statistics of the area-under-the-curve with respect to increase (AUCi relative to baseline) per participant, we calculated Cohen's d, a measure of effect

size difference between males and females (Fig 4b), showing that males generally responded to stress with higher levels of cortisol than females, although there is a high degree of heterogeneity across studies.

At the deepest level of data information, individual participant data enables the full re-analysis of previous experiments. This can be used to confirm existing hypotheses, or test new ones. In Fig 4c, we selected experiments with at least 10 participants of the following groups: males, females using oral contraceptives, and females not using oral contraceptives. In our database, 4 studies met these criteria. For each participant, we calculated the increase in peak cortisol concentration relative to baseline, with peaks identified for each study independently. Across the identified studies, females using oral contraceptives had a smaller increase in cortisol peak compared to females without oral contraceptives (Fig. 4c). Use of oral contraceptives may therefore partially explain the high variability observed in Fig. 4b – an analysis that would not be possible without the individual participant data.

Data access and contribution

The STRESS-NL database is governed by a consortium agreement, allowing anonymized individual participant data to be accessed by consortium members. External parties with ownership of human acute stress data can apply to become formal member of the STRESS-NL database, also outside the Netherlands, if they sign and adhere to the consortium agreement. The STRESS-NL database is open for new human acute stress studies, and the consortium agreement is suited and compatible with EU countries. External parties who cannot or do not wish to become member, can gain access to anonymized individual participant data via an analysis plan submitted to the STRESS-NL Steering Committee (there will be a limited fee to maintain and update the STRESS-NL database). There, data plans and data release are governed via a consortium agreement with an opt-in principle.

STRESS-NL data can be accessed in multiple ways. Meta-data of individual studies and, in time, summary statistics will be made available via a web portal (www.stressdatabase.eu). Summary statistics can be used, for example, for meta-analyses, Bayesian evidence synthesis, power calculations or the definition of informative priors.). Moreover, an analysis plan can be created and submitted to the Steering Committee of the STRESS-NL database. At the website, an interactive user interface is available where researchers can explore experimental design characteristics and their frequencies, and where estimates of the sample size available in the STRESS-NL database are provided. The information of interest is selected and directly transferred to a predefined analysis plan that can then be edited. After central approval, PIs of studies that can and want to contribute to the analysis plan can be approached for the necessary data. This allows direct interaction with the data, without direct contact or storage with identifiable or privacy-sensitive information.

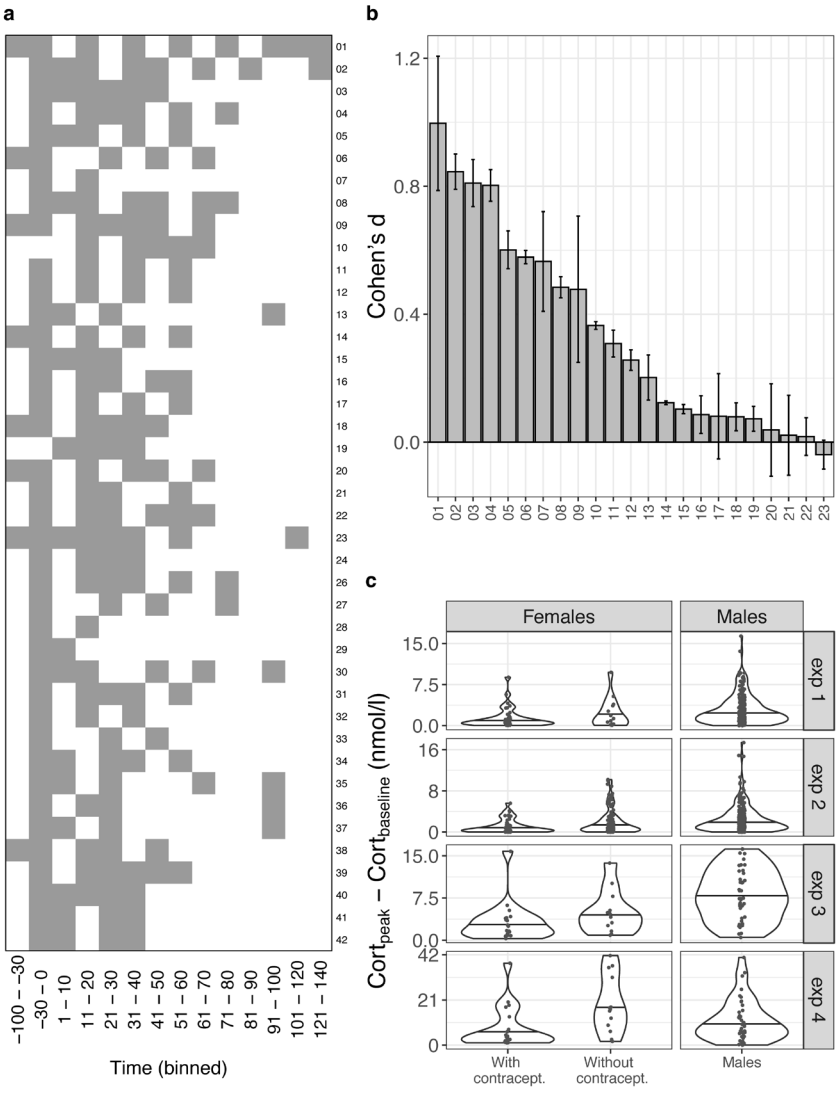


Fig. 4 Examples of analyses for cortisol measurements. a) Heatmap of cortisol timepoints across experiments. Grey = measurement present; white = measurement absent. Numbers on the y axis correspond to the experiments in Supplementary Table S2. b) Effect size difference between males and females in cortisol concentration after acute stress measured with the area under the curve (AUC_I). Positive effect sizes indicate higher values for males, negative effect sizes for females. The results are shown per study. Numbers on the x axis correspond to the experiments in Supplementary Table S2. c) Difference between peak cortisol and baseline cortisol in studies with males, and females with/without contraceptives. Exp 1 to exp 4 represent the four independent studies in the STRESS-NL database reporting all experimental groups (males, females using contraceptives, females not using contraceptives) in one study, with at least 10 participants per group. Thick horizontal line line corresponds to the median. Each dot corresponds to a participant. Cort = cortisol

Discussion

To promote the reuse and combining of existing data, we established a collaborative interdisciplinary database that combines (neuro)biological, psychological, and behavioral data across many acute stress studies in the Netherlands. Although currently all studies included in the database measured cortisol levels after stress, this will no longer be considered a strict requirement in the future. With 12 Dutch research groups from 6 Dutch universities, we created the STRESS-NL database with information on 5529 individuals (2281 females and 3348 males, age range 6-99 years) stemming from 57 experiments described in 42 independent studies.

This inventory of (neuro)biological, physiological, and behavioral data from laboratory-based human studies employing acute stress tests has the potential to accelerate our understanding of the human acute stress response. The STRESS-NL database contains data that allow meta-analytical as well as proof-of-principle analyses, enabling human stress research to take new avenues in both explorative and hypothesis-driven data analyses with high statistical power (see for example[62]). Such collaboration and combining of studies can lead to novel opportunities for scientific endeavors, for example to disentangle how humans respond to stress in health and disease.

The STRESS-NL database not only facilitates access to existing acute stress data in humans, but also allows a converging consensus on future acute stress studies, for example by harmonizing and summarizing terminology, methodology and data structure across human stress studies. A combined database not only quickly gives insight in the data available nation-wide and whom to contact for data access, but also allows analyses on large sets of data, to validate and replicate previous findings. Data sharing generally increases the sample size and results in a concomitant increase in statistical power, and can lead to more awareness of methodological differences. For instance, one could test hypotheses across populations with a collective large number of participants (e.g. difference between males and females). As stress studies typically have small samples, combining data from different studies also allows for a more optimal analysis of moderating factors that can explain heterogeneity in results. As the database includes various parameters and outputs—from stress markers to genetics and brain imaging to cognitive and behavioral measures and other relevant data — this also enables the integration of stress outcomes at different levels, from physiology, behavior, neuroimaging, to cortisol levels. This may accelerate a 'multi-layer' understanding of stress across relevant outcomes, rather than only focusing on one or two outcome domains.

From our preliminary analyses, it is obvious that quite a large heterogeneity with regard to methodology, population, and outcomes exists. Studies often did not use the same stress paradigm, and outcomes were vastly different and measured at different time points following stress. Moreover, there is currently a distinct bimodal

age distribution due to the nature and goals of the studies included so far, which may prevent firm conclusions about age-related changes in stress reactivity. Although this can be regarded a limitation, combining data from multiple studies using different paradigms might offer a better understanding of task-related differences in findings. The next challenge will be analytical: integration of this heterogeneous data requires a thorough and robust analysis plan. Previous research highlighted that simple data aggregation may not always be appropriate for neuroendocrine data[63,64]. Future analyses should be therefore based on state-of-the-art individual participant data methodology (for an overview: https://www.ipdma.co.uk/) or Bayesian evidence synthesis[65]. These methodologies are not limited to performing a statistical test, but they include 1) a thorough assessment of the methodological quality of included studies, 2) an assessment of the risk of bias, 3) a check of the validity of assumptions, 4) they address methodological differences in missing values, time points, assays by accounting for study-specific effects, 5) they use sound statistical models to obtain pooled effects, which can also be used to assess heterogeneity. The flexibility of individual participant data therefore comes with the necessity of increased statistical expertise. Although using percentage change could be at times possible, this can overlook study-specific effects. For this reason, we recommend that the planning and performance of each analysis is supervised by a statistician.

So far, details on storage and analysis of outcomes (e.g. method of cortisol assessment such as LIA/RIA, inter- and intra-assay variability, and single or duplicate outcomes) have not been taken into account. To further increase the size and scope of the STRESS-NL database, future steps could be to incorporate new studies beyond the borders of the Netherlands; or build similar databases for experimental stress studies in rodents. In our view, there are major scientific benefits in initiating and maintaining such international efforts.

References

1 Joels, M. & Baram, T. Z. The neuro-symphony of stress. *Nat Rev Neurosci* 10, 459-466 (2009).
2 de Kloet, E. R., Joels, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nat Rev Neurosci* 6, 463-475 (2005).
3 Godoy, L. D., Rossignoli, M. T., Delfino-Pereira, P., Garcia-Cairasco, N. & de Lima Umeoka, E. H. A Comprehensive Overview on Stress Neurobiology: Basic Concepts and Clinical Implications. *Frontiers in behavioral neuroscience* 12, 127, doi:10.3389/fnbeh.2018.00127 (2018).
4 Joels, M., Sarabdjitsingh, R. A. & Karst, H. Unraveling the time domains of corticosteroid hormone influences on brain activity: rapid, slow, and chronic modes. *Pharmacol Rev* 64, 901-938, doi:10.1124/pr.112.005892 (2012).
5 Hermans, E. J., Henckens, M. J., Joels, M. & Fernandez, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends Neurosci* 37, 304-314, doi:10.1016/j.tins.2014.03.006 (2014).
6 van Leeuwen, J. M. C. et al. At-risk individuals display altered brain activity following stress. *Neuropsychopharmacology* 43, 1954-1960, doi:10.1038/s41386-018-0026-8 (2018).
7 Lupien, S. J., McEwen, B. S., Gunnar, M. R. & Heim, C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci* 10, 434-445, doi:10.1038/nrn2639 (2009).
8 Ulrich-Lai, Y. M. & Herman, J. P. Neural regulation of endocrine and autonomic stress responses. *Nat Rev Neurosci* 10, 397-409 (2009).
9 Zorn, J. V. et al. Cortisol stress reactivity across psychiatric disorders: A systematic review and meta-analysis. *Psychoneuroendocrinology* 77, 25-36, doi:10.1016/j.psyneuen.2016.11.036 (2017).
10 Vinkers, C. H. et al. Stress exposure across the life span cumulatively increases depression risk and is moderated by neuroticism. *Depress Anxiety* 31, 737-745, doi:10.1002/da.22262 (2014).
11 Kirschbaum, C., Pirke, K. M. & Hellhammer, D. H. The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76-81 (1993).
12 von Dawans, B., Kirschbaum, C. & Heinrichs, M. The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology* 36, 514-522, doi:10.1016/j.psyneuen.2010.08.004 (2011).
13 Silverthorn, D. U. & Michael, J. Cold stress and the cold pressor test. *Adv Physiol Educ* 37, 93-96, doi:10.1152/advan.00002.2013 (2013).
14 Schwabe, L. & Schachinger, H. Ten years of research with the Socially Evaluated Cold Pressor Test: Data from the past and guidelines for the future. *Psychoneuroendocrinology* 92, 155-161, doi:10.1016/j.psyneuen.2018.03.010 (2018).
15 Smeets, T. et al. Introducing the Maastricht Acute Stress Test (MAST): a quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology* 37, 1998-2008, doi:10.1016/j.psyneuen.2012.04.012 (2012).
16 Gunnar, M. R. et al. Validation of an online version of the Trier Social Stress Test in a study of adolescents. *Psychoneuroendocrinology* 125, 105111, doi:10.1016/j.psyneuen.2020.105111 (2021).
17 Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E. & Domes, G. Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology* 101, 186-192, doi:10.1016/j.psyneuen.2018.11.010 (2019).
18 Shiban, Y. et al. Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *Int J Psychophysiol* 110, 47-55, doi:10.1016/j.ijpsycho.2016.10.008 (2016).
19 Vinkers, C. H. et al. Time-dependent changes in altruistic punishment following stress. *Psychoneuroendocrinology* 38, 1467-1475, doi:10.1016/j.psyneuen.2012.12.012 (2013).
20 Schlotz, W. et al. Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: a question of timing. *Psychosom Med* 70, 787-796, doi:10.1097/PSY.0b013e3181810658 (2008).
21 Kudielka, B. M., Schommer, N. C., Hellhammer, D. H. & Kirschbaum, C. Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology* 29, 983-992 (2004).
22 Kudielka, B. M., Buske-Kirschbaum, a., Hellhammer, D. H. & Kirschbaum, C. HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: impact of age and gender. *Psychoneuroendocrinology* 29, 83-98, doi:10.1016/S0306-4530(02)00146-4 (2004).
23 Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C. & Hellhammer, D. H. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic medicine* 61, 154-162 (1999).
24 de Brouwer, S. J. et al. The psychophysiological stress response in psoriasis and rheumatoid arthritis. *The British journal of dermatology* 170, 824-831, doi:10.1111/bjd.12697 (2014).
25 de Brouwer, S. J. et al. Psychophysiological responses to stress after stress management training in patients with rheumatoid arthritis. *PLoS One* 6, e27432, doi:10.1371/journal.pone.0027432 (2011).
26 Houtepen, L. C., Boks, M. P., Kahn, R. S., Joels, M. & Vinkers, C. H. Antipsychotic use is associated with a blunted cortisol stress response: a study in euthymic bipolar disorder patients and their unaffected siblings. *Eur Neuropsychopharmacol* 25, 77-84, doi:10.1016/j.euroneuro.2014.10.005 (2015).
27 Hermans, E. J. et al. Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* 334, 1151-1153, doi:10.1126/science.1209603 (2011).

28 Klumpers, F. et al. Dorsomedial Prefrontal Cortex Mediates the Impact of Serotonin Transporter Linked Polymorphic Region Genotype on Anticipatory Threat Reactions. *Biol Psychiatry* 78, 582-589, doi:10.1016/j.biopsych.2014.07.034 (2015).
29 Roelofs, K. et al. Hypothalamus-pituitary-adrenal axis hyperresponsiveness is associated with increased social avoidance behavior in social phobia. *Biol Psychiatry* 65, 336-343, doi:10.1016/j.biopsych.2008.08.022 (2009).
30 Bakvis, P. et al. Basal hypercortisolism and trauma in patients with psychogenic nonepileptic seizures. Epilepsia 51, 752-759, doi:10.1111/j.1528-1167.2009.02394.x (2010).
31 Bakvis, P., Spinhoven, P. & Roelofs, K. Basal cortisol is positively correlated to threat vigilance in patients with psychogenic nonepileptic seizures. Epilepsy Behav 16, 558-560, doi:10.1016/j.yebeh.2009.09.006 (2009).
32 Zhang, W. et al. Acute stress alters the 'default' brain processing. *Neuroimage* 189, 870-877, doi:10.1016/j.neuroimage.2019.01.063 (2019).
33 Kaldewaij, R. et al. Frontal Control Over Automatic Emotional Action Tendencies Predicts Acute Stress Responsivity. Biological psychiatry. Cognitive neuroscience and neuroimaging 4, 975-983, doi:10.1016/j.bpsc.2019.06.011 (2019).
34 Cornelisse, S., Joels, M. & Smeets, T. A randomized trial on mineralocorticoid receptor blockade in men: effects on stress responses, selective attention, and memory. *Neuropsychopharmacology* 36, 2720-2728, doi:10.1038/npp.2011.162 (2011).
35 van Campen, J. S. et al. Relation between stress-precipitated seizures and the stress response in childhood epilepsy. *Brain* 138, 2234-2248, doi:10.1093/brain/awv157 (2015).
36 Tollenaar, M. S., Elzinga, B. M., Spinhoven, P. & Everaerd, W. Autobiographical memory after acute stress in healthy young men. Memory 17, 301-310, doi:10.1080/09658210802665845 (2009).
37 Oei, N. Y., Everaerd, W. T., Elzinga, B. M., van Well, S. & Bermond, B. Psychosocial stress impairs working memory at high loads: an association with cortisol levels and memory retrieval. *Stress* 9, 133-141, doi:10.1080/10253890600965773 (2006).
38 Oei, N. Y. et al. Stress shifts brain activation towards ventral 'affective' areas during emotional distraction. Soc Cogn Affect Neurosci 7, 403-412, doi:10.1093/scan/nsr024 (2012).
39 Oei, N. Y. L., Both, S., van Heemst, D. & van der Grond, J. Acute stress-induced cortisol elevations mediate reward system activity during subconscious processing of sexual stimuli. *Psychoneuroendocrinology* 39, 111-120, doi:10.1016/j.psyneuen.2013.10.005 (2014).
40 de Rooij, S. R. et al. Cortisol responses to psychological stress in adults after prenatal exposure to the Dutch famine. *Psychoneuroendocrinology* 31, 1257-1265, doi:10.1016/j.psyneuen.2006.09.007 (2006).
41 Hartman, C. A., Hermanns, V. W., de Jong, P. J. & Ormel, J. Self- or parent report of (co-occurring) internalizing and externalizing problems, and basal or reactivity measures of HPA-axis functioning: a systematic evaluation of the internalizing-hyperresponsivity versus externalizing-hyporesponsivity HPA-axis hypothesis. *Biol Psychol* 94, 175-184, doi:10.1016/j.biopsycho.2013.05.009 (2013).
42 Bouma, E. M., Riese, H., Ormel, J., Verhulst, F. C. & Oldehinkel, A. J. Adolescents' cortisol responses to awakening and social stress; effects of gender, menstrual phase and oral contraceptives. The TRAILS study. *Psychoneuroendocrinology* 34, 884-893, doi:10.1016/j.psyneuen.2009.01.003 (2009).
43 Smeets, T., Jelicic, M. & Merckelbach, H. The effect of acute stress on memory depends on word valence. *Int J Psychophysiol* 62, 30-37, doi:10.1016/j.ijpsycho.2005.11.007 (2006).
44 Smeets, T. Autonomic and hypothalamic-pituitary-adrenal stress resilience: Impact of cardiac vagal tone. *Biol Psychol* 84, 290-295, doi:10.1016/j.biopsycho.2010.02.015 (2010).
45 Smeets, T. Acute stress impairs memory retrieval independent of time of day. *Psychoneuroendocrinology* 36, 495-501, doi:10.1016/j.psyneuen.2010.08.001 (2011).
46 Smeets, T., Otgaar, H., Raymaekers, L., Peters, M. J. & Merckelbach, H. Survival processing in times of stress. *Psychon Bull Rev* 19, 113-118, doi:10.3758/s13423-011-0180-z (2012).
47 Quaedflieg, C. W., Schwabe, L., Meyer, T. & Smeets, T. Time dependent effects of stress prior to encoding on event-related potentials and 24 h delayed retrieval. *Psychoneuroendocrinology* 38, 3057-3069, doi:10.1016/j.psyneuen.2013.09.002 (2013).
48 Quaedflieg, C. W., Meyer, T. & Smeets, T. The imaging Maastricht Acute Stress Test (iMAST): a neuroimaging compatible psychophysiological stressor. *Psychophysiology* 50, 758-766, doi:10.1111/psyp.12058 (2013).
49 Smeets, T. et al. Enhanced memory performance on an internal-internal source monitoring test following acute psychosocial stress. *Behav Neurosci* 120, 1204-1210, doi:10.1037/0735-7044.120.6.1204 (2006).
50 Quaedflieg, C. W., Meyer, T., Smulders, F. T. & Smeets, T. The functional role of individual-alpha based frontal asymmetry in stress responding. *Biol Psychol* 104, 75-81, doi:10.1016/j.biopsycho.2014.11.014 (2015).
51 Smeets, T., Jelicic, M. & Merckelbach, H. Stress-induced cortisol responses, sex differences, and false recollections in a DRM paradigm. *Biol Psychol* 72, 164-172, doi:10.1016/j.biopsycho.2005.09.004 (2006).
52 Smeets, T., Giesbrecht, T., Jelicic, M. & Merckelbach, H. Context-dependent enhancement of declarative memory performance following acute psychosocial stress. *Biol Psychol* 76, 116-123, doi:10.1016/j.biopsycho.2007.07.001 (2007).
53 Giesbrecht, T., Smeets, T., Merckelbach, H. & Jelicic, M. Depersonalization experiences in undergraduates are related to heightened stress cortisol responses. *J Nerv Ment Dis* 195, 282-287, doi:10.1097/01.nmd.0000253822.60618.60 (2007).
54 Smeets, T. et al. Stress selectively and lastingly promotes learning of context-related high arousing information. *Psychoneuroendocrinology* 34, 1152-1161, doi:10.1016/j.psyneuen.2009.03.001 (2009).
55 Smeets, T., Dziobek, I. & Wolf, O. T. Social cognition under stress: differential effects of stress-induced cortisol elevations in healthy young men and women. *Horm Behav* 55, 507-513, doi:10.1016/j.yhbeh.2009.01.011 (2009).

56 Nelemans, S. A. et al. The role of stress reactivity in the long-term persistence of adolescent social anxiety symptoms. Biol Psychol 125, 91-104, doi:10.1016/j.biopsycho.2017.03.003 (2017).
57 Vogel, S. et al. Blocking the mineralocorticoid receptor in humans prevents the stress-induced enhancement of centromedial amygdala connectivity with the dorsal striatum. *Neuropsychopharmacology* 40, 947-956, doi:10.1038/npp.2014.271 (2015).
58 Tollenaar, M. S. & Overgaauw, S. Empathy and mentalizing abilities in relation to psychosocial stress in healthy adult men and women. *Heliyon* 6, e04488, doi:10.1016/j.heliyon.2020.e04488 (2020).
59 Tollenaar, M. S., Elzinga, B. M., Spinhoven, P. & Everaerd, W. A. The effects of cortisol increase on long-term memory retrieval during and after acute psychosocial stress. *Acta psychologica* 127, 542-552, doi:10.1016/j.actpsy.2007.10.007 (2008).
60 Jansen, S. W., van Heemst, D., van der Grond, J., Westendorp, R. & Oei, N. Y. Physiological responding to stress in middle-aged males enriched for longevity: a social stress study. *Stress* 19, 28-36, doi:10.3109/10253890.2015.1105213 (2016).
61 Oei, N. Y. L. et al. Stress evokes stronger medial posterior cingulate deactivations during emotional distraction in slower paced aging. *Biol Psychol* 135, 84-92, doi:10.1016/j.biopsycho.2018.02.018 (2018).
62 Harris, B. N. Stress hypothesis overload: 131 hypotheses exploring the role of stress in tradeoffs, transitions, and health. *Gen Comp Endocrinol* 288, 113355, doi:10.1016/j.ygcen.2019.113355 (2020).
63 Miller, R. et al. The CIRCORT database: Reference ranges and seasonal changes in diurnal salivary cortisol derived from a meta-dataset comprised of 15 field studies. *Psychoneuroendocrinology* 73, 16-23, doi:10.1016/j.psyneuen.2016.07.201 (2016).
64 Miller, R. & Plessow, F. Transformation techniques for cross-sectional and longitudinal endocrine data: application to salivary cortisol concentrations. *Psychoneuroendocrinology* 38, 941-946, doi:10.1016/j.psyneuen.2012.09.013 (2013).
65 Kuiper, R. M., Buskens, V., Raub, W. & Hoijtink, H. Combining Statistical Evidence From Several Studies: A Method Using Bayesian Updating and an Example From Research on Trust Problems in Social and Economic Exchange *Sociological Methos & Research* 42, 60-81 (2012).
66 Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2), 163–171. https://doi.org/10.1016/j.psyneuen.2008.10.026

Part B

CHAPTER 4

The behavioral phenotype of early life adversity: a 3-level meta-analysis of rodent studies

Neurosci and Biobehav Rev. (2019), doi: 10.1016/j.neubiorev.2019.04.021

Bonapersona V[1]
Kentrop J[1]
Van Lissa CJ[2]
van der Veen R[3]
Joëls M[1,4]
Sarabdjitsingh RA[1]

1 Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, The Netherlands
2 Department of Methodology and Statistics, Utrecht University, The Netherlands
3 Centre for Child and Family Studies, Leiden University, The Netherlands.
4 University Medical Center Groningen, Groningen University, The Netherlands

Abstract

Altered cognitive performance is considered an intermediate phenotype mediating early life adversity (ELA) effects on later-life development of mental disorders, e.g. depression. Whereas most human studies are limited to correlational conclusions, rodent studies can prospectively investigate how ELA alters cognitive performance in several domains. Despite the volume of reports, there is no consensus on i) the behavioral domains being affected by ELA and ii) the extent of these effects. To test how ELA (here: aberrant maternal care) affects specific behavioral domains, we used a 3-level mixed-effect meta-analysis, and thoroughly explored heterogeneity with MetaForest, a novel machine-learning approach. Our results are based on >400 independent experiments, involving ~8600 animals. Especially in males, ELA promotes memory formation during stressful learning but impairs non-stressful learning. Furthermore, ELA increases anxiety-like and decreases social behavior. The ELA phenotype was strongest when i) combined with other negative experiences ("hits"); ii) in rats; iii) in ELA models of ~10 days duration. All data is easily accessible with MaBapp (https://vbonapersona.shinyapps.io/MaBapp/), allowing researchers to run tailor-made meta-analyses, thereby revealing the optimal choice of experimental protocols and study power.

Introduction

Early life adversity (ELA) is a consistent risk factor of psychiatric disorders[1,2], and it is regularly associated with poorer cognitive outcomes later in life[3–5]. Indeed, impaired cognitive processing is a prominent feature of psychopathologies[3,6,7], e.g. dysregulated contextual memory in post-traumatic stress disorder[8] or social cognition in schizophrenia[9]. ELA may therefore alter cognitive development, thereby resulting in behavioral abnormalities that may render individuals more vulnerable to psychiatric disorders[10].

To investigate exactly how ELA affects cognitive processing, rodent models are a valuable resource: they complement human studies by in-depth and thorough investigations of otherwise hard-to-study mechanisms. In animal experiments, genetic and environmental influences can be more precisely controlled and experimentally varied than in humans[11]. Furthermore, prospective designs are more feasible. For example, rodent studies have disentangled the different components of mother-pup interaction, a critical factor of early development across mammalian species[12–14]. This has helped uncover links between disturbed maternal care and disturbed emotional and cognitive functioning later in life, implicating the stress system[15] and "hidden regulators"[16].

Rodent studies have also highlighted paradoxical ELA effects on cognitive abilities. For instance, Benetti *et al.*[17] reported that rats with a history of ELA had impaired memory in the object recognition task. Conversely, Champagne *et al.*[18] reported that ELA mice display increased memory in a fear conditioning paradigm. Both tests have historically been used as memory tasks, albeit in a non-stressful and stressful context respectively. Possibly, the equivocal results are due to different underlying biological mechanisms (e.g. learning in stressful versus non-stressful situations) or pertain to the divergent methodology used (e.g. type of test or ELA model, species, experimenters, labs). A few studies have investigated the latter by testing the same animals in different memory tasks[19–22]. Although these studies favor the former explanation, the limited amount of animals used[23] – alongside the heterogeneous methodology – prohibits firm conclusions.

To address this conundrum, we here carried out a large-scale 3-level meta-analysis of all peer-reviewed preclinical literature on the subject, and tested the hypothesis that ELA (here defined as aberrant maternal care, i.e. differing from care seen in undisturbed, standard housed laboratory mice and rats) differentially affects specific behavioral domains in adulthood. We focused on memory formation after stressful or non-stressful learning, anxiety-like and social behavior, given their relevance for psychopathologies. We addressed (potential) sex-differences by investigating males and females separately. Furthermore, we tested whether the presence of multiple hits (e.g. other negative life experiences, independent of the developmental stage, see S1.4)[24] amplified ELA effects. Finally, we applied the novel, machine-learning based

analysis MetaForest[25] to identify the most important moderators of ELA effects on behavior.

Based on this comprehensive analysis, we evaluate the translational potential of ELA rodent models. With the aid of a specially developed web-based tool MaBapp (Meta-Analysis of Behavior application) (https://vbonapersona.shinyapps.io/MaBapp/), interested researchers can perform their own meta-analysis and retrieve valuable ad hoc information for experimental design and power calculations.

Methods

We adhered to SYRCLE's guidelines[26,27], and to the PRISMA[28] reporting checklist. To ease reading of the methodology, definitions of technical terms are provided in Supplemental Methods (S1.1). A summary of the general approach can be found in Figure 1.

Search strategy

The electronic databases PubMed and Web of Science (Medline) were used to conduct a comprehensive literature search on *the effects of ELA on behavior* on December 6th 2017. The search string was constructed with the terms "behavioral tests", "ELA" (as aberrant postnatal maternal care) and "rodents" (S1.2).

Prior to the beginning of the study, four experts were consulted. After elaborate discussions they agreed upon *i)* the selection of tests and related outcomes (S1.3), *ii)* their classification into behavioral domains (S1.3) and *iii)* the definition of multiple hits (S1.4). The results of each individual test, independent of categorization, are available for consultation on MaBapp (Section 5.1). Studies' titles and abstracts were screened independently by two researchers (VB & JK) and selected if the inclusion criteria were met (S1.5). Studies' inclusion was performed blinded to the studies' results. In case of doubt, the full text was inspected. Any disagreement was resolved by greater scrutiny and discussion.

To limit subjectivity in the data gathering and entry process, data from eligible studies were extracted in a standardized dataset alongside its explanatory codebook (https://osf.io/ra947/).

For each individual comparison, we calculated *Hedge's* G[29], a standardized mean difference with a correction for small samples[30]. S1.6 details the extraction of statistical information as well as handling of missing values. We estimated the summary statistics of data presented only graphically with *Ruler for Windows* (https://a-ruler-for-windows.en.softonic.com/), of which we previously validated the accuracy[31]. If the data was not reported in any format (or other crucial information was missing e.g. sex), we contacted two authors per manuscript published after 2008 (response rate 52.6%). If no answer was received within two months and after a reminder, the authors were considered not reachable, and the comparison was excluded.

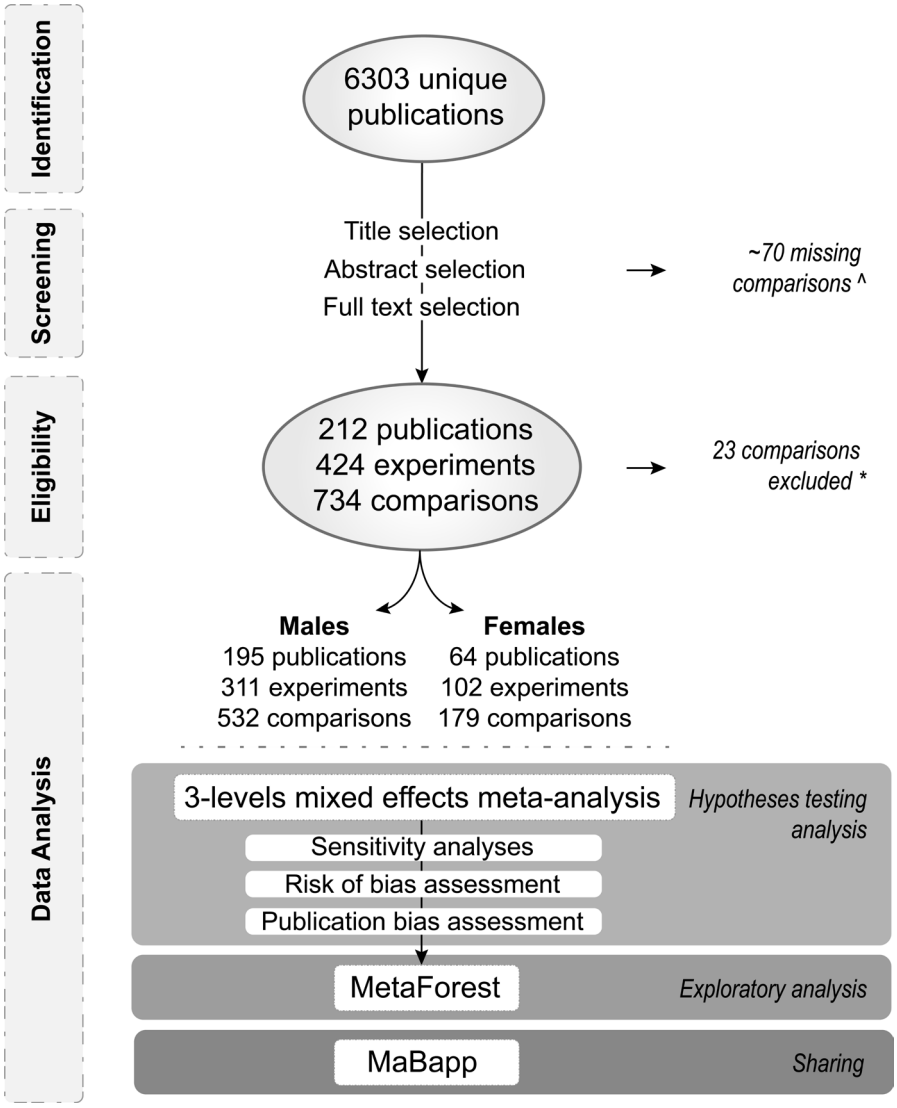


Figure 1 Flow chart of study selection and analysis. Of note: in 47 publications, both males and females were tested. ^ = estimation of missing comparisons (S2.1); * = comparisons excluded from the meta-analysis due to controversial behavioral domain categorization (S2.2).

Meta-analysis: research questions and statistical approach

To avoid possible biases, the experimenter (VB) was blinded to the ELA effects while coding the analysis. This was achieved by randomly multiplying half of the effect sizes by -1.

Hypothesis-testing We built a 3-level mixed effect meta-analysis with restricted maximum likelihood estimation. In our experimental design, the 3 levels correspond to variance of effect size between 1) animals, 2) outcomes and 3) experiments. This approach accounts for the violation of the assumption of independency when the data is collected from the same animals[31–33], thereby improving the robustness of the conclusions drawn. We included "domains" and "hits" as moderators in order to address the following two research question: 1) what are the effects of ELA on each behavioral domain?; 2) are the effects enhanced if the animals experienced multiple hits?. Since both questions were answered with the same model, effect sizes were estimated only once.

We ensured that all behavioral measurements were in the same theoretical direction by multiplying – whenever necessary – the effect sizes by -1 [34](S1.3). Although this was essential for the model estimation, we here report effect sizes in a more interpretable manner: an increase in Hedge's G signifies an enhancement of the behavioral domain under study (e.g. more anxiety-like behavior, better memory).

We conducted several sensitivity analyses (S1.7) to assess the robustness and consistency of our conclusions. We examined whether the quality of the studies affected the estimation of the results by dissecting the influence of reporting bias, blinding, randomization and study power. Furthermore, we thoroughly investigated influential and outlying cases[35] according to multiple definitions (S1.7).

To compensate for methodological limitations, we tested the presence of publication bias with various qualitative/quantitative methods (S1.8), and quantified its influence with fail-safe N[36] and trim-and-fill analyses[37](S1.8).

Risk of bias was evaluated with SYRCLE's assessment tool[38], where we distinguished between study-level and outcome-level biases[28]. Lack of reporting of experimental details was scored as an unclear risk of bias.

Heterogeneity was assessed with Cochrane Q-test[33] and I^2, which was estimated at each of the 3-levels of the model to determine how much variance could be attributed to differences within (level 2) or between experiments (level 3)[39]. Estimates of explained variance can be positively biased when based on the data used to estimate the model[40]. For this reason, we used 10-fold cross-validation to obtain an estimate of how much variance our model might explain in new data. This cross-validated estimate of R^2 (R_{cv}^2) is robust to overfitting and provides evidence for the results' generalizability.

P-values were corrected with Bonferroni for family-wise error rate (each research question considered as a separate family of tests) to limit capitalization on chance. Since we expected the amplitude of effect sizes to differ between sexes[41,42], we considered males and females as two separate datasets.

Exploratory analysis We used MetaForest[25], a novel exploratory approach to identify the most important moderators of the ELA effects on behavioral domains. This

innovative, data-driven technique adapts random forests (a machine learning algorithm) for meta-analysis, by means of bootstrap sampling. MetaForest ranks moderators based on their influence on the effect size.

Preclinical experiments often adopt diverse protocols. Although this can be an advantage[43], in a meta-analysis it induces substantial heterogeneity. Therefore, we classified the published experimental protocols in >30 standardized variables with the intent to identify potential methodological sources of heterogeneity. Based on theoretical importance, we selected 26 of these moderators for inclusion in the MetaForest analysis. We used 10-fold cross-validation (S1.9) to determine the optimal tuning parameters that minimized RMSE: uniform weighting, 4 candidate moderators at each split, and a minimum node size of 2. The marginal bivariate relationship of each moderator with effect size was averaged over the values of all other moderators (S1.9). Residual heterogeneity was estimated with τ^2 (S1.9).

Lastly, we created MaBapp (https://vbonapersona.shinyapps.io/MaBapp/) for anyone to perform their own meta-analysis on the topic by selecting their favorite characteristics (Section 5.1).

Analyses were conducted in R (version 3.5.1)[44], using the following packages: 1) *metafor*[29] for conducting the analysis, 2) *metaforest*[45] for data exploration, 3) *shiny*[46] to create MaBapp, and 4) *dplyr*[47] for general data handling. For further specifications about the analysis, the R script is available (https://osf.io/ra947/).

Results

Studies selection and characteristics

In total ~8600 animals (age_{weeks} median[IQR] = 12[4]; proportion rats = 68%) were included in the analysis, 77.7% of which were males. Anxiety-like behavior was the domain most investigated (48.8%), elevated plus maze the most popular test (14.3%), and maternal separation the ELA paradigm most often used (48.9%). For additional descriptive information on study characteristics, see S2.3.

Although no publication reported on all SYRCLE's potential bias items, 41 publications (19.3%) were blinded as well as randomized, and overall we estimated a risk of bias of 3[1] (median[IQR]) on a 10 points scale (S2.4). Lastly, at a systematic review level (S2.5), 68.5% of comparisons were either not-significant (n_{comp}=386) or the result could not be directly interpreted from the information provided (n_{comp}=117).

ELA effects are pronounced in males and with "multiple hits"

The effect sizes included ranged between -6.4 and 6.1 (mean[SD] = 0.29[1.06]), with 95% of comparisons between -2 and 2. Sample size ranged between 6 and 59 animals (mean[SD] = 22[7.8]), and differed <20% between control and ELA groups in 90% of the cases (estimation).

When qualitatively comparing sexes, the effects of ELA were more evident in

males than in females. Male rodents with a history of ELA displayed increased anxiety-like (HedgesG [95%CI] = 0.278 [0.165,0.39], z = 4.819, p<.000), improved memory after stressful learning (HedgesG [95%CI] = 0.283 [0.141,0.425], z = 3.9, p<.000), impaired memory after non-stressful learning (HedgesG [95%CI] = -0.594 [-0.792,-0.395], z = -5.86, p<.000) and decreased social behavior (HedgesG [95%CI] = -0.614[-0.88,-0.348], z = -4.521, p<.000, Figure 2A, S2.6). We were unable to confirm any effect of ELA on behavior in females, although directionality was generally comparable in both sexes (Figure 2B, S2.7).

Overall, the presence of multiple hits (for our definition, see S1.4) intensified the effects of ELA in males (HedgesG [95%CI] = 0.222 [0.018,0.426], z = 2.131, p = 0.033) yet marginally in females (HedgesG [95%CI] = 0.297 [-0.003,0.596], z = 1.939, p = 0.052). Although these enhancing effects were not significant at a single-domain level (posthoc analysis, Figure 2C-D, S2.6/S2.7), memory after non-stressful learning was the most impacted domain in males (difference in *Hedge's* G = 0.435, z = 2.156, p = 0.124) as well as in females (difference in *Hedge's* G = 0.565, z = 2.234, p = 0.102).

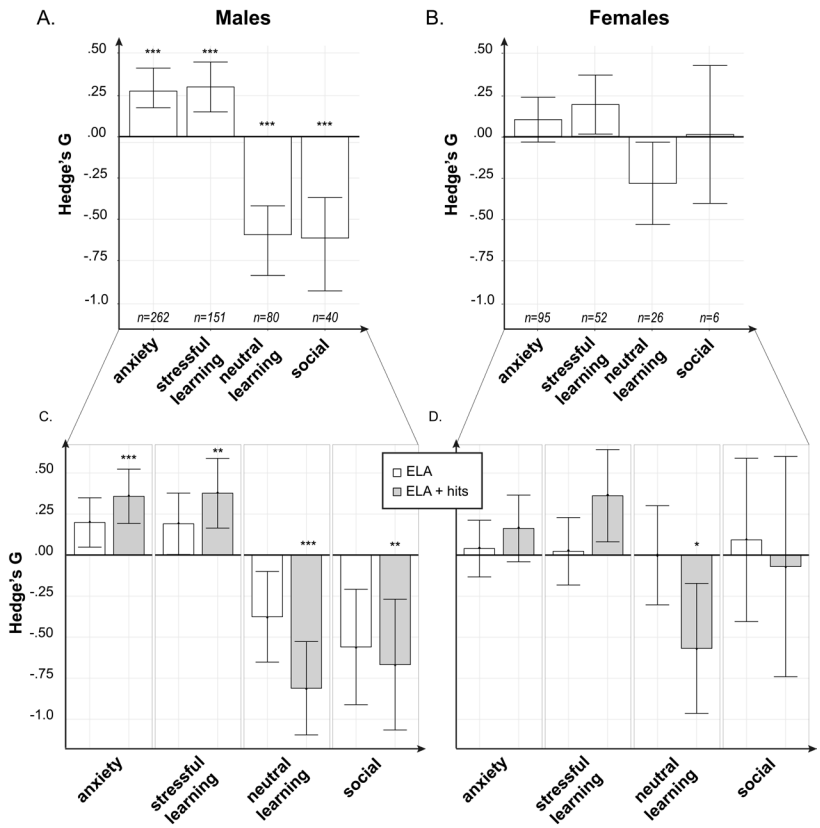


Figure 2 Effects of ELA on behavioral domains in males (A) and females (B), and the role of multiple hits (in addition to ELA, grey bars) compared to only ELA (white bar) in mediating these effects (males: C, females: D). Each bar represents the size of the effect (*Hedge's* G, standardized mean difference) of the ELA manipulation when comparing a control and an experimental group. * = p<.05, ** = p<.01, *** = p<.001

Sensitivity analyses and publication bias. Qualitative evaluation of funnel plot asymmetry suggested the presence of publication bias, which was confirmed by Egger's regression and Begg's test (S2.8). Nonetheless, fail-safe N as well as trim-and-fill analyses confirmed that – albeit present – publication bias is unlikely to distort the interpretation of the results (S2.8). Furthermore, the robustness of the male and female models was confirmed by several sensitivity analyses (S2.9).

Exploration of moderators

Although the models of the hypotheses-testing analysis described a significant proportion of variance ($R_{cv\ males}^2$=0.026, $R_{cv\ females}^2$=0.03), substantial heterogeneity was recorded in both models (males: Q(524) = 1763.118, p<.000; females: Q(171) = 326.93, p<.000, S2.10). This was not surprising due to the diversity of publications included in the meta-analysis.

To investigate the source of the heterogeneity, we used MetaForest, a new statistical technique that ranks moderators (Figure 3A) based on their predictive value. These can roughly be divided in 4 groups, describing: *i)* characteristics of the animals (e.g. origin of the breeding animals (Figure 3B) and species investigated (Figure 3C)), *ii)* ELA model used (e.g. type of model and duration of ELA (Figure 3D-E)), *iii)* outcome measures (e.g. domain and test used), and *iv)* potential bias (e.g. blinding and randomization). MetaForest confirmed that the selected moderators account for a substantial portion of the variance (R_{cv}^2[SD]=0.12[0.09]).

Offspring of dams purchased pregnant had larger effect sizes than offspring bred in the own facility (Figure 3B). Rats had overall larger effect sizes than mice (Figure 3C). Concerning ELA models (Figure 3D), selecting the extremes of natural variation (licking-and-grooming model) yielded the strongest phenotype. Lastly, effect sizes appeared to be maximal with a 10 days' ELA duration (Figure 3E).

Discussion

In this study, we substantiate that adversities early in life profoundly and lastingly change rodent behavior. Due to low power[23] and heterogeneous methodologies, results at a single-study level are often inconclusive and difficult to interpret. Here, by adopting a meta-analytic approach, we provide extensive evidence that ELA (due to maternal care that differs from that provided by undisturbed, standard-housed dams) has differential effects on memory: it enhances memory if learning occurs in a stressful situation, but it hampers learning under non-stressful circumstances. Furthermore, ELA increases anxiety-like and decreases social behavior, particularly in males. In line with the multiple-hits hypotheses[24,48], the effects are amplified if the animals experience other stressful life events (e.g. prenatal stress due to transport of pregnant females),

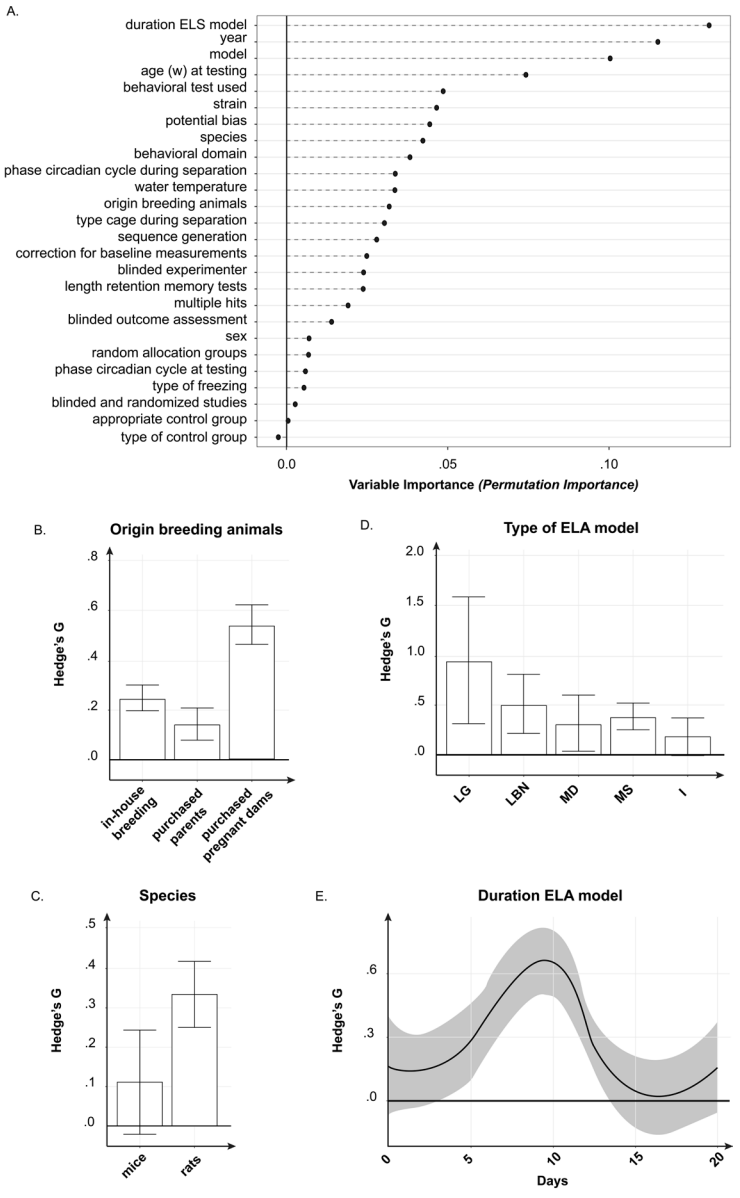


Figure 3 Exploratory MetaForest analysis. (A) Rank moderators' importance. Variable/permutation importance is a measure of how strongly each moderator explains differences in effect size, capturing (non-)linear relationships as well as higher order interactions. For information about MetaForest's partial dependence plots, see S2.11. Effect sizes distinguished by origin of the breeding animals (B), species (C), type of ELA model (D) and duration of ELA (E). Results are expressed as Hedge's G[95%CI]. The usefulness of this exploration can be best appreciated with the aid of MaBapp. For example, the overall estimate of the effects of ELA on anxiety-like behavior is Hedge's G=.24. However, if we select only the LBN model, the effect size rises to .37. If we combine LBN and rats, the effect size further rises to .60. If we then select only elevated plus maze as respectively behavioral test, the effect size rises to .81. LG = licking-and-grooming, LBN = limited bedding and nesting, MD = maternal deprivation, MS = maternal separation, I = isolation.

independent of the developmental period during which these occur (S1.4). These results are independent of the type of ELA or behavioral test used, and are remarkably similar to what has been reported at a correlational level in humans[49,50]. Altogether, our results point to a clear and robust phenotype of ELA in four behavioral domains and complement the human literature by supporting a causative role of ELA in altering behavior, which may predispose individuals to precipitate symptoms of psychiatric disorders.

Methodological considerations

The lack of sufficient power to detect experimental effects is an emerging issue in preclinical literature[23,31] that seriously hampers research interpretation[23]. As a consequence, results from single-studies are useful for hypotheses generation but do require replication. The ability to recreate experiments (replication) and/or to reach similar conclusions via different methods (reproducibility) are fundamental aspects of scientific inquiry. Underpowered research undermines both aspects, as the conclusions drawn are likely to be uncertain[51].

Indeed, in our study the majority of comparisons (68.5%) was not-significant at a systematic review level, but the effects were significant when analyzed meta-analytically. In addition to study preregistration, realistic power calculations, and testing by several independent teams[23,51], statistical tools such as meta-analyses can therefore be very useful to substantiate conclusions from animal data and translate them more reliably to patients[52]. Furthermore, our study showcases how "negative" research is also useful, and reminds how (lack of) formal statistical significance (typically p-value <0.05) must not be a decisive requirement to publish research.

In this project, we intertwine these concepts with state-of the-art statistical methodology, adopting an approach never used in preclinical studies. Firstly, our meta-analysis was built with a 3-level model[33], which allows for a more robust estimation of the effects by accounting for the dependency of same-animal's data[31,53]. Secondly, a leading strength of preclinical meta-analyses is the systematic exploration of heterogeneity[52]. Instead of the standard subgroup/meta-regression approach, we opted for an exploratory analysis using *MetaForest*[25], a newly developed technique that ranks moderators' importance by adapting the machine learning algorithm random forests to summary-statistics' data. A major strength of *MetaForest* is its robustness to overfitting, and its ability to accommodate non-linear effects[25], as shown by the impact of ELA duration on effect sizes.

Thirdly, we extensively coded potential (biological and experimental) moderators. Although possibly relevant moderators were not included due to insufficient reporting (e.g. temperature during separation[54], cross-fostering[55], culling[56]), this dataset treasures relevant information for future experimental designs. To facilitate others to exploit this dataset, we created *MaBapp* (https://vbonapersona.shinyapps.io/MaBapp/), a web-

based app with a user-friendly interface through which anyone can perform his/her own meta-analysis on the topic of ELA and behavioral domains. Within the app, a wide variety of features can be selected, such as ELA models and their components (e.g. type, timing, predictability), behavioral tests used, age and sex of the animals, etc. Based on the characteristics indicated, the app reports forest, funnel and cumulative plots, as well as a list of relevant publications. The app is a useful resource, which can be used to i) comprehensively retrieve relevant publications, ii) explore the literature at an individual researcher's needs' level, iii) define new hypotheses, iv) evaluate publication bias and replicability of findings, and v) estimate realistic effect sizes on which to ground future research.

The validity of our conclusions is not limited to the robustness of the models used but grounded on the vast primary evidence included (>200 publications). As a consequence, accidental findings have little weight. Although the methods and approach we adopt are rigorous and reasonably conservative, the quality of the conclusions critically depends on the quality of the studies and data included. From our qualitative bias assessment, the risk for potential bias was lower than previously reported in Neuroscience[31,57,58]; yet, only ~20% of studies stated being blinded as well as randomized. Furthermore, any meta-analytic dataset is burdened with missing data, due to publication bias or to the preferred investigation of certain factors over others[59]. Our models did display evidence of publication bias, yet they were robust to several corrections and sensitivity analyses. Although we cannot fully exclude that the above-mentioned limitations may affect the outcome, it is unlikely that the conclusions drawn would be substantially impacted. Nevertheless, we have attempted to address these methodological issues as comprehensively as possible in our analysis.

Considerations on ELA models

ELA encompasses a wide range of pre- and post-natal experiences, but we here focused on altered maternal care (relative to care provided by undisturbed, standard-housed dams). Although this definition limits the generalizability of the conclusions, it is essential to enable the comparability (thus meaningful quantitative synthesis) of the studies incorporated in our meta-analysis.

The behavioral changes we report are presumably a convergent phenotype of distinct, model-dependent, underlying biological mechanisms. An organism's development is not linear nor simultaneous for every component, but it occurs in critical periods[60]. For example, postnatal day (P)2-P5 is a sensitive period for the maturation of the adrenal glands[61], P9 for prepulse inhibition[62], and ~P10 for adrenal responsiveness[63]. Furthermore, higher cognitive functions develop as multistage processes of sequential nature[60]. Accordingly, ELA may particularly disrupt the development of competences whose critical period is active during the time of stress, thereby heightening the variability of the ELA phenotype.

Evidence supporting these notions derives from studies using a single 24h maternal deprivation paradigm, which show a persistent yet paradoxical hypo- and hyper-responsiveness of juvenile ACTH if deprivation occurred at P3 or P11 respectively[64]. Thus, while meta-analyses may serve to discern patterns among vast amounts of studies, exploratory studies experimentally dissecting components of ELA in rodents remain indispensable for addressing the underlying mechanisms of action of ELA to the brain (for example:[65,66]).

Suggestions for future ELA research Given that the criteria for construct and face validity of ELA models have been met[67], our results provide a practical framework where researchers can anticipate the ELA effect on cognitive outcomes and/or build their own ELA model accordingly. Our exploratory analysis gives insights in the suitability of the models and tests to choose, depending on the question.

Based on this analysis, we tentatively conclude that i) rats seem overall more sensitive to ELA-induced changes than mice. Moreover, ii) elements such as transporting pregnant dams appear to amplify the effects of ELA. Such stressful life events may have substantial impact on the system, in line with the multiple-hit theory 24. As evident from Figure 3, iii) a duration of 10 days ELA produced the most robust phenotype. Finally, iv) the limited bedding and nesting (LBN) model produced the largest effect sizes when compared to separation/deprivation models. Given this reliability, in combination with the feasibility and translational validity, LBN seems an influential paradigm to investigate the mechanisms of chronic stress early in life[41,68].

According to the rank of moderators by MetaForest, publication year, age of testing, strain and behavioral test used account for a substantial portion of the variance. The impact of publication year has previously been reported in several areas of biology[69] , and could be the result of the Winner's curse[23]. In brief, the first published studies on any topic are likely to be biased towards extreme effect sizes. This bias tends to disappear as evidence accumulates, thereby providing an explanation for the influence of publication year in our dataset.

Conversely, age of testing, strain and behavioral test used did not show any theory-interpretable pattern. One explanation could be that there is no preferable age/strain/test, but that the different elements of the study design interact with one another. For example, the open field (OF) and the elevated plus maze (EPM) are behavioral tests used to assess anxiety-like behavior. Conceptually, they both aim to create a conflict between the rodents' exploratory drive and their fear of exposed spaces[70]. With MaBapp, we can explore the confidence interval of these two tests following the LBN model (OF: HedgesG [95%CI] = 0.12 [-0.21,0.44]; EPM: HedgesG [95%CI] = 0.49 [0.22,0.75]) or maternal separation (OF: HedgesG [95%CI] = 0.32[0.14, 0.5]; EPM: HedgesG [95%CI] = 0.4[0.15, 0.65]). Tentatively, the EPM appears more sensitive than the OF to represent the effects of the LBN model, while rather similar when investigating

the effects of maternal separation. Similarly, we can explore the interaction between these tests and any specific strain. For example, C57Bl/6 mice appear more sensitive to the EPM (HedgesG [95%CI] = 0.38 [0.07,0.68]) than to the OF (HedgesG [95%CI] = 0.00 [-0.27,0.28]), independent of the ELA model used. These examples illustrate the complexity of these interactions. Unfortunately, the information so far available is insufficient to conduct meaningful quantitative analyses. Nonetheless, researchers can now make more informed decision on experimental designs by exploring with MaBapp (feasible) possibilities that fit their needs. Alternatively, we refer researchers to primary publications in which the effects of age[71] or strain[72] were experimentally investigated.

To reduce variability and improve comparability across studies, ELA should be preferably applied with consistent protocols (S1.5), unless manipulation of particular aspect(s) of the model is under investigation. Obviously, the importance of individual variation is a factor that should not be overlooked. In our analysis, the paradigm of licking-and-grooming – which is not experimentally induced but based on natural variation in maternal care – consistently evoked the largest effect sizes, although these were based on fewer publications than the other models.

Translational potential

ELA is one of the most consistent environmental risk factors for the development of psychopathology[2]. Although the effects of ELA on the brain can be adaptive, they may evolve into dysfunctional elements in genetically predisposed individuals[2]. Behavioral performance in specific cognitive domains seems to be a relevant intermediate phenotype[10], as it may mediate the effects of ELA on psychopathology. For example, in post-traumatic stress disorder, enhanced memory of stressful events becomes pathological after a later-life trauma[8].

In humans, the concept of ELA is extremely varied. Even when considering solely maltreatment, this can be characterized by repeated or sustained episodes of various forms of neglect and abuse[73]. Furthermore, the environmental variation is intertwined with socio-economic status, complex relations (e.g. family, neighborhoods, peers, school), and intergenerational transmissions[73]. Rodent paradigms do not capture the complexity of human ELA, but they can model specific aspects of the human variability in a well-controlled setting. For example, LBN is based on the erratic and unpredictability of maternal care[41,68,74] that has been established as a hallmark in childhood abuse situations[75]. Similarly, cognitive performance (e.g. memory after stressful learning) can be modelled in rodents, albeit with clear restraints: the tasks are obviously different, should be interpreted in relation to the animal's normal behavior, and cannot investigate a range of outcomes such as verbal abilities, critical for social interaction and psychopathology[76], also in relation to ELA[77].

Explaining how ELA increases psychopathology risk requires the understanding of its complex interplay with other susceptibility/resilience factors, such as genetic

background and later life stressors[78]. This mechanistic investigation is difficult to achieve in humans, where limited material, difficulty of prospective and longitudinal designs, complexity and lack of control over the environment and genetic variation hamper causal inferences of ELA to later life cognitive performance. To this end, animal studies can be of considerable added value[41].

An interesting issue in evaluating the translational potential of ELA rodent models is sex differences. In our analysis, males showed larger effect sizes (albeit in the same direction) than females to the effects of ELA on all outcomes, thereby confirming previous preclinical literature[42]. Conversely, in clinical populations, females appear more sensitive to childhood trauma as well as to the development of stress-related psychopathologies[41], although sex differences depend on the type of disorder[79]. A plausible biological explanation for this discrepancy is the developmental timing during which stress occurs. Although humans and rodents are altricial species, the brain of newborn rats corresponds roughly to 23/24-week old human fetuses[80]. Interestingly, the sensitivity to adversities in the last trimester of gestation in humans has been suggested to affect males more than females[81]. Experimentally manipulating the timing of ELA exposure may further elucidate 'female' stress-sensitive periods. It therefore remains to be established whether the effects of ELA on cognitive domains are truly different between sexes. Our analyses suggest that the effects may not be sexually dysmorphic in nature but may result from the experimental designs used. For example, ELA models and behavioral tests were originally developed for males: maternal care shows clear sex-specific differences[82,83], and females perform poorly in behavior tests such as object recognition and object-in-location[41,42]. Consequently these paradigms may not be sensitive enough for a female's phenotype. Indeed, the recorded effects were in the same direction across sexes, and MetaForest attributed to sex a relatively modest variable importance. Our results showcase the necessity to study sex as a biological variable[81,84], which requires the development of tests and models that are female-specific. This step is required for a more meaningful comparison between rodent and humans, and a delineation of the underlying sex-dependent mechanisms of ELA.

Despite these drawbacks, our meta-analysis confirms and importantly extends standing hypotheses on ELA based on exploratory studies. To aid future investigations in this field, we provide a online tool to evaluate existing literature and direct the experimental design of new studies.

References

1 Kessler, R. C. et al. Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *The British journal of psychiatry : the journal of mental science* 197, 378–85 (2010).

2. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience* 17, 652–666 (2016).

3. Masson, M., East-Richard, C. & Cellard, C. A meta-analysis on the impact of psychiatric disorders and maltreatment on cognition. *Neuropsychology* 30, 143–156 (2016).

4. Nelson, C. A. et al. Cognitive Recovery in Socially Deprived Young Children: The Bucharest Early Intervention Project. *Science* 318, 1937–1940 (2007).
5. Vargas, T. et al. Childhood Trauma and Neurocognition in Adults With Psychotic Disorders: A Systematic Review and Meta-analysis. *Schizophrenia Bulletin* 1–14 (2018) doi:10.1093/schbul/sby150.
6. Millan, M. J. et al. Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery* 11, 141–168 (2012).
7. Monfils, M. H. & Holmes, E. A. Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry* 0366, (2018).
8. Liberzon, I. & Abelson, J. L. Context processing and the neurobiology of post-traumatic stress disorder. *Neuron* 92, 14–30 (2016).
9. Green, M. F., Horan, W. P. & Lee, J. Social cognition in schizophrenia. *Nature Reviews Neuroscience* 16, 620–631 (2015).
10. Ammerman, R. T., Cassisi, J. E., Hersen, M. & Van Hasselt, V. B. Consequences of physical abuse and neglect in children. *Clinical Psychology Review* 6, 291–310 (1986).
11. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Current Opinion in Psychology* 15, 174–181 (2017).
12. Bowlby, J. Maternal care and mental health. *World Health Organization Monograph Series* 179 (1951).
13. Harlow, H. & Harlow, M. The affectional systems. *in Behavior of nonhuman primates* (eds. Schrier, A., Harlow, H. & Stollnitz, F.) 287–334 (Academic Press, 1965).
14. Meaney, M. J. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* 24, 1161–1192 (2001).
15. Levine, S. Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology* 30, 939–946 (2005).
16. Hofer, M. A. Hidden Regulatory Processes in Early Social Relationships. *in Perspectives in Ethology: social behavior* (eds. Bateson, P. & Klopfer, P.) 135–201 (Springer, 1978).
17. Benetti, F. et al. Early postnatal maternal deprivation in rats induces memory deficits in adult life that can be reversed by donepezil and galantamine. *International Journal of Developmental Neuroscience* 27, 59–64 (2009).
18. Champagne, D. L. et al. Maternal Care and Hippocampal Plasticity: Evidence for Experience-Dependent Structural Plasticity, Altered Synaptic Functioning, and Differential Responsiveness to Glucocorticoids and Stress. *Journal of Neuroscience* 28, 6037–6045 (2008).
19. Kanatsou, S. et al. Overexpression of Mineralocorticoid Receptors in the Mouse Forebrain Partly Alleviates the Effects of Chronic Early Life Stress on Spatial Memory, Neurogenesis and Synaptic Function in the Dentate Gyrus. *Frontiers in Cellular Neuroscience* 11, 1–13 (2017).
20. Ivy, A. S. et al. Hippocampal Dysfunction and Cognitive Impairments Provoked by Chronic Early-Life Stress Involve Excessive Activation of CRH Receptors. *Journal of Neuroscience* 30, 13005–13015 (2010).
21. Bredy, T. W., Humpartzoomian, R. A., Cain, D. P. & Meaney, M. J. Partial reversal of the effect of maternal care on cognitive function through environmental enrichment. *Neuroscience* 118, 571–576 (2003).
22. Mello, P. B., Benetti, F., Cammarota, M. & Izquierdo, I. Physical exercise can reverse the deficit in fear memory induced by maternal deprivation. *Neurobiology of Learning and Memory* 92, 364–369 (2009).
23. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14, 365–376 (2013).
24. Daskalakis, N. P. et al. The three-hit concept of vulnerability and resilience: Toward understanding adaptation to early-life adversity outcome. *Psychoneuroendocrinology* 38, 1858–1873 (2013).
25. van Lissa, C. J. MetaForest: Exploring heterogeneity in meta-analysis using random forests. *PsyArXiv* (2018) doi:10.31234/osf.io/myg6s.
26. De Vries, R. B. M. et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclinical Medicine* 1, 1–9 (2015).
27. Leenaars, M. et al. A step-by-step guide to systematically identify all relevant animal studies. *Laboratory Animals* 46, 24–31 (2012).
28. Moher, D. et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6, (2009).
29. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 36, 1–48 (2010).
30. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221, 92–102 (2014).

31. Bonapersona, V., Joels, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: A 3 level meta-analysis of rodent studies. *Neuroscience and biobehavioral reviews* 95, 1–16 (2018).
32. Aarts, E., Verhage, M., Veenvliet, J. V, Dolan, C. V & van der Sluis, S. A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience* 17, 491–496 (2014).
33. Cheung, M. W. L. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods* 19, 211–229 (2014).
34. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221, 92–102 (2014).
35. Viechtbauer, W. & Cheung, M. W.-L. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* 1, 112–125 (2010).
36. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86, 638–641 (1979).
37. Duval, S. & Tweedie, R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463 (2000).
38. Hooijmans, C. R. et al. SYRCLE ' s risk of bias tool for animal studies. *BMC Medical Research Methodology* 14, 1–9 (2014).
39. Assink, M. & Wibbelink, C. J. M. Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology* 12, 154–174 (2016).
40. Hastie, T., Friedman, J. & Tibshirani, R. *The elements of statistical learning: Data Mining, Inference and Prediction.* (Springer Series in Statistics, 2009).
41. Walker, C.-D. et al. Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* 0, 1–28 (2017).
42. Loi, M. et al. Effects of early-life stress on cognitive function and hippocampal structure in female rodents. *Neuroscience* 342, 101–119 (2017).
43. Karp, N. A. Reproducible preclinical research—Is embracing variability the answer? *PLoS Biology* 16, 1–5 (2018).
44. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2015).
45. van Lissa, C. J. Package ' metaforest '. *Cran* (2018).
46. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for R. (2017).
47. Wickham, H., François, R., Henry, L., Müller, K. & RSudio. Package ' dplyr ' Version 0.7.6. (2018).
48. Walker, A. K. et al. Neonatal lipopolysaccharide and adult stress exposure predisposes rats to anxiety-like behaviour and blunted corticosterone responses: Implications for the double-hit hypothesis. *Psychoneuroendocrinology* 34, 1515–1525 (2009).
49. Pechtel, P. & Pizzagalli, D. A. Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology* 214, 55–70 (2011).
50. Suor, J. H., Sturge-Apple, M. L., Davies, P. T., Cicchetti, D. & Manning, L. G. Tracing Differential Pathways of Risk: Associations Among Family Adversity, Cortisol, and Cognitive Functioning in Childhood. *Child Development* 86, 1142–1158 (2015).
51. Ioannidis, J. P. A. Why most published research findings are false. *PLoS medicine* 2, e124 (2005).
52. Hooijmans, C. R. & Ritskes-Hoitinga, M. Progress in using systematic reviews of animal studies to improve translational research. *PLoS Medicine* 10, 1–4 (2013).
53. Rosenthal, R. *Meta-analytic procedures for social research.* (Sage, 1991).
54. Pryce, C. R., Bettschen, D., Nanz-Bahr, N. I. & Feldon, J. Comparison of the effects of early handling and early deprivation on conditioned stimulus, context, and spatial learning and memory in adult rats. *Behavioral Neuroscience* 117, 883–893 (2003).
55. Penke, Z. et al. Postnatal maternal deprivation produces long-lasting modifications of the stress response, feeding and stress-related behaviour in the rat. *The European journal of neuroscience* 14, 747–755 (2001).
56. Veenema, A. H., Bredewold, R. & Neumann, I. D. Opposite effects of maternal separation on intermale and maternal aggression in C57BL/6 mice: link to hypothalamic vasopressin and oxytocin immunoreactivity. *Psychoneuroendocrinology* 32, 437–450 (2007).
57. Antonic, A. et al. Stem Cell Transplantation in Traumatic Spinal Cord Injury: A Systematic Review and Meta-Analysis of Animal Studies. *PLoS Biology* 11, (2013).
58. Egan, K. J., Sena, E. S. & Vesterinen, H. M. Making the most of animal data - Improving the prospect of success in pragmatic trials in the neurosciences. *Trials* 12, no pagination (2011).
59. Cooper, H., Hedges, L. V & Valentine, J. C. *The handbook of research synthesis and meta-analysis 2nd edition.*

(Russel Sage Foundation, 2009).

60. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience* 6, 877–888 (2005).
61. Levine, S. & Lewis, G. W. Critical period for effects of infantile experience on maturation of stress response. *Science* 129, 42–43 (1959).
62. Ellenbroek, B. A. & Cools, A. R. Early maternal deprivation and prepulse inhibition: The role of the postdeprivation environment. *Pharmacology Biochemistry and Behavior* 73, 177–184 (2002).
63. Witek-Janusek, L. Pituitary-adrenal response to bacterial endotoxin in developing rats. *American Journal Physiology* 255, E525–E530 (1988).
64. Van Oers, H. J. j., De Kloet, E. R. & Levine, S. Early vs. late maternal deprivation differentially alters the endocrine and hypothalamic responses to stress. *Developmental Brain Research* 111, 245–252 (1998).
65. Peña, C. J. et al. Early life stress confers lifelong stress scsceptibility in mice via ventral tegmental area OTX2. *Science* 1, 1185–1188 (2017).
66. Singh-Taylor, A. et al. NRSF-dependent epigenetic mechanisms contribute to programming of stress-sensitive neurons by neonatal experience, promoting resilience. *Molecular Psychiatry* 23, 648–657 (2018).
67. Suchecki, D. Maternal regulation of the infant's hypothalamic-pituitary-adrenal axis stress response: Seymour 'Gig' Levine's legacy to neuroendocrinology. *Journal of Neuroendocrinology* 30, 1–17 (2018).
68. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* 149, 4892–4900 (2008).
69. Jennions, M. D. & Møller, A. P. Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences* 269, 43–48 (2002).
70. Wigger, A. & Neumann, I. D. Periodic maternal deprivation induces gender-dependent alterations in behavioral and neuroendocrine responses to emotional stress in adult rats. *Physiology and Behavior* 66, 293–302 (1999).
71. Oitzl, M., Workel, J., Fluttert, M., Frösch, F. & Ronde Kloet, E. Maternal deprivation affects behaviour from youth to senescence: amplication of individual differences in spatial learning and memory in senescent Brown Norway rats. *European Journal of Neuroscience* 12, 3771–3780 (2000).
72. Millstein, R. A. & Holmes, A. Effects of repeated maternal separation on anxiety- and depression-related phenotypes in different mouse strains. *Neuroscience and Biobehavioral Reviews* 31, 3–17 (2007).
73. Teicher, M. H. & Samson, J. A. Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *American Journal of Psychiatry* 170, 1114–1133 (2013).
74. Baram, T. Z. et al. Fragmentation and unpredictability of early-life experience in mental disorders. *American Journal of Psychiatry* 169, 907–915 (2012).
75. Whipple, E. E. & Webster-Stratton, C. The role of parental stress in physically abusive families. Child Abuse & Neglect 15, 279–291 (1991).
76. Cohen, N. *Language impairment and psychopathology in infants, children, and adolescnets.* (Sage Publications, 2001).
77. Miller, A. B. et al. Dimensions of deprivation and threat, psychopathology, and potential mediators: A multi-year longitudinal analysis. *Journal of Abnormal Psychology* 127, 160–170 (2018).
78. Bale, T. L. et al. Early life programming and neurodevelopmental disorders. *Biological Psychiatry* 68, 314–319 (2010).
79. Riecher-Rössler, A. Sex and gender differences in mental disorders. *The Lancet Psychiatry* 4, 8–9 (2017).
80. Plotsky, P., Bradley, C. & Anand, K. Behavioral and neuroendocrine consequences of neonatal stress. *Pain Research and Clinical Management* 10, 77–100 (2000).
81. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. *Nature neuroscience* 18, 1413–20 (2015).
82. Oomen, C. A. et al. Opposite effects of early maternal deprivation on neurogenesis in male versus female rats. *PLoS ONE* 4, (2009).
83. van Hasselt, F. N., Boudewijns, Z. S. R. M., Van Der Knaap, N. J. F., Krugers, H. J. & Joëls, M. Maternal Care Received by Individual Pups Correlates with Adult CA1 Dendritic Morphology and Synaptic Plasticity in a Sex-Dependent Manner. *Journal of Neuroendocrinology* 24, 331–340 (2012).
84. McCarthy, M. M. Multifaceted origins of sex differences in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, (2016).

Chapter 5

Effects of early life stress on biochemical indicators of the dopaminergic system: a 3 level meta-analysis of rodent studies

Neurosci and Biobehav Rev. (2018), doi: 10.1016/j.neubiorev.2018.09.003

Bonapersona V [1]
Joëls M[1,2]
Sarabdjitsingh RA[1]

1 Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, The Netherlands
2 University of Groningen, University Medical Center Groningen, The Netherlands

Abstract

Adverse early life events are a well-established risk factor for the precipitation of behavioral disorders characterized by anomalies in the dopaminergic system, such as schizophrenia and addiction. The correlation between early life conditions and the dopaminergic system has been causally investigated in more than 90 rodent publications. Here, we tested the validity of the hypothesis that early life stress (ELS) alters dopamine signaling by performing an extensive 3-level mixed effect meta-analysis. We included several ELS models and biochemical indicators of the dopaminergic system in a variety of brain areas, for a total of 1009 comparisons. Contrary to our expectations, only a few comparisons displayed a significant effect. Specifically, the striatal area was the most vulnerable, displaying decreased dopamine precursor and increased metabolites after ELS. To make all data openly accessible, we created MaDEapp, a tool to explore data of the meta-analysis with the intent to guide future (pre)clinical research and allow power calculations. All in all, ELS induces a few yet robust changes on biochemical indicators of the dopaminergic system.

Introduction

During the perinatal period, the brain matures and is rapidly wired[1], rendering it particularly vulnerable to negative life experiences that might lastingly impact brain function and behavior[2]. This may contribute to the well-established observation that exposure to adverse conditions during childhood is a major risk factor for the later development of psychopathologies[3], including schizophrenia and substance abuse[4–7].

Prevailing evidence highlights that the dopamine (DA) system may be a prime candidate in mediating the influence of adverse events early in life on vulnerability to psychopathology[8]. The DA system develops early during the embryonic period, matures throughout adolescence, and forms stable patterns during young adulthood[9]. This prolonged development provides an extensive window of time in which adverse conditions early in life can tip the balance towards dysfunction[9]. Indeed, alterations in this system have been consistently associated with mental disorders (for a review: [9]). For example, genetic variations of the DA degradation enzyme COMT are associated with schizophrenia and bipolar disorder as well as an increased risk for psychosis, autism and anxiety[9]. In line, the DA receptor[2] is a major target for antipsychotics.

Overall, the associative studies in humans have led to the assumption that childhood adversities result in developmental alterations of the dopaminergic system. To investigate causality, preclinical studies using animal models have adopted behavioral early life stress (ELS) paradigms to mimic negative childhood conditions, aiming to understand the neurobiological substrate by which ELS adds to the development of DA system dysfunction. Although extensive, the existing literature is quite heterogeneous: it uses disparate models and outcome measures, and each study focuses on only a limited number of variables; moreover, preclinical studies are frequently underpowered[10]. The resulting findings are rather incoherent and difficult to interpret. This limitation hinders our understanding of the entire biological system and its development, and delays translational applicability.

To overcome these limitations, we performed a meta-analysis, a powerful method still sparsely applied to preclinical research which allows to systematically synthesize the scientific knowledge of a specific topic. Recent advances in the field of statistics such as the 3-level approach[11,12] along with their implementation in R packages[13,14] now enable researchers to use more sophisticated and robust methodology when analyzing meta-data. This method allows to include multiple data-points from a single study (nesting), without necessarily knowing their (often unreported) covariance. Ultimately, this substantially increases the flexibility of meta-analysis applications and improves the validity of the conclusions drawn.

Here, we aimed to investigate whether preclinical studies support an effect of ELS on dopaminergic signaling. We included diverse types and timings of ELS models (Fig 1), and we operationalized the dopaminergic system by quantifying several biochemical markers in mice and rats (Fig 2), across brain areas (Fig 3), considering possible confounders.

We determined whether the quality of the studies affected the estimation of the results. To make our knowledge readily available to others, we organized all information in a freely accessible open-source dataset and created a user-friendly web-app as a tool to guide future (pre)clinical research (e.g. power analysis calculation), thereby avoiding unnecessary replication and limit animal experimentation.

Materials and Methods

The review adhered to SYRCLE (Systematic Review Center for Laboratory animal Experimentation) guidelines for protocol[15], search strategy[16], and risk of bias assessment[17].

Theoretical definitions and assumptions

We defined as *individual comparison* each test performed within a published study between a control group and an experimental group with a history of ELS. As often occurs in experimental studies[18], multiple outcomes (*individual comparisons*) were collected from the same groups of animals (*nesting*).

We defined as *experiment* the ensemble of outcome measures from the same animals. According to this definition, each published study can report multiple

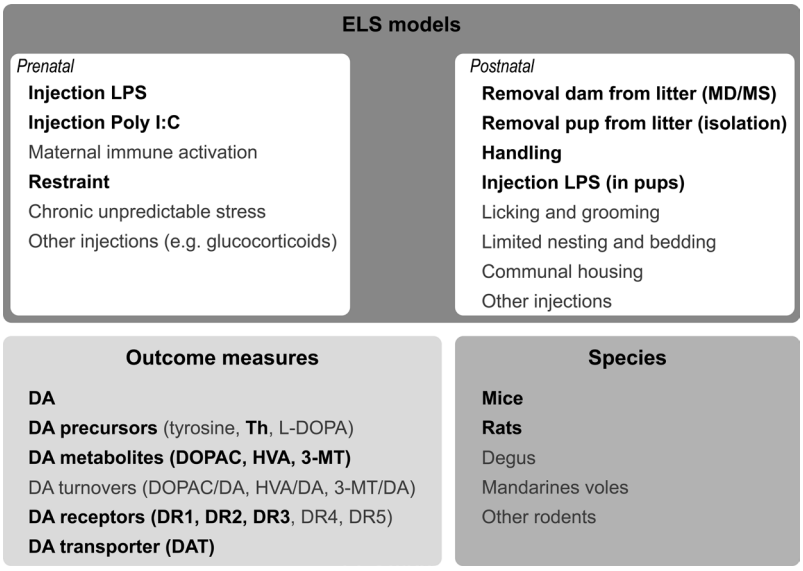


Figure 1. Search string and inclusion criteria. Graphical representation of the three main components of the search string. Items highlighted in bold were ultimately included in the analysis; other items were not included in the final analysis as they were reported in only a limited amount of publications (< 4 comparisons from 3 different publications from at least two different laboratories, see also Methods).

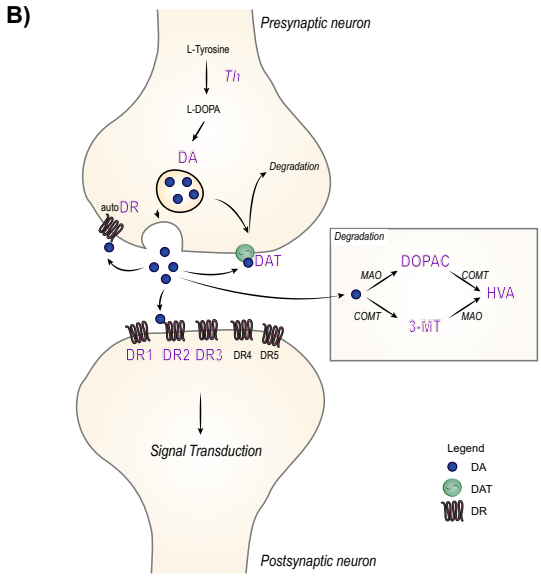


Figure 2. Graphical representation of signaling pathway. Dopamine is synthesized by the enzyme tyrosine hydroxylase (Th). When released in the synaptic cleft, DA can 1) bind post-synaptic receptors (DR1-DR5), 2) bind auto-receptors, 3) bind dopamine transporters (DAT), 4) be converted to the metabolites DOPAC, 3-MT and HVA by the action of the enzymes *MAO* and *COMT*[151]. Items in large (purple) font were included in the meta-analysis.

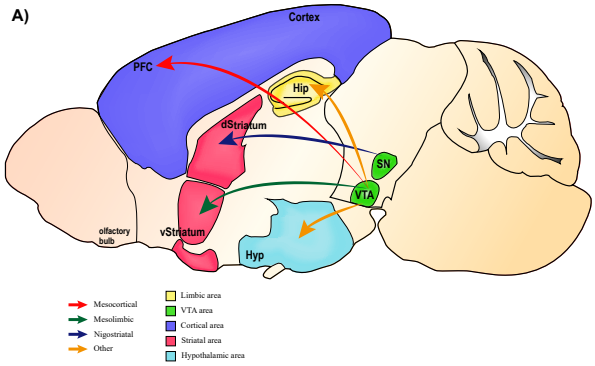


Figure 3. Graphical representation DA system projections. DA neurons are mainly situated in the midbrain, and can be subdivided with respect to their projection site[128]. In particular, DA neurons define separate populations of neurons that project to specific brain regions[152,153]. The major DA pathways are 1) mesocortical pathway, which defines projections from the VTA to the prefrontal cortex (PFC); 2) mesolimbic, from VTA to limbic system; 3) and nigrostriatal pathway, from substantia nigra (SN) to dorsal striatum, caudate nucleus and putamen. Other projections connect VTA to the hypothalamus, hippocampus and amygdala. Besides hosting dopaminergic neurons, these brain areas are involved in the feedback response to stress[154].

experiments when conducted on different sets of animals. Similarly, experiments conducted on different sets of animals could potentially be reported in separate publications. For these reasons, we nested multiple individual comparisons belonging to the same animals within the same experiments, but considered experiments from the same publications as independent from each other.

Search strategy

A comprehensive literature search was conducted regarding *the effects of early life stress on biochemical indicators of dopaminergic signaling* on February 14th 2017. The search string was composed by the factors "dopamine", "early life stress" and "rodents" (Fig 1 and supplementary appendix S1). The search was conducted on the electronic databases PubMed (www.pubmed.com) and Web of Science (www.webofknowledge.com). For a flow chart of the entire methodology, see Fig 4.

Studies' titles and abstracts were screened, and selected if the inclusion criteria were met (supplementary Table S2-1). In case of doubt, the full text was inspected. Eligible studies were evaluated by two independent reviewers (VB and RAS).

Study characteristics

Selection and data extraction To limit subjectivity in the data gathering and entry process, data extracted from eligible studies were recorded in a standardized database[19]. The following information was included: species, strain, sex, type and timing (relative to age) of the ELS model, outcome, time (relative to age) of outcome, technique used for outcome, brain area investigated, number of animals used, mean, standard deviation (SD) and standard error of the mean (SEM). If only SEM was reported, SD was calculated as SEM*√n, where n = amount of animals per group. If number of animals were reported as a range (e.g. 6-8 animals per group), we used the mean of this number (e.g. 7 animals per group). If a single control group was used to compare experimental groups in which ELS was induced with different models (e.g. handling and maternal deprivation), the sample size of the control group was equally divided as control for each experimental group (e.g. n=10 in not handled control becomes n=5 for control of maternal deprivation and n=5 for control of handling)[20].

When the data was not reported numerically in the publication, we contacted two authors per manuscript. If no answer was received within three weeks and after a reminder, the authors were considered not reachable. Only 5 out of 56 contacted authors replied to our request. Given the low response rate, we estimated most of the data presented only in graphs with Ruler for Windows (https://a-ruler-for-windows.en.softonic.com/). We tested the accuracy of this method by comparing effect sizes calculated from either supplied data or evaluated with the ruler, and verified that they were highly correlated (R2 = 0.74, supplementary Figure S2-1).

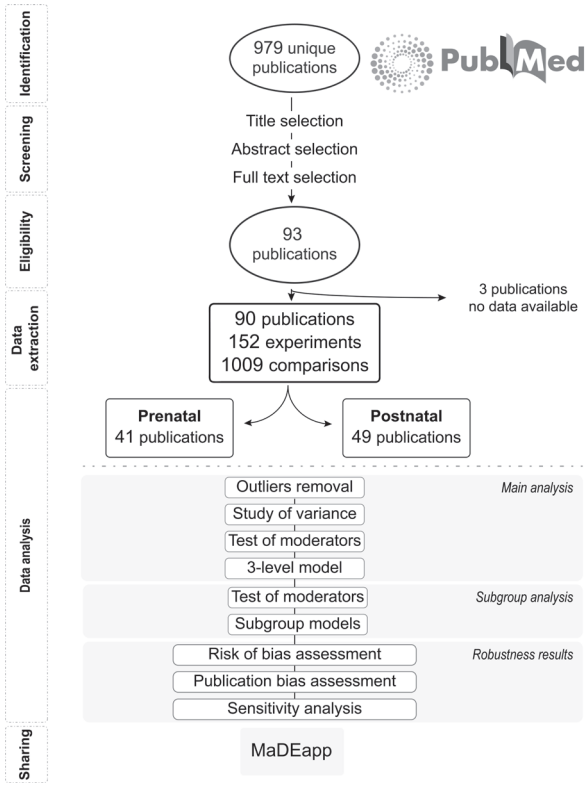


Figure 4. Flow-chart of study selection and analysis.

Concerning metabolites, some papers reported either concentrations, turnovers or both. In 97.5% of cases it was possible to calculate concentrations from turnovers with Pythagoras. Since concentrations and turnovers are related to the same information (though not identical), only concentrations were included in the analysis in order to avoid multi-collinearity. Turnover data-points are available in the MaDEapp (Meta-Analysis on Dopamine and Early life stress) for consultation.

Assessment of risk of bias in included studies

We used the SYRCLE tool to assess the risk of bias (supplementary Table S2-2)[17]. The criteria are based on the possible presence of selection bias (items 1, 2 and 3), performance bias (items 4 and 5), detection bias (items 7, 8, 9) and attrition bias (item 10). Furthermore, we added the item "quality of control" (item 6) to the category performance bias.

Since poor reporting of experimental details plays a role in heightening the quantified risk of bias, lack of reporting was scored as unclear risk of bias.

For quantitative inclusion in the analysis, amount of potential bias was operationalized by summing the risk of bias of each item according to the definition: "yes" = 0, "unclear" = 1, "no" = 2. This produced a continuous variable of integer increment between 0 (no risk bias) and 20 (maximum risk of bias), which was then scaled (*mean* = 0) to interpret the studies as of average risk of bias.

Data synthesis and statistical analysis

Effect size We estimated the effect size for each individual comparison with *escalc* (R package *metafor*) as standardized mean difference with Hedge's G method, which includes a correction factor for small sample sizes[20].

Study of heterogeneity Heterogeneity was tested with Cochran Q-test[21] and I^2 statistics[11]. Study of the distribution of variance was conducted for models without moderators to determine how much variance could be attributed to differences between effect sizes within experiments (level 2) and to differences between experiments (level 3). Substantial distribution of heterogeneity at these levels further encouraged the use of moderators.

Model We used a 3-level mixed effect model, which accounts for the anticipated heterogeneity of the studies as well as the dependency of effects within experiments. In our experimental design, the 3 levels correspond to variance of effect size between 1) animals, 2) outcomes and 3) experiments.

Since pre- and post-natal models act on times of development that are particularly disparate regarding the array of environmental factors, we considered them as different datasets and consequently analyzed them separately.

Effect sizes were considered outliers if their z score was above +3.29 or below -3.29[22], and removed from the analysis.

Since we hypothesized that the effect of ELS on DA signaling may not be evident from an overall estimate, we defined a priori possible moderators of this effect. These belonged to two different categories: biological and technical. The biological moderators were: outcome measure used (e.g. DA and metabolites), brain area investigated, sex, species, age as a continuous variable, and whether the outcome was at a RNA level/protein level/functional (referred to as method of assessment). Specific regions within the brain areas were investigated only in subgroup analysis due to the limited amount of observations. We considered the type of ELS model and amount of potential bias as technical moderators. These moderators may not underlie a biological difference, but can nevertheless explain heterogeneity across studies. The postnatal ELS model 'handling' has been reported repeatedly to cause effects in the opposite direction to those induced by other ELS models[2]. We therefore multiplied each calculated effect size for handling by -1[20], so that the overall estimate would be in the same direction. We

verified that the same conclusions would have been drawn if handling was excluded as a model (supplementary Figure S2-2).

To avoid multicollinearity among moderators, we firstly assessed each biological moderator univariately. We set the significance level at $p<0.10$ to test whether a moderator significantly reduced the previously quantified heterogeneity. A less restrictive p-value was chosen to assure the inclusion of moderators that have a multivariate but not univariate effect[23]. Only interactions with at least 4 comparisons from 3 different publications from at least two different laboratories were quantitatively assessed. This explains why some of the keywords in our search string were not included in the final analysis (Fig 1).

Subgroup analysis As the 3-level models revealed significant heterogeneity, we conducted subgroup analyses to further investigate its source. In particular, we tested the influence of the technical moderators (type of ELS model and amount of potential bias) as well as of the brain regions within the brain areas previously described (Fig 3), in subsets of the dataset with sufficient observations. For information on the type of ELS model used, please see supplementary appendix S2-1.

Sensitivity analysis and publication bias According to the standards of meta-analyses, we should investigate the robustness of our effect sizes by performing analysis only on those studies that were blinded and randomized. Unfortunately, the amount of blinded and randomized studies was insufficient to proceed with this approach. As an alternative, we performed the analysis by including the amount of potential bias as a moderator. The results of this sensitivity analysis should be interpreted as the effects of ELS on biochemical markers of the dopaminergic system on studies of average risk of bias.

To detect publication bias, funnel plots' asymmetry for each outcome variable was qualitatively evaluated. To the best of our knowledge, there are no available methods to quantify missing data (due for example to publication bias) in a multi-level setting[12]. Nonetheless, we evaluated publication bias with Egger's regression[24]. However, these results should be interpreted with caution as they are not based on a 3-level model. Lastly, we excluded those studies responsible for funnel plot asymmetry and conducted sensitivity analysis on the remaining dataset in the attempt to evaluate the influence of publication bias in the meta-analysis.

Data are presented as Hedge's G and 95% C.I. Data analysis was conducted with the computer program R (version 3.5.1)[25], with the aid of the following R packages: 1) *metafor*[13] for conducting the analysis, 2) *ggplot2*[26] for graphics, and 3) *shiny*[27] to create the MaDEapp.

Results

Study selection and characteristics

Study selection and data extraction The process of study selection is illustrated in the flow chart (Fig 4). The search string identified a total of 979 unique research papers. Statistical measurements (e.g. mean, SD and N) for quantitative analysis were extracted from 90 peer-reviewed publications that met our pre-specified inclusion criteria as described in the methods section. Three publications[28–30] were excluded from the analysis as it was not possible to extract nor infer any statistical measurement. Similarly, information was lacking from 23 comparisons of three other publications[31–33].

The included studies dated between 1996 and 2016, used ~2600 animals yielding a total of 1009 comparisons from 152 experiments. The publications were analyzed in two separate datasets, respectively using prenatal (41 publications) and postnatal (49 publications) ELS models. For a summary of experimental characteristics across studies see supplementary Table S3-1.

Four observations of the prenatal dataset (striatal Th[34], striatal DA[35], striatal HVA[36], striatal DR2[37]) and 6 observations of the postnatal dataset (striatal DR1[38], striatal DR2[39], DAT in the VTA area[40], striatal DOPAC[32], cortical DA[41] and limbic HVA[42]) were excluded from the analysis as outliers.

List of included publications The publications included in the analysis were[33,34,36,37,42–127].

Meta-analysis: prenatal ELS

Heterogeneity Substantial heterogeneity was recorded in the prenatal dataset ($Q(378) = 954.969$, $p < 0.001$), indicating that our search string identified a diverse range of experiments. In particular, 34.7% of variance could be attributed to within-sampling variance, 18.7% to within-experiment variance and 46.6% to between-experiment variance. These results suggested the use of moderators.

Moderators Potential moderators (supplementary Table S3-2) – such as brain area, sex, and species – were selected prior the beginning of the study based on hypotheses of the ELS literature.

As identified with univariate analysis of potential moderators (supplementary Table S3-3), outcome measure ($F(7, 371) = 3.956$, $p<0.001$) and brain area investigated ($F(4, 374) = 6.144$, $p<0.001$) were significant moderators in the prenatal dataset, explaining 4.3% and 6% of variance respectively. There was no detectable moderating effect of sex, species, age used, and method of assessment (RNA, protein or functional level).

Model The moderators that were univariately identified were included in the 3-level model to investigate the effects of ELS on markers of dopaminergic signaling. We

hypothesized that the effects were dependent on the outcome analyzed and that they differed across brain areas.

Of the 20 interactions between outcome measure and brain area with enough data-points, only 2 reached statistical significance (Fig 5). For a summary of all interactions, see supplementary Table S3-4. In particular, in the striatal zone, *Th* was decreased (Hedge's G(se) = -1.164(.295), p<.001, supplementary Fig S3-1) while DOPAC was increased (Hedge's G(se) = .323(.136), p = .018, Fig 6A) following prenatal ELS.

Fig6B shows a cumulative forest plot of striatal DOPAC to exemplify that the chronological combining of the experiments shows consistency since 2010, and that subsequent experiments have not contributed to the direction nor the size of the effect. The cumulative forest plot does not correct for the multi-level structure of the model.

The interaction between outcome measure and brain area explained 22.2% of variance in the prenatal dataset. We identified 17 interactions with enough comparisons to further address heterogeneity (Q(346) = 742.97, p<.001).

Subgroup analysis Subgroup analysis was used to further investigate the unexplained heterogeneity deriving from the 3-level model. The type of ELS model and sub-brain areas were univariately evaluated as potential moderators for each interaction between outcome measure and brain area.

In the prenatal dataset, 15 interactions had enough observations to be considered for further subgroup analysis. Of these, 5 had a significant test of ELS model as moderator (supplementary Table S3-5), namely DOPAC and D1R in the striatal area, DA in the cortical area, and *Th* in the VTA area.

Subgroup analysis revealed that injection of LPS or Polyl:C has consistently different effects than maternal restraint. In particular, LPS and Polyl:C significantly increased DOPAC in the striatum (LPS: Hedge's G(se) = 0.941(.229), p<0.001; Polyl:C: Hedge's G(se) =0.608 (0.238), p=0.016), while restraint did not (Hedge's G(se)= 0.012(0.207), p = 0.956, Fig 6A). Conversely, restraint decreased *Th* in the VTA area (Hedge's G(se) = -0.85(0.348), p=0.03) whilst LPS injection did not (LPS: Hedge's G(se) = 0.026(0.188), p =0 .889, supplementary Fig S3-2). Concerning cortical DA, the effects of Polyl:C and restraint had opposite directions but did not reach statistical significance, whilst the LPS model could not be quantitatively evaluated. Concerning D1R, no significant effects of the subgroup analyses were recorded.

Sub-brain area was a significant moderator only for DA in the striatal area (including nucleus accumbens, dorsal and central striatum and nucleus caudatus; supplementary Fig S3-3). In particular, DA was increased in the nucleus accumbens after ELS (Hedge's G(se) = 0.392 (0.159), p = 0.01) but it was unaffected in other parts of the striatum (Hedge's G(se) = -0.122 (0.146), p = 0.40). For a summary of all subgroup analyses, see supplementary Table S3-6.

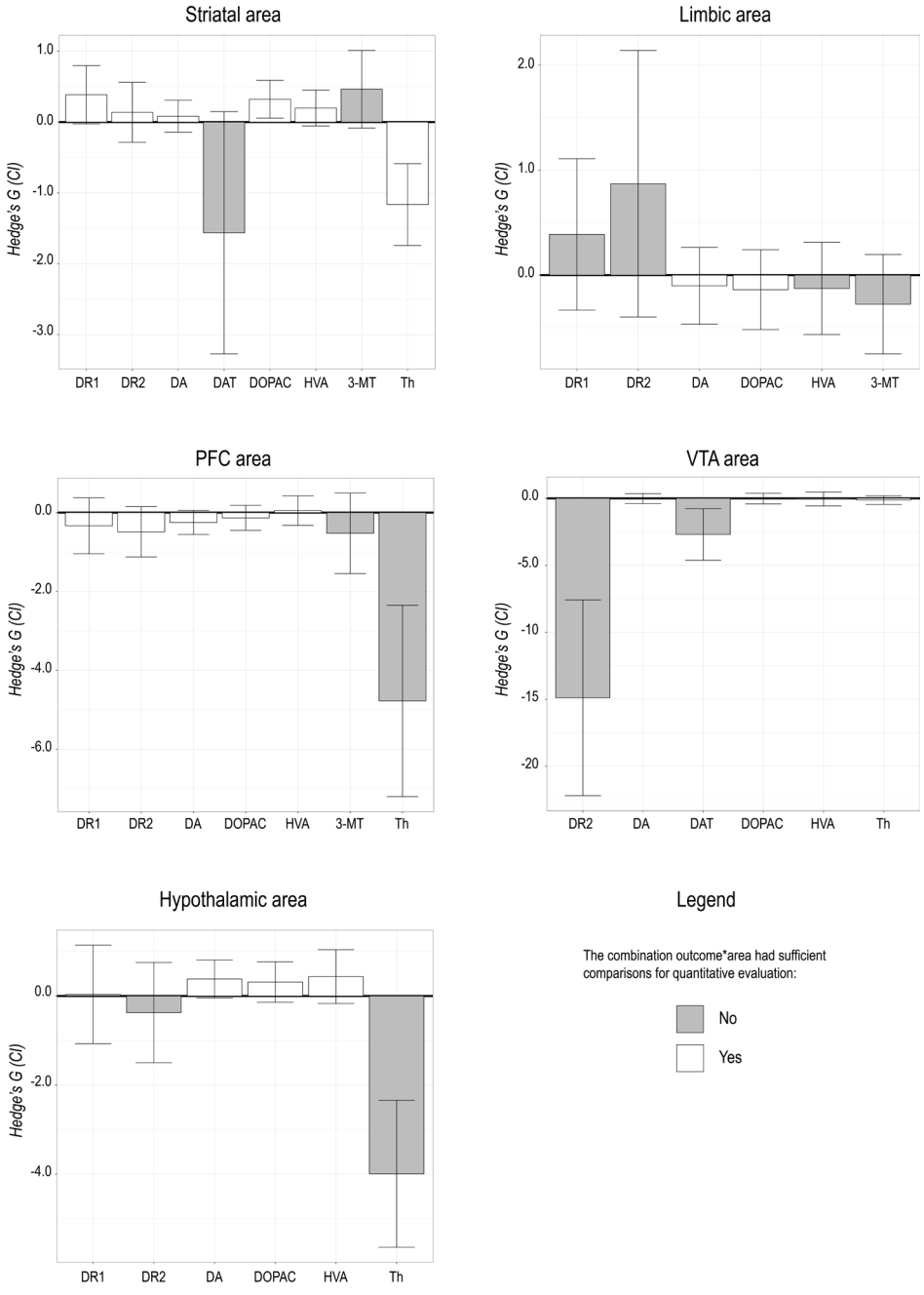
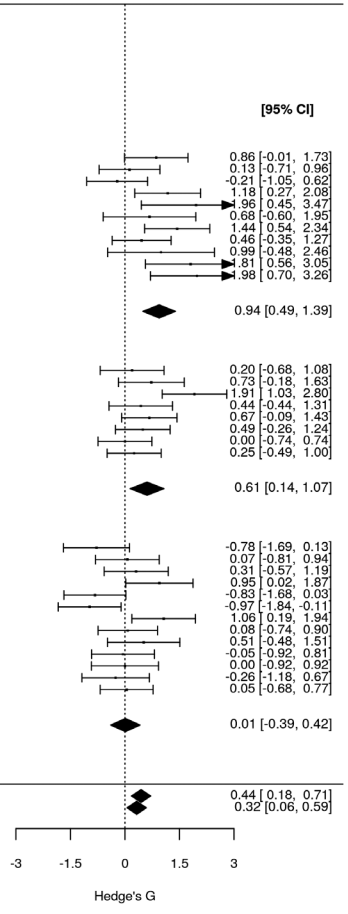


Figure 5. Summary of effects: prenatal dataset. Boxplot representing the summary of effect estimates for every combination between outcome variable (biochemical markers) and brain area. White bars = enough comparisons for meaningful quantification (rule of thumb: at least 4 comparisons from 3 papers from 2 research groups), black bars = number of comparisons insufficient for analysis.

prenatal striatal DOPAC

Study	Sex	Method	Area	Control N	Control Mean	Control SEM	Experimental N	Experimental Mean	Experimental SEM	[95% CI]
1. LPS										
Romero 2010 _ 3.2	M	protein	striatum	11	40.7	4.3	11	60.7	8.5	0.86 [-0.01, 1.73]
Romero 2010 _ 3.1	M	protein	NAc	11	16.8	2.9	11	17.9	2.1	0.13 [-0.71, 0.96]
Romero 2010 _ 2.2	M	protein	striatum	11	43.6	5.8	11	40	3.7	-0.21 [-1.05, 0.62]
Romero 2010 _ 2.1	M	protein	NAc	11	21.9	3.9	11	47.4	8	1.18 [0.27, 2.08]
Romero 2010 _ 1.2	M	protein	striatum	5	81.6	6.1	5	113.7	7.1	1.96 [0.45, 3.47]
Romero 2010 _ 1.1	M	protein	NAc	5	57.6	9.4	5	71.2	6.6	0.68 [-0.60, 1.95]
Romero 2007 _ 1.2	M	protein	striatum	12	118	12	12	175	10	1.44 [0.54, 2.34]
Romero 2007 _ 1.1	M	protein	NAc	12	138	23	12	172	18	0.46 [-0.35, 1.27]
Delattre 2016 _ 2	M	protein	striatum	4	152	4	4	174	13	0.99 [-0.48, 2.46]
Basta 2011 _ 2	M	protein	striatum	7	1229.2	23.3	7	1409.3	44.1	1.81 [0.56, 3.05]
Basta 2011 _ 1	F	protein	striatum	7	1049	54.3	7	1313.7	39	1.98 [0.70, 3.26]
Subgroup: LPS										0.94 [0.49, 1.39]
2. Poly I:C										
Winter 2009 _ 1.2	M	protein	NAc	10	295	20	10	308	20	0.20 [-0.68, 1.08]
Winter 2009 _ 1.1	M	protein	caudate	10	304	27	10	362	21	0.73 [-0.18, 1.63]
Ozawa 2006 _ 2	P	protein	striatum	12	89	16	17	163	4	1.91 [1.03, 2.80]
Luchicchi 2016 _ 1	M	function	NAc	9	1627	236	12	2015	274	0.44 [-0.44, 1.31]
Bitanihirwe 2010 _ 2.2	F	protein	NAc	14	1.02	0.16	14	2.1	0.57	0.67 [-0.09, 1.43]
Bitanihirwe 2010 _ 1.2	M	protein	NAc	14	1.07	0.14	14	1.74	0.48	0.49 [-0.26, 1.24]
Bitanihirwe 2010 _ 2.1	F	protein	caudate	14	2.21	0.3	14	2.21	0.43	0.00 [-0.74, 0.74]
Bitanihirwe 2010 _ 1.1	M	protein	caudate	14	0.96	0.12	14	1.05	0.05	0.25 [-0.49, 1.00]
Subgroup: Poly										0.61 [0.14, 1.07]
3. restraint										
Weston 2014 _ 2.4	F	protein	striatum	10	114	10	10	84	13	-0.78 [-1.69, 0.13]
Weston 2014 _ 2.3	F	protein	NAc	10	136	15	10	139	12.5	0.07 [-0.81, 0.94]
Weston 2014 _ 1.4	M	protein	striatum	10	117	10	10	130	15	0.31 [-0.57, 1.19]
Weston 2014 _ 1.3	M	protein	NAc	10	83	9	10	118	13	0.95 [0.02, 1.87]
Weston 2014 _ 2.2	F	protein	striatum	12	40.6	4.2	11	30.6	2.2	-0.83 [-1.68, 0.03]
Weston 2014 _ 2.1	F	protein	NAc	12	40.1	3.1	11	25	5.6	-0.97 [-1.84, -0.11]
Weston 2014 _ 1.2	M	protein	striatum	12	30.16	3.3	11	50.53	7.2	1.06 [0.19, 1.94]
Weston 2014 _ 1.1	M	protein	NAc	12	25.5	2.6	11	26.45	4.2	0.08 [-0.74, 0.90]
Gerardin 2005 _ 1	M	protein	striatum	8	691.6	124.76	8	868.6	104.22	0.51 [-0.48, 1.51]
Cory 2013 _ 3	M	protein	striatum	12	18430	333.13	9	18352	562.21	-0.05 [-0.92, 0.81]
Cory 2009 _ 2	M	protein	NAc	9	54	11	9	54	9	0.00 [-0.92, 0.92]
Cory 2009 _ 1	F	protein	NAc	9	95	18	9	82	14	-0.26 [-1.18, 0.67]
ALONSO 1994 _ 1	F	protein	striatum	15	0.41	0.11	14	0.43	0.11	0.05 [-0.68, 0.77]
Subgroup: restraint										0.01 [-0.39, 0.42]
Univariate model										0.44 [0.18, 0.71]
Effect 3-level										0.32 [0.06, 0.59]



B

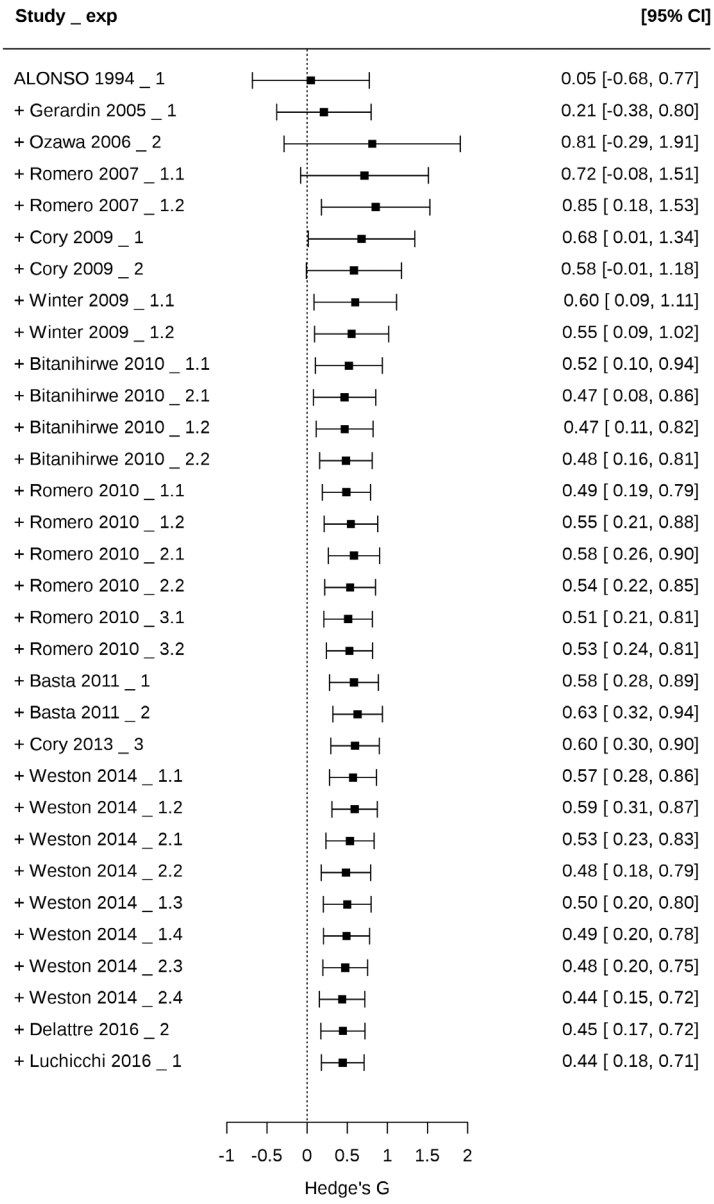


Figure 6. Striatal DOPAC in the prenatal dataset. (A) Forest plot. Results of univariate and 3-level meta-analysis are reported (bottom diamonds), as well as subgroup analysis on the ELS model used. (B) Cumulative forest plot. This plot displays the accumulation of evidence over time as the individual comparisons are added in chronological order. M = males, F = females, Method = method of assessment, hip = hippocampus, N = amount of animals, SEM = standard error of the mean, CI = confidence interval. Following the name of the study, __n represents the number of the experiment, .n = represents which comparison within that experiment.

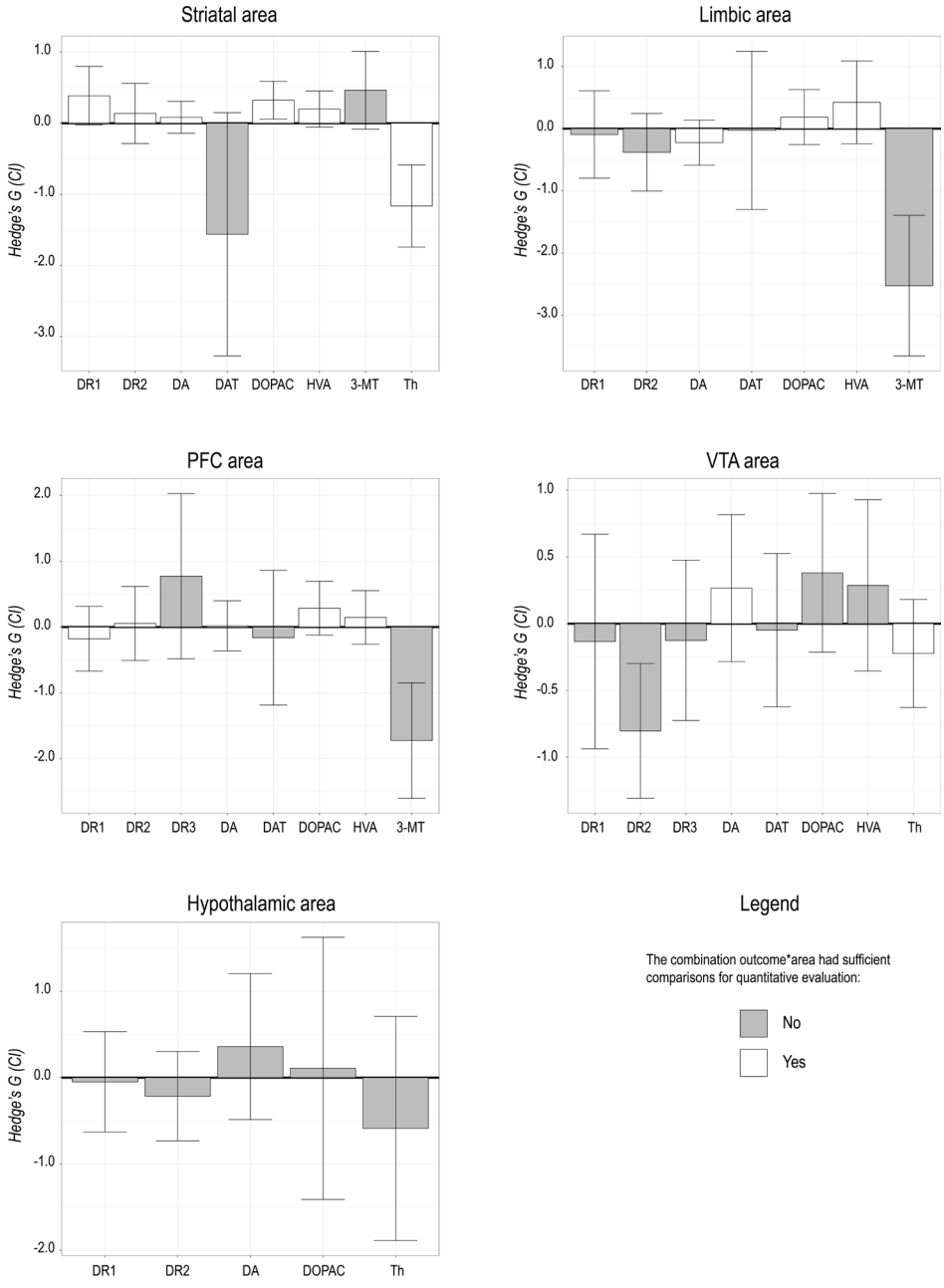


Figure 7. Summary of effects: postnatal dataset. Boxplot representing the summary of effect estimates for every combination between outcome variable (biochemical markers) and brain area. White bars = enough comparisons for meaningful quantification (rule of thumb: at least 4 comparisons from 3 papers from 2 research groups), black bars = number of comparisons insufficient for analysis.

Meta-analysis: postnatal ELS

Heterogeneity Our search identified a diverse range of experiments in the postnatal dataset, as shown by the substantial heterogeneity recorded (*Q*(*381*) = *1061.278, p<.001*). In particular, 33.1% to within-sampling, 46.9% to within-experiment, and 20.1% to between-experiment variance.

Moderators Univariate analysis of potential moderators (supplementary Table S3-7, p-value significance set at <.10) identified outcome measure (*F*(*8, 373*) = *9.139, p <0 .001*) and brain area investigated (*F*(*4,377*) = *2.035, p =0 .089*) as significant moderators, explaining 10.3% and 0.8% of variance respectively. Other moderators, such as species and sex of the animal, had no detectable moderating effect.

Model The moderators that were univariately identified were included in the 3-level model to investigate the effects of ELS on markers of dopaminergic signaling. We hypothesized that the effects were dependent on the outcome analyzed and that they differed across brain areas.

Of the 17 interactions with sufficient comparisons, 3 reached statistical significance (Fig 7, supplementary Table S3-8). In particular, in the striatal zone, DOPAC (*Hedge's G*(*se*) = *0.541*(*0.207*), p = 0.009, Fig 8), HVA (*Hedge's G*(*se*) = *0.555*(*.22*), p=0.012, supplementary Fig S3-4) and DA (*Hedge's G*(*se*) = *0.307*(*0.147*), p = .038, supplementary Fig S3-5) were increased.

The interaction between outcome measure and brain area explained 15.3% of variance in the postnatal dataset. We identified 16 interactions with enough comparisons to further address heterogeneity (*Q*(*345*) = *898.4, p<0.001*).

Subgroup analysis The moderator effects of the ELS model used and sub-brain areas were evaluated with a subgroup analysis.

In the postnatal dataset, 16 interactions had sufficient observations to be further analyzed. Of these, 7 revealed a significant impact of the ELS model used (supplementary Table S3-9): HVA, DOPAC and DA in the cortical area, HVA and DOPAC in the striatal area, DOPAC in the limbic area, and DA in the VTA area.

Subgroup analysis showed that the effects of ELS model as moderator in striatal DOPAC (Fig 8) and HVA (supplementary Fig S3-4) were mainly due to handling. In particular, handling decreased HVA (*Hedge's G*(*se*) = *-0.778*(*0.295*), p = 0.03) as well as DOPAC (*Hedge's G*(*se*) = *-0.77*(*0.301*), p = 0.029) in the striatum, whilst separation of the mother from the pups had no effect (*HVA: Hedge's G*(*se*) = *0.08*(*0.227*), p = 0.735; *DOPAC: Hedge's G*(*se*) = *-0.205*(*0.239*), p = 0.411).

Sub-brain area was not a potential moderator in any of the interactions evaluated.

For a summary of all subgroup analyses, see supplementary Table S3-10.

Figure 8. Striatal DOPAC in the postnatal dataset. Forest plot. Results of univariate as well as 3-level meta-analysis are reported (bottom diamonds), as well as subgroup analysis on the ELS model used. * = effect sizes of handling were multiplied by -1 to maintain consistency directionality of the other models. A decrease in the graph identifies an increase in DOPAC in ELS animals. M = males, F = females, Method = method of assessment, hip = hippocampus, N = amount of animals, SEM = standard error of the mean, CI = confidence interval, MD = mother separated from the pups, LPS = injection of LPS. Following the name of the study, __n represents the number of the experiment, .n = represents which comparison within that experiment.

Sensitivity analysis

Sensitivity analysis was conducted to test the robustness of our findings. We examined whether the quality of the studies included had an impact on the interpretation of our results.

Quality of the studies: SYRCLE bias report No publication reported information on all SYRCLE potential bias items. Overall, of the 90 publications, 37 (41%) reported randomization sequence of the animals in the experiments, 3 (3.3%) random housing allocation, 59 (65%) random group allocation, 49 (54.4%) random selection of the animals (Fig 9). In 11 (12%) publications the caretaker were reported blinded to the experimental condition, in 20 (22.2%) the experimenters blinded. Handling of incomplete data was reported in 42 publications (46.7%). 11 studies (12.2%) did not report sufficient information to evaluate the quality of the control group. Only 11 studies yielding a total of 117 comparisons reported being blinded and randomized.

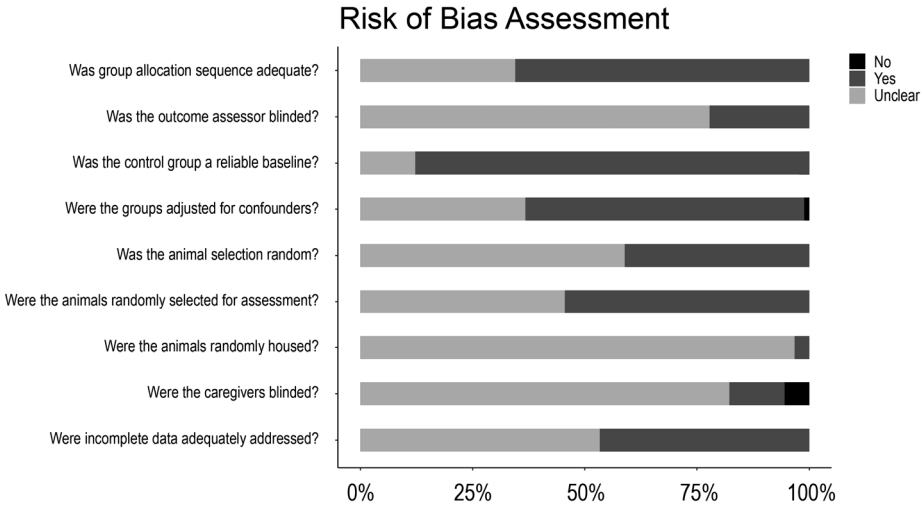


Figure 9. Risk of bias assessment. Each bar represents a different risk of bias item. Yes = measurements have been taken to avoid bias; no = no measurements were taken to avoid bias; unclear = not enough information provided in the paper to determine the risk of bias.

Sensitivity analysis for potential bias Since the amount of publications was insufficient to evaluate the robustness of our effects in a blinded and randomized dataset, we operationalized the amount of potential bias and performed the analysis again by including this factor as a moderator. Therefore, the results of this sensitivity analysis were interpreted as the effects of ELS on markers of DA signaling on studies of average bias.

The amount of potential bias was a significant moderator in the prenatal dataset ($F(1, 377) = 3.536$, $p = 0.061$); yet, this did not affect the qualitative interpretation of the meta-analysis.

In the postnatal dataset, the test of moderators for amount of potential bias was not significant ($F(1, 380) = 0.500$, $p = 0.480$). The interpretation of the results did not change, with the exception of DA in the striatal area, of which the effect size was decreased and the effect at a trend level (Hedge's G(se)=-0.289(0.15), $p = 0.057$).

Publication bias Due to the lack of methods to quantitatively evaluate publication bias in a multi-level setting, we qualitatively estimated the risk for publication bias with funnel plots (Fig 10). Publication bias was more pronounced in the prenatal than the postnatal dataset. The same conclusion was reached when performing Egger's regression (no multi-level regression models): there was evidence for publication bias in the prenatal (**$z = -5.014$, $p < 0.001$**) but not in the postnatal (**$z = -0.612$, $p = 0.54$**) datasets. The presence of publication bias in the prenatal dataset may indicate an overestimation of the reported effect sizes.

Furthermore, we conducted an analysis of influential cases by removing studies with large standard error as well as residual values. Since the results did not change qualitatively, publication bias was considered low-to-moderate.

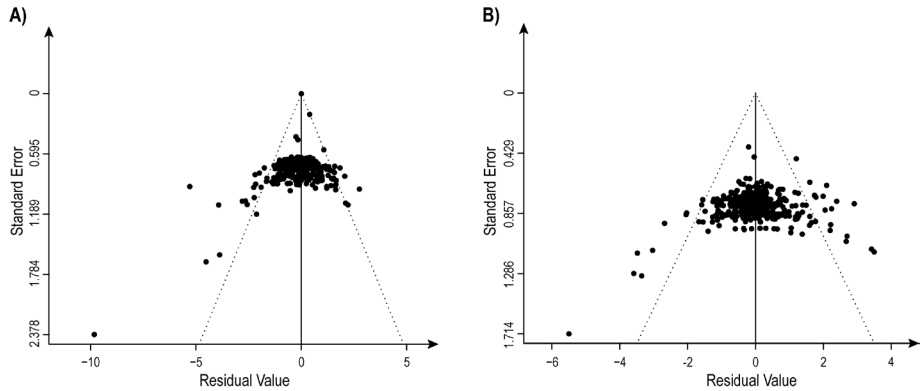


Figure 10. Funnel plots. Publication bias was evaluated by qualitatively assessing symmetry in funnel plot in the (A) prenatal and (B) postnatal datasets.

MaDEapp

Finally, we created a MaDEapp (https://vbonapersona.shinyapps.io/MaDEapp/), a web-based app with a user-friendly interface in which each researcher can perform his/her own meta-analysis on the topic of ELS and biochemical indicators of the dopaminergic signaling. The app offers the possibility to choose across a wide variety of options, such as outcome measures, brain areas, sex of the animals, type and timing

of the ELS model. Based on the characteristics indicated, the app reports forest, funnel and cumulative plots. The forest plot includes a 3-level effect estimate (Hedge's G and CI), which can be used for future power calculation.

For example, a researcher is interested in the effects of postnatal ELS on DR1 in the striatal area. In MaDEapp, the researcher selects the "postnatal" dataset, with "DR1" as outcome measure in the "striatal area". The resulting forest plot reports the estimated Hedge's G (CI) = -0.5 [-0.91, -0.1]. The estimated effect size is smaller than 0. From this exploration, the researcher hypothesizes that postnatal ELS decreases DR1 expression in the striatal area. The effect size -0.5 may be an overestimation of the real size of the effect due to potential (publication) bias. The researcher would then use an effect size of -0.45 for power calculation for his/her future experiments.

Discussion

Schizophrenia and addiction are examples of psychiatric disorders reported to be linked to DA dysfunction. Childhood trauma is a well-documented risk factor[3,8,9,128]. This clinical observation led to the hypothesis that the dopaminergic system mediates the risk of ELS. Although this link has been causally investigated in more than 90 rodent publications over 20 years, no consensus has yet been reached on the extent, directionality and specificity of this effect. Therefore, we performed a meta-analysis to question: Do rodent studies support long-lasting effects of ELS on biochemical indicators of the dopaminergic system? Overall, our results indicate that only a limited number of comparisons were significant suggesting that the effects of ELS on the dopaminergic system may not be apparent on a biochemical level at baseline conditions.

Methodological considerations

Dopaminergic signaling involves multiple interdependent elements (e.g. precursors, metabolites, receptors), which altogether contribute to the system as a whole. Data on these elements are sometimes gathered from the same animals, and are therefore dependent on each other[18]. In a meta-analysis setting, dependency implies overlap of information, which ultimately leads to an erroneous interpretation of the results [129,130]. To deal with this obstacle, several strategies have been adopted: from selecting only one effect size per study to ignoring the problem altogether [11]. Although sophisticated methods such as multivariate and multilevel analysis exist, these have the strong limitation that the needed covariance between the dependent effects is rarely reported in publications[23]. The 3-level approach that we used overcomes both limitations: it corrects for dependency of observations, without the use of covariances[12]. To the best of our knowledge, this approach has never been used before in rodent literature. Although this powerful and practical method was initially created for human studies[11],

its applicability in preclinical research is warranted due to the multiple-outcome nature of such studies. The method is already available in the R packages *metafor* and metaSEM.

Together with the application of a 3-level mixed effect meta-analysis to preclinical literature, we here promote the use of tools to facilitate data exploration and advocate open science. We created MaDEapp, a freely available user-friendly app that allows to run a tailor-made meta-analysis on ELS and the dopaminergic system, depending on the specific question one has. Each individual can select a set of characteristics (e.g. prenatal/postnatal models, sex, age). The app returns a forest plot in which the total univariate 3-level estimate is presented, as well as a funnel plot to evaluate publication bias. This can be used to generate hypotheses, evaluate estimated sample sizes for power analysis, and explore which outcomes/brain areas have received most attention and which did not. We believe this app is a useful tool to guide future research on the topic.

MaDEapp and the analysis here presented are complementary. Meta-analysis is a statistical test, and it is limited by the frequency (*power*) and quality (*potential bias*) of the published data. In our analysis there was no significant effect of postnatal ELS on DR1 in the striatum ($p = 0.053$); however, when we analyzed the same outcome univariately with MaDEapp, the confidence interval of the estimate did not include 0 and could be interpreted as "significant". This discrepancy may be due to a lack of power to confirm the effect in our analysis or due to an increased bias not corrected for in the more specific model used by the app. Therefore, non-significant results should be interpreted as lack of confirmation of an effect, not as evidence of no effect, since the meta-analysis could be underpowered to detect a specific marker in a specific brain area. Alongside, the use of the app should be intended as exploratory only and not as confirmatory.

Quality of the studies

Meta-analysis as a methodology is not simply the collection of statistical methods used to achieve integration of available evidence. Its power lies in the application of systematic scientific strategies to the literature review[131]. In addition to summarizing effects' estimates, it allows to evaluate the extent to which conclusions are at risk of bias.

In our analysis, surprisingly few studies (12%) reported being randomized as well as blinded. On the other hand, random allocation to group (41%) as well as blinded assessor (22.2%) was comparable[132] or better[133] than previous publications in neuroscience. Although it is likely that investigators did take measures to reduce bias, lack of their reporting induced an unclear risk[134,135] and hindered estimation of the value of the publications[134]. The importance of quality of reporting has been an emerging issue in preclinical research[134,135]. Despite the increased awareness, the

quality of reporting of the publications included in this meta-analysis has not improved since 2005 (supplementary Fig S4-1). Such evidence should encourage preclinical researchers as well as reviewers to adhere to reporting guidelines such as ARRIVE[134].

Although imprecise reporting does not necessarily imply poor study quality, underpowered experiments seriously hamper research interpretation[10]. From the reported amount of animals included per experiment, we back-calculated the power at the beginning of the study, assuming at least one true positive effect per publication. We performed this analysis considering small (Hedge's G = 0.5), medium (Hedge's G = 0.8) or large (Hedge's G = 1) effect sizes (supplementary Fig S4-2). When considering a large effect size, 391 comparisons (38.7%) had power below chance level, and only 63 (6.2%) had power >0.8, a cut-off value136 generally aimed at in preclinical research. Although 43 papers (47%) had at least one comparison with power >0.5, only 5 papers (5.5%) had at least one comparison with power >0.8. This means that the vast majority of the experiments was not sufficiently powered to reliably detect an effect – an (already dramatic) best-case scenario given that the percentages were calculated assuming that the studies compared only two groups (t-tests) as well as a large and truly existing effect. Future preclinical studies should be grounded in power calculations based on realistically estimated effects. Although for each single study the amount of animals will be larger, overall higher power will lead to more reliable, reproducible and therefore higher quality research.

ELS causes limited alterations on biochemical markers of the DA system

In our analysis, we evaluated the dopaminergic system by including numerous biochemical markers across brain areas as well as potential moderators. These gave rise to a myriad of viable comparisons. Despite the extent, only a handful of significant effects were identified, thereby suggesting that biochemical indicators of the DA system well adapt to ELS interventions. Clearly, we cannot exclude the possibility that other indicators of the DA system (e.g. electrophysiological parameters or behavioral tests that critically depend on DA function) would have yielded clearer results. This awaits future investigation.

Prenatal and postnatal ELS were treated separately because the prenatal environment differs substantially from that postnatally. Nonetheless, both datasets shared consistent findings. Specifically, the striatal area was the most vulnerable: following prenatal ELS, *Th* was decreased and DOPAC increased; while postnatal ELS caused an increased in DOPAC as well as HVA. These changes were stable and reliable: the analysis used is adequately conservative and robust, and the effects survived sensitivity analysis and publication bias corrections. The stability of the effects is also qualitatively substantiated by the cumulative plots, which operationalize how subsequent experiments update our knowledge of the previously estimated effect size. Our results display that these were durable over time, and that replication after the

initial 3-5 studies might not be very informative on these variables (except as a positive control in a study investigating another variable), as additional experiments did not alter the estimated effect. All in all, the sparse effects here reported are reliable and of medium size, suggesting that the system is damaged, which may in turn contribute to the vulnerability of ELS-dependent disorders.

The results can be interpreted as either hyper- or hypo-activation. For instance, the increase in metabolites can indicate an increase in the available amount of the substrate DA (hyperactivation) as well as an increased conversion rate causing less DA (hypoactivation). Similarly, the decrease in *Th* (precursor conversion enzyme) is not necessarily indicative of a decrease in DA function. Although our analysis suggests that postnatal ELS increases DA levels in the striatum, the effect is small in size and less robust that the other effects mentioned above. The mismatch between DA precursor and metabolism may suggest changes in the intermediate stage of DA conversion. For example, the DA converting enzyme COMT has been repeatedly shown to interact with ELS for the later development of psychiatric disorders[137–139]. On the other hand, L-DOPA – product of *Th* and precursor of DA – has been suggested to act as a novel transmitter itself or may have neurotropic functions, and thereby be transiently involved in perinatal developmental processes[140]. Since the interaction between ELS and L-DOPA has not been further investigated, the link remains circumstantial.

We defined a priori several factors established in preclinical literature to be potential moderators of ELS effects. To our surprise, species, sex and method of assessment were not significant moderators. Although males, mice and protein as method of assessment were the most described conditions, plenty of observations were present for all groups. Nonetheless, the lack of evidence for a moderator effect should not be interpreted as evidence for absence: mice/rats should be chosen according to standard practice, both sexes should always be considered[141], and there is substantial evidence that a decrease in RNA level does not automatically result in a decrease in protein and therefore in function, as e.g. shown in a systems approach[142].

Lastly, the unexplained variance may not only indicate methodological differences, but also underlie additional biological moderators, such as sub-brain areas or differences across hemispheres (lateralization).

Translational potential?

The translational applicability of preclinical studies depends on the understanding of psychopathological clinical and intermediate phenotypes. For example, ELS is a main risk factor for schizophrenia as well as substance abuse disorder. However, these diseases have opposite intermediate phenotypes: while schizophrenia is supposed to be characterized by hyperreactivity of the DA system although presumably to its afferent control[128], substance abuse may be linked to DA hypo(re)activation[143]. To what extent do ELS studies in rodents accurately model these two conditions?

Three factors currently limit answering this question. Firstly, ELS in humans is a complex concept, generally involving low socio-economic status, physical and/or psychological abuse, poor living conditions and high caloric food$_3$. Conversely, animal models are extremely controlled and standardized pre- and postnatally. Although this facilitates the definition of "traumatic early life" as well as the deriving caused effects, one can question its ecological validity. Secondly, the dual hypo- / hyper- interpretation of the ELS-induced phenotype in rodents prevents a whole-system level comparison, and restricts it to a micro level, focusing on a particular compound in a particular brain area. Thirdly, the DA-dependent changes in schizophrenia and addiction are most likely far more complex than the ones observed following ELS in rodents. For example, our analysis failed to confirm any effect of ELS on DA receptors. This was surprising, as changes in the availability of DA receptors is a consistent characteristic across different types of addictions[144] as well as in schizophrenia[145]. Although the discrepancy could partly be due to a power problem of the meta-analysis, these limitations challenge the reliability of ELS models for translational purposes, at least with regard to these specific aspects of the abovementioned human disorders. It cannot be excluded that more relevant models may become apparent in light of different ELS theories[146], for example after acute or chronic stress.

Lastly, although our study supports that ELS causes some changes in the DA system, these associations remain at a correlational level in humans and should be interpreted as such.

Limitations of the study

Despite our efforts to be as comprehensive as possible in the description of the effects of ELS on the DA system, we encountered several important limitations. Firstly, we investigated the DA system by evaluating the effects of ELS on biochemical markers. Although this provides a thorough conceptualization of the system, it does not supply a comprehensive functional evaluation. For example, the approach here reported is unable to operationalize DA innervations, projections and tone, nor potential epigenetic mechanisms. ELS has been reported to alter DR3-signalling and neuronal activity in the lateral septum[147]. Chronic stress in adulthood has been reported to change DA neurons' activity in a stressor-dependent manner[148]. These reports suggest that spontaneous activity, bursting and timing of dopaminergic firing may be susceptible to ELS action, yet they are not apparent from assessment of ligands, receptors and metabolites. Despite the high relevance of such measurements, these were excluded from the analysis as the publications on the topic are scarce and their integration not straightforward in a meta-analytic setting.

Secondly, the classification of "timing" of ELS to either prenatal or postnatal may be too reductionist, since neuronal circuits are shaped by experiences during critical periods of development of variable length (from days to years depending on the

species)[149]. The interested researcher can further explore this avenue by combining our dataset with RNA expression of *Th* or DA receptors found in the Allen Developing Mouse Brain Atlas[150]. Unfortunately, the literature so far published is insufficient to investigate how stress during specific postnatal days in which a certain RNA *X* is highly expressed uniquely alters its functioning later in life.

Thirdly, due to insufficient data-points per outcome per brain area in several cases, a meaningful quantitative estimation was not possible for all combinations of outcome*area. All currently available measurements are reported as supplementary material and can be further investigated via MaDEapp.

Fourthly, we included data only from published studies. Especially in the prenatal dataset, there is evidence of low-to-medium publication bias as qualitatively estimated with a funnel plot and sensitivity analysis, which may result in an overestimation of the effect sizes. Despite the robustness of our methodology, this limitation should be considered in future power calculations.

Lastly, we limited our analysis to baseline (i.e. unchallenged) conditions. Future studies should focus on conditions where the DA system is challenged, as ELS manipulations can interact with later life challenges to result in a pathological phenotype[146].

To conclude, ELS induces a few yet robust effects on biochemical indicators of the DA system, with – based on the currently available studies – the striatum being the brain area most affected. Although the changes observed can be interpreted as both hypo- and hyper- activation of the DA system, the effects were consistent across prenatal and postnatal ELS models, sex, species and method of assessment.

References

1. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Progress in Neurobiology* 106–107, 1–16 (2013).
2. Levine, S. Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology* 30, 939–946 (2005).
3. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience* 17, 652–666 (2016).
4. Agid, O. et al. Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar disorder and schizophrenia [see comments]. *Molecular Psychiatry* 4, 163–172 (1999).
5. Enoch, M.-A. The role of early life stress as a predictor for alcohol and drug dependence. *Psychopharmacology* 214, 17–31 (2011).
6. Dube, S. R. et al. Childhood Abuse , Neglect , and Household Dysfunction and the Risk of Illicit Drug Use : The Adverse Childhood Experiences Study. 111, (2003).
7. Scheller-Gilkey, G., Thomas, S. M., Woolwine, B. J. & Miller, A. H. Increased early life stress and depressive symptoms in patients with comorbid substance abuse and schizophrenia. *Schizophrenia bulletin* 28, 223–231 (2002).
8. Gatzke-Kopp, L. M. The canary in the coalmine: The sensitivity of mesolimbic dopamine to environmental adversity during development. *Neuroscience and Biobehavioral Reviews* 35, 794–803 (2011).
9. Money, K. M. & Stanwood, G. D. Developmental origins of brain disorders: roles for dopamine. *Frontiers in Cellular Neuroscience* 7, 1–14 (2013).
10. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14, 365–376 (2013).
11. Cheung, M. W. L. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling

approach. *Psychological Methods* 19, 211–229 (2014).
12. Assink, M. & Wibbelink, C. J. M. Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology* 12, 154–174 (2016).
13. Viechtbauer, M. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36, 1–48 (2010).
14. Cheung, M. W. L. metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology* 5, 1–7 (2014).
15. De Vries, R. B. M. et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclinical Medicine* 1, 1–9 (2015).
16. Leenaars, M. et al. A step-by-step guide to systematically identify all relevant animal studies. *Laboratory Animals* 46, 24–31 (2012).
17. Hooijmans, C. R. et al. SYRCLE ' s risk of bias tool for animal studies. *BMC Medical Research Methodology* 14, 1–9 (2014).
18. Aarts, E., Verhage, M., Veenvliet, J. V, Dolan, C. V & van der Sluis, S. A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience* 17, 491–496 (2014).
19. Bonapersona, V. Dataset: Effect of early life stress on biochemical indicators of the dopaminergic system. *Mendeley data* (2018) doi:10.17632/9x9fc3v26k.1.
20. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221, 92–102 (2014).
21. Moreira, P. S., Almeida, P. R., Leite-Almeida, H., Sousa, N. & Costa, P. Impact of Chronic Stress Protocols in Learning and Memory in Rodents: Systematic Review and Meta-Analysis. *Plos One* 11, e0163245 (2016).
22. Tabachnick, B. G. & Fidell, L. S. *Using multivariate statistics.* (Pearson Education,2013, 2013).
23. Hox, J. J. *Multilevel analysis: techniques and applications.* (Routledge, 2010).
24. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* (Clinical research ed.) 315, 629–34 (1997).
25. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2015).
26. Wickham, H. ggplot2: elegant graphics for data analysis. (2009).
27. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for R. (2017).
28. Arborelius, L. & Eklund, M. B. Both long and brief maternal separation produces persistent changes in tissue levels of brain monoamines in middle-aged female rats. *Neuroscience* 145, 738–750 (2007).
29. Kippin, T. E., Szumlinski, K. K., Kapasova, Z., Rezner, B. & See, R. E. Prenatal stress enhances responsiveness to cocaine. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 33, 769–782 (2008).
30. Kosten, T. A., Zhang, X. Y. & Kehoe, P. Neurochemical and behavioral responses to cocaine in adult male rats with neonatal isolation experience. *The Journal of pharmacology and experimental therapeutics* 314, 661–7 (2005).
31. Kikusui, T., Faccidomo, S. & Miczek, K. A. Repeated maternal separation: Differences in cocaine-induced behavioral sensitization in adult male and female mice. *Psychopharmacology* 178, 202–210 (2005).
32. Ognibene, E. et al. The effect of early maternal separation on brain derived neurotrophic factor and monoamine levels in adult heterozygous reeler mice. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 32, 1269–1276 (2008).
33. Kirsten, T. B. et al. Hypoactivity of the central dopaminergic system and autistic-like behavior induced by a single early prenatal exposure to lipopolysaccharide. *Journal of Neuroscience Research* 90, 1903–1912 (2012).
34. Baier, C. J. et al. Intrastriatal 6-OHDA lesion differentially affects dopaminergic neurons in the ventral tegmental area of prenatally stressed rats. *Neurotoxicity research* 26, 274–84 (2014).
35. Basta-Kaim, A. et al. Prenatal lipopolysaccharide treatment enhances MK-801-induced psychotomimetic effects in rats. *Pharmacology Biochemistry and Behavior* 98, 241–249 (2011).
36. Ling, Z. D. et al. Combined toxicity of prenatal bacterial endotoxin exposure and postnatal 6-hydroxydopamine in the adult rat midbrain. *Neuroscience* 124, 619–628 (2004).
37. Son, G. H. et al. Hyperactivity and alteration of the midbrain dopaminergic system in maternally stressed male mice offspring. *Biochemical and Biophysical Research Communications* 352, 823–829 (2007).
38. Zhang, Y. et al. Maternal Deprivation Enhances Behavioral Vulnerability to Stress Associated with miR-504 Expression in Nucleus Accumbens of Rats. *PLoS ONE* 8, (2013).
39. Zhang, Y. et al. Dopamine receptor D2 and associated microRNAs are involved in stress susceptibility and resistance to escitalopram treatment. *International Journal of Neuropsychopharmacology* 18, 1–10 (2015).
40. Tien, L. T. et al. Neonatal exposure to lipopolysaccharide enhances accumulation of ??-synuclein aggregation and dopamine transporter protein expression in the substantia nigra in responses to rotenone challenge in later life. *Toxicology* 308, 96–103 (2013).
41. Matthews, K., Dalley, J. W., Matthews, C., Tsai, T. H. & Robbins, T. W. Periodic maternal separation of neonatal rats produces region- and gender-specific effects on biogenic amine content in postmortem adult brain. *Synapse* 40, 1–10 (2001).
42. Kubesova, A. et al. Biochemical, histopathological and morphological profiling of a rat model of early immune stimulation: Relation to psychopathology. *PLoS ONE* 10, 1–16 (2015).

43. Adrover, E., Berger, M. A., Pérez, A. A., Tarazi, F. I. & Antonelli, M. C. Effects of prenatal stress on dopamine D2 receptor asymmetry in rat brain. *Synapse* 61, 459–462 (2007).
44. Alonso, S. J., Navarro, E. & Rodriguez, M. Permanent dopaminergic alterations in the n. accumbens after prenatal stress. *Pharmacology, Biochemistry and Behavio*r 49, 353–358 (1994).
45. Baharnoori, M., Bhardwaj, S. K. & Srivastava, L. K. Effect of Maternal Lipopolysaccharide Administration on the Development of Dopaminergic Receptors and Transporter in the Rat Offspring. PLoS ONE 8, 18–23 (2013).
46. Bakos, J. et al. Prenatal immune challenge affects growth, behavior, and brain dopamine in offspring. *Annals of the New York Academy of Sciences* 1018, 281–287 (2004).
47. Basta-Kaim, A. et al. Prenatal lipopolysaccharide treatment enhances MK-801-induced psychotomimetic effects in rats. *Pharmacology Biochemistry and Behavior* 98, 241–249 (2011).
48. Berger, M. A., Barros, V. G., Sarchi, M. I., Tarazi, F. I. & Antonelli, M. C. Long-term effects of prenatal stress on dopamine and glutamate receptors in adult rat brain. *Neurochem Res* 27, 1525–1533 (2002).
49. Bingham, B. C., Sheela Rani, C. S., Frazer, A., Strong, R. & Morilak, D. A. Exogenous prenatal corticosterone exposure mimics the effects of prenatal stress on adult brain stress response systems and fear extinction behavior. *Psychoneuroendocrinology* 38, 2746–2757 (2013).
50. Bitanihirwe, B. K. Y., Peleg-Raibstein, D., Mouttet, F., Feldon, J. & Meyer, U. Late prenatal immune activation in mice leads to behavioral and neurochemical abnormalities relevant to the negative symptoms of schizophrenia. *Neuropsychopharmacology* 35, 2462–78 (2010).
51. Brenhouse, H., Lukkes, J. & Andersen, S. Early Life Adversity Alters the Developmental Profiles of Addiction-Related Prefrontal Cortex Circuitry. *Brain Sciences* 3, 143–158 (2013).
52. Cabib, S., Puglisi-Allegra, S. & D'Amato, F. R. Effects of postnatal stress on dopamine mesolimbic system responses to aversive experiences in adult life. *Brain Research* 604, 232–239 (1993).
53. Cai, Z. et al. Neonatal Systemic Exposure to Lipopolysaccharide Enhances Susceptibility of Nigrostriatal Dopaminergic Neurons to Rotenone Neurotoxicity in Later Life. Developmental Neuroscience 35, 155–171 (2013).
54. Camp, D. M., Robinson, T. E. & Becker, J. B. Sex-differences in the effects of early experience on the development of behavioral and brain asymmetries in rats. *Physiology & Behavior* 33, 433–439 (1984).
55. Carboni, E. et al. Prenatal restraint stress: An in vivo microdialysis study on catecholamine release in the rat prefrontal cortex. *Neuroscience* 168, 156–166 (2010).
56. Chocyk, A. et al. The impact of maternal separation on the number of tyrosine hydroxylase-expressing midbrain neurons during different stages of ontogenesis. *Neuroscience* 182, 43–61 (2011).
57. Choy, K. H. C., De Visser, Y. P. & Van Den Buuse, M. The effect of 'two hit' neonatal and young-adult stress on dopaminergic modulation of prepulse inhibition and dopamine receptor density. *British Journal of Pharmacology* 156, 388–396 (2009).
58. Choy, K. H. C. & van den Buuse, M. Attenuated disruption of prepulse inhibition by dopaminergic stimulation after maternal deprivation and adolescent corticosterone treatment in rats. *European Neuropsychopharmacology* 18, 1–13 (2008).
59. Cory-Slechta, D. A., Virgolini, M. B., Rossi-George, A., Weston, D. & Thiruchelvam, M. Experimental manipulations blunt time-induced changes in brain monoamine levels and completely reverse stress, but not Pb+/-stress-related modifications to these trajectories. *Behavioural Brain Research* 205, 76–87 (2009).
60. Cory-slechta, D. A., Weston, D., Liu, S. & Allen, J. L. Brain hemispheric differences in the neurochemical effects of lead, prenatal stress, and the combination and their amelioration by behavioral experience. *Toxicological Sciences* 132, 419–430 (2013).
61. Dallé, E., Daniels, W. M. U. & Mabandla, M. V. Anti-parkinsonian effects of fluvoxamine maleate in maternally separated rats. *International Journal of Developmental Neuroscience* 53, 26–34 (2016).
62. Delattre, A. M. et al. Maternal Omega-3 Supplement Improves Dopaminergic System in Pre- and Postnatal Inflammation-Induced Neurotoxicity in Parkinson???s Disease Model. *Molecular Neurobiology* 1–17 (2016) doi:10.1007/s12035-016-9803-8.
63. Fan, L. W. et al. Neonatal exposure to lipopolysaccharide enhances vulnerability of nigrostriatal dopaminergic neurons to rotenone neurotoxicity in later life. *Neurobiology of Disease* 44, 304–316 (2011).
64. Fan, L.-W. et al. Dopaminergic neuronal injury in the adult rat brain following neonatal exposure to lipopolysaccharide and the silent neurotoxicity. *Brain, behavior, and immunity* 25, 286–297 (2011).
65. Gerardin, D. C. C. et al. Sexual behavior, neuroendocrine, and neurochemical aspects in male rats exposed prenatally to stress. *Physiology and Behavior* 84, 97–104 (2005).
66. Gondré-Lewis, M. C., Darius, P. J., Wang, H. & Allard, J. S. Stereological analyses of reward system nuclei in maternally deprived/separated alcohol drinking rats. *Journal of Chemical Neuroanatomy* (2015) doi:10.1016/j.jchemneu.2016.02.004.
67. Gracia-Rubio, I. et al. Maternal Separation Impairs Cocaine-Induced Behavioural Sensitization in Adolescent Mice. *Plos One* 11, e0167483 (2016).
68. Granholm, A. C. et al. Prenatal LPS increases inflammation in the substantia nigra of Gdnf heterozygous mice. *Brain Pathology* 21, 330–348 (2011).
69. Hall, F. S., Wilkinson, L. S., Humby, T. & Robbins, T. W. Maternal deprivation of neonatal rats produces enduring changes in dopamine function. *Synapse* 32, 37–43 (1999).
70. Hensleigh, E. & Pritchard, L. M. Maternal separation increases methamphetamine-induced damage in the striatum in male, but not female rats. *Behavioural Brain Research* 295, 3–8 (2015).

71. Hensleigh, E. & Pritchard, L. M. The effect of early environmental manipulation on locomotor sensitivity and methamphetamine conditioned place preference reward. *Behavioural Brain Research* 268, 66–71 (2014).
72. Hermel, E. E. S. et al. Neonatal handling and the expression of immunoreactivity to tyrosine hydroxylase in the hypothalamus of adult male rats. Brazilian Journal of Medical and Biological Research 34, 1191–1195 (2001).
73. Hill, R. A. et al. Long-term effects of combined neonatal and adolescent stress on brain-derived neurotrophic factor and dopamine receptor expression in the rat forebrain. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1842, 2126–2135 (2014).
74. Jahng, J. W. et al. Mesolimbic dopaminergic activity responding to acute stress is blunted in adolescent rats that experienced neonatal maternal separation. *Neuroscience* 171, 144–152 (2010).
75. Jahng, J. W., Yoo, S. B., Kim, J. Y., Kim, B. T. & Lee, J. H. Increased mesohippocampal dopaminergic activity and improved depression-like behaviors in maternally separated rats following repeated fasting/refeeding cycles. *Journal of Obesity* 2012, (2012).
76. Katunar, M. R., Saez, T., Brusco, A. & Antonelli, M. C. Ontogenetic expression of dopamine-related transcription factors and tyrosine hydroxylase in prenatally stressed rats. *Neurotoxicity research* 18, 69–81 (2010).
77. Kawakami, S. E., Quadros, I. M. H., Machado, R. B. & Suchecki, D. Sex-dependent effects of maternal separation on plasma corticosterone and brain monoamines in response to chronic ethanol administration. *Neuroscience* 253, 55–66 (2013).
78. Kikusui, T., Faccidomo, S. & Miczek, K. A. Repeated maternal separation: Differences in cocaine-induced behavioral sensitization in adult male and female mice. *Psychopharmacology* 178, 202–210 (2005).
79. Lazzaretti, C., Kincheski, G. C., Pandolfo, P. & S, N. De. Neonatal handling causes impulsive behavior and decreased pharmacological response to methylphenidate in male adult wistar rats. *Journal of Integrative Neuroscience* 14, 1–15 (2015).
80. Lejeune, S. et al. The dopamine D1 receptor agonist SKF 38393 improves temporal order memory performance in maternally deprived rats. *Neurobiology of Learning and Memory* 106, 268–273 (2013).
81. Li, W. Y., Chang, Y. C., Lee, L. J. H. & Lee, L. J. Prenatal infection affects the neuronal architecture and cognitive function in adult mice. *Developmental Neuroscience* 36, 359–370 (2014).
82. Ling, Z. et al. Rotenone potentiates dopamine neuron loss in animals exposed to lipopolysaccharide prenatally. *Experimental Neurology* 190, 373–383 (2004).
83. Ling, Z. et al. Prenatal lipopolysaccharide does not accelerate progressive dopamine neuron loss in the rat as a result of normal aging. *Experimental Neurology* 216, 312–320 (2009).
84. Liu, Y. W. et al. Psychotropic effects of Lactobacillus plantarum PS128 in early life-stressed and naïve adult mice. *Brain Research* 1631, 1–12 (2016).
85. Lovic, V. et al. Early postnatal experience and DRD2 genotype affect dopamine receptor expression in the rat ventral striatum. *Behavioural brain research* 237, 278–82 (2013).
86. Luchicchi, A. et al. Maternal immune activation disrupts dopamine system in the offspring. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum* (CINP) 19, pyw007 (2016).
87. Madruga, C., Xavier, L. L., Achaval, M., Sanvitto, G. L. & Lucion, A. B. Early handling, but not maternal separation, decreases emotional responses in two paradigms of fear without changes in mesolimbic dopamine. *Behavioural Brain Research* 166, 241–246 (2006).
88. Matthews, K., Dalley, J. W., Matthews, C., Tsai, T. H. & Robbins, T. W. Periodic maternal separation of neonatal rats produces region- and gender-specific effects on biogenic amine content in postmortem adult brain. *Synapse* 40, 1–10 (2001).
89. Meyer, U. et al. Relative prenatal and postnatal maternal contributions to schizophrenia-related neurochemical dysfunction after in utero immune challenge. *Neuropsychopharmacology* 33, 441–456 (2008).
90. Meyer, U., Nyffeler, M., Yee, B. K., Knuesel, I. & Feldon, J. Adult brain and behavioral pathological markers of prenatal immune challenge during early/middle and late fetal development in mice. *Brain, Behavior, and Immunity* 22, 469–486 (2008).
91. Novak, G., Fan, T., O'Dowd, B. F. & George, S. R. Postnatal maternal deprivation and pubertal stress have additive effects on dopamine D2 receptor and CaMKII beta expression in the striatum. *International Journal of Developmental Neuroscience* 31, 189–195 (2013).
92. Ognibene, E. et al. The effect of early maternal separation on brain derived neurotrophic factor and monoamine levels in adult heterozygous reeler mice. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 32, 1269–1276 (2008).
93. Oreland, S. et al. Ethanol-induced effects on the dopamine and serotonin systems in adult Wistar rats are dependent on early-life experiences. *Brain Research* 1405, 57–68 (2011).
94. Ozawa, K. et al. Immune activation during pregnancy in mice leads to dopaminergic hyperfunction and cognitive impairment in the offspring: A neurodevelopmental animal model of schizophrenia. *Biological Psychiatry* 59, 546–554 (2006).
95. Pallarés, M. E. et al. Age-dependent effects of prenatal stress on the corticolimbic dopaminergic system development in the rat male offspring. *Neurochemical Research* 38, 2323–2335 (2013).
96. Panagiotaropoulos, T. et al. Regulation of Corticotropin and Stress Neonatal Handling and Gender Modulate Brain Monoamines and Plasma Corticosterone Levels following Repeated Stressors in Adulthood. 181–191 (2004)

doi:10.1159/000082516.
97. Pang, Y. et al. Interleukin-1 receptor antagonist reduces neonatal lipopolysaccharide-induced long-lasting neurobehavioral deficits and dopaminergic neuronal injury in adult rats. *International Journal of Molecular Sciences* 16, 8635–8654 (2015).
98. Papaioannou, A., Dafni, U., Alikaridis, F., Bolaris, S. & Stylianopoulou, F. Effects of neonatal handling on basal and stress-induced monoamine levels in the male and female rat brain. *Neuroscience* 114, 195–206 (2002).
99. Ploj, K., Roman, E. & Nylander, I. Long-term effects of maternal separation on ethanol intake and brain opioid and dopamine receptors in male wistar rats. *Neuroscience* 121, 787–799 (2003).
100. Rentesi, G. et al. Early maternal deprivation-induced modifications in the neurobiological, neurochemical and behavioral profile of adult rats. *Behavioural Brain Research* 244, 29–37 (2013).
101. Reynaert, M. L. et al. Hedonic sensitivity to natural rewards is affected by prenatal stress in a sex-dependent manner. *Addiction Biology* 21, 1072–1085 (2016).
102. Reznikov, A. G. & Nosenko, N. D. Catecholamines in steroid-dependent brain development. *The Journal of steroid biochemistry and molecular biology* 53, 349–53 (1995).
103. Romano-López, A. et al. Maternal separation and early stress cause long-lasting effects on dopaminergic and endocannabinergic systems and alters dendritic morphology in the nucleus accumbens and frontal cortex in rats. *Developmental Neurobiology* (2015) doi:10.1002/dneu.22361.
104. Romero, E. et al. Neurobehavioral and immunological consequences of prenatal immune activation in rats. Influence of antipsychotics. *Neuropsychopharmacology* 32, 1791–1804 (2007).
105. Romero, E., Guaza, C., Castellano, B. & Borrell, J. Ontogeny of sensorimotor gating and immune impairment induced by prenatal immune challenge in rats: implications for the etiopathology of schizophrenia. *Molecular Psychiatry* 15, 372–83 (2010).
106. Rossi-George, A., Virgolini, M. B., Weston, D., Thiruchelvam, M. & Cory-Slechta, D. A. Interactions of lifetime lead exposure and stress: Behavioral, neurochemical and HPA axis effects. *NeuroToxicology* 32, 83–99 (2011).
107. Rots, N. Y. et al. Neonatal maternally deprived rats have as adults elevated basal pituitary-adrenal activity and enhanced susceptibility to apomorphine. *Journal of Neuroendocrinology* 8, 501–506 (1996).
108. Sasagawa, T. et al. Long-term effects of maternal separation coupled with social isolation on reward seeking and changes in dopamine D1 receptor expression in the nucleus accumbens via DNA methylation in mice. *Neuroscience Letters* 641, 33–39 (2017).
109. Silvagni, A., Barros, V. G., Mura, C., Antonelli, M. C. & Carboni, E. Prenatal restraint stress differentially modifies basal and stimulated dopamine and noradrenaline release in the nucleus accumbens shell: An 'in vivo' microdialysis study in adolescent and young adult rats. *European Journal of Neuroscience* 28, 744–758 (2008).
110. Silveira, P. P. et al. Early life experience alters behavioral responses to sweet food and accumbal dopamine metabolism. *International Journal of Developmental Neuroscience* 28, 111–118 (2010).
111. Tien, L. T. et al. Neonatal exposure to lipopolysaccharide enhances accumulation of ??-synuclein aggregation and dopamine transporter protein expression in the substantia nigra in responses to rotenone challenge in later life. *Toxicology* 308, 96–103 (2013).
112. Vazquez, V., Weiss, S., Giros, B., Martres, M. P. & Daug??, V. Maternal deprivation and handling modify the effect of the dopamine D3 receptor agonist, BP 897 on morphine-conditioned place preference in rats. *Psychopharmacology* 193, 475–486 (2007).
113. Vuillermot, S. et al. Prenatal Immune Activation Interacts with Genetic Nurr1 Deficiency in the Development of Attentional Impairments. *Journal of Neuroscience* 32, 436–451 (2012).
114. Wang, S., Yan, J. Y., Lo, Y. K., Carvey, P. M. & Ling, Z. Dopaminergic and serotoninergic deficiencies in young adult rats prenatally exposed to the bacterial lipopolysaccharide. *Brain Research* 1265, 196–204 (2009).
115. Weston, H. I. et al. Sex-dependent and non-monotonic enhancement and unmasking of methylmercury neurotoxicity by prenatal stress. *NeuroToxicology* 41, 123–140 (2014).
116. Weston, H. I., Weston, D. D., Allen, J. L. & Cory-Slechta, D. A. Sex-dependent impacts of low-level lead exposure and prenatal stress on impulsive choice behavior and associated biochemical and neurochemical manifestations. *NeuroToxicology* 44, 169–183 (2014).
117. Winter, C. et al. Prenatal immune activation leads to multiple changes in basal neurotransmitter levels in the adult brain: implications for brain disorders of neurodevelopmental origin such as schizophrenia. *The International Journal of Neuropsychopharmacology* 12, 513–24 (2009).
118. Womersley, J. S., Hsieh, J. H., Kellaway, L. a, Gerhardt, G. a & Russell, V. a. Maternal separation affects dopamine transporter function in the spontaneously hypertensive rat: an in vivo electrochemical study. *Behavioral and brain functions* : BBF 7, 49 (2011).
119. Zager, A., Mennecier, G. & Palermo-Neto, J. Maternal immune activation in late gestation enhances locomotor response to acute but not chronic amphetamine treatment in male mice offspring: Role of the D1 receptor. *Behavioural Brain Research* 232, 30–36 (2012).
120. Zavitsanou, K. et al. Neonatal lipopolysaccharide treatment has long-term effects on monoaminergic and cannabinoid receptors in the rat. *Synapse* 67, 290–299 (2013).
121. Zhang, Y. et al. Dopamine receptor D2 and associated microRNAs are involved in stress susceptibility and resistance to escitalopram treatment. *International Journal of Neuropsychopharmacology* 18, 1–10 (2015).
122. Zhang, X. Y., Kehoe, P. & Kosten, T. A. Neonatal isolation alters estrous cycle effects on ventral striatal extracellular monoamine levels. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 30, 504–511 (2006).

123. Zhang, Y. et al. Maternal Deprivation Enhances Behavioral Vulnerability to Stress Associated with miR-504 Expression in Nucleus Accumbens of Rats. *PLoS ONE* 8, (2013).
124. Zhu, X. et al. Maternal deprivation-caused behavioral abnormalities in adult rats relate to a non-methylation-regulated D2 receptor levels in the nucleus accumbens. *Behavioural Brain Research* 209, 281–288 (2010).
125. Zhu, X., Peng, S., Zhang, S. & Zhang, X. Stress-induced depressive behaviors are correlated with Par-4 and DRD2 expression in rat striatum. *Behavioural Brain Research* 223, 329–335 (2011).
126. Zhu, Y., Carvey, P. M. & Ling, Z. Altered glutathione homeostasis in animals prenatally exposed to lipopolysaccharide. *Neurochemistry International* 50, 671–680 (2007).
127. Zuckerman, L., Rehavi, M., Nachman, R. & Weiner, I. Immune activation during pregnancy in rats leads to a postpubertal emergence of disrupted latent inhibition, dopaminergic hyperfunction, and altered limbic morphology in the offspring: a novel neurodevelopmental model of schizophrenia. *Neuropsychopharmacology* 28, 1778–89 (2003).
128. Grace, A. A. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience* 17, 524–532 (2016).
129. Rosenthal, R. Meta-analytic procedures for social research. (Sage, 1991).
130. Van den Noortgate, W., López-López, J. A., Marín-Martínez, F. & Sánchez-Meca, J. Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods* 45, 576–594 (2013).
131. Cornell, J. & Mulrow, C. *Meta-analysis.* (Sage, 1999).
132. Antonic, A. et al. Stem Cell Transplantation in Traumatic Spinal Cord Injury: A Systematic Review and Meta-Analysis of Animal Studies. *PLoS Biology* 11, (2013).
133. Egan, K. J., Sena, E. S. & Vesterinen, H. M. Making the most of animal data - Improving the prospect of success in pragmatic trials in the neurosciences. *Trials* 12, no pagination (2011).
134. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The arrive guidelines for reporting animal research. *Animals* 4, 35–44 (2013).
135. Landis, S. C. et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490, 187–191 (2012).
136. Cohen, J. A power primer. *Psychological Bulletin* 112, 155–159 (1992).
137. Lovallo, W. R. et al. Joint Impact of Early Life Adversity and COMT Val158Met (rs4680) Genotypes on the Adult Cortisol Response to Psychological Stress. *Psychosomatic Medicine* 79, 631–637 (2017).
138. Klaus, K. et al. The effect of COMT Val158Met and DRD2 C957T polymorphisms on executive function and the impact of early life stress. *Brain and Behavior* 7, 1–12 (2017).
139. Sheikh, H. I., Kryski, K. R., Kotelnikova, Y., Hayden, E. P. & Singh, S. M. Catechol-O-Methyltransferase gene (val158met) polymorphisms and anxious symptoms in early childhood: The roles of hypothalamus-pituitary-adrenal axis reactivity and life stress. *Neuroscience Letters* 659, 86–91 (2017).
140. Braun, K., Lange, E., Metzger, M. & Poeggel, G. Maternal separation followed by early social deprivation affects the development of monoaminergic fiber systems in the medial prefrontal cortex of Octodon degus. *Neuroscience* 95, 309–318 (1999).
141. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. Nature neuroscience 18, 1413–20 (2015).
142. Williams, E. G. et al. Systems proteomics of liver mitochondria function. *Science* 352, aad0189–aad0189 (2016).
143. Wise, R. A. & Koob, G. F. The Development and Maintenance of Drug Addiction. *Neuropsychopharmacology* 39, 254–262 (2014).
144. Nutt, D. J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P. R. A. The dopamine theory of addiction: 40 years of highs and lows. *Nature Reviews Neuroscience* 16, 305–312 (2015).
145. Sanyal, S. & Van Tol, H. H. M. Review the role of dopamine D4 receptors in schizophrenia and antipsychotic action. *Journal of Psychiatric Research* 31, 219–232 (1997).
146. Daskalakis, N. P., Bagot, R. C., Parker, K. J., Vinkers, C. H. & de Kloet, E. R. The three-hit concept of vulnerability and resilience: Toward understanding adaptation to early-life adversity outcome. *Psychoneuroendocrinology* 38, 1858–1873 (2013).
147. Shin, S. et al. Drd3 Signaling in the Lateral Septum Mediates Early Life Stress-Induced Social Dysfunction. *Neuron* 97, 195-208.e6 (2018).
148. Valenti, O., Gill, K. M. & Grace, A. A. Different stressors produce excitation or inhibition of mesolimbic dopamine neuron activity: Response alteration by stress pre-exposure. *European Journal of Neuroscience* 35, 1312–1321 (2012).
149. Hensch, T. K. Critical period plasticity in local cortical circuits. Nature Reviews Neuroscience 6, 877–888 (2005).
150. Thompson, C. L. et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* 83, 309–323 (2014).
151. Meiser, J., Weindl, D. & Hiller, K. Complexity of dopamine metabolism. *Cell Communication and Signaling* 11, 34 (2013).
152. Swanson, L. W. The projections of the ventral tegmental area and adjacent regions: a combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain research bulletin* 9, 321–53 (1982).
153. Menegas, W. et al. Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *eLife* 4, 1–30 (2015).
154. Keller-Wood, M. Hypothalamic-Pituitary-Adrenal Axis-Feedback Control. *Comprehensive Physiology* 5, 1161–1182 (2015).

Chapter 6

Changes in monoamine systems after postnatal early life adversity in rodents: a systematic review with meta-analysis

in preparation

Bonapersona V[1]
Joëls M[1,2]
van der Veen R[3]

1 Dept Translational Neuroscience, UMC Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands
2 University Medical Center Groningen, Groningen, The Netherlands
3 Brain plasticity group, SILS, University of Amsterdam, The Netherlands

Abstract

Early life adversity (ELA) is a well-characterised risk factor for mental health disorders related to monoaminergic dysfuction, most notably depression and addiction. The causal consequences of postnatal ELA on monoaminergic systems have been reported for adult rodents in 47 publications, here systematically reviewed and meta-analysed. The effects of ELA on monoaminergic systems in adulthood were limited in male animals tested at rest, although consistent with a hypo-activation. With respect to the dopaminergic system, the striatum was the brain area most sensitive to ELA, while for the serotonergic system this was the prefrontal cortex. Noradrenaline was investigated in only a limited number of publications. Exploratory analyses suggest that the effects of ELA on monoaminergic systems are consistently larger in female than in male rodents. Furthermore, the effects on monoamine concentration appeared to be stronger when the (ELA and control) animals were tested under aroused/stressed circumstances rather than at rest. While this review synthetises the current status of the literature and overall points to relatively mild effects of ELA on rodents' monoaminergic systems in adulthood, the number of outcome-by-brain area comparisons is still limited, which asks for careful interpretation.

Introduction

Exposure to adverse experiences during early life (i.e. early life adversity, ELA) sets off changes in brain functioning and behavior[1,2] and increases the susceptibility to develop psychopathologies later in life[3–5]. This is presumably due to the increased vulnerability and malleability of the brain during early life, when it undergoes rapid development and important maturation[6,7].

The monoaminergic systems are important neuromodulators that coordinate cognitive and emotional reactions to the environment, guide coping behavior under stressful circumstances[8] as well as emotional states and mood of the individual[9,10]. Monoamines' synthesis is conserved in ventebrate and invertebrate species, where they play a significant role in the nervous system's function and plasticity[11]. The most common monoamines are dopamine (DA), serotonin (5HT) and noradrenaline (NE). These are also a central focus of neuropsychiatric research, since their hypo- or hyper-functioning has been related with several mental health diagnoses, including depression, obsessive-compulsive disorder, attention-deficit hyperactivity disorder and psychosis[12–14]. For example, single nucleotide polymorphisms (SNPs) in close proximity of the dopamine receptor D2 (DRD2) gene has been associated (among others) with an increased risk of developing schizophrenia[15] as well as major depression disorder[16]. Serotonergic gene "sets" have been related with bipolar disorder and major depression[17]. Furthermore, most medications currently used in psychiatry target (directly or indirectly) the monoaminergic systems[14,18]. Specifically, most of the antipsychotic treatments (e.g. chlorpromazine and haloperidol) block DRD2, and they are often combined with serotonin-2A receptor agonists (e.g. clozapine and risperidone)[17]. These results suggest that there is a relationship between mental health phenotypes and a general dysfunctioning of monoaminergic systems.

Early life adversity (ELA) could play an important role in the relationship between altered monoaminergic function and mental health disorders in humans[19,20]. For example, ELA events were found to increase the risk to develop an anxiety disorder in the presence of specific genotypes related to *COMT* and *MAO A*[21], two enzymes involved in the degradation of monoamines. ELA has also been correlated with decreased levels of the serotonin transporter, presumably via an epigenetic link[22]. However, the evidence from human studies is confined to correlations. Many have therefore used rodent models to investigate how ELA causes changes in indicators of the brain's functionality of monoaminergic systems. Therefore, animal models of ELA are widely used to study these long-lasting consequences of ELA on brain structure and function, that might underlie changes in adult behavior and coping (e.g. reviewed in[23,24]). The bulk of these studies is performed on laboratory rats and mice, with a focus on the early postnatal period of development. Although these studies adopt a variety of ELA models, it is still possible to systematically identify relevant studies and to quantify the most robust changes, for example by using systematic review and meta-

analysis. This approach is increasingly common[25], and can help to uncover knowledge gaps and to guide future studies in the field[26,27]. Here, we present a systematic review and meta-analysis to identify robust changes in rodent models of postnatal ELA (as alteration of maternal care) on biochemical indicators of monoaminergic systems in the adult brain. This includes an update and expansion of our earlier work on ELA-induced dopaminergic changes[28], but now also includes the reports on the serotoninergic and noradrenergic systems. We focused on postnatal ELA models that are based on an alteration of maternal care, i.e. maternal separation and deprivation, isolation, limited nesting and bedding, and natural variation in the amount of licking and grooming. We included outcomes in adult animals (>8 weeks of age, less than one year old), to investigate long-term effects on the monoaminergic systems. In all, we aimed to provide a comprehensive and systematic review of the effects of postnatal ELA on monoaminergic systems in adult rodents.

Materials and methods

This review adhere's to the SYRCLE's guidelines for protocol[29], search strategy[30], and risk of bias assessment[31]. Reporting is done in accordance with the PRISMA reporting checklist[32]. The analytic strategy is based on earlier work of our own lab[23,28]. Materials, data and scripts used for this project are available via the open science framework (https://osf.io/4ngu3/).

This manuscript builds upon a similar publication of our own lab, on the effects of prenatal and postnatal ELA on biochemical indicators of the dopaminergic system[28]. From that dataset, we included 18 publications using ELA models as alterations of maternal care, which matched our current inclusion criteria. This was complemented with an additional 9 publications on dopaminergic outcomes, which were published after 2018; and comprehensive reports on the serotonergic and noradrenergic systems.

Search strategy

The search strategy of this study was conducted in parallel with other studies of our own lab[33,34]. Briefly, on April 3rd 2019, we conducted a systematic literature search on the electronic databases PubMed and WebOfScience, which included the terms 'mice and rats' and 'postnatal ELA' (Supplementary note 1). ELA was defined as postnatal models that are based on an alteration of maternal care, either naturally varying (i.e., licking and grooming[35]) or experimentally induced in (at least) the first two postnatal weeks (i.e., maternal deprivation[36] and separation[37]; isolation; limited nesting and bedding[38]).

Study selection was performed in two stages. In the first stage, titles and abstracts were excluded if: 1) the articles were not a primary publication, 2) experiments were not conducted in mice or rats, 3) the study did not concern early life adversity. This stage of study selection was performed in Rayyan[39] by at least three (out of 5, see

Acknowledgements) researchers. During study selection, the order of the articles differed between researchers. During the second stage, full text was screened and studies were included if: 1) the outcomes were related to monoamines (Supplementary Table 1, prespecified), 2) the outcomes were measured in adulthood (older than 8 weeks but younger than 1 year), 3) the animals did not experience other pharmachological/dietary/genetic interventions, 4) the sex of animals was known (through the publication or after contacting authors). The second step of study selection was performed by two researchers (VB and EK, see Acknowledgements), and in case of disagreements these were resolved by discussion. The full list of inclusion/exclusion criteria is available in Supplementary note 2. An overview of the study procedure is shown in Figure 1.

From the eligible studies, we selected those experiments where control and ELA groups differed only in the experience of the early life condition, i.e. adversity or not. All other variables within each experiments were identical for the control and ELA groups. Each individual comparison between a control and an experimental group was organized in a standardized database. This database summarizes information about 1) the publications (author, year, reference), 2) the experimental design (e.g. species, sex, model, other life events, age of the animals at the time of testing, state of the animal just before death), 3) information about the outcomes extracted (e.g. brain area, technique used), and 4) summary statistics of the data measured (e.g. sample size, mean, and deviation (SEM or SD)). Data that was reported exclusively in graphs was digitalized with Web Plot Digitalizer[40]. For all remaining missing information, we contacted the corresponding author of each manuscript (response rate 83%). If the information could not be retrieved, it was considered missing and excluded from the analysis.

Data preprocessing

Effect sizes for each individual comparison (i.e. the difference between control and ELA on each specific outcome) were calculated with escalc (R package metafor,[41]) as standardized mean difference with Hedge's g (g) method, which includes a correction factor for small sample sizes[42]. To use *escalc*, we harmonized the reported measures of variation and of sample size. If only SEM was reported, SD was calculated as SEM * n , where n = number of animals per group. If the number of animals used was reported as a range (e.g. 6–8 animals per group), we used the lower boundary (e.g. 6 animals per group), taking a conservative approach.

Since experimental designs are heterogeneous, we categorized elements of the experimental design to ease interpretation and allow for data synthesis. Specifically, we grouped brain areas according to the embryological origin offered by the Allen Brain Reference Atlas[43]; the grouping is available in Supplementary Table 2. Furthermore, we categorized the life experiences of the animals, distinguishing between negative life events with (possible) long-lasting effects and the acute situation at testing. The

categorization is available in Supplementary Table 3. If a study reported multiple sub-brain areas within one of our categorizations, these were combined for the quantitative synthesis to limit heterogeneity and avoid paper-specific biasing. Similarly, if a study reported multiple sub-outcomes within one of our categorization, these were combined into one measure. Specifically, dopamine receptors 1 and 5 were combined into dopamine-receptor 1-like; dopamine receptors 2 and 3 were combined into dopamine-receptors 2-like. The aggregation occurred at the level of effect sizes, with the function aggregate (rho = 0.6) from the R package *metafor*[41].

Data synthesis and statistical analysis

Initially, we planned to conduct our analysis on data from males and females separately. We also aimed to investigate the experience of other life events as potential moderators, specifically distinguishing between events with long-lasting effects (i.e. multiple "hits") and effects with acute effects at testing (i.e. state of the animal at testing, Supplementary Table 3). However, the number of included publications was much lower than expected. Therefore, we maintained the same research questions as reported in our study protocol, but we changed the statistical approach. Specifically, we conducted one main analysis, one subgroup analysis and two exploratory analyses on a subset of the data.

Main and subgroup analyses The main research question aimed to investigate the effects of ELA on monoaminergic outcomes. We focused specifically on data obtained in male rodents at rest at the time of testing (i.e. before death), to maintain the dataset as homogeneous as possible. In other words, since it was not possible to systematically investigate the interaction between multiple hits and the situation of the animals at testing, we limited the analysis on a subset of the data (i.e. rest at testing) which was the most abundant. From previous studies of our own lab[34], we observed that the acute status of the animals pushed the ELA effects in opposite directions; therefore we preferred to select a subset of the data rather than keeping all data and assuming that the acute status of the animal would have not had any effect. We used a 3-level mixed effect model[23,44], which accounts for the anticipated heterogeneity of the studies as well as the dependency of effects within experiments. Here, brain areas were added to the model as moderator. In our experimental design, the 3 levels corresponded to variance of effect size between 1) animals, 2) experiments and 3) publications. The selected outcomes relevant to the monoamines' systems are summarized in Supplementary Table 1. Prior to the beginning of the study it was not possible to estimate which of the outcomes were sufficiently explored in the literature to further focus on. We therefore decided to analyze all outcomes reported by at least 3 independent publications, and to lower the significance threshold to 0.01 rather than to decrease the probability of committing a Type I error (i.e. false positives).

Lastly, we conducted a subgroup analysis on this data to test whether the experience of other major negative life events would impact effect sizes, as previously observed in relation to behavior[23]. On each subset of the data (with or without other negative life events), we used the same 3-level mixed effect model as per the main analysis. Of note, this analysis was not performed for each brain area separately due to the availability of the data, but it included brain area as a covariate. Then, we analyzed whether the estimates were different from each other with a Wald-type test. This approach is in agreement with the recommendations by metafor[41].

Exploratory analysis: life experiences To test the potential moderating effect of life experiences (i.e., state of the animals just before death, and experience of other major negative life events), we performed an exploratory analysis on our data. For this analysis, we included only a subset of the data. Our research question was whether the effects of ELA interact with acute and chronic life experiences, as we previously observed in a study related to immediate early genes[34]. Therefore, we selected within the male dataset only those outcomes per brain area that were investigated for each of the following groups: 1) at rest with no other negative life experiences (see [23] for the definition), 2) aroused/stressed with no other negative life experiences, 3) at rest with other negative life experiences, and 4) aroused/stressed with other negative life experiences. This was done to limit as much as possible study-related confounding effects. Of note, the dataset used for this subgroup analysis partially overlapped with that used for the main analysis. With the dataset for the subgroup analysis, we built a 3-level mixed effect meta-analysis similar to the main analysis, but now the state of the animals just before death was in interaction with the experience of other major negative life events was included as moderators. Although brain areas and outcomes were used as covariates, we tested only the main effects (i.e., aroused/stress vs rest and no other vs other negative life events) and the resulting 4 groups of the interaction (i.e. rest with no other major life event, rest with other major life event, aroused/stress with no other major life event and lastly aroused/stressed with other major life event) against 0 (i.e. no effect).

Exploratory analyses: sex differences The effects of stress on the brain are often sexually dysmorphic, meaning that they can differ between sexes. For this reason, we planned to analyze males and females separately. Although the number of female studies has increased in recent years, the evidence is not yet sufficient for a quantitative analysis. We aimed nonetheless to explore evidence in favour/against sex differences in relation to the effects of ELA on monoamines' outcomes. To test this, we selected only those publications with experiments performed in males as well as females. We used the same 3-level mixed effect model as the main analysis, but this time we added sex as a moderator of the effects. Specifically, we tested whether the

effects of ELA on each sex was different from 0 (i.e. no difference), once correcting for outcome and brain areas. Then, we tested whether the effects of ELA in females were significantly different from those of males with a Wald-type test. This analysis would answer whether in this specific dataset either sex is more sensitive to ELA, irrespective of the outcome. We performed the same analysis also on absolute hedge's g values rather than including outcome-by-brain area as potential moderators, to evaluate the robustness of the effects.

Study of heterogeneity

Heterogeneity was tested with a Cochrane Q-test[45] and I^2 statistics[46]. A significant test of heterogeneity means that there is still unexplained variance in the data, despite the used moderators. We therefore performed a machine-learning-based exploratory approach to identify potential moderators using MetaForest[47]. Metaforest applies random forests to meta-analysis data by means of bootstrap sampling, thereby ranking moderators based on their (non)linear influence on the effect size. Of the variables classifying the experimental design, we selected 9 for metaforest analysis. After identifying a convergence range, we built our MetaForest model and conducted a 10x cross validation to determine the optimal tuning parameters that minimized RMSE (fixed weights, 2 candidate moderators at each split, minimum node size = 6). To estimate how much variance is explained by our model, we calculated the cross-validated R^2 (R_{cv}^2), which is robust to overfitting and provides evidence for the results' generalizability.

Bias assessment

Publication bias was assessed by qualitative inspection of funnel plot asymmetry. To the best of our knowledge no quantitative method is available for the inspection of publication bias for a multi-level setting. Despite this limitation, we inspected funnel plots based on multivariate models without moderators, for each analysis separately. The funnel plots were adapted by using a measure of pooled standard deviation in the formula for precision (1/variance)[42]. Contrary to our initial study protocol, we did not conduct an Egger's regression. This test would have likely been underpowered due to the limited number of publications included[48]. Rather, we interpreted the probable influence of publication bias by qualitatively estimating the extent of the asymmetry based on areas of significance, as recommended by[48].

Software

The analyses were conducted in R (version 3.5.1) (R Core Team, 2015), using the following packages: 1) *metafor*[41] for conducting the analysis, 2) *metaforest*[47] for heterogeneity exploration, and 3) *dplyr* version 1.0.7[49] for general data handling. The

R scripts, data, and all materials are available on our Open Science Framework page (https://osf.io/4ngu3/).

Results

Study selection and characteristics An overview of the study design is summarized in the flow chart (Figure 1). Our pre-specified inclusion criteria (Methods) were met by 47 publications, published between 1996 and 2019. The included publications contributed 81 unique experiments for a total of 572 comparisons, from which we extracted statistical measurements (e.g. mean or median, standard error (SEM), deviation (SD) or interquantile range (IQR), and sample size (N)). After pre-processing, the number of comparisons used for analysis were 368. 1 publication[50] and 30 other comparisons (n_{study} = 5) were excluded from the analysis as it was not possible to obtain any statistical measurement.

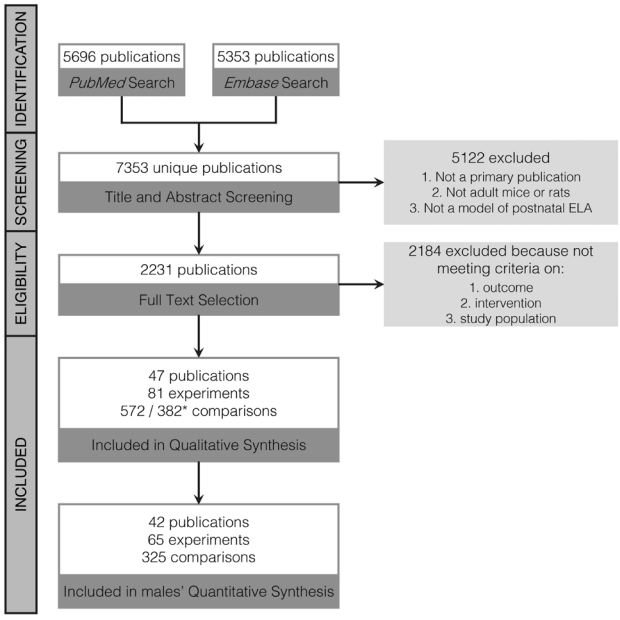


Figure 1. PRISMA flowchart of study selection process. * = 572 is the number of comparisons before preprocessing; 382 refers to the number included after processing. These values are for males and females combined. For information about preprocessing, see Methods.

One comparison was excluded from the analysis because the outcome was >3.29 SD away from the mean and it was also identified as an influential case. In this comparison, noradrenaline protein was measured with HPLC in female control and ELA animals from dams transported pregnant and who experienced the forced swim test 15 minutes prior. Although no other elements of the experimental design particularly stood out, the animals had several major traumatic life experiences during life, and we therefore

reasoned that the population studied in this publication was not comparable with the others.

Description of the population

The included publications used mainly rats (n_{publ} = 89.4%, Figure 2A) and the maternal separation ELA model (n_{publ} = 70.2%), followed by isolation (n_{publ} = 14.9%), material deprivation (n_{publ} = 10.6%) and lastly limited nesting and bedding (n_{publ} = 4.3%, Figure 2B). We included only adult animals (definition: older than 8 weeks but younger than 1 year), which in this particular dataset were aged between 8 and 22 weeks old (Figure 2C).

The experimental design of only 28 experiments (35%, n_{study} = 17) did not use elements that may be considered additional negative life events (Figure 2D). Specifically, they scored "no" in "major life events", did not score "purchased pregnant dams" in origin, and did not score "stressful" in behavior. Of these experiments, 22 experiments (n_{study} = 15) measured monoamines outcomes in naive animals, i.e. animals that did not experience any other event (neither prenatally nor postnatally) besides the experience of early life adversity or control early life condition.

The brain areas reported by publications were categorized in 10 unique brain areas (Figure 2E, Supplementary Table 2 for categorization) for our analysis, according to embryonic development specified by the Allen Brain Reference Atlas[43]. The brain areas with the majority of comparisons were: striatum, hippocampus and prefrontal cortex, for both sexes.

A total of ~1530 animals were used, 82.3% of which were males (Figure 2D). Seven publications performed experiments in both male and female rodents. Similarly to previous studies[23,28,34], we aimed to analyze males and females separately, as two different biological systems, since sex-dismorphic characteristics have been frequently observed in stress research[51–53]. However, due to the paucity of observations, it was possible to conduct meaningful quantitative synthesis only on the male dataset. The female dataset is available for exploration at https://osf.io/4ngu3/.

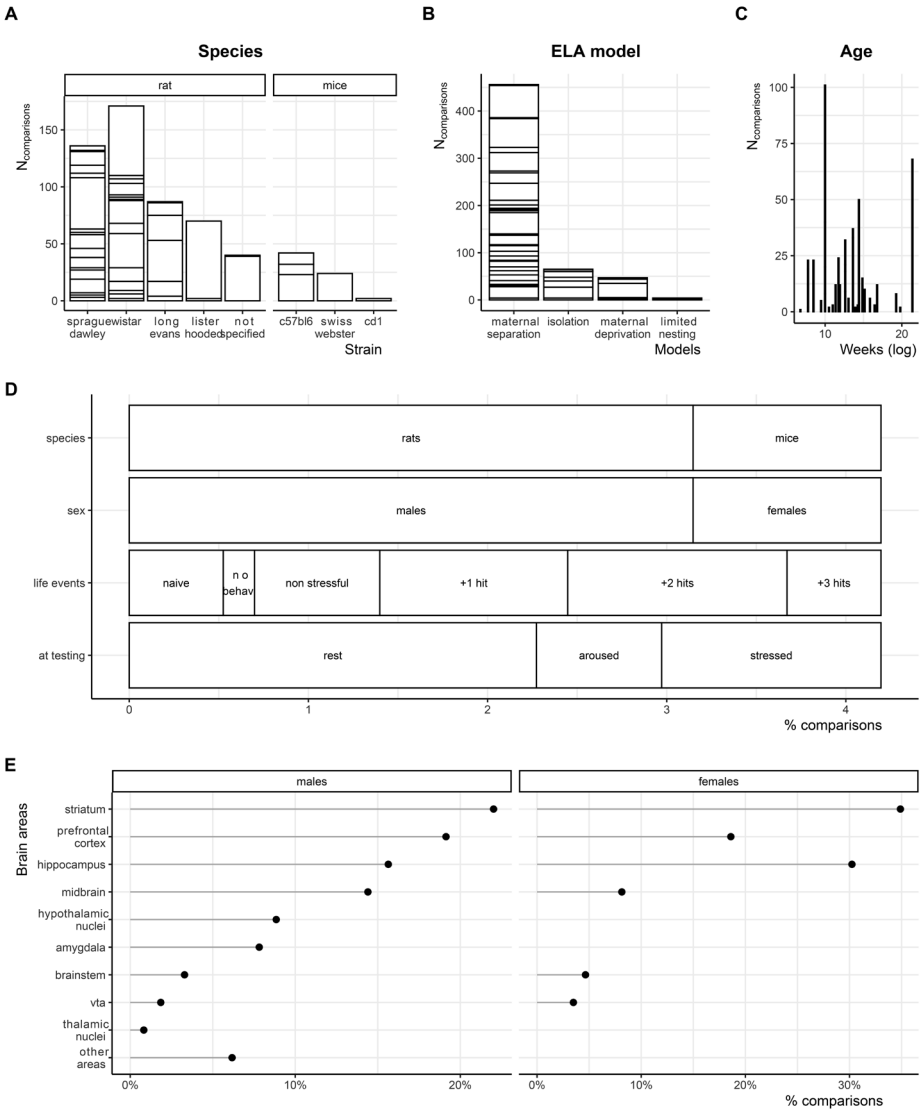


Figure 2. Study population. Desciptives of the number of comparisons (572 comparisons from 47 publications, see flowchart in Figure 1) for A) species, B) ELA model, and C) age (weeks). Concerning age, 5 studies only described the animals as "adults" but did not report the exact age at testing, and are therefore not visible in the graph. D) Percentage of comparisons for different independent variables. "Life events" describes which other experiences the animal had during life. Of note, for each comparison, both the control and ELA animals must have had the same experiences in order to be included in the study. Control and ELA animals always only differed in the experience of ELA. For example, "non stressful" means that both control and ELA animals performed non stressful behavior tasks (e.g. open field test) at least one day prior to measuring monoamines. Of note, in the final analysis, we only distinguished between animals that did not experience other major life events (i.e., here categorized as "naïve", "no behavior(al tests)" and "non stressful") and those who did (i.e., here categorized as "+1 hit", "+2 hits", "+3 hits" following the cumulative stress theory). "At testing" refers to the acute state (i.e. situation) of the animal just before death. In the final analysis with respect to acute state prior to death, we distinguished only between "rest" and "not rest", i.e. aroused and stressed together. E) Percentage of comparisons for each brain area, for males (n_{comp} =) and females (n_{comp} =) separately. behav. = behavior; vta = ventral tegmental area,

Quantitative effects of ELA on biochemical indicators of monoamines in males

In the quantitative synthesis, we included data from male animals at rest (n_{comp} = 239, n_{exp} = 56, n_{study} = 38, see Methods for preprocessing). Data on turnover were not included in the quantitative analysis to avoid redundancy (see study protocol); they are however reported in Supplementary Table 4.

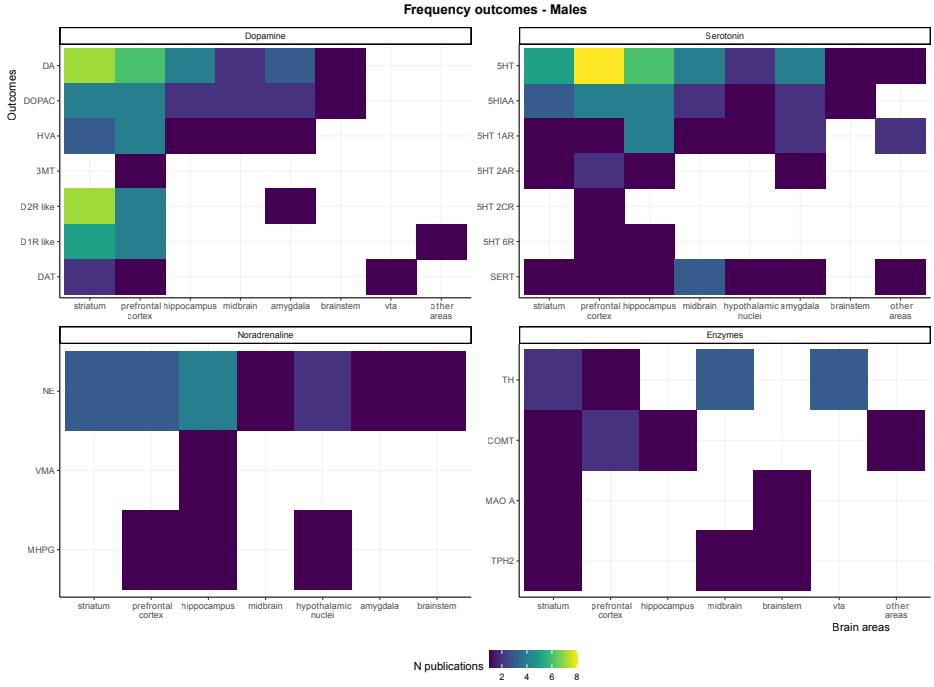


Figure 3. Distribution of the publications across monoamine measures and brain areas in males tested at rest. The most investigated outcomes are dopamine and its receptors in the striatum; and serotonin in the prefrontal cortex.

Figure 3 visualizes the distribution of male outcomes at rest across brain areas. Dopamine outcomes have been investigated mainly in the striatum and prefrontal cortex. Serotonin has been frequently investigated in several brain areas; however, its metabolites mainly in the prefrontal cortex and hippocampus. Information on the noradrenergic system is remarkably limited, with noradrenaline being mostly studied in hippocampus. White spaces indicate outcome-by-brain area combinations that have never been investigated, and are effectively "gaps" in the literature.

In our 3-level mixed effect model, we tested all possible (>3 publications, $n_{eligible\ comp}$ = 147) outcomes per brain area. To limit the probability of false positives, we set our alpha = 0.01. In this analysis, we performed 27 tests: 4 had p (non adjusted) below 0.05, of which 3 with $p < 0.01$. Specifically, the significant comparisons pointed

towards a decrease in dopamine receptors in the striatum (D1R-like$_{striatum}$: n_{study} = 6, $g(se)$ = -0.93(0.24), p < 0.001; D2R-like$_{striatum}$: n_{study} = 7, $g(se)$ = -0.671(0.22), p = 0.002, Figure 4) and in the prefrontal cortex (D2R-likeprefrontal: n_{study} = 4, g(se) = -1.235(0.34), p < 0.001, Figure 4). Concerning the serotonergic system, the current analysis points towards possible changes in expression of serotonin and its metabolites specifically in the prefrontal cortex (5HT$_{prefrontal}$: n_{study} = 8, $g(se)$ = -0.44(0.23), p = 0.057; 5HIAA$_{prefrontal}$: n_{study} = 4, $g(se)$ = 0.58(0.26), p = 0.026, Figure 4), although these outcomes do not meet the significance criterion chosen for the current study. All other results are reported in Supplementary Figures 1 to 4 and numerically in Supplementary Table 5.

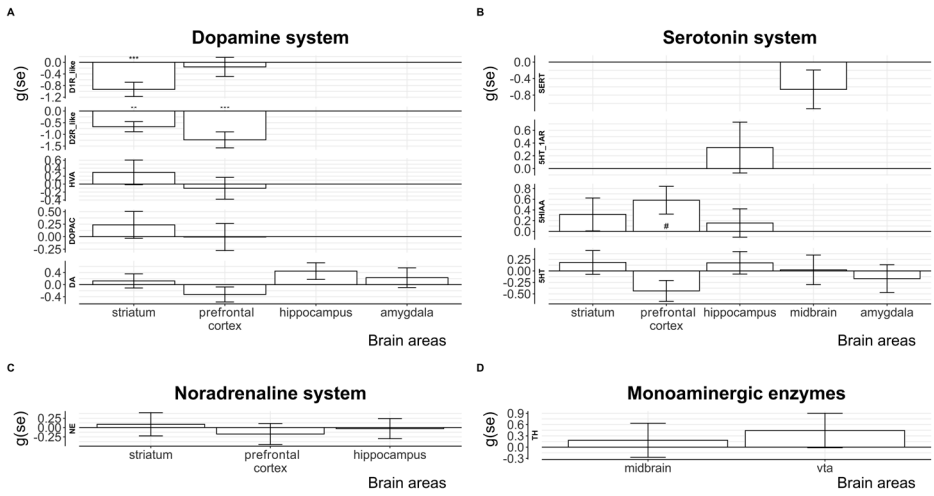


Figure 4. The effects of ELA in different brain areas in males at rest on A) dopamine system, B) Serotonin system, C) Noradrenaline system and D) enzymes involved in monoaminergic systems. Only comparisons with at least 3 publications were analyzed and here visualized. For all other comparisons, see Supplementary Figure 1 to 4 and Supplementary Table 5. *g = effect size Hedge's g; se = standard error ; DA = dopamine; DOPAC = 3,4-Dihydroxyphenylacetic acid (dopamine metabolite); HVA = homovanillic acid (dopamine metabolite); D1R = Dopamine receptors of the D1 type (i.e. D1-like); D2R = Dopamine receptors of the D2 type (i.e. D2-like); 5HT = serotonin; 5HIAA = 5-hydroxyindoleacetic acid (serotonin metabolite); 5HT 1AR = subtype of serotonin receptor; SERT = serotonin transporter; NE = noradrenaline; TH = tyrosine hydroxylase; VTA = ventral tegmental area; # = p< 0.05; ** = p < 0.01; *** p < 0.001.*

From subgroup analysis, overall the effects of ELA were strongest when the (control and ELA) animals experienced other traumatic life experiences (events$_{no\ other\ negative}$: n_{study} = 15, $g(se)$ = 0.66(0.08), p < 0.001; events$_{with\ other\ negative}$: n_{study} = 24, $g(se)$ = 0.838(0.11), p <0.001), however, the two groups were not significantly different from each other ($g(se)$ = 0.177(0.138), p = 0.201, Wald test).

Our 3-level model on the effects of ELA on male rodents at rest had a moderate remaining unexplained heterogeneity ($Q_E(152)$ = 393.99, p < .0001, I^2 = 66.08%; for interpretation, see[54] and Chapter 9). We therefore conducted an exploratory analysis to identify the most important moderators of the ELA effects on monoamines in male animals at rest. We used the machine learning algorhythm *Metaforest*, which uses random forests and bootstrap sampling to rank moderators based on how much

variance they can explain. However, none of the 9 potential moderators particularly stood out in explaining variance (Supplementary Figure 5). This was reflected in a non-significant cross validated R_{CV}^2 (R_{CV}^2 (SD) = 0.25 (0.34)), thereby suggesting a limited role of these moderators in explaining heterogeneity.

Exploratory analyses

Due to the limited availability of the data, we could not investigate in the main analysis two important factors: 1) life events interacting with the state of the animals just before death, and 2) sex differences. We therefore conducted two exploratory analyses to evaluate the effects of these two factors on the overall dataset. Of note, the data of this analysis partially overlaps with that used for the main analysis on male rodents at rest.

Life events From previous studies from our own lab[23,33,34], we hypothesized that there could be interactions between the state of the animal just before death and the experience of other negative life events. We selected a subset of the male data, i.e. only those outcomes per brain areas with observations for each of the following groups: 1) at rest with no other negative life experiences, 2) aroused/stressed with no other negative life experiences, 3) at rest with other negative life experiences, and 4) aroused/stressed with other negative life experiences. Importantly, we used only those outcomes arguably able to change in a short time frame, i.e. dopamine, serotonin and noradrenaline concentrations. The final dataset contained 64 observations (n_{study} = 16, n_{exp} = 2), containing 3 outcomes distributed in 5 brain areas.

In this subgroup, the effects of ELA were strongest when the animals were in an aroused/stressed state just prior to death, independently of the experience of other negative life events ($rest_{no\ other\ negative\ events}$: n_{study} = 4, $g(se)$ = -0.03(0.26), p = 0.9; $rest_{with\ other\ negative\ events}$: n_{study} = 10, $g(se)$ = 0.02(0.17), p = 0.91; not $rest_{no\ other\ negative\ events}$: n_{study} = 2, $g(se)$ = -1.2(0.46), p = 0.014; not $rest_{with\ other\ negative\ events}$: n_{study} = 3, $g(se)$ = -0.87(0.29), p = 0.004, Figure 5A).

Sex differences To compare male and female data, we performed an exploratory analysis only on those publications investigating outcomes in males and females at rest. This subset of the data consisted of 6 publications investigating 25 matched outcomes in males and females. We used the same model as for the main analysis, but this time we added sex as an additional moderator. In this matched dataset, the effects of ELA were more pronounced in females than in males (females: g(se) = -0.43(0.13), p = 0.007; males: $g(se)$ = -0.04(0.12), Figure 9B), although the direct comparison was not significant according to our set alpha of 0.01 ($g(se)$ = -0.39(0.18), p = 0.03). These effects were stable also if the absolute values of g were considered, and if sex was used as the only univariate moderator. These analyses are available at https://osf.io/4ngu3/.

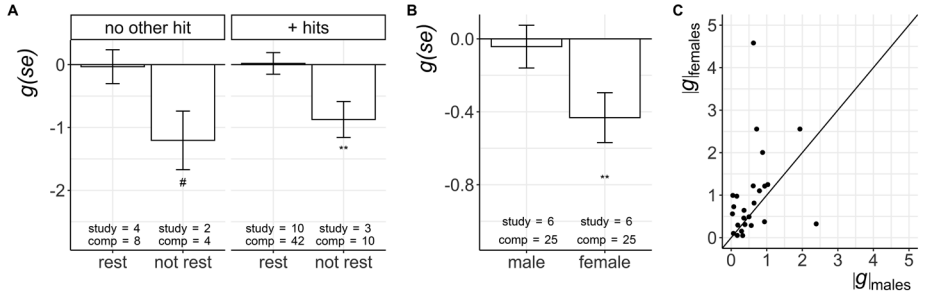


Figure 5. Exploratory analyses. A) Exploratory analysis on the interaction between ELA, state of the animal just before death (here: rest vs not rest) and the experience of other negative life events (here: no other hit vs + hits). The effects of ELA are increased when animals are tested in an aroused or stressed state (i.e. not rest) regardless of having experienced additional major life events. B) The effects of ELA are stronger in females than in males, in a subset of studies that investigated both sexes at rest on the same outcomes. C) Effect size estimate (g in absolute values for visualization) of males and females. Each dot corresponds to an outcome of a specific paper that investigated both sexed. If all estimates would be identical in males and females, they would follow the line.

w

Publication bias assessment

We evaluated publication bias by assessing funnel plot asymmetry (Figure 6) separately for each analysis conducted, specifically for the main analysis, and the two subgroup analyses (i.e. other life events and sex differences). Concerning the main analysis (i.e. males at rest), several comparisons between ELA and control animals are in the highest area of significance; however, from qualitative assessment, there is a fair symmetry, suggesting that the effect estimates are unlikely to be pushed towards one particular direction due to publication bias. Concerning the subgroup analyses, the funnel plots (Figure 6 B and C) have similar characteristics to that of the main analysis, but here the violations are less severe. We therefore conclude that there is a limited evidence of publication bias.

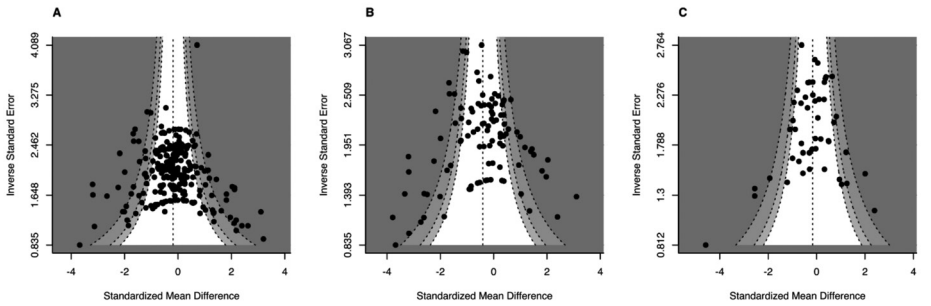


Figure 6. Funnel plot for assessment of publication bias for effects of ELA on (A) males at rest, (B) the influence of other life events and (C) sex differences analyses. Each dot corresponds to a comparison between control and ELA groups. Dark gray = areas of highest significance (p-value). The darker the grey the higher the significance.

Discussion

We conducted a systematic review with meta-analysis to quantify the effects of ELA on monoaminergic systems in adult mice and rats. We focused on dopaminergic, serotonergic and noradrenergic systems, whose biochemical indicators were investigated in several brain regions. Our review is restricted to postnatal ELA models that are based on altered maternal care during the first three postnatal weeks. The results show that dopamine was mostly investigated in striatum and prefrontal cortex, while studies involving serotonin were mostly performed in the prefrontal cortex and hippocampus. Surprisingly, noradrenergic markers were investigated in only 11 publications. Despite its important role especially in relation to acute stress, noradrenaline is therefore understudied. The literature on males was far more extensively available than that of females, and the former group was therefore the primary focus of our quantitative synthesis with meta-analysis. Specifically, we focused on monoaminergic outcomes in adult male rodents tested at rest. The choice of selection of the acute situation was justified by previous work from our own lab, where the acute situation interacted with life events in the expression of the ELA phenotype on immediate early genes' expression[34].

In our analysis in male rodents at rest, the emerging picture supports a mildly hypofunctional dopaminergic system especially in the striatum, particularly at the level of dopaminergic receptors. While not all monoaminergic outcomes could be meta-analysed due to the limited number of publications, the hypoactive state of the monoaminergic systems of the brain of male ELA animals is also supported by qualitative systematic review of the literature. Specifically, several studies point towards a lower level of serotonin[55–61] and dopamine[62–68,68,69] receptors in several other brain areas. The serotonin transporter was found to be increased in the striatum[61] of ELA animals when compared to controls, which suggests a higher re-uptake. Concurrently, ELA animals were found to have higher expression levels of the *MAO A* enzyme in the striatum[70], which suggests a faster degradation of monoamines. ELA animals also had a lower expression of tryptophan hydroxylase, the rate limiting enzyme in the biosynthesis of serotonin, an effect previously highlighted in other reviews[71]. While these studies were conducted with heterogeneous designs in different brain areas, they point towards a hypo-functioning of the monoaminergic systems. These results are in line with human observations, where ELA is a well-documented risk factor for the development of major depression[4], and imbalances in monoaminergic function in limbic areas are thought to underlie depressive symptoms[72]. Furthermore, higher serotonin turover rate was identified in depressed patients when compared to healthy volunteeers[71,73]. Future studies should extend the current insights and concurrently investigate multiple brain areas, in rodents as well as humans. This could also reveal whether the hypoactive monoaminergic tone is a property of the entire system or specific to distinct brain areas and cognitive/emotional functions.

The results of our main analysis were limited to male rodents at rest. It is however

likely that the results are different when the animals are in an "activated" state (for example, while performing a task), be it aroused or stressed. These differences in acute states have been observed at a behavioral level in previous meta-analyses, where ELA was found to increase memory after stressful learning but decrease memory related to non-stressful learning[23,24]. It is therefore likely that these major behavioral differences in rodents with an ELA history are matched by changes in neurotransmission, as previously highlighted in relation to immediate early genes' expression[34]. There is only a limited number of publications investigating the effects of ELA on animals in a aroused/stressed state, so that meta-analyses on each monoaminergic outcome was not possible. We here performed a limited and exploratory analysis evaluating the moderating effects of the acute state of the animal just before death (at testing), in relation to the presence or absence of other major negative life events. In this exploratory analysis, we selected only those outcomes likely to change on a short time-frame, i.e. dopamine, serotonin and noradrenaline concentrations: rapid state-dependent shifts in transmitter concentrations intuitively seem more likely to occur than rapid shifts in e.g. receptor expression, although the latter cannot be entirely ruled out. Of note, both control and ELA animals were in the same state / other life event category, therefore the results should be interpreted as the effects of ELA related to a specific state/category. Our exploratory analysis suggests that monoamine concentrations are lower in ELA animals compared to controls when the animals are in an aroused/stressed situations. Rather than reflecting an absolute reduction in dopamine concentration, this result should be interpreted as a smaller increase from rest/baseline in the ELA group compared to the control group. These results are in discordance with what was observed in humans. PET studies in humans demonstrated that the release of is associated with psychosocial stress[74]; this phenotype was most evident in individuals with a history of early life adversity[75] or psychosis[76]. In line with this reasoning, selective serotonin and noradrenaline re-uptake inhibitors as well as NASSAs (noradrenergic and specific serotonergic antidepressants) are drugs frequently used in the treatment of depressive symptoms. These drugs have in common that – in the short term – they increase the concentration of serotonin and noradrenaline in the synaptic cleft and serotonin-mediated neurotransmission[18]. All in all, our systematic review and meta-analytic results support the existence of a (mild) hypo-active state of the monoaminergic systems in rodents with a history of ELA. It remains to be proven of course whether this hypo-active state is a direct consequence of ELA or involves intermediate steps at timepoints between early life and the (young) adult state reported in the current dataset. Also, while there are parallels with the human literature, there are also important discrepancies. These discrepancies might be linked to methodological (e.g. techniques, population, brain areas) elements but also to species-dependent properties. Finally, the present conclusions are based on only a few publications; the emerging pictures require regular updating and might still

change in the future. Future experiments would definitely benefit from taking the acute situation of the individual at testing into account.

We started this review by assuming differences between male and female rodents with regard to long-term consequences of ELA. This was primarily based on clinical populations, where females seem more susceptible than males to childhood adversities[77]. Incidentally, this was not replicated in rodent models of ELA so far[23,24]. One could even reason that ELA models were originally developed for males[78,79], plausibly rendering males more susceptible than females to ELA models. In the current study, the number of female publications was insufficient for a complete meta-analysis of female data, in line with what was previously observed[33,34]. We therefore opted for a different approach, and performed an exploratory analysis to test whether there are intrinsic sex-specific differences of ELA on monoaminergic outcomes. To our surprise (and contrary to what was reported before for behavioral endpoints[23], the effects of ELA were more pronounced in females than in males when considering only those publications that used both sexes, although the direct comparison with Wald-type test was not significant. It remains to be established whether these differences truly reflect an ELA sex-specific vulnerability, or are linked to other experimental or reporting biases. Possibly, the results are linked to intrinsic sex differences in monoamine systems. For instance, both in rats and non-human primates females have higher expression of dopamine receptors than males[80], which – in humans – may increase the predisposition for depressive-like and anxiety-like behavior[80]. Similarly, the expression of monoaminergic transporters was increased in females with depression compared to males with the same diagnosis[81]. The higher expression levels could make measurements more sensitive, and thereby ELA effects more evident.

The current analysis also has important limitations. In our study, we included only ELA models of altered maternal care in the first 3 postnatal weeks. This does not capture the entire period during which monoaminergic systems are remodelled throughout development, which occurs from the prenatal period until adolescence (for a review, see[82]). This remodelling is likely to occur in critical periods[83], with different parts of the system developing at transient stages. As a consequence, although the ELA models used have been considered translationally valid[38,84,85], it is difficult to establish which exact monoaminergic critical periods they directly impact, and which are changed as adaptive mechanisms. Secondly, our main analysis had a moderate remaining heterogeneity, which could not be further explained by our exploratory analysis with MetaForest. Qualitatively unexplained heterogeneity reduces the certainty in the evidence[86]. The present observation is in line with our previous analysis[28], where species and assessment method were not significant moderators of the effects of prenatal nor postnatal ELA on biochemical indicators of the dopaminergic system. The current absence of additional potential moderators does not prove that such

moderators do not exist. Future studies should evaluate potential subgroup effects, which may become clearer with an increased pool of available publications and improved reporting (e.g. of the situation at death and other potential biases). Such analyses would be essential to better evaluate the replicability of findings across a wide range of methodological differences.

Despite these limitations, the current study highlights that ELA may induce a hypo-activity of monoaminergic systems at rest. Consistent with previous studies, the biochemical effects observed are rather limited, when compared to the behavioral phenotype to which they supposedly contribute. Our exploratory analyses suggest interesting sex differences, as well as potential interactions between the effects of ELA and the acute situation of the animal at testing. These observations should be further explored in future studies.

References

1. Glynn, L. M. & Baram, T. Z. The Influence of Unpredictable, Fragmented Parental Signals on the Developing Brain. *Front Neuroendocrinol* 53, 100736 (2019).
2. Miguel, P. M., Pereira, L. O., Silveira, P. P. & Meaney, M. J. Early environmental influences on the development of children's brain structure and function. *Dev Med Child Neurol* 61, 1127–1133 (2019).
3. Corney, K. B. et al. Relationship between adverse childhood experiences and Alzheimer's disease: a systematic review and meta-analysis protocol. *BMJ Open* 11, e049768 (2021).
4. Heim, C. & Binder, E. B. Current research trends in early life stress and depression: review of human studies on sensitive periods, gene-environment interactions, and epigenetics. *Exp Neurol* 233, 102–111 (2012).
5. Teicher, M. H., Gordon, J. B. & Nemeroff, C. B. Recognizing the importance of childhood maltreatment as a critical factor in psychiatric diagnoses, treatment, research, prevention, and education. *Mol Psychiatry* (2021) doi:10.1038/s41380-021-01367-9.
6. Herzberg, M. P. & Gunnar, M. R. Early life stress and brain function: Activity and connectivity associated with processing emotion and reward. *Neuroimage* 209, 116493 (2020).
7. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog Neurobiol* 106–107, 1–16 (2013).
8. Cools, R. & Arnsten, A. F. T. Neuromodulation of prefrontal cortex cognitive function in primates: the powerful roles of monoamines and acetylcholine. *Neuropsychopharmacology* 47, 309–328 (2022).
9. Bell, E., Boyce, P., Porter, R. J., Bryant, R. A. & Malhi, G. S. Irritability in Mood Disorders: Neurobiological Underpinnings and Implications for Pharmacological Intervention. *CNS Drugs* 35, 619–641 (2021).
10. Tanaka, M., Yoshida, M., Emoto, H. & Ishii, H. Noradrenaline systems in the hypothalamus, amygdala and locus coeruleus are involved in the provocation of anxiety: basic studies. *Eur J Pharmacol* 405, 397–406 (2000).
11. Azizi, S. A. Monoamines: Dopamine, Norepinephrine, and Serotonin, Beyond Modulation, "Switches" That Alter the State of Target Networks. *Neuroscientist* 1073858420974336 (2020) doi:10.1177/1073858420974336.
12. Lin, Z. et al. Monoamine transporters: vulnerable and vital doorkeepers. Prog Mol Biol Transl Sci 98, 1–46 (2011).
13. Martin, E. I., Ressler, K. J., Binder, E. & Nemeroff, C. B. The Neurobiology of Anxiety Disorders: Brain Imaging, Genetics, and Psychoneuroendocrinology. *Psychiatric Clinics of North America* 32, 549–575 (2009).
14. Yamamoto, K., Shinba, T. & Yoshii, M. Psychiatric symptoms of noradrenergic dysfunction: A pathophysiological view. *Psychiatry and Clinical Neurosciences* 68, 1–20 (2014).
15. Ripke, S. et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. Nature 511, 421–427 (2014).
16. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* 22, 343–352 (2019).
17. Cabana-Domínguez, J., Torrico, B., Reif, A., Fernàndez-Castillo, N. & Cormand, B. Comprehensive exploration of the genetic contribution of the dopaminergic and serotonergic pathways to psychiatric disorders. *Transl Psychiatry* 12, 11 (2022).
18. Aggarwal, S. & Mortensen, O. V. Overview of Monoamine Transporters. *Current Protocols in Pharmacology* 79, 12.16.1-12.16.17 (2017).
19. Kasanova, Z. et al. Early-Life Stress Affects Stress-Related Prefrontal Dopamine Activity in Healthy Adults, but Not in Individuals with Psychotic Disorder. *PLOS ONE* 11, e0150746 (2016).
20. Pani, L., Porcella, A. & Gessa, G. L. The role of stress in the pathophysiology of the dopaminergic system. *Mol Psychiatry* 5, 14–21 (2000).
21. Baumann, C. et al. The interaction of early life experiences with COMT val158met affects anxiety sensitivity. *Genes Brain Behav* 12, 821–829 (2013).
22. Kang, H.-J. et al. Association of SLC6A4 methylation with early adversity, characteristics and outcomes in depression. *Prog Neuropsychopharmacol Biol Psychiatry* 44, 23–28 (2013).
23. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neuroscience & Biobehavioral Reviews* 102, 299–307 (2019).
24. Rocha, M., Wang, D., Avila-Quintero, V., Bloch, M. H. & Kaffman, A. Deficits in hippocampal-dependent memory across different rodent models of early life stress: systematic review and meta-analysis. *Transl Psychiatry* 11, 231 (2021).
25. Luijk, J. van et al. Systematic Reviews of Animal Studies; Missing Link in Translational Research? *PLOS ONE* 9, e89981 (2014).
26. Miake-Lye, I. M., Hempel, S., Shanman, R. & Shekelle, P. G. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Reviews* 5, 28 (2016).
27. Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R. & Howells, D. W. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. *J Cereb Blood Flow Metab* 34, 737–742 (2014).
28. Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: A 3 level meta-analysis of rodent studies. *Neuroscience & Biobehavioral Reviews* 95, 1–16 (2018).
29. de Vries, R. B. M. et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclinical Medicine* 2, e00007 (2015).

30. Leenaars, M. et al. A step-by-step guide to systematically identify all relevant animal studies. *Lab Anim* 46, 24–31 (2012).
31. Hooijmans, C. R. et al. SYRCLE's risk of bias tool for animal studies. BMC Medical Research Methodology 14, 43 (2014).
32. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372, n71 (2021).
33. Joëls, M., Kraaijenvanger, E., Sarabdjitsingh, R. A. & Bonapersona, V. Structural changes after early life adversity in rodents: a systematic review with meta-analysis. 2022.01.17.476580 (2022) doi:10.1101/2022.01.17.476580.
34. Schuler, H., Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. Effects of early life adversity on immediate early gene expression: Systematic review and 3-level meta-analysis of rodent studies. *PLOS ONE* 17, e0253406 (2022).
35. Champagne, F. A., Francis, D. D., Mar, A. & Meaney, M. J. Variations in maternal care in the rat as a mediating influence for the effects of environment on development. *Physiology & Behavior* 79, 359–371 (2003).
36. Levine, S. Regulation of the hypothalamic-pituitary-adrenal axis in the neonatal rat: the role of maternal behavior. *Neurotox Res* 4, 557–564 (2002).
37. Sánchez, M. M., Ladd, C. O. & Plotsky, P. M. Early adverse experience as a developmental risk factor for later psychopathology: evidence from rodent and primate models. *Dev Psychopathol* 13, 419–449 (2001).
38. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A Novel Mouse Model for Acute and Long-Lasting Consequences of Early Life Stress. *Endocrinology* 149, 4892–4900 (2008).
39. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews 5, 210 (2016).
40. Rohatgi, A. WebPlotDigitizer. https://automeris.io/WebPlotDigitizer (2020).
41. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 36, 1–48 (2010).
42. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221, 92–102 (2014).
43. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953.e20 (2020).
44. Cheung, M. W.-L. A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes. *Neuropsychol Rev* 29, 387–396 (2019).
45. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21, 1539–1558 (2002).
46. Cheung, M. W.-L. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol Methods* 19, 211–229 (2014).
47. Lissa, C. J. van. Small Sample Meta-Analyses: Exploring heterogeneity using MetaForest. in *Small Sample Size Solutions* (Routledge, 2020).
48. Sterne, J. A. C. et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343, d4002 (2011).
49. Wickham, H., François, R., Henry, L., Müller, K. & RStudio. *dplyr: A Grammar of Data Manipulation*. (2021).
50. Shin, S. et al. Drd3 Signaling in the Lateral Septum Mediates Early Life Stress-Induced Social Dysfunction. *Neuron* 97, 195-208.e6 (2018).
51. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. *Nat Neurosci* 18, 1413–1420 (2015).
52. Bangasser, D. A. & Cuarenta, A. Sex differences in anxiety and depression: circuits and mechanisms. *Nat Rev Neurosci* 1–11 (2021) doi:10.1038/s41583-021-00513-0.
53. Hodes, G. E. & Epperson, C. N. Sex Differences in Vulnerability and Resilience to Stress Across the Life Span. *Biol Psychiatry* 86, 421–432 (2019).
54. Higgins, Julian PT et al. Cochrane Handbook for Systematic Reviews of Interventions. in *Cochrane Handbook for Systematic Reviews of Interventions* i–xxi (John Wiley & Sons, Ltd, 2008). doi:10.1002/9780470712184.fmatter.
55. Benekareddy, M., Goodfellow, N. M., Lambe, E. K. & Vaidya, V. A. Enhanced function of prefrontal serotonin 5-HT(2) receptors in a rat model of psychiatric vulnerability. *J Neurosci* 30, 12138–12150 (2010).
56. Bravo, J. A., Dinan, T. G. & Cryan, J. F. Early-life stress induces persistent alterations in 5-HT1A receptor and serotonin transporter mRNA expression in the adult rat brain. *Front Mol Neurosci* 7, 24 (2014).
57. Gartside, S. E., Johnson, D. A., Leitch, M. M., Troakes, C. & Ingram, C. D. Early life adversity programs changes in central 5-HT neuronal function in adulthood. *Eur J Neurosci* 17, 2401–2408 (2003).
58. Maniam, J., Antoniadis, C. P., Le, V. & Morris, M. J. A diet high in fat and sugar reverses anxiety-like behaviour induced by limited nesting in male rats: Impacts on hippocampal markers. *Psychoneuroendocrinology* 68, 202–209 (2016).
59. Marcos, B., Aisa, B. & Ramírez, M. J. Functional interaction between 5-HT(6) receptors and hypothalamic-pituitary-adrenal axis: cognitive implications. *Neuropharmacology* 54, 708–714 (2008).
60. Rentesi, G. et al. Early maternal deprivation-induced modifications in the neurobiological, neurochemical and behavioral profile of adult rats. *Behav Brain Res* 244, 29–37 (2013).
61. Vicentic, A. et al. Maternal separation alters serotonergic transporter densities and serotonergic 1A receptors in rat brain. *Neuroscience* 140, 355–365 (2006).
62. Brake, W. G., Zhang, T. Y., Diorio, J., Meaney, M. J. & Gratton, A. Influence of early postnatal rearing conditions on

mesocorticolimbic dopamine and behavioural responses to psychostimulants and stressors in adult rats. *Eur J Neurosci* 19, 1863–1874 (2004).
63. Brenhouse, H. C., Lukkes, J. L. & Andersen, S. L. Early life adversity alters the developmental profiles of addiction-related prefrontal cortex circuitry. *Brain Sci* 3, 143–158 (2013).
64. Hill, R. A. et al. Long-term effects of combined neonatal and adolescent stress on brain-derived neurotrophic factor and dopamine receptor expression in the rat forebrain. *Biochim Biophys Acta* 1842, 2126–2135 (2014).
65. Lejeune, S. et al. The dopamine D1 receptor agonist SKF 38393 improves temporal order memory performance in maternally deprived rats. *Neurobiol Learn Mem* 106, 268–273 (2013).
66. Ploj, K., Roman, E. & Nylander, I. Long-term effects of maternal separation on ethanol intake and brain opioid and dopamine receptors in male Wistar rats. *Neuroscience* 121, 787–799 (2003).
67. Romano-López, A. et al. Maternal separation and early stress cause long-lasting effects on dopaminergic and endocannabinergic systems and alters dendritic morphology in the nucleus accumbens and frontal cortex in rats. *Dev Neurobiol* 76, 819–831 (2016).
68. Zhang, Y. et al. Maternal deprivation enhances behavioral vulnerability to stress associated with miR-504 expression in nucleus accumbens of rats. *PLoS One* 8, e69934 (2013).
69. Zhang, Y. et al. Dopamine Receptor D2 and Associated microRNAs Are Involved in Stress Susceptibility and Resistance to Escitalopram Treatment. *Int J Neuropsychopharmacol* 18, pyv025 (2015).
70. Wong, P. et al. Early life environmental and pharmacological stressors result in persistent dysregulations of the serotonergic system. *Front Behav Neurosci* 9, 94 (2015).
71. Hale, M. W., Shekhar, A. & Lowry, C. A. Development by environment interactions controlling tryptophan hydroxylase expression. *J Chem Neuroanat* 41, 219–226 (2011).
72. Edinoff, A. N. et al. Selective Serotonin Reuptake Inhibitors and Clozapine: Clinically Relevant Interactions and Considerations. *Neurol Int* 13, 445–463 (2021).
73. Barton, D. A. et al. Elevated brain serotonin turnover in patients with depression: effect of genotype and therapy. *Arch Gen Psychiatry* 65, 38–46 (2008).
74. Lataster, J. et al. Psychosocial stress is associated with in vivo dopamine release in human ventromedial prefrontal cortex: a positron emission tomography study using [18F]fallypride. *Neuroimage* 58, 1081–1089 (2011).
75. Pruessner, J. C., Champagne, F., Meaney, M. J. & Dagher, A. Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: a positron emission tomography study using [11C] raclopride. *J Neurosci* 24, 2825–2831 (2004).
76. Soliman, A. et al. Stress-induced dopamine release in humans at risk of psychosis: a [11C]raclopride PET study. *Neuropsychopharmacology* 33, 2033–2041 (2008).
77. Walker, C.-D. et al. Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* 20, 421–448 (2017).
78. Oomen, C. A. et al. Opposite Effects of Early Maternal Deprivation on Neurogenesis in Male versus Female Rats. *PLOS ONE* 4, e3675 (2009).
79. van Hasselt, F. N., Boudewijns, Z. S. R. M., van der Knaap, N. J. F., Krugers, H. J. & Joëls, M. Maternal Care Received by Individual Pups Correlates with Adult CA1 Dendritic Morphology and Synaptic Plasticity in a Sex-Dependent Manner. *Journal of Neuroendocrinology* 24, 331–340 (2012).
80. Hasbi, A. et al. Sex difference in dopamine D1-D2 receptor complex expression and signaling affects depression- and anxiety-like behaviors. *Biol Sex Differ* 11, 8 (2020).
81. Bristow, G. C., Eisenlohr-Moul, T., Lotesto, K. & Sodhi, M. S. Sex differences in the transcription of monoamine transporters in major depression. *J Affect Disord* 295, 1215–1219 (2021).
82. Suri, D., Teixeira, C. M., Cagliostro, M. K. C., Mahadevia, D. & Ansorge, M. S. Monoamine-Sensitive Developmental Periods Impacting Adult Emotional and Cognitive Behaviors. *Neuropsychopharmacol* 40, 88–112 (2015).
83. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nat Rev Neurosci* 6, 877–888 (2005).
84. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr Opin Psychol* 15, 174–181 (2017).
85. Suchecki, D. Maternal regulation of the infant's hypothalamic-pituitary-adrenal axis stress response: Seymour 'Gig' Levine's legacy to neuroendocrinology. *Journal of Neuroendocrinology* 30, e12610 (2018).
86. Guyatt, G. H. et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336, 924–926 (2008).

CHAPTER 7

Effects of early life adversity on immediate early gene expression: systematic review and 3-level meta-analysis of rodent studies

Plos One (2022), doi: 10.1371/journal.pone.0253406

Schuler H[1,2,¶]
Bonapersona V[1,¶]
Joëls M[1,3,&]
Sarabdjitsingh RA[1,&]

1 Dept Translational Neuroscience, UMC Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands
2 Integrated Program in Neuroscience, McGill University, Montréal, QC, Canada.
3 University Medical Center Groningen, Groningen, The Netherlands

¶ Contributed equally;
& Contributed equally

Abstract

Early-life adversity (ELA) causes long-lasting structural and functional changes to the brain, rendering affected individuals vulnerable to the development of psychopathologies later in life. Immediate-early genes (IEGs) provide a potential marker for the observed alterations, bridging the gap between activity-regulated transcription and long-lasting effects on brain structure and function. Several heterogeneous studies have used IEGs to identify differences in cellular activity after ELA; systematically investigating the literature is therefore crucial for comprehensive conclusions. Here, we performed a systematic review on 39 pre-clinical studies in rodents to study the effects of ELA (alteration of maternal care) on IEG expression. Females and IEGs other than cFos were investigated in only a handful of publications. We meta-analyzed publications investigating specifically cFos expression. ELA increased cFos expression after an acute stressor only if the animals (control and ELA) had experienced additional hits. At rest, ELA increased cFos expression irrespective of other life events, suggesting that ELA creates a phenotype similar to naïve, acutely stressed animals. We present a conceptual theoretical framework to interpret the unexpected results. Overall, ELA likely alters IEG expression across the brain, especially in interaction with other negative life events. The present review highlights current knowledge gaps and provides guidance to aid the design of future studies.

Introduction

Synaptic connections in the brain are continuously altered, including via gene expression, to accommodate experiences, thereby preparing the organism to deal with future events [1–3]. This potential for adaptation, called neuronal or synaptic plasticity, is prominently present during critical periods early in life[4]. For this reason, adverse experiences throughout childhood – such as physical, sexual or emotional abuse – have far-reaching effects on an individual's brain function and structure, and consequently on cognition and behavior [5–7]. It is therefore not surprising that *early-life adversity* (ELA) is consistently associated with an increased risk for psychopathologies later in life, including major depressive disorder (MDD), post-traumatic stress disorder (PTSD), and schizophrenia [8,9].

To investigate the mechanisms underlying the effects of ELA on brain and behavior, several models of alteration of maternal care in rodents have been developed [10,11]. These models consistently show that ELA leads to fundamental remodeling of stress-sensitive brain regions, which in turn may be linked to altered function [12,13]. For example, ELA has been reported to modify the regulatory response of the *hypothalamic-pituitary-adrenal* (HPA) axis, an essential part of the organism's stress response system [14,15]. Furthermore, rodents exposed to ELA display a robust behavioral phenotype characterized by enhanced anxiety-like behavior, changes in memory formation, and decreased social behavior [16–19]. Overall, this evidence highlights that ELA leads to structural, functional and behavioral alterations in the rodent brain, yet the events giving rise to the said alterations remain unclear.

Immediate-early genes (IEGs), such as *cFos* (alias *Fos*), *Egr1* (alias *Zif-268*, *NGFI-A, Krox-24*) and *Arc* (alias *Arg3.1*), provide a potential link between experience-induced cellular activity in the brain and the resulting long-term changes in neurons and synapses. IEGs are immediately and transiently expressed in response to extracellular calcium influx, as occurs when an action potential is fired [20]. Among the IEGs, *cFos* is most often studied; it forms the activator protein-1 (AP1) by dimerization with a *Jun*-family transcription factor [21]. The AP1 complex initiates the transcription of other late genes, which result in long-lasting changes of cellular physiology. Consequently, a strong relationship between IEG expression and neuronal activity is observed, with increases in neuronal activity being accompanied by increased IEG expression [20]. For decades, IEGs have been a prominent tool for mapping neuronal activity in rodents by means of immunohistochemistry (IHC) and in-situ hybridization (ISH) due to their brain-wide expression. More recently, IEGs have been increasingly investigated for their protein properties, in particular with respect to synaptic plasticity [22].

Whereas the downstream products of IEGs are diverse (e.g., transcription factors, postsynaptic proteins, secretory factors), their functions are surprisingly homogeneous and can mostly be related to cellular processes, such as dendrite and spine development; synapse formation, strength and elimination; and regulation of the

excitatory/inhibitory balance[3]. In line with this functional similarity, knockouts (KOs) of several different IEGs affect behavior and synaptic plasticity in a similar manner. More specifically, system-wide *Arc*-KO and *Egr1*-KO, as well as central nervous system-specific cFos-KO mice all display behavioral impairments in learning and memory as well as deficits in long-term potentiation or depression, underscoring the necessity of IEGs for memory formation and retention [23–25]. In addition, many neuropsychiatric disorders characterized by memory impairments, such as major depressive disorder, post-traumatic stress disorder and schizophrenia, have also been shown to feature a dysregulation of activity-dependent transcription [26]. Interestingly, the risk to develop any of these disorders is increased by exposure to ELA, further indicating a potential causal interaction between ELA, IEGs and mental health [8,9].

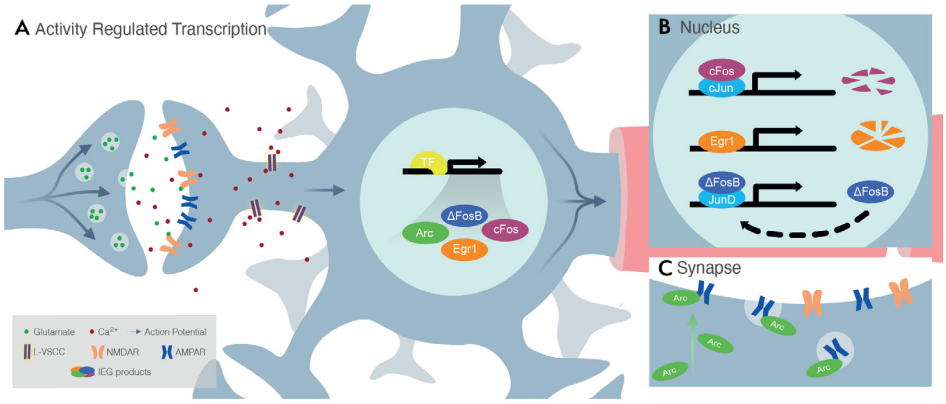


Fig 1. Mechanisms of IEG activation. A) Action potentials induced by glutamate signaling result in membrane depolarization, which in turn results in opening of L-type voltage calcium channels (LVCCs). The resulting Ca2+ influx induces calcium-dependent signaling pathways. These cascades further result in the recruitment of existing transcription factor, such as CREB, which in turn lead to the expression of IEGs. Once transcribed, IEGs act as B) transcription factors in the nucleus or C) regulators of synaptic plasticity at the synapse as, for example, post-synaptic proteins. B) The transcription factors of the Fos family bind to a transcription factor of the Jun family to form the AP1 complex, whereas Egr1 acts independently. Egr1 and cFos are transiently expressed, whereas FosB accumulates over time in the nucleus. C) Arc acts at the post-synaptic density by reducing the number of surface AMPA receptors. Therefore, increased Arc expression results in reduced synaptic strength by AMPA receptor endocytosis.

While numerous studies have used IEGs to identify differences in cellular activity after ELA, the study designs are heterogeneous, and findings are seemingly discrepant. Reviewing the available literature will provide a clearer picture of the effects of ELA on IEG expression and will aid future development of study designs by identifying sources of heterogeneity within and between experiments. To that end, we performed a systematic review to synthesize the available evidence and explore outcomes in a sex-, gene- and region-specific manner. A meta-analysis was then conducted on a subset of the data based on *a priori* determined thresholds. We hypothesized that ELA

as alteration of maternal care leads to an exaggerated increase in IEG expression after an acute stress challenge, further amplified by exposure to additional hits in life, in line with the multiple-hit concept of vulnerability [27].

Methods

Search strategy, protocol and risk of bias assessment of the present review were performed in line with SYRCLE (Systematic Review Center for Laboratory animal Experimentation) guidelines [28–30]. We adhered to the PRISMA checklist for reporting [31] (Supporting Information). The protocol (S1.1) and the PRISMA checklist are openly accessible at https://osf.io/qkyvd/.

Study selection and data extraction

We conducted a systematic literature search with the search engines PubMed and Embase on the 3rd of April 2019 to select experiments investigating differences in IEG expression between control and ELA exposed rodents. The terms *'mice and rats'* and *'postnatal ELA'* were used to construct the search string (S1.2). For the purpose of this review, ELA was defined as models altering maternal care. We included the ELA models of maternal separation and deprivation, isolation, limited bedding and nesting, as well as licking and grooming. Study selection was performed in Rayyan [32] in alphabetical order and any disagreements between investigators were resolved by discussion until unison was reached. An overview of the study selection procedure is displayed in Fig 2.

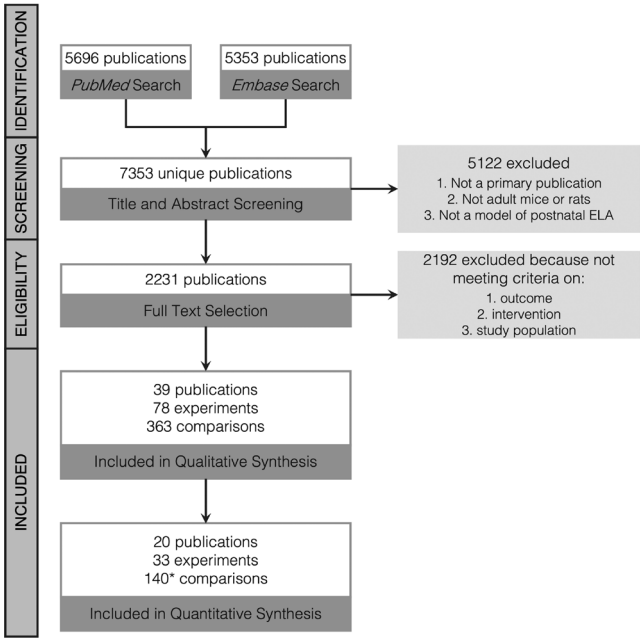


Fig 2. Flow-chart of study selection process. ^ = not included in pre-specified inclusion/exclusion criteria (S1.1).

A complete list of final inclusion and exclusion criteria can be found in the protocol (S1.1). First, titles and abstracts were screened by at least three blinded investigators (HS, VB, EK, DvN, LvM) for the following exclusion criteria: 1) not a primary experimental publication, 2) not using adult (>8 weeks) mice or rats which are younger than 1 year, 3) not using a postnatal model of ELA as specified in *S1.1.* Eligibility was then determined by full-text screening of the remaining studies by at least two blinded investigators (HS, EK, LvM), with a random subset screening performed by a fourth (VB), blinded investigator to confirm agreement. Publications were deemed non-eligible based on the following criteria: 1) not measuring an IEG product in the brain, 2) deviation from *a priori* determined criteria concerning the background of the animals, interventions, or outcomes, 3) control and experimental groups differed at more aspects than just ELA exposure. Lastly, reference sections of eligible publications were screened for articles missed by the search string, but none were added through this procedure.

Data from eligible studies were extracted into a combined dataset using *a priori* determined sets of variables to comprehensively capture experimental design, methods and results with minimal subjectivity (S1.3). Differently from the original protocol, we extracted also measurements without acute stress to have an appropriate control, baseline condition. Outcome data for each comparison (*i.e.* group-based mean and variance) were extracted in the following order of preference: 1) from numbers provided in the text or tables; 2) from graphs by using WebPlotDigitizer (v4.3 [33]; or 3) from statistical test results. A comparison is defined as the difference in expression of a specific IEG in a specific brain area at rest or after acute stress exposure in ELA-exposed animals and controls. To compare the results on a systematic review level, we performed an independent samples t-tests on the extracted summary statistics. The results were interpreted dichotomously as significant / not significant, with $p<0.05$ used as a criterion. We chose this approach to equalize the statistical method used for analysis across publications.

Meta-analysis

Data selection. We performed a meta-analysis on outcomes that were assessed by at least three independent comparisons (*i.e.*, at least one comparison from three independent publications). During analysis coding, the investigators were blinded to the outcome by randomly multiplying half of the effect sizes by -1.

To account for potential sex differences, we planned to perform separate meta-analyses for males and females. However, only few comparisons were reported for female rodents, and their study designs were strongly heterogeneous. We therefore restricted our quantitative synthesis to outcomes from male rodents, with female data being evaluated qualitatively only. Furthermore, only comparisons using either IHC, immunocytochemistry (ICC), or ISH to quantify IEG expression were included on the meta-analytic level. While both methods differ in the type of molecule being assessed,

quantification and analysis procedures largely overlap. To confirm this, we investigated whether the choice of quantification method affects the outcome. PCR based methods and western blots were evaluated qualitatively only.

Based on the aforementioned threshold and restrictions, the meta-analysis was performed on comparisons of cFos expression in the amygdala, thalamus, hippocampal formation, hypothalamus, prefrontal cortex and midbrain at rest and after acute stress experiences. Smaller subregions were grouped into larger structures (*S1.4*) in line with the Allen Mouse Brain Atlas (©2004, Allen Institute for Brain Science) to allow for comparisons between studies.

Statistical analysis. For comparisons included in the meta-analysis, we calculated the standardized mean difference Hedge's *g* as a measure of effect size. If only the standard error of the mean (SEM) was reported, the standard deviation (SD) was calculated as $SEM^* \sqrt{n}$, where n = the number of animals per group. If the total number of animals was reported, this was distributed equally across groups. If the number of animals was reported as a range (e.g.6-8 animals/group), we used the mean (e.g. 7 animals/group). If the same control group was used as control of multiple experimental groups (e.g. different ELA models), the sample size of the control group was divided by the number of experimental groups and the adjusted sample size was used for the calculation of the effect size [34]. Heterogeneity was assessed with Cochran's *Q*-test [35]. Influential outliers were determined in accordance with Viechtbauer and Cheung [36] and removed from quantitative synthesis. Of such comparisons, we explored whether elements of the experimental design could explain the deviation of these comparisons from the mean.

A three-level mixed-effects model was built to capture variance not only between publications (Level 1), but also between experiments (Level 2) and outcomes (Level 3), thereby taking into account the statistical dependency of outcomes acquired from the same animals within the same publication [37–39]. Moderators of the multilevel model were i) presence of an acute stress challenge, ii) presence of additional hits and iii) brain area.

We tested whether ELA effect sizes at rest or after acute stress challenges are significantly different from zero to understand the effects of ELA on *cFos* expression under each of these conditions. Subsequently, a subgroup analysis was performed to investigate whether the effects are moderated by the experience of multiple negative life experiences (additional hits). The presence of additional hits was classified with previously determined criteria [16]. Finally, we explored the effects of type of acute stressor (*i.e. mild* versus *severe*, S1.4), novelty of stress experience, and brain region using subgroup analyses.

Bias assessment and sensitivity analyses. We followed SYRCLE guidelines on risk

of bias assessment, with items not reported being coded as 'unclear' [30]. To detect publication bias, funnel plot asymmetries for each outcome variable were evaluated [30]. Due to the uneven frequency of the number of studies, we performed sensitivity analyses (rather than subgroup as specified in the protocol) on the type of ELA model, and difference between mRNA and protein. Since these analyses were not initially included, the results were only qualitatively assessed and were in line with the interpretation of the main results. All analyses can be found at our repository (https://osf.io/qkyvd/).

Software

All analyses were performed in R[40]. The following R packages were used: metafor (v 2.1.0; [41]), tidyverse (v1.2.1; [42]). Data are presented as the standardized mean difference Hedge's *g* and standard error of the mean (*g*[±SEM]). The significance level α was set to 0.05. Multiple testing correction on the planned analysis was performed using the Holm-Bonferroni method [43]. The code for analysis is openly accessible at https://osf.io/qkyvd/.

Results

Study selection and characteristics

A total of 1019 animals reported in 39 publications were included in the review. The animals were predominantly male (72.5%); rats (76.3%) were used more often than mice; and protein (77.4%) rather than mRNA was more frequently assessed as outcome. The IEG *cFos* was investigated in the majority of studies (88.7%), and maternal separation was the most frequently used ELA model (90.6%). Fig 3 shows a graphical overview of the study characteristics.

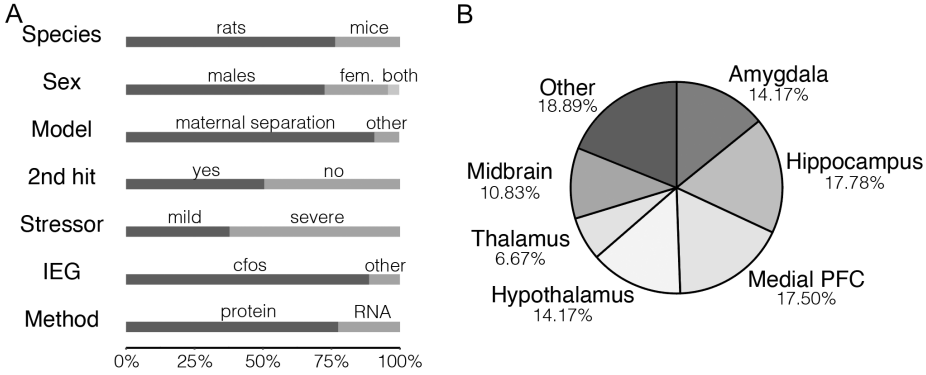


Fig 3. A) Study characteristics and B) Investigated brain areas reported as percentage of experiments. Fem = females.

Table 1. Overview of study designs and findings of reviewed publications reporting on cFos expression in ELA and control animals.

Author (Year)	MA	Model (PNDs)	Species	Sex	Exp. Design details	AS	Effect	Area(s)
Auth (2018) [44]		MS (2-15)	Mouse	F	Dark-light box	√	↔	BLA, LA, CEA, PVN
					Open-field test	√	↔	BLA, LA, CEA, PVN, dlPAG, vlPAG
					Two independent naïve cohorts	x	↑	dlPAG
							↔	BLA, CEA, PVN, vlPAG, LA
						x	↑	LA
							↔	BLA, LA, CEA, PVN
Banqueri (2018) [45]		MS (1-10)	Rat	F	Morris water maze	√	↑	CA1, DG
							↓	ACA
							↔	avTN, amTN, IL, PL
		MS (1-21)	Rat	F	Morris water maze	√	↑	DG
							↓	IL, PL, ACA
							↔	avTN, amTN
Benner (2014) [46]		MS (2-15)	Mouse	M	Competitive dominance task	√	↑	BLA
							↓	CA1
							↔	ACA, CEA, DG, IL, PL
Chung (2007) [47]	√	MS (2-14)	Rat	M	Colorectal distension	√	↑	ACA
							↔	CEA, cmTN, PAG, PVT, vmHN
					-	x	↑	ACA
							↔	CEA, cmTN, PAG, PVT, vmHN
Clarke (2013)[48]	√	MS (10-11)	Rat	M	Small litter (12 pups); Restraint stress	√		mPPVN
							↔	vBNST, MGPVN, lPPVN, dPPVN
					Small litter	x	-	dPPVN
							↔	mPPVN, MGPVN, lPPVN, vBNST
					Large litter (20 pups); Restraint stress	√	↔	mPPVN, MGPVN, lPPVN, dPPVN, vBNST
					Large litter	x	↔	mPPVN, MGPVN, lPPVN, dPPVN, vBNST
Cohen (2013) [49]	√	LBN (2-21)	Mouse	M	Novel environment	√	↔	BLA
Daskalakis (2014) [50]		MS (3-5)	Rat	M	MS pups remained in HC; re-exposure to fearful context	√		MEA
							↔	BLA, CEA
					MS pups placed in NC; re-exposure to fearful context	√		BLA, MEA
							↔	CEA

Author (Year)	MA	Model (PNDs)	Species	Sex	Exp. Design details	AS	Effect	Area(s)
Desbonnet (2008) [51]	√	MS (2-14)	Rat	M	Forced swim test	√	↔	PVT, CEA, PVN, BNST, DG
					-	x	↔	PVT, CEA, PVN, BNST, DG
				F	Forced swim test	√	↔	PVT, CEA, PVN, BNST, DG
					-	x	↔	PVT, CEA, PVN, BNST, DG
Felice (2014) [52]	√	MS (2-12)	Rat	M	Open-field test	√	↔	BLA, CEA, rostral & caudal ACA, IL, PL
					Colorectal distension	√	↑	rostral & caudal ACA, IL, PL
							↔	BLA, CEA
					-	x	↔	BLA, CEA, rostral & caudal ACA, IL, PL
Gardner (2005) [53]		MS (2-14)	Rat	M	Social defeat paradigm; cFos counts summed	√	↔	DRN
					across 4 slices	x	↔	DRN
		Handling (2-14)	Rat	M	Social defeat paradigm; cFos counts summed	√	↔	DRN
					across 4 slices	x	↔	DRN
Gaszner (2009) [54]		MS (8-14)	Rat	M	Restraint stress	√	↔	EW
					-	x	↔	EW
				F	Restraint stress	√	↔	EW
					-	x	↔	EW
		Handling (8-14)	Rat	M	Restraint stress	√	↔	EW
					-	x	↔	EW
				F	Restraint stress	√	↑	EW
					-	x	↔	EW
Genest (2004) [55]	√	MS (3-12)	Rat	M	Novel environment	√	↑	PVN
				F	Novel environment	√	↔	PVN
Hidaka (2018) [56]		MS (2-14)	Mouse	M	Three chamber test	√	↔	ACA, IL, PL
James (2014) [57]	√	MS (2-14)	Rat	M	Restraint stress	√	↓	mPPVN
							↔	PVT
				F	Restraint stress	√	↔	mPPVN, PVT
Loi (2017) [58]		MS (3-4)	Rat	M	Rodent Iowa gambling task	√	↓	rCA1, rCA3, leAI, lelL

Author (Year)	MA	Model (PNDs)	Species	Sex	Exp. Design details	AS	Effect	Area(s)
							↔	r&leDG, r&leACA, r&lePL, le CA1, le CA3, r&le dlSX, r&le mlSX, r&le AI, r&le NAcc Shell&Core, rIL, r&le vOFC, r&le mOFC, r&le cOFC
Menard (2004)[59]	√	LG	Rat	M	Shock-probe burial task with electrified probe	√	↓ ↔	dlSX, vlSX, vSUB, dPAG, vPAG vDG, dDG, mSX, CA1, CA3, aHN, CEA, BLA, IC, NAcc shell
O'Leary (2014)[60]		MS (1-14)	Mouse	F	Restraint stress	√	↓ ↔	dDG, vCA3 dCA1, dCA2, PVN, dCA3, vdG, NAcc, VTA, IL, PL, ACA, LA, BLA, CEA, DRN
Ren (2007)[61]		MS (2-21)	Rat	M	Colorectal distension	√	↔	DRN
Renard (2010)[62]	√	MS (1-21)	Rat	M F M F	Perfusion 24h after last day of chronic variable stress -	× ×	↔ ↔ ↔ ↔	mPPVN mPPVN mPPVN mPPVN
Rincel (2016)[63]	√	MS (2-14)	Rat	M	Open-field test	√	↓	PVN
Rivarola (2008)[64]		MS (1-21)	Rat	F	Perfusion 24h after last day of chronic variable stress -	× ×	↑ ↑	adTN adTN
Rivarola (2009)[65]		MS (1-21)	Rat	F	Perfusion 24h after last day of chronic variable stress -	× ×	↑ ↔ ↔	RSP adTN, MMN adTN, RSP MMN
Shin (2018)[66]	√	MS (1-14)	Mouse	M	Social interaction after 1d social isolation -	√ ×	 ↔ ↔	ISX, VTA mPfC, NAcc, vPAL, AHA, VH ISX, VTA, mPfC, NAcc, vPAL, AHA, VH
Tenorio-Lopes (2017)[67]	√	MS (3-12)	Rat	M	Novel Environment	√	↔	BLA, CEA, MEA, DMH, PVN
Troakes (2009)[68]	√	MS (5-21)	Rat	M	Elevated plus maze	√	↓ ↔	PIR ACA, SSb, ISX, PVN, CEA, MEA, dCA1, vCA1, dCA2, vCA2, dCA3, vCA3, dDG, vDG, CP, DRN, Pontine region, CB

Author (Year)	MA	Model (PNDs)	Species	Sex	Exp. Design details	AS	Effect	Area(s)
					-	x	↔	ACA, SSb, PIR, ISX, PVN, CEA, MEA, dCA1, vCA1, dCA2, vCA2, dCA3, vCA3, dDG, vDG, CP, DRN, Pontine region, CB
Trujillo (2016) [69]	√	MS (1-21)	Rat	M	Perfusion 24h after last day of chronic variable stress	x		MEA
							↔	CA1, CA2, CA3, PVN
					-	x		CA1, CA2, CA3, MEA
							↔	PVN
van Hasselt (2012) [70]		LG	Rat	P	Rodent Iowa gambling task; results reported as correlation with %LG	√	↑*	NAcc Shell, AI
							↔*	mOFC, vOFC, lOFC, ACA, PL, IL, dlSTR, dmStR, NAcc Core, CEA, BLA, DG, CA1
Vivinetto (2013) [71]	√	MS (1-21)	Rat	M	Foot shock in step-down inhibitory avoidance task	√	↔	CA1, CA3, DG
Yajima (2018) [72]		MS (2-14)	Mouse	M	-	x	-	HPF
Zhang (2009) [73]	√	MS (2-14)	Rat	M	Colorectal distension	√		cmTN
							↔	ACA, vplTN, PVT
					-	x	↑	vplTN
							↔	ACA, cmTN, PVT
Zhao (2013) [74]	√	MS (2-14)	Rat	M	Chinese language publication	x	↑	PVN

Header: MA – whether some or all comparisons from this study are included in the meta-analysis (√) or on systematic review level only; *Model(PNDs)* – which ELA model (MS – maternal separation, LBN – limited bedding and nesting, LG – licking and grooming) was applied during which postnatal days (PNDs); *Sex* – animals were female (F) or male (M); *Exp. design details* – indicates how experiments (nests) differed, if – then rest/no manipulation; *AS* – if acute stress challenge was present (√) or not (x); *Effect* – if ELA significantly increased (↑), decreased (↓) or did not alter (↔) IEG expression as based on independent t-tests; * = t-test could not be performed and effects are shown as reported in the original publication; *Areas* – brain areas as identified in publication, with position (lowercase, if identified) and area acronym as follows:

Area acronyms (in alphabetical order): ACA – anterior cingulate area; AHA – anterior hypothalamic nucleus; AI – agranular insular cortex; BLA – basolateral amygdala; BNST – bed nuclei of the stria terminalis; CB – cerebellum; CEA – central amygdala; CP – caudate putamen; CTX – cortex; DG – dentate gyrus; DRN – dorsal raphe nucleus; EW – Edinger-Westphal nucleus; HN – hypothalamic nucleus; HPF – hippocampal formation; IL – infralimbic area; DMH – dorsomedial hypothalamic nucleus; LA – lateral amygdala; LC – locus coeruleus; MEA – medial amygdala; MGPVN – magnocellular part of the PVN; MMN – mammillary nucleus; NAcc – nucleus accumbens; OFC – orbital-frontal cortex; PAG – periaqueductal gray; PAL – Pallidum; PFC – prefrontal cortex; PIR – Piriform cortex; PL – prelimbic area; PPVN – parvocellular part of the PVN; PVN – paraventricular nucleus of the hypothalamus; PVT – paraventricular nucleus of the thalamus; RSP – retrosplenial cortex; SSb – somatosensory barrel cortex; STR – striatum; SUB – subiculum; SX – septum; TN – thalamic nucleus; VH – ventral hypothalamic nucleus; VTA – ventral tegmental area.

Position: a – anterior; c – central; d – dorsal; l – lateral; le – left; m – medial; p – posterior; r – right; v – ventral.

Research synthesis

Systematic review of cFos and ELA. A total of 31 publications reported cFos expression in control and ELA animals (*Table 1*). IEG expression was reported to be significantly affected by ELA in 72 (45.8%) comparisons, of which 33 (59.6%) displayed upregulation and 39 comparisons (54.2%) reported downregulation.

Overall, of the 322 comparisons within these studies, 140 comparisons (n_{pub} = 20) qualified for further meta-analysis in male rodents after removal of influential outliers (n_{comp} = 1); these are analyzed quantitatively in the following section. No element of the experimental design pointed towards a biological origin of the outlying value, nor was its publication published in a predatory journal [75]. Comparisons were excluded from quantitative review because of brain area (n_{comp} = 40), acute stressor type (n_{comp} = 49; S2.4.3) or unspecified or pooled sex (n_{comp} = 15). The excluded comparisons are subject to a qualitative review in the Supporting Information.

Given fundamental biological differences between males and females [76], we *a priori* chose to evaluate female cFos data separately from males'. However, only ten publications reported on cFos expression in female rodents (n_{comp} = 77). Given the limited number of studies, with variable designs, we had to abandon the separate meta-analytical evaluation of female rodents. Qualitatively, the majority of the studies with females found no significant differences between cFos levels of ELA versus controls at rest or after an acute stress challenge (n_{comp} = 55 [51,54,55,57,62]). A more detailed description is supplied in the Supporting Information.

Systematic review of ELA and other IEGs. We here only summarize the main findings on IEGs other than cFos. In general, the number of studies on these IEGs compared to cFos was very limited. For a more elaborate description and discussion we refer to the Supporting Information.

Arc is a post-synaptic protein, which plays an essential role in regulating the homeostatic scaling of AMPA receptors, thereby directly modifying plasticity at the synapse [77]. *Arc* expression was investigated in only five publications under varying conditions in male and female mice and rats (see Table 2 and Supporting Information).

Early-growth response (Egr) proteins are a family of transcription factors with a zinc-finger motif, which allows all Egr factors to connect to identical DNA binding sites[82]. We identified only three studies investigating Egr expression after ELA exposure at rest (Table 3 and Supporting Information); specifically, one investigated Egr-1 [83], another investigated Egr-4 only [80], and one other investigated Egr-2 and Egr-4 [79].

FosB is an IEG of the *Fos family*, and - similarly to *cFos* - it binds to members of the *Jun* family to form the AP1 transcription factor [84]. Of particular interest in stress research is its isoform Δ*FosB*, whose extended half-life makes it an exceptional marker for chronic stress [84]. Three publications reporting on the expression of Δ*FosB* at rest in ELA and control animals were identified (see Table 4 and Supporting Information).

Table 2. Overview of study designs and findings of reviewed publications reporting on Arc expression in ELA and control animals.

Author (Year)	Model (PNDs)	Species	Sex	Exp. design details	AS	Effect	Area(s)
Benekareddy (2010) [78]	MS (2-14)	Rat	M	-	x	↔	mPFC
Benner (2014) [46]	MS (2-15)	Mouse	M	Competitive dominance task	√	↔	ACA, BLA, CEA, CA1, DG, IL, PL
McGregor (2018) [79]	MS (2-14)	Rat	M	Juvenile restraint stress	x	↑*	dSTR
				-	x	↑*	dSTR
Rincel (2019) [80]	MS (2-14)	Mouse	M	-	x	↓*	mPFC
			F	-	x	↑*	mPFC
Solas (2010) [81]	MS (2-21)	Rat	M	-	x	↓	CA1, CA3, DG

Header: *Model(PNDs)* – which ELA model (MS – maternal separation) was applied during which postnatal days (PNDs); *Sex* – animals were female (F) or male (M) or not specified (NS); *Exp. design details* – indicates how experiments (nests) differed, if – then rest/no manipulation; *AS* – if acute stress challenge as present (√) or not (x); *Effect* – if ELA significantly increase (↑), decreased (↓) or did not alter (↔) IEG expression as based on independent t-tests; * = t-test could not be performed and effects are shown as reported in the original publication; *Areas* – brain areas as identified in publication, with area acronym as follows:
Area acronyms (in alphabetical order): ACA – anterior cingulate area; BLA – basolateral amygdala; CEA – central amygdala; DG – dentate gyrus; IL – infralimbic area; mPFC – medial prefrontal cortex; PL – prelimbic area; dSTR – dorsal striatum.

Table 3. Overview of study designs and findings of reviewed publications reporting on expression of the Egr-family in ELA and control animals.

Author (Year)	Model (PNDs)	Species	IEGs	Sex	Exp. design details	AS	Effect	Area(s)
McGregor (2018) [79]	MS (2-14)	Rat	Egr-2	M	Juvenile restraint stress	x	↔*	dSTR
					-	x	↑*	dSTR
			Egr-4	M	Juvenile restraint stress	x	↑*	dSTR
					-	x	↑*	dSTR
Navailles (2010) [83]	MS (2-15)	Mouse	Egr-1	M	Balb/c strain	x	↓	CTX
							↔	DG, CA1, CA2, CA3
					C57BL/6 strain	x	↔	CTX
Rincel (2019) [80]	MS (2-14)	Mouse	Egr-4	M	-	x	↓*	mPFC
				F	-	x	↑*	mPFC

Header: *Model(PNDs)* – which ELA model (MS – maternal separation) was applied during which postnatal days (PNDs); *Sex* – animals were female (F) or male (M) or not specified (NS); *Exp. design details* – indicates how experiments (nests) differed, if – then rest/no manipulation; *AS* – if acute stress challenge as present (√) or not (x); *Effect* – if ELA significantly increase (↑), decreased (↓) or did not alter (↔) IEG expression; * = t-test could not be performed and effects are shown as reported in the original publication; *Areas* – brain areas as identified in publication, with area acronym as follows:
Area acronyms (in alphabetical order): CTX – cortex; DG – dentate gyrus; dSTR – dorsal striatum; mPFC – medial prefrontal cortex.

Table 4. Overview of study designs and findings of reviewed publications reporting on DFosB expression in ELA and control animals.

Author (Year)	Model (PNDs)	Species	Sex	Exp. design details	AS	Effect	Area(s)
Kim (2015) [85]	MS (1-14)	Rat	F	-	x	↓	NAcc
Lippmann (2007) [86]	MS (2-14)	Rat	M	-	x	↔*	CTX, NAcc, STR
	Handling (2-14)	Rat	M	-	x	↔*	CTX, NAcc, STR
Wang (2016) [87]	MS (1-15)	Rat	NS	-	x	↑	mPFC

Header: *Model(PNDs)* – which ELA model (MS – maternal separation) was applied during which postnatal days (PNDs); *Sex* – animals were female (F) or male (M) or not specified (NS); *Exp. design details* – indicates how experiments (nests) differed, if – then rest/no manipulation; *AS* – if acute stress challenge as present (√) or not (x); *Effect* – if ELA significantly increase (↑), decreased (↓) or did not alter (↔) IEG expression; * = t-test could not be performed and effects are shown as reported in the original publication; *Areas* – brain areas as identified in publication, with area acronym as follows:
Area acronyms (in alphabetical order): CTX – cortex; mPFC – medial prefrontal cortex; NAcc – nucleus accumbens; STR – striatum.

Meta-analysis of cFos in male rodents. For cFos, our survey yielded sufficient data to carry out a meta-analysis, next to the systematic review. In comparison to control animals, rodents with a history of ELA displayed significantly increased cFos levels at rest (*g[SEM]* = 0.421[±0.18], t = 2.35, p_{adj} = 0.041), but not after acute stress exposure (*g[SEM]* = 0.133[±0.166], t = 0.805, p_{adj} = 0.422; Fig 4a). To gain a deeper understanding of these findings, we performed subgroup analyses to investigate the experience of additional hits, i.e. an additional negative life event. Of note, the control and experimental groups always differed only in the presence/absence of ELA. Therefore, in the 'additional hits' comparisons, both control and ELA animals experienced multiple negative life events. This was important for cFos expression after acute stress, where the effects of ELA were pronounced only in synergy with additional hits (Fig 4b, $acute_{no\ hit}$: *g[SEM]* = -0.193[±0.135], *z* = -1.436, p_{adj} = 0.151; $acute_{mult\ hits}$: *g[SEM]* = 0.442[±0.159], *z* = 2.784, *p* = 0.016; at restno hit: *g[SEM]* = 0.475 [±0.16], *z* = 2.976, *p* < .012; at $rest_{mult\ hits}$: *g[SEM]* = 0.344[±0.153], *z* = 2.253, *p* = 0.049; the analyses were conducted comparing the effect size between control and ELA animals against 0). Lastly, we performed an exploratory analysis to investigate potential interactions with acute stressor severity on the effect sizes. For the categorization of acute stressor severity, please see Supporting Information S1.4. Acute stressor severity was not a significant moderator ($Q_M(3)$ = 4.35, *p* = 0.226). Of note, no publication investigated cFos expression after a mild acute stressor in animals that experienced additional hits (n_{comp} = 0). Of the 20 publications included in the meta-analysis, only two did not use maternal separation as an ELA model[49,59]. Nevertheless, these studies

adhere to the above findings with no significant differences found after acute stress in the areas meta-analytically investigated. The findings of our main analysis do not confirm our hypothesis that cFos expression is higher in ELA animals compared to control particularly after acute stress; rather, the results indicate that cFos expression is increased after ELA already at baseline, i.e. at rest. Moreover, the results highlight the relevance of including the presence of additional hits in the analysis.

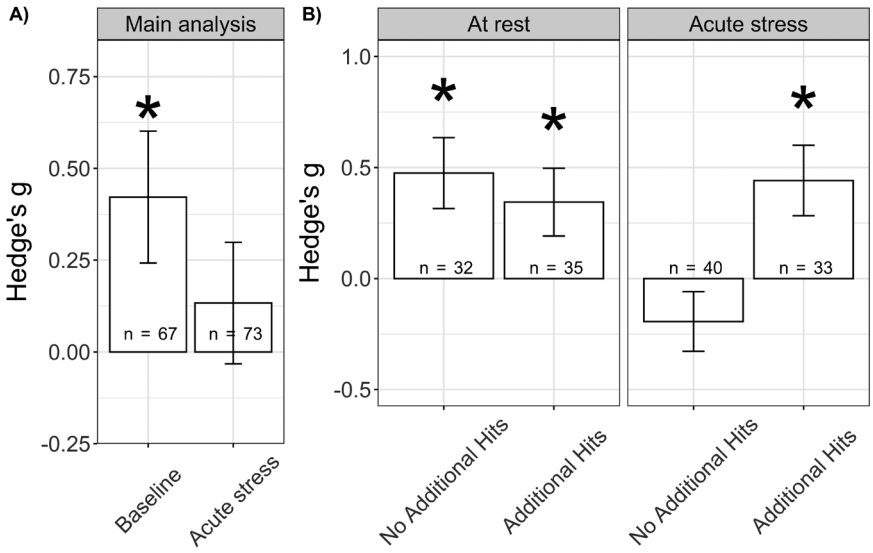


Fig 4. Main and subgroup analyses. A) Effects of ELA on cFos expression in male rodents at rest and after an acute stressor. B) Subgroup analysis for absence (No Additional Hits) or presence (Additional Hits) of additional negative life events. Of note, control and experimental animals always differed only in the presence/absence of ELA. Therefore, in the 'Additional Hits' comparison, also control animal experienced the additional negative life events. * $p < 0.05$.

Next, we tested whether the effects of ELA on cFos expression differed across brain regions important for the stress reaction (Fig 5), when only considering those datasets with sufficient observations ($n_{publications}>3$). Brain region was not a significant moderator ($Q_M(12) = 13.908$, $p = 0.307$) of the effects of ELA on cFos expression. Exploratory subgroup analysis suggests that at rest all brain areas show a comparable increase in cFos expression. After an acute stress challenge, the effects appeared more variable across brain areas than at rest. We then performed an additional exploratory analysis to investigate whether brain areas after acute stress differed after ELA with / without the experience of additional hits. The results of this analysis suggested that the prefrontal cortex may be specifically affected; however, since this effect was supported by those studies unevenly represented in the funnel plot, these results may not be reliable due to presumed publication bias.

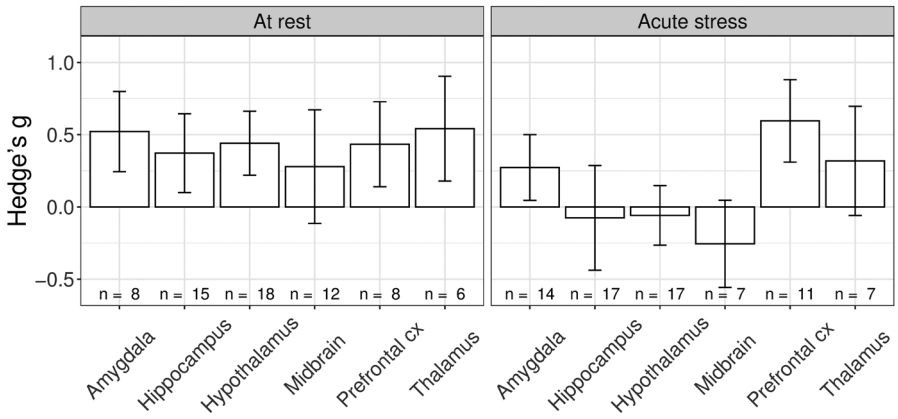


Fig 5. Effects of ELA on cFos expression across brain areas at rest and after an acute stressor.

Despite significant contribution of the moderators ($Q_M(23) = 40.089$, $p = 0.015$), residual heterogeneity between studies remained significant ($Q_E(117) = 167.95$, $p = 0.001$). Study of the distribution of variance showed that remaining variance is mainly attributable to differences between experiments (Level 2) and not to differences within experiments (Level 3). Concerning potential bias, while reporting risk of bias was incomplete in all publications (*S2.1a*), 46% of studies reported adequate randomization and blinding procedures ($n_{publications} = 10$). Visual assessment of the funnel plot for the studies qualifying for quantitative synthesis suggests the presence of publication bias (*S2.1b*), which was also supported by a significant Eggers' test ($z = 4.6903$, $p < .0001$). We identified two studies[52,69] which were mainly responsible for the bias.

Discussion

In this review, we synthesized the evidence of 39 publications investigating the effects of ELA on IEG expression in mice and rats. Due to low number of animals used in preclinical research, studies are commonly underpowered[88], rendering results of individual studies vulnerable to confounding effects of the chosen study design. In order to circumvent this limitation, we systematically reviewed the available literature on several IEGs in males and females. We meta-analyzed a subset of our male data to quantify cFos expression following ELA exposure and to identify potential moderators of the observed effects. Using a three-level mixed effects model, we observed an increase in cFos expression after an acute stress exposure due to ELA only in combination with one or more other negative life events. This suggests that ELA creates a vulnerable phenotype that is manifested only when sufficiently triggered. If rodents had 'only' experienced ELA, we report – contrary to our expectations – an increase in cFos expression already at rest, suggesting that the situation normally seen (in naïve rodents) after acute stress is already visible at rest when the animals have

been exposed to early life adversity. These findings led us to propose a new model as outlined in Fig 6.

At rest, ELA animals compared to controls show increased IEG expression. Since raw values of IEG expression are either not reported or of incomparable scales, we could only investigate effect sizes and not absolute values of IEG expression. This has a direct effect on the interpretation of the results. Specifically, if IEG levels in control animals were low, effect sizes could be inflated. If IEG levels in control animals were high, the results should be interpreted not as "rest" but rather as "mildly aroused", since IEG levels are expected to be minimal for control, naïve animals. Nonetheless, we observed a consistent, positive standardized mean difference in cFos expression after ELA across five out of the six quantitatively investigated brain regions. This suggests a small, but stable brain-wide effect. Previous studies showed that IEG expression matches the transcriptional activity from early environment and experiences[89]. In control animals, this is likely to result in a minimal IEG expression. However, in ELA animals, the expression observed may be the result of long-lasting ELA effects on brain structure and chemistry[90]. Indeed, the transcriptional activity of ELA mice at rest is comparable to that of acutely-stressed control mice[91]. Increased activity-regulated transcription at rest after ELA could be indicative of an overall synaptic alteration, in accordance with increased anxiety-like behavior and reduced memory performance under neutral conditions[16]. Functionally, increased IEG expression at rest could reflect a differential, less adaptive way of processing previous experiences and could potentially hint towards an overall increased transcriptional activity as a result of synaptic sensitization. Intuitively, considering the relationship between IEGs and synaptic plasticity, we would suspect that ELA results in increased synaptic plasticity. In line with this idea, it has been shown that ELA leads to increased LTP in freely-behaving adult, male rats as compared to controls[92].Taken together, this evidence suggests that differences we report in IEG expression after ELA at rest may underlie long-lasting effects on transcriptional activity, pushing the system towards an "activated" state similar to acute stress.

The model proposed in Fig 6 relies primarily on the quantitative and qualitative analysis of *cFos* data, as there are only few publications investigating the effects of ELA on the expression of the IEGs *Arc*, Δ*FosB*, and IEGs of the *Egr*-family. Nonetheless, the available evidence suggests a sensitization effect of ELA on IEG expression (and, more generally, synaptic plasticity) at rest. Although IEGs overlap in function and overall expression pattern, they have specific and independent roles[3,20,23–25,93]. *cFos* and *Egr*-family members are transcription factors, while Arc is a post-synaptic protein modifying dendritic AMPA receptors, and Δ*FosB* is a less transient marker of neuronal activity[84,94,95]. With technological advances in the field of immunohistochemistry and bioinformatics it becomes increasingly feasible to investigate and interpret multiple IEGs within one animal, thereby also allowing for the investigation of interactions

between IEGs and their downstream effects. In the future, the study of different IEGs could be used as a proxy to more thoroughly understand ELA-induced changes in gene-regulated synaptic plasticity[96].

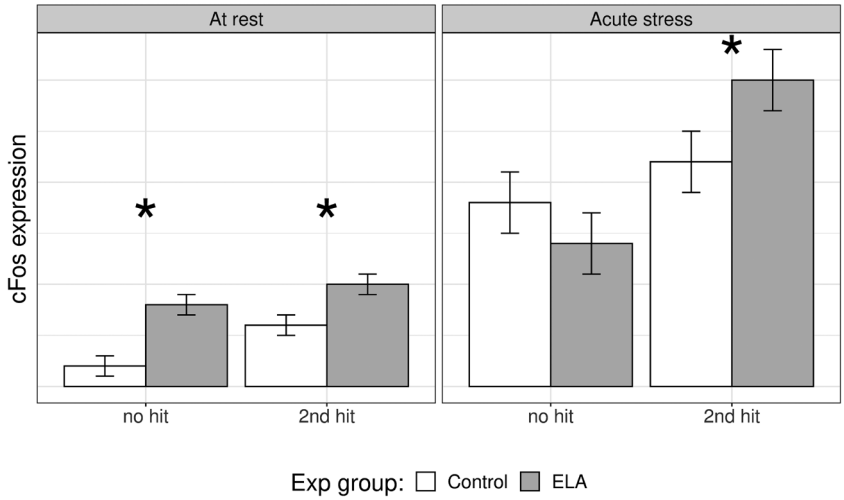


Fig 6. Summary interpretation of the results. Cartoon image of how to interpret effect sizes in absolute terms (y-axis, cFos expression, e.g. number of cFos+ cells). Significance levels identify the difference between control and ELA groups that we identified in our analysis (Fig 3). Of note, cFos expression levels are expected to be higher after acute stress than at rest, although this cannot be studied in the current meta-analysis.

On a systematic review level, effects in females appear more limited than in males. Whether this is a true biological effect is unclear. For example, it could be due to the comparatively low number of female publications, or to a male-biased experimental methodology[16,76]. While it has been shown that acute stress exposure increases IEG expression in both sexes in the hippocampus[97], it is possible that effects of ELA on IEG expression will be more subtle in females than in males due to model characteristics. Of note, among the 39 publications included in this review, only 5 investigated males and females under the same experimental conditions.

Lastly, given the expected heterogeneity in study designs, we restricted our meta-analysis to adult animals only, and – at this stage – it cannot be generalized to other age groups. It is possible that different patterns of IEG expression associated to ELA exposure may emerge in juvenile or adolescent animals. Future experiments investigating the longitudinal effects of ELA on IEG expression over the course of development can shed light on the interaction between ELA, development and IEG-related brain activity.

To conclude, we systematically provided a general overview on the relationship between ELA and IEG expression and highlighted current knowledge gaps. Despite

subject-specific and methodological limitations, the outcomes of the meta-analysis were robust and suggest a sensitization of activity-regulated transcription in ELA rodents at rest while changes after acute stress only became apparent in combination with additional hits. Recent advances in the fields of immunostaining, live cell imaging and bioinformatics may help close the described voids, yielding a more comprehensive picture on the complex relationship between IEGs, ELA and psychopathologies.

References

1. Burns, S. B., Szyszkowicz, J. K., Luheshi, G. N., Lutz, P.-E. & Turecki, G. Plasticity of the epigenome during early-life stress. *Seminars in Cell & Developmental Biology* 77, 115–132 (2018).
2. Citri, A. & Malenka, R. C. Synaptic plasticity: Multiple forms, functions, and mechanisms. *Neuropsychopharmacology* 33, 18–41 (2008).
3. Yap, E. L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* 100, 330–348 (2018).
4. Hübener, M. & Bonhoeffer, T. Neuronal plasticity: Beyond the critical period. *Cell* 159, 727–737 (2014).
5. Lang, J. *et al.* The hallmarks of childhood abuse and neglect: A systematic review. *PloS one* 15, e0243639 (2020).
6. Lippard, E. T. C. & Nemeroff, C. B. The devastating clinical consequences of child abuse and neglect: Increased disease vulnerability and poor treatment response in mood disorders. *American Journal of Psychiatry* 177, 20–36 (2020).
7. Nemeroff, C. B. Paradise Lost: The Neurobiological and Clinical Consequences of Child Abuse and Neglect. *Neuron* 89, 892–909 (2016).
8. Herzog, J. I. & Schmahl, C. Adverse Childhood Experiences and the Consequences on Neurobiological, Psychosocial, and Somatic Conditions Across the Lifespan. *Frontiers in Psychiatry* 9, 420 (2018).
9. Merrick, M. T. et al. Unpacking the impact of adverse childhood experiences on adult mental health. *Child Abuse and Neglect* 69, 10–19 (2017).
10. Liu, D. *et al.* Maternal care, hippocampal glucocorticoid receptors, and hypothalamic- pituitary-adrenal responses to stress. *Science* 277, 1659–1662 (1997).
11. Plotsky, P. M. & Meaney, M. J. Early, postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. *Molecular Brain Research* 18, 195–200 (1993).
12. McEwen, B. S., Nasca, C., Gray, J. D. & Milliken, M. Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. *Neuropsychopharmacology Reviews* 41, 3–23 (2016).
13. Molet, J. *et al.* MRI uncovers disrupted hippocampal microstructure that underlies memory impairments after early-life adversity. *Hippocampus* 26, 1618–1632 (2016).
14. Chen, Y. & Baram, T. Z. Toward understanding how early-life stress reprograms cognitive and emotional brain networks. *Neuropsychopharmacology* 41, 197–206 (2016).
15. Van Bodegom, M., Homberg, J. R. & Henckens, M. J. A. G. Modulation of the hypothalamic-pituitary-adrenal axis by early life stress exposure. *Frontiers in Cellular Neuroscience* 11, 87 (2017).
16. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neuroscience and Biobehavioral Reviews* 102, 299–307 (2019).
17. Sandi, C. & Haller, J. Stress and the social brain: behavioural effects and neurobiological mechanisms. *Nature Reviews Neuroscience* 2015 16:5 16, 290–304 (2015).
18. Molet, J., Maras, P. M., Avishai-Eliner, S. & Baram, T. Z. Naturalistic rodent models of chronic early-life stress. *Developmental psychobiology* 56, 1675–88 (2014).
19. Rocha, M., Wang, D., Avila-Quintero, V., Bloch, M. H. & Kaffman, A. Deficits in hippocampal-dependent memory across different rodent models of early life stress: systematic review and meta-analysis. *Translational Psychiatry* 2021 11:1 11, 1–12 (2021).
20. Okuno, H. Regulation and function of immediate-early genes in the brain: Beyond neuronal activity markers. *Neuroscience Research* 69, 175–186 (2011).
21. Kovács, K. J. c-Fos as a transcription factor: A stressful (re)view from a functional map. *Neurochemistry International* 33, 287–297 (1998).
22. Yap, E. L. et al. Bidirectional perisomatic inhibitory plasticity of a Fos neuronal network. *Nature* 590, 115–121 (2021).
23. Fleischmann, A. et al. Impaired long-term memory and NR2A-type NMDA receptor-dependent synaptic plasticity in mice lacking c-fos in the CNS. *Journal of Neuroscience* 23, 9116–9122 (2003).
24. Plath, N. *et al.* Arc/Arg3.1 Is Essential for the Consolidation of Synaptic Plasticity and Memories. Neuron 52, 437–444 (2006).
25. Jones, M. W. *et al.* A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. *Nature Neuroscience* 4, 289–296 (2001).
26. Gallo, F. T., Katche, C., Morici, J. F., Medina, J. H. & Weisstaub, N. V. Immediate Early Genes, Memory and Psychiatric Disorders: Focus on c-Fos, Egr1 and Arc. *Frontiers in Behavioral Neuroscience* 12, 79 (2018).
27. Daskalakis, N. P., Bagot, R. C., Parker, K. J., Vinkers, C. H. & de Kloet, E. R. The three-hit concept of vulnerability and resilience: Toward understanding adaptation to early-life adversity outcome. *Psychoneuroendocrinology* 38, 1858–1873 (2013).
28. de Vries, R. B. M. *et* al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclinical Medicine* 2, e00007 (2015).
29. Leenaars, M. *et* al. A step-by-step guide to systematically identify all relevant animal studies. *Laboratory Animals* 46, 24–31 (2012).
30. Hooijmans, C. R. *et* al. SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology* 14, (2014).

31. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery* 88, (2021).
32. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews* 5, 210 (2016).
33. Rohatgi, A. WebPlotDigitizer. https://automeris.io/WebPlotDigitizer (2020).
34. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221, 92–102 (2014).
35. Chochran, W. G. The comparison of percentages in matched samples. *Biometrika* 37, 256–266 (1950).
36. Viechtbauer, W. & Cheung, M. W. L. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*2 1, 112–125 (2010).
37. Assink, M. & Wibbelink, C. J. M. Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology* 12, 154–174 (2016).
38. Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V. & Van Der Sluis, S. A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience* 17, 491–496 (2014).
39. Cheung, M. W. L. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods* 19, 211–229 (2014).
40. R Core Team. R: A Language and Environment for Statistica. (2019).
41. Viechtbauer, W. Conducting meta-analyses in {R} with the {metafor} package. *Journal of Statistical Software* 36, 1–48 (2010).
42. Wickham, H. tidyverse: Easily Install and Load the 'Tidyverse'. (2017).
43. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 65–70 (1979).
44. Auth, C. S. et al. Differential anxiety-related behaviours and brain activation in Tph2-deficient female mice exposed to adverse early environment. *European Neuropsychopharmacology* 28, 1270–1283 (2018).
45. Banqueri, M., Méndez, M. & Arias, J. L. Why are maternally separated females inflexible? Brain activity pattern of COx and c-Fos. *Neurobiology of Learning and Memory* 155, 30–41 (2018).
46. Benner, S., Endo, T., Endo, N., Kakeyama, M. & Tohyama, C. Early deprivation induces competitive subordinance in C57BL/6 male mice. *Physiology and Behavior* 137, 42–52 (2014).
47. Chung, E. K. Y. et al. Neonatal maternal separation enhances central sensitivity to noxious colorectal distention in rat. *Brain Research* 1153, 68–77 (2007).
48. Clarke, M., Cai, G., Saleh, S., Buller, K. M. & Spencer, S. J. Being suckled in a large litter mitigates the effects of early-life stress on hypothalamic-pituitary-adrenal axis function in the male rat. *Journal of Neuroendocrinology* 25, 792–802 (2013).
49. Cohen, M. M. et al. Early-life stress has persistent effects on amygdala function and development in mice and humans. *Proceedings of the National Academy of Sciences of the United States of America* 110, 18274–18278 (2013).
50. Daskalakis, N. P. et al. Early experience of a novel-environment in isolation primes a fearful phenotype characterized by persistent amygdala activation. *Psychoneuroendocrinology* 39, 39–57 (2014).
51. Desbonnet, L., Garrett, L., Daly, E., McDermott, K. W. & Dinan, T. G. Sexually dimorphic effects of maternal separation stress on corticotrophin-releasing factor and vasopressin systems in the adult rat brain. *International Journal of Developmental Neuroscience* 26, 259–268 (2008).
52. Felice, V. D. et al. Differential activation of the prefrontal cortex and amygdala following psychological stress and colorectal distension in the maternally separated rat. *Neuroscience* 267, 252–262 (2014).
53. Gardner, K. L., Thrivikraman, K. V, Lightman, S. L., Plotsky, P. M. & Lowry, C. A. Early life experience alters behavior during social defeat: Focus on serotonergic systems. *Neuroscience* 136, 181–191 (2005).
54. Gaszner, B. et al. Effects of maternal separation on dynamics of urocortin 1 and brain-derived neurotrophic factor in the rat non-preganglionic Edinger-Westphal nucleus. *International Journal of Developmental Neuroscience* 27, 439–451 (2009).
55. Genest, S. E., Gulemetova, R., Laforest, S., Drolet, G. & Kinkead, R. Neonatal maternal separation and sex-specific plasticity of the hypoxic ventilatory response in awake rat. *Journal of Physiology* 554, 543–557 (2004).
56. Hidaka, C., Kashio, T., Uchigaki, D. & Mitsui, S. Vulnerability or resilience of motopsin knockout mice to maternal separation stress depending on adulthood behaviors. *Neuropsychiatric Disease and Treatment* 14, 2255–2268 (2018).
57. James, M. H. et al. Exercise reverses the effects of early life stress on orexin cell reactivity in male but not female rats. *Frontiers in Behavioral Neuroscience* 8, 244 (2014).
58. Loi, M. et al. Transient prepubertal mifepristone treatment normalizes deficits in contextual memory and neuronal activity of adult male rats exposed to maternal deprivation. eNeuro 4, (2017).
59. Menard, J. L., Champagne, D. L. & Meaney, M. J. P. Variations of maternal care differentially influence 'fear' reactivity and regional patterns of cFOS immunoreactivity in response to the shock-probe burying test. *Neuroscience* 129, 297–308 (2004).
60. O'Leary, O. F. et al. GABAB(1) receptor subunit isoforms differentially regulate stress resilience. *Proceedings of the National Academy of Sciences of the United States of America* 111, 15232–15237 (2014).
61. Ren, T. H. et al. Effects of neonatal maternal separation on neurochemical and sensory response to colonic

distension in a rat model of irritable bowel syndrome. *American Journal of Physiology - Gastrointestinal and Liver Physiology* 292, 849–856 (2007).
62. Renard, G. M., Rivarola, M. A. & Suárez, M. M. Gender-dependent effects of early maternal separation and variable chronic stress on vasopressinergic activity and glucocorticoid receptor expression in adult rats. *Developmental Neuroscience* 32, 71–80 (2010).
63. Rincel, M. *et* al. Maternal high-fat diet prevents developmental programming by early-life stress. T*ranslational Psychiatry* 6, e966–e966 (2016).
64. Rivarola, M. A., Dalmasso, C., Valdez, D. J., Vivas, L. M. & Suárez, M. M. Early maternal separation and chronic variable stress as adults differentially affect fos expression in the anterodorsal thalami nuclei. *International Journal of Neuroscience* 118, 735–748 (2008).
65. Rivarola, M. A. & Suárez, M. M. Early maternal separation and chronic variable stress in adulthood changes the neural activity and the expression of glucocorticoid receptor in limbic structures. *International Journal of Developmental Neuroscience* 27, 567–574 (2009).
66. Shin, S. *et* al. Drd3 Signaling in the Lateral Septum Mediates Early Life Stress-Induced Social Dysfunction. *Neuron* 97, 195–208 (2018).
67. Tenorio-Lopes, L. *et* al. Neonatal maternal separation opposes the facilitatory effect of castration on the respiratory response to hypercapnia of the adult male rat: Evidence for the involvement of the medial amygdala. *Journal of Neuroendocrinology* 29, e12550 (2017).
68. Troakes, C. & Ingram, C. D. Anxiety behaviour of the male rat on the elevated plus maze: Associated regional increase in c-fos mRNA expression and modulation by early maternal separation. *Stress* 12, 362–369 (2009).
69. Trujillo, V., Durando, P. E. & Suárez, M. M. Maternal separation in early life modifies anxious behavior and Fos and glucocorticoid receptor expression in limbic neurons after chronic stress in rats: Effects of tianeptine. *Stress* 19, 91–103 (2016).
70. van Hasselt, F. N. et al. Individual variations in maternal care early in life correlate with later life decision-making and c-Fos expression in prefrontal subregions of rats. *PLoS ONE* 7, e37820 (2012).
71. Vivinetto, A. L., Suárez, M. M. & Rivarola, M. A. Neurobiological effects of neonatal maternal separation and post-weaning environmental enrichment. *Behavioural Brain Research* 240, 110–118 (2013).
72. Yajima, H. et al. Early-life stress induces cognitive disorder in middle-aged mice. *Neurobiology of Aging* 64, 139–146 (2018).
73. Zhang, X.-J. et al. Activation of Extracellular Signal-Regulated Protein Kinase is Associated with Colorectal Distension-Induced Spinal and Supraspinal Neuronal Response and Neonatal Maternal Separation-Induced Visceral Hyperalgesia in Rats. *Journal of Molecular Neuroscience* 37, 274–287 (2009).
74. Zhao, X. P., Yu, L., Chen, Z. Y., Hua, R. & Zhang, Y. M. Effect of early-life stress on chronic functional visceral pain and CRH R1 expression in the paraventricular nucleus in adult rats. *World Chinese Journal of Digestology* 21, 3344–3355 (2013).
75. Grudniewicz, A. et al. Predatory journals: no definition, no defence. *Nature* vol. 576 210–212 (2019).
76. Shansky, R. M. & Murphy, A. Z. Considering sex as a biological variable will require a global shift in science culture. *Nature Neuroscience* 24, 457–464 (2021).
77. Bramham, C. R., Worley, P. F., Moore, M. J. & Guzowski, J. F. The Immediate Early Gene Arc/Arg3.1: Regulation, Mechanisms, and Function. *Journal of Neuroscience* 28, 11760–11767 (2008).
78. Benekareddy, M., Goodfellow, N. M., Lambe, E. K. & Vaidya, V. A. Enhanced function of prefrontal serotonin 5-HT2 receptors in a rat model of psychiatric vulnerability. *Journal of Neuroscienc*e 30, 12138–12150 (2010).
79. McGregor, N. W. et al. A translational approach to the genetics of anxiety disorders. *Behavioural Brain Research* 341, 91–97 (2018).
80. Rincel, M. et al. Multi-hit early life adversity affects gut microbiota, brain and behavior in a sex-dependent manner. *Brain, Behavior, and Immunity* 80, 179–192 (2019).
81. Solas, M. et al. Interactions between age, stress and insulin on cognition: Implications for alzheimer's disease. *Neuropsychopharmacology* 35, 1664–1673 (2010).
82. Beckmann, A. M. & Wilce, P. A. Egr transcription factors in the nervous system. *Neurochemistry International* 31, 477–510 (1997).
83. Navailles, S., Zimnisky, R. & Schmauss, C. Expression of glucocorticoid receptor and early growth response gene 1 during postnatal development of two inbred strains of mice exposed to early life stress. *Developmental Neuroscience* 32, 139–148 (2010).
84. Nestler, E. J. δfosB: A transcriptional regulator of stress and antidepressant responses. *European Journal of Pharmacology* 753, 66–72 (2015).
85. Kim, J. Y. et al. Beneficial effects of highly palatable food on the behavioral and neural adversities induced by early life stress experience in female rats. *International Journal of Biological Sciences* 11, 1150–1159 (2015).
86. Lippmann, M., Bress, A., Nemeroff, C. B., Plotsky, P. M. & Monteggia, L. M. Long-term behavioural and molecular alterations associated with maternal separation in rats. *European Journal of Neuroscience* 25, 3091–3098 (2007).
87. Wang, H. *et* al. Chronic stress is associated with pain precipitation and elevation in DeltaFosB expression. *Frontiers in Pharmacology* 7, 138 (2016).
88. Bonapersona, V. *et* al. Increasing the statistical power of animal experiments with historical control data. *Nature*

Neuroscience 24, 470–477 (2021).
89. Marrone, D. F., Schaner, M. J., McNaughton, B. L., Worley, P. F. & Barnes, C. A. Immediate-early gene expression at rest recapitulates recent experience. *Journal of Neuroscience* 28, 1030–1033 (2008).
90. De Kloet, E. R., Joëls, M. & Holsboer, F. Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience* vol. 6 463–475 (2005).
91. Marrocco, J. et al. Early life stress restricts translational reactivity in CA3 neurons associated with altered stress responses in adulthood. *Frontiers in Behavioral Neuroscience* 13, 157 (2019).
92. Blaise, J. H., Koranda, J. L., Chow, U., Haines, K. E. & Dorward, E. C. Neonatal isolation stress alters bidirectional long-term synaptic plasticity in amygdalo-hippocampal synapses in freely behaving adult rats. *Brain Research* 1193, 25–33 (2008).
93. Guzowski, J. F., Setlow, B., Wagner, E. K. & McGaugh, J. L. Experience-dependent gene expression in the rat hippocampus after spatial learning: A comparison of the immediate-early genes Arc, c-fos, and zif268. *Journal of Neuroscience* 21, 5089–5098 (2001).
94. Vialou, V. et al. Prefrontal Cortical Circuit for Depression- and Anxiety-Related Behaviors Mediated by Cholecystokinin: Role of ΔFosB. *The Journal of Neuroscience* 34, 3878 LP – 3887 (2014).
95. Vialou, V. et al. ΔFosB in brain reward circuits mediates resilience to stress and antidepressant responses. *Nature Neuroscience* 13, 745–752 (2010).
96. Li, Y. et al. A critical evaluation of the activity-regulated cytoskeleton-associated protein (Arc/Arg3.1) 's putative role in regulating dendritic plasticity, cognitive processes, and mood in animal models of depression. *Frontiers in Neuroscience* 9, 279 (2015).
97. Marrocco, J. et al. A sexually dimorphic pre-stressed translational signature in CA3 pyramidal neurons of BDNF Val66Met mice. *Nature Communications* 8, 808 (2017).

CHAPTER 8

Structural changes after early life adversity in rodents: a systematic review with meta-analysis

biorXiv (2022), doi: 10.1101/2022.01.17.476580

Joëls M[1,2]
Kraaijenvanger E[1]
Sarabdjitsingh RA[1]
Bonapersona V[1]

1 University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands.
2 University Medical Center Groningen, Groningen, The Netherlands.

Abstract

Early life adversity (ELA) is a well-documented risk factor for psychiatric illnesses in humans. This risk may, in part, be conferred by structural changes induced by ELA, lasting into adulthood. We here review the evidence for such lasting structural changes in rodent models for ELA involving altered maternal care during the first two postnatal weeks. In total, we extracted data from 64 studies reporting on 260 comparisons in adult rats or mice which experienced ELA or control treatment. Most of the observations concerned structural changes in the hippocampus of adult male rats earlier exposed to maternal separation. A 3-level meta-analysis revealed that ELA reduced hippocampal volume and the number of dendritic branches as well as dendritic length of principal hippocampal cells. No differences were observed across the hippocampal subfields. In terms of adult neurogenesis in the dentate subgranular zone, both staining for BrdU and the early neuronal marker DCX were significantly reduced, while the general proliferation marker Ki67 remained unchanged. The neuronal growth factor BDNF did not show significant changes, although the unexplained heterogeneity was moderate. Generally, the effect of ELA compared to control on structural markers was not affected by additional stressors experienced in life. Overall, the data available support the notion that ELA, at least in the hippocampus of male rats, lastingly reduces volume, hampers dendritic growth and suppresses adult neurogenesis.

Introduction

Adversity experienced early in life, when the brain is still developing, is a well-documented trigger for lasting changes in brain connectivity and behavior in humans, presumably increasing the risk of individuals to develop psychopathology later in life[6,7]. To study the long-lasting consequences of early life adversity (ELA) on brain structure and function, researchers have often reverted to animal models[8,9]. These offer multiple advantages over human investigations, e.g. that i) early life environment is known and can be specifically altered; ii) genetic variation, especially in in-bred mice, is relatively low; iii) housing conditions can be kept constant; and iv) the lifespan is quite short, so that consequences of ELA for the adult brain can be studied over the courses of months rather than decades[10].

Recently, we and others have shown in meta-analyses that in animal models too, ELA results in very consistent changes in behavioral function[11–13]. Shifts in behavior likely result from alterations in the underlying neuronal substrate. This can relate to many ELA-induced changes, including the connectivity between areas, structural modifications within specific areas and cell types, but also functional changes related to neurotransmitter actions and/or the cellular responses downstream of neurotransmitters and their receptors, such as second messenger systems or gene transcription. Also, the response of the individual animal to stressful circumstances may be altered in ELA-exposed compared to control animals[14], which may affect behavioral outcome particularly in challenging tasks, e.g. contextual fear conditioning.

To date, the effects of ELA on neuronal structure and structural plasticity in adult rodents have not been examined meta-analytically. We here focused on studies in rats and mice describing the effects of postnatal adversity (i.e., starting during the first 2 weeks of life) on structural outcome; all models involved altered maternal care, an important environmental determinant of adversity during this developmental window. We focused on three sets of outcomes: First, volume of adult brain areas and morphology of neurons after ELA, including reports on dendritic tree morphology and spine density. Secondly, adult neurogenesis in the dentate gyrus[15] of animals with an ELA history, summarizing data on cell proliferation, differentiation and survival. And thirdly, studies reporting on brain derived nerve growth factor (BDNF), which might give insight in potential mechanisms by which structure could be changed. For the main analysis, a 3-level mixed effect model was applied[16,17]. In case of significant unexplained heterogeneity, we next performed an exploratory random forest analysis[18] to identify the most promising potential moderators. Since the effects of stress have often been shown to differ between males and females[19,20], we planned in advance to conduct our analyses for males and females separately.

Methods

This review adheres to SYRCLE's guidelines for protocol (De Vries 2015), search strategy (Leenaars et al., 2012), and risk of bias assessment (Hooijmans et al., 2014). Reporting is in accordance with the PRISMA reporting checklist (Moher et al., 2009, Supplementary Information S2). The analytic strategy is based on earlier work of our own lab[11,16]. Materials, data and scripts used for this project are available via the open science framework (https://osf.io/9gru2/).

Search strategy and data gathering

To investigate the effects of ELA on structural plasticity, we conducted a systematic literature search on April 3rd 2019 on the electronic databases PubMed and WebOfScience. The search string included the terms 'mice and rats' and 'postnatal ELA' (Supplementary note 1), which was previously already used by our own lab[21]. For this particular study, ELA was defined as all postnatal models that are based on alterations in maternal care[22], either experimentally induced (through maternal deprivation of pups[23]; separation or isolation of dam and pups; or exposing dams and their litter to limited nesting and bedding material [24]), or naturally (i.e. variations in the amount of licking and grooming of pups by the dam[25], for a review [8]). Study selection was performed in Rayyan[26] by at least two researchers (see also acknowledgements, HS, DvN, LvM). The order in which the publications were assessed differed across researchers, and it occurred in two stages. In the first stage, studies were excluded based on titles and abstracts if they: 1) were not a primary publication, 2) did not use mice or rats, 3) were not related to early life adversity. During the second stage, full text was screened and studies were included if: 1) structural plasticity outcomes were measured; 2) the outcomes were measured in adulthood (older than 8 weeks but younger than 1 year of age); 3) the animals and previous generations did not experience other pharmacological / dietary / genetic interventions; 4) the animals were not germ free, were not specifically bred for certain traits and were not reported in split groups (e.g. high/low performance); 6) the sex of animals was known (either based on the report or after contacting authors); 7) in the intervention models, the control group was separated from the mothers for less than 5 min (i.e. the "handling" model was excluded). Disagreements in study selection were resolved by involving an independent scientist (MJ). An overview of the study procedure is shown in Figure 1.

Data from eligible studies was organized in a standardized database, which is available via the open science framework (https://osf.io/9gru2/). Two reviewers (VB and EK) shared the data collection task. Papers considered unclear were evaluated by both reviewers independently, and subsequently discussed with a third reviewer (MJ). It includes details about 1) the publications (author, year, reference), 2) the experimental design (e.g. species, sex, model, other life events, age and state of the animals at the time of testing), 3) information about the outcome extracted (e.g. brain

area, technique used), and 4) summary statistics of the data measured (e.g. sample size, mean or median, and deviation or interquantile range (IQR)). According to this structure, we organized the information of each individual comparison between a control and an experimental group. Of note, the groups always differed only in the experience of ELA. All other variables (e.g. additional 'life events') were comparable between the control and ELA group.

If only SEM was reported, SD was calculated as SEM * n, where n is the number of animals per group. If median and IQR were reported rather than mean and SD, we evaluated whether the median could be an approximation of the mean, i.e. the median was in the approximate center of the IQR range. If this condition was not met ($n_{comp\ excluded} = 2$), the comparison was excluded as it was not possible to obtain an effect size measure comparable to that of the other publications. If the condition was met, the median was transformed to mean according to[27] formula:

$$mean = (iqr_{low} + 2 * median + iqr_{high})/4$$

$$\sigma^2 = \frac{1}{12} * \left(\frac{(iqr_{low} - 2 * median + iqr_{high})^2}{4} + (iqr_{high} - iqr_{low})^2\right)$$

If the number of animals were reported as a range (e.g. 6–8 animals per group), we used the lower boundary of this number (e.g. 6 animals per group) as a conservative estimate. Data that was reported exclusively in graphs was digitalized with WebPlotDigitizer[28]. For all remaining missing information, we contacted the corresponding author of each manuscript after 2008 (response rate 80%). If no answer was received after 2 months and a reminder, we considered the data as not retrievable.

Data synthesis and statistical analysis

Effect sizes for each individual comparison (i.e. the standardized mean difference between control and ELA on each specific outcome) were calculated with escalc (R package metaphor, version 3.0-2[29]) using the Hedge's g (g) method, which includes a correction factor for small sample sizes[30].

For the main analysis, we used a 3-level mixed effect model which accounts for the anticipated heterogeneity of the studies as well as the dependency of effects within experiments[17]. In our experimental design, the 3 levels correspond to variance of effect size between 1) animals, 2) outcomes and 3) publications. Structural plasticity was broadly classified in three sets of outcome: a) (neuronal) morphology, b) neurogenesis, and c) growth factors, specifically BDNF. Given their different biological meaning, these were analyzed in separate models. Prior to the start of the study, we defined potential moderators of the effects of ELA on structural plasticity, namely: 1) specific outcome

parameters, 2) brain area(s), 3) experience of other traumatic events, 4) product measured (mRNA or protein, only for the outcome BDNF), 5) state of the animal at death (only for BDNF and neurogenesis, i.e. rest or not), 6) delay between the start of the experimental manipulation and measuring the outcome (only for the neurogenesis-related parameter BrdU where delay between injection and measurement gives rise to a specific interpretation of the data). The final moderators were selected based on the frequency of the available literature, to maximize interpretability and robustness of the results. Supplementary Table 1 summarizes the final models and the considerations taken.

Since most of the analytical decisions were based on frequencies, some categorizations were modified after data collection (but before analysis). These changes were based uniquely on the frequencies, with the intent to maximize the balance between subcategories, while maintaining interpretability. Generally, we used as a rule of thumb that a category could be analyzed only if it was investigated by at least 4 independent publications. Categorizations were therefore adapted to meet this requirement. Specifically, the state of the animals at death was initially coded as rest, aroused or stressed. However, due to the limited number of comparisons in the aroused/stressed categories, these were merged into a "not rest" category. Furthermore, if a study reported multiple sub-brain areas within one of our categorizations, these were combined for the quantitative synthesis to limit heterogeneity and over representation of a certain outcome in the analysis. Similarly, if a study reported multiple outcomes (e.g. multiple BDNF exons), these were combined into one measure. For volumes, this was achieved by adding together summary statistics. Given $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, $Z = X + Y$. Then, $Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

For all other outcomes, effect sizes were merged with the *metafor's* function *aggregate*[29]. For all analyses, p-values were adjusted for multiple comparisons using the *Holm* correction[31].

Lastly, we performed an additional exploratory subgroup analysis to compare specifically basal vs apical dendrites in the CA1 and CA3 hippocampal areas. In the main analysis, basal and apical dendrites were merged together in a unique effect size. For this analysis, we build two 3-levels mixed effect models (one for basal and one for apical) and compared them with a Wald test. Of note, we used all publications reporting on either basal and apical, because only 2 publications reported both within the same animals.

Heterogeneity was tested with Cochran Q-test[32] and I^2 statistics[17]. A significant test of heterogeneity or a large I^2 (see rules of thumb in [17]) signifies that the data still has variance that cannot be explained by chance alone, despite the used moderators. For those models with significant unexplained heterogeneity (i.e., neurogenesis and BDNF), we next performed an exploratory analysis to explore the source of the unexplained heterogeneity. This was conducted using *metaforest*[18], a novel exploratory approach to identify the most promising potential moderators to explain heterogeneity.

This method is an application of random forests to meta-analysis data by means of bootstrap sampling. *Metaforest* ranks moderators based on their (non)linear influence on the effect size. Here, this analysis was conducted for neurogenesis and BDNF separately. We selected 15 potential moderators for *metaforest* analysis. After identifying a convergence range, we conducted a recursive preselection based on 100 replications, and selected only those moderators that were selected in at least 50% of the replications. With these variables, we built our *metaforest* model and conducted a 10x cross validation to determine the optimal tuning parameters that minimized root-mean-square deviation (for neurogenesis: random weights, 2 candidate moderators at each split, minimum node size = 4; for BDNF: fixed weights, 4 candidate moderators at each split, minimum node size = 5). To estimate how much variance was explained by our model, we calculated the cross-validated R^2 (R_{cv}^2), which is robust to overfitting and provides evidence for the results' generalizability.

Sex differences

Prior to the beginning of the study, we planned to conduct our analyses for males and females separately, since the effects of stress on brain and behavior have often been shown to differ across sexes[19,20], at least regarding effect size[11]. However, due to the limited number of publications in females, quantitative analysis was feasible only in the males' dataset. As an alternative, we focused on investigating sex differences in a subset of publications reporting data on both sexes. Although with this sex-matched dataset it still was not possible to explore sex differences related to specific outcomes, we can investigate whether there are fundamental sex differences in the effect sizes, for example due to male-developed ELA models[11].

We calculated the effect sizes (Hedge's *g*) for males and females separately on a subset of studies, i.e. the sex-matched dataset. Two identical models were built for the data subsets, one for each sex, without moderators due to the limited amount of evidence available. In these models, we used the absolute value of the effect size since the different outcomes may have opposing effects thereby cancelling each other out in the meta-analytic model. We then performed a Wald test to compare the female versus male models.

Bias assessment

To assess risk of bias, we followed SYRCLE's risk of bias guidelines[33]. Two reviewers (EK and VB) assessed risk of bias independently on the whole dataset, and resolved disagreements with discussion. To the best of our knowledge no quantitative method is available for the inspection of publication bias for a multi-level setting. Publication bias was therefore assessed on the univariate models for each of the functional outcomes (morphology, neurogenesis and BDNF) by qualitative inspection of funnel plot asymmetry, adapted using a measure of pooled standard deviation in the formula

for precision (1/variance) as suggested by Vesterinen and colleagues[30]. Contrary to our initial study protocol, we did not conduct a Egger's regression, because the number of publications was low and Egger's regression would have most likely been underpowered[34]. Rather, we interpret the probable influence of publication bias based on the areas of significance, following[34].

Software

The analyses were conducted in R (version 3.5.1) (R Core Team, 2015), using the following packages: 1) *metafor* [29] version 3.0-2 for conducting the analysis, 2) *metaforest*[18] version 0.1.3 for data exploration, 3) dplyr [35] version 1.0.7 for general data handling. The R script and data are available (https://osf.io/9gru2/).

Results

Study selection and qualitative analysis

An overview of the study design is summarized in the flow chart (Figure 1). Our pre-specified inclusion criteria (see Methods) were met by 64 publications, published between 2002 and 2018. The included publications contributed 110 unique experiments, with a total of 260 comparisons from which we extracted statistical measurements (e.g. mean, standard deviation (SD) and sample size (N)). 9 comparisons from 3 publications were excluded from the analysis, as it was not possible to extract nor infer any statistical measurement.

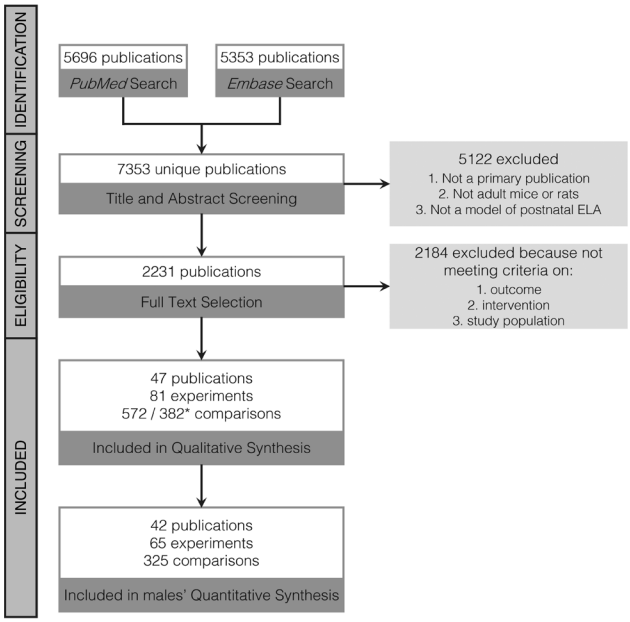


Figure 1. PRIMA flowchart for study selection and inclusion. * = 538 is the number of comparisons before preprocessing; 260 refers to the number included after processing. For information about preprocessing, see Methods.

The included publications used mainly rats (n_{publ} = 83%) and the maternal separation model to mimic ELA (n_{publ} = 66%), followed by the limited nesting and bedding model (n_{publ} = 17%), maternal deprivation (n_{publ} = 11%), observation of natural variations in licking and grooming (n_{publ} = 5%) and isolation (n_{publ} = 1-2%).

To study structural plasticity after ELA, we focused on morphology (i.e., changes in the size of brain areas, morphology of dendrites or spine density), neurogenesis (i.e., staining for BrdU, DCX and Ki67) and the growth factor BDNF. Table 1 summarizes the frequency of each outcome.

Table 1. Outcome frequencies in both sexes. The functional categorizations correspond to the classifications for the analyses. # = number

Functional interpretation	outcome	#studies	#experiments	#comparisons
morphology	volume	10	17	35
morphology	dendritic changes	14	17	49
morphology	spines	6	8	18
neurogenesis	BrdU + cells	16	28	30
neurogenesis	DCX staining	13	17	17
neurogenesis	Ki67 staining	11	14	14
growth factor	BDNF	28	44	97

In total, more than 10 brain areas were investigated, with most studies describing the hippocampus (67% of all comparisons in males and females, Figure 2). Within the hippocampus, 81 comparisons were from the dentate gyrus, 22 from the CA1, 22 from the CA3, 3 from the CA4, while 47 measured the whole hippocampus (unspecified); CA4 was excluded from further analysis because of a too low number.

A total of 3336 animals were used, of which the majority (79%) was male. Thirteen publications performed experiments in both male and female rodents. Similar to previous studies, we aimed to analyze males and females separately, i.e. as two different biological systems, since sex-dependent characteristics have been frequently observed in stress research[19,20]. However, data on females was too scarce to be analyzed quantitatively at a meta-level. We therefore focused our quantitative analyses on males, and subsequently performed an exploratory analysis on a subset of the data to explore potential sex differences.

Based on the frequencies reported above, we included only the outcomes of the hippocampus in subsequent quantitative synthesis, in male mice. Similarly, the number of dendrites (n_{publ} = 2) and spine density (n_{publ} = 3) were not included in the meta-analyses due to the limited number of publications. The descriptive results of these two parameters are summarized in Table 2.

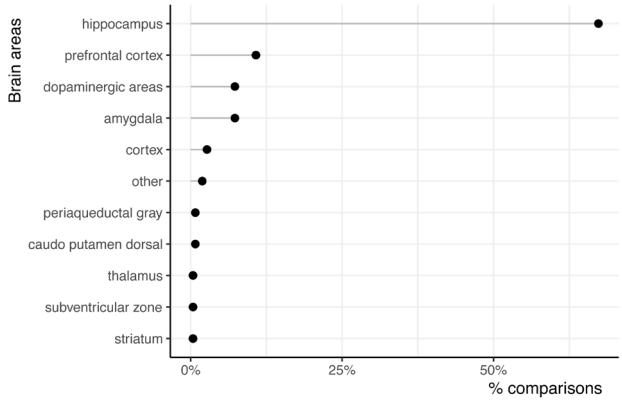


Figure 2. Distribution of the brain areas investigated in relationship to ELA and structural plasticity. The categorizations follow the Allen Brain Atlas embryological classification.

Table 2. Summary evidence on spine density and number of dendrites. MS = Maternal Separation, MD = Maternal Deprivation, LNB = limited nesting and bedding; in 'Origin', we consider purchasing pregnant dams as a stressful experience due to transportation stress of the dams; in 'Behavior', naïve = animals left undisturbed (regardless of ELA or control treatment), non stressful = animals that performed non-stressful behavior tests (e.g. open field), stressful = animals that performed stressful behavior tests (e.g. fear conditioning), no behavior = the animals were not naïve, but did not perform behavioral tests (e.g. saline injections); in 'Major Life Events', experiments score "yes" if the animals experienced other (besides ELA, prenatal transport stress and stressful behavior tests) traumatic life events e.g. chronic stress during adolescence; in 'at death', we defined the status of the animal at death, i.e. at rest or aroused/stressed; g = Hedge's g; SE = Sampling Error; sig = systematic review significance, where "+" means increase, "-" means decrease and "ns" means "not significant".

study	model	origin	behavior	major life events	at death	outcome	brain area	g	SE	sig
Bathalta (2013)[36]	MS	purchased pregnant dams	stressful	yes	rest	#basal dendrite	CA1	1.21	0.44	+
Oomen (2010)[37]	MD	own breeding	naïve	no	rest	#primary dendrites	dentate gyrus	-2.01	0.49	-
De Melo (2018)[38]	MS	not specified	non stressful	no	rest	spines	CA1	5.34	1.55	+
De Melo (2018)[38]	MS	not specified	non stressful	no	rest	spines	dentate gyrus	-0.45	0.35	-
Liu (2016)[39]	LNB	purchased naive parents	no behavior	no	rest	spines	CA3	-3.21	0.85	-
Wang (2011)[40]	LNB	own breeding	no behavior	yes	rest	spines	CA1	0.37	0.51	ns
Wang (2011)[40]	LNB	own breeding	no behavior	yes	rest	spines	CA3	-2.39	1.14	-
Wang (2011)[40]	LNB	own breeding	no behavior	yes	rest	spines	dentate gyrus	-0.28	0.51	ns

Quantitative analysis of morphology

A 3-level model was used to investigate whether ELA significantly impacted morphology of the adult hippocampus. In particular, we analyzed i) whether the effects differed across outcomes (volume of the brain area, number of dendritic branches and length of dendrites); and ii) whether other traumatic life experiences interacted with the effects. Of note, the groups compared always differed only in the experience (or not) of ELA. Therefore, effects of other traumatic life events should be considered as "enhancing" (or not) the effects of early life adversity.

Overall, ELA significantly reduced the volume of the hippocampus (g(se) = -0.819 (0. 185), t = -4. 424, p = 0.001) and decreased the total dendritic length (g(se) = -1. 66 (0.303), t = -5.473, p < 0.001). The decrease in the number of branches per dendritic tree (g(se) = -0.699(0.262), t = -2.663, p = 0.053) was just not significant (Figure 3a). The effects were largest when both the ELA and control groups experienced no other traumatic events, (g(se) = 1.113(0.268), t = 4.16, p = 0.002, Figure 3c); qualitatively, this was consistent across all outcomes (Supplementary Figure 1).

The effects appeared similar in all sub-fields of the hippocampus (Table 3). The effects on apical dendrites were more pronounced than on basal dendrites (apical: g(se) = -1.08(0.299), z = -3.61, p < 0.001; basal: g(se) = -0.246 (0.237), z = -1.037, p = 0.3), as highlighted by a subgroup analysis (g(se) = 0.835(0.382), z = 2.185, p = 0.029, Figure 3b).

Table 3. No differences across hippocampal sub-brain areas.

comparison	g	se	t	p.adj
dentate vs CA1 + CA3	0.158	0.312	0.508	0.875
CA1 vs dentate + CA3	0.226	0.286	0.789	0.875
CA3 vs dentate + CA1	-0.571	0.290	-1.968	0.181

Quantitative analysis of neurogenesis

Neurogenesis was determined in the dentate gyrus, more specifically in the subgranular zone; one paper[41] was excluded since it reported on the whole hippocampus. Although this may reflect what happens in the dentate gyrus, we excluded the paper to maintain consistency of our sample.

Concerning neurogenesis, ELA significantly decreased the expression of DCX (g(se) = -0.825 (0.299), t = -2.764, p = 0.039), a marker for neuronal differentiation thought to stain immature neurons[42] (Figure 4). BrdU staining was suppressed after ELA with short (g(se) = -1.335(0.327), t = -4.081, p = 0.002) but not long (g(se) = -1.119(0.435), t = -2.57, p = 0.046) delay since injection. These are generally considered as markers of proliferation and survival, respectively[43]. By contrast, Ki67

expression (a marker of proliferation[44]) was unaffected (g(se) = -0.309(0.388), t = -0.797, p = 0.863). Overall, the results were comparable both for (ELA and control) groups that experienced traumatic life events and those that did not (g(se) = 0.182(0.29), t = 0.63, p = 0.863, Figure 4b).

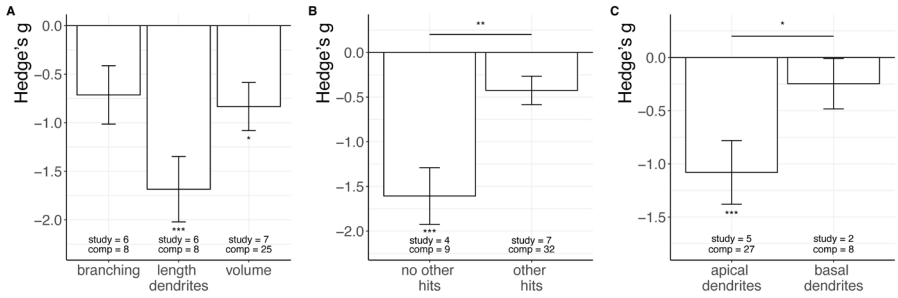


Figure 3. Results on males' morphology in the hippocampus. A) ELA decreases the volume of the hippocampus and the length of hippocampal dendrites. B) ELA significantly reduces the overall outcome of morphology, both for animals that did not and did experience other major life events. However, the effects are more pronounced when animals (ELA compared to control) did not experience other major life events ("hits"). C) The effects of ELA are more pronounced for apical than for basal dendrites. For each bar the numbers at the bottom refer to the number of studies (study) and comparisons (comp) respectively on which the mean is based. *** = $p_{adj} < 0.001$; ** = $p_{adj} < 0.01$; * = $p_{adj} < 0.05$; study = number of independent publications; comp = number of comparisons (ie difference between ELA and control groups); hits = experience of major life events.

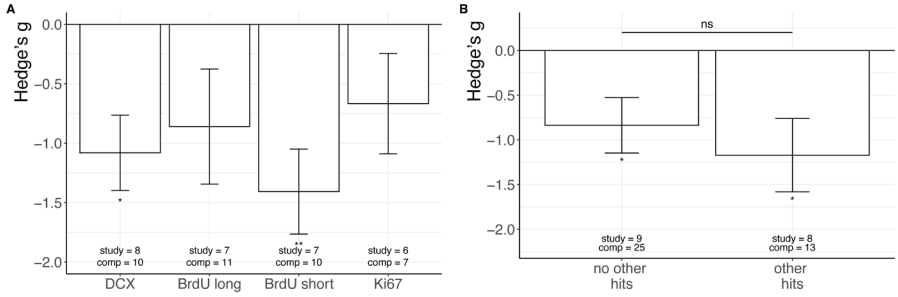


Figure 4. Results of males' neurogenesis in the hippocampus. A) ELA decreases dcx expression and BrdU expression after a short time since BrdU injection. ELA and control groups are not significantly different in Ki67 expression and in BrdU expression with a long time since injection. B) The effects of ELA are comparable whether or not (ELA and control) animals experienced other major life events ("hits"). *** = $p_{adj} < 0.001$; ** = $p_{adj} < 0.01$; * = $p_{adj} < 0.05$; study = number of independent publications; comp = number of comparisons (ie difference between ELA and control groups); hits = experience of major life events.

Quantitative analysis of BDNF

Overall, ELA did not alter BDNF expression (g(se) = -0.32(0.23), t = -1.412, p = 0.994), and the pre-specified moderators (type of outcome investigated (mRNA or protein), experience of other traumatic life events, and status of the animal at death (rest / not rest)) did not explain a significant portion of the variance ($Q_m(8)$ = 8.437, p = 0.392). Qualitative exploration of the effect sizes (Supplementary Figure 2) suggests that there may be complex 3-way interactions between the factors considered.

Exploratory analysis with *metaforest*

With regard to morphology, in our 3-level model the moderators cumulatively explained a significant portion of the variance ($Q_m(16)$ = 45.42, p<0.001). The remaining heterogeneity was not significant ($Q_h(25)$ = 21.065, p = 0.689, I^2 = 27.98%), thereby suggesting that no other moderators are necessary to explain the effects of ELA on morphology.

This was different in the case of (all markers of) neurogenesis. Here the moderators did explain a significant portion of the variance ($Q_m(8)$=21.762, p=0.005); however, there was still unexplained heterogeneity in the model ($Q_h(27)$=100.154, p < 0.001), suggesting that additional moderators may be relevant to explain the effects. This was next assessed with an exploratory moderator analysis with *metaforest*. After a thresholded preselection, *metaforest* ranks moderators based on how much variance they can explain using random forests. Of the 13 variables investigated, 6 were selected as having a positive variable importance in at least 50% of the 100 bootstrap replications (Figure 5a). Specifically, the factor 'own breeding of the experimental animals' yielded larger effect sizes compared to animals of different origins (e.g. purchasing naïve parents or pregnant females). Besides this, qualitative exploration of the partial dependency plots (Supplementary Figure 3) suggests that the other factors may not be biologically relevant, as supported by the fairly low explained variance (R_{cv}^2(SD) = 0.385(0.33), R_{oob}^2(MSE) = 0.038(1.56)).

The model on BDNF also had moderate remaining unexplained variance ($Q_h(25)$ = 114.145, p < 0.001, I^2 = 67.33), and from pre-selected confirmatory analysis none of the moderators explained a significant portion of the variance (see section "Quantitative analysis on BDNF"). We therefore chose to use *metaforest* to explore other potential moderators of the effects. Of the 13 potential moderators selected for *metaforest* analysis, 3 were selected because they had a positive variable importance in at least 50% of the 100 bootstrap replications (Figure 5b). Specifically, partial dependency plots suggest that 1) animals that did not perform behavior tasks and 2) animals that experienced major other life events had larger effect sizes (Supplementary Figure 4). However, the explained variance was still low (R_{cv}^2(SD) = 0.366(0.35), R_{oob}^2(MSE) = -0.0256 (1.46)).

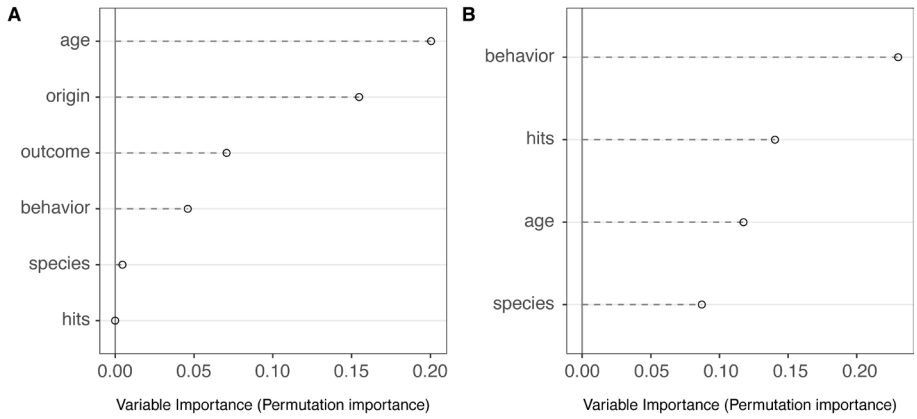


Figure 5. Metaforest variable importance plots after preselection for (A) neurogenesis and (B) BDNF.

Comparing males and females

As argued, the number of studies reporting on female animals was quite low, prohibiting a full meta-analysis. However, we performed an exploratory analysis in those studies (n_{publ} = 12, n_{comp} = 39) that reported on both sexes. The obtained dataset contained comparisons from all outcomes and brain areas. Figure 6 visualizes the relationship between male and female data.

Using a Wald test on this subset of the data no evidence for any sex difference was discerned in effect sizes ($g_{males\ vs\ females}$(se) = 0.14(0.14), z = 0.996, p = 0.32), thereby suggesting that there is no evidence for an overall difference between males and females regarding the effects of ELA on structural plasticity. Given the limited dataset, though, sex differences on specific outcomes and/or brain areas can certainly not be excluded.

Bias assessment

Risk of bias (Figure 7) was assessed using SYRCLE's risk of bias tool[33]. Although no publication reported on all items, only two publications did not report being blinded and randomized. However, only four publications provided sufficient information to interpret how randomization was performed. Most importantly, no publication took measures to reduce bias in selective outcome reporting, and this may have been a potential bias in 67.8% of the publications.

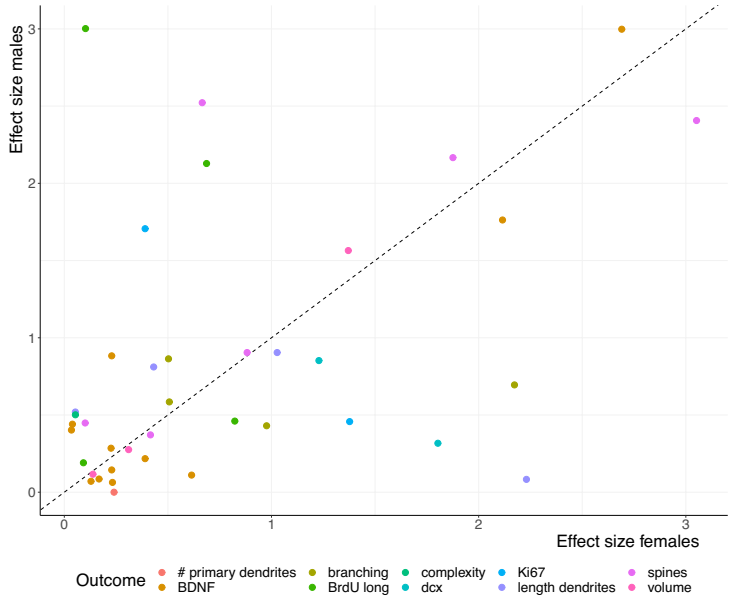


Figure 6. Relationship between male and female effect sizes. Each dot corresponds to the effect size (*g*) in males and females, obtained in the same publication for the same parameter. The dotted line corresponds to the 45 degrees line where all dots should be if males and females had identical effect sizes. Deviations from the line could be due to error or to real sex differences.

Figure 7. Risk of bias assessment. Each row corresponds to a separate item of the SYRCLE's risk of bias assessment. Each pubmed ID refers to an independent publication. For the full data, see https://osf.io/9gru2/.

We evaluated publication bias by assessing funnel plot asymmetry (Figure 8) separately for morphology, neurogenesis and BDNF. Concerning morphology (Figure 8a), there is evidence of asymmetry in the funnel plot. However, only one comparison was present in the highest significance area. Although this may be due to reporting bias, it is unlikely to affect the interpretation of the results. For the neurogenesis and BDNF analyses, the same asymmetry in the funnel plot is observed; yet, here several comparisons are in the high significance area. This suggests that there may be some publication bias, which could lead to an overestimation of effect sizes in the current study. This potential bias could be an important factor when considering the remaining unexplained heterogeneity in our models.

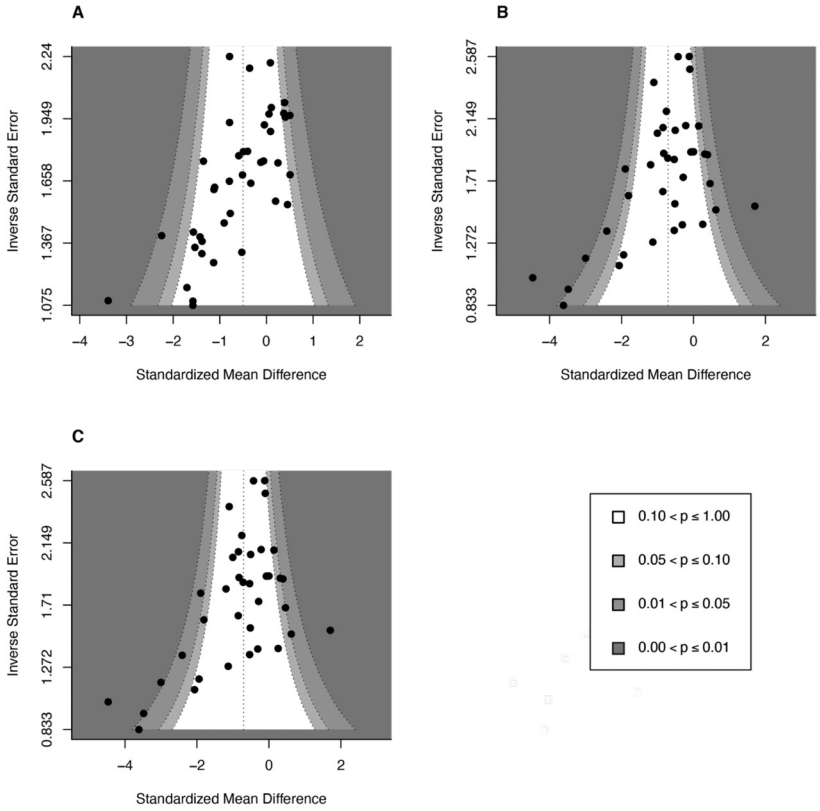


Figure 8. Funnel plot for assessment of publication bias for effects of ELA on (A) morphology, (B) neurogenesis and (C) BDNF analyses. Each dot corresponds to a comparison between control and ELA male groups. *Dark gray = areas of highest significance (p-value). The darker the grey the higher the significance as reported by the legend.*

Discussion

We set out to review the effects of ELA on neuronal structure and structural plasticity in adult rodents. The survey was restricted to studies in rats and mice, and adversity (here limited to altered maternal care) experienced during the first two weeks of life, focusing on i) the volume of adult brain areas and the morphology of neurons; ii) adult neurogenesis in the dentate gyrus; and iii) expression of BDNF. The descriptive analysis showed that most studies reported on structural changes in the hippocampus of male rats, in most cases using maternal separation as model of ELA. The bias towards the hippocampus is to some extent – but not entirely, i.e. not for morphology – explained by the fact that adult neurogenesis is restricted to a limited number of brain areas, including the dentate gyrus[45]. Given the distribution of papers, subsequent quantitative analyses were restricted to observations in the hippocampus of male rodents.

Nearly all structural markers were found to be significantly suppressed after

ELA compared to control treatment. This was generally not affected by exposure to additional stressors in life or by the state (at rest or aroused / stressed) of the animal immediately before the experiment. Surprisingly, no effects were identified for BDNF, while there was still remaining unexplained heterogeneity. This could possibly be due to data preprocessing (e.g. merging BDNF exons of a subset of papers to maintain consistency across publications), which may have "diluted" the results. Also, we cannot exclude that alterations in BDNF expression took place in the time that elapsed between ELA and the measurements in adulthood. Finally, some caution regarding these conclusions is necessary, since there are indications for publication bias, particularly in the case of neurogenesis and BDNF.

As argued, it is slightly surprising – except in the case of adult neurogenesis – that nearly all studies focused on the hippocampus, with substantially lower numbers of reports on the prefrontal cortex, midbrain dopaminergic areas and the amygdala and (near)absence of studies in the remainder of the brain. It is likely that structural effects of ELA are not restricted to pyramidal neurons in the hippocampus and may well occur e.g. in other pyramidal neurons of the cortex. However, at this stage findings in the hippocampus cannot be simply extrapolated to other regions. This is underlined by, for instance, one group of investigators showing reduced dendritic length in the dentate gyrus of adult male rats earlier exposed to 24 h of maternal deprivation on postnatal day 3[37], while no such changes were observed in the basolateral amygdala[46]. There is a clear need for extension of the current literature to areas other than the hippocampus.

We started out by investigating male and female rodents separately, expecting differences based on earlier reports[19,20,47]. The number of reports on female rats or mice, however, was so low that a solid quantitative analysis was not possible. We therefore only carried out an exploratory analysis, using those studies that investigated both sexes. This allowed a comparison in effect sizes in a presumably less heterogeneous sub-group, sharing at least within-study conditions like the experimental procedures, the experimenters carrying out the study and housing conditions of the animals. Although the number of observations was low and varied, there was no evidence that effect sizes were consistently smaller (or larger) in females than in males. Nevertheless, the sparsity of studies in female rodents underlines the message that females are heavily understudied, which may leave potential differences undiscerned[48].

The descriptive analysis also underlined that most studies to date have been carried out in rats rather mice, despite the fact that reliable models for ELA are available in mice too[49,50]. Clearly, there are substantial differences within and between species, e.g. with regard to anxiety-proneness[51]. To what extent this affects the way in which ELA causes lasting changes in brain structure and structural plasticity remains an unresolved issue until more studies in mice models have been carried out. This also holds true for the type of early life adversity employed in the models, which is currently dominated by maternal separation for several hours during the first 2 postnatal weeks.

This model is characterized by a large degree of predictability for the pups[52], in contrast to e.g. the limited bedding/ nesting material model 24 or a single (24 h) period of maternal deprivation. The latter model has revealed that the exact (postnatal) day of deprivation is crucial for the consequences in adulthood (e.g.[53]), most likely related to (among other things) the development of the brain and hypothalamus-pituitary-adrenal system at the time of deprivation.

The reduction in volume and dendritic characteristics after ELA were quite robust. The two moderators included in the model, i.e. outcome and cumulative life experiences, explained a significant part of the variation. Of note, we did not consider the possibility that these outcomes would be affected by the state of the animal (at rest versus aroused or stressed) just before the experiment, arguing that changes in volume and particularly dendritic complexity require at least hours to develop. This is different from markers involved in proliferation or the expression of growth factors, which could also be influencing volume. *Metaforest* analyses suggested that other life experiences too could be important moderators of the effects of ELA on neurogenesis and outcomes. Specifically, origin of the breeding animals and cumulative life experiences were identified for neurogenesis and BDNF, respectively. For BDNF, the presence of other cumulative life experiences in this list may appear as a surprise, since confirmatory moderator analysis was not significant. This is due to the underlying assumptions of the analysis: either due to the selection method (a pre-specified p-value in moderator analysis vs a permutation approach in *metaforest*), or to non-linear effects that can be established with *metaforest* but not with the moderator approach. Future research is required to disentangle these two possibilities. Interestingly, origin of the breeding animals and cumulative life experiences ("hits") were also important moderators for the effects of ELA on behavioral outcomes[11]. For instance, transporting pregnant dams resulted in stronger effects of ELA on behavioral phenotype than seen with in-house bred dams. In the cases of neurogenesis, also the time elapsed between injecting BrdU and immunohistochemical analysis turned out to be a moderator, likely related to BrdU being an index for proliferation or cell survival, depending on the time elapsed[43]. Interestingly, while BrdU staining shortly after injection was significantly reduced after ELA, no significant change was observed for the proliferation marker Ki67. One explanation could be that effects of ELA are most apparent in the S-phase of dividing cells[54], for which BrdU is a more specific marker than Ki67. Since Ki67 is also present during the G1, G2 and M-phases of cell proliferation, this could have obscured potential effects of ELA in the S-phase.

All in all, we observed a consistent suppressive effect of ELA during the first postnatal weeks on adult structural markers in the hippocampus, specifically on volume, dendritic characteristics and neurogenesis. Possibly, ELA-dependent changes in the activity of growth factors like BDNF could explain such structural changes, although there may be many other pathways through which ELA can lastingly affect

structural markers in adulthood. Given the limitation of the vast majority of current reports to the hippocampal area, to one model of early life adversity only (maternal separation) and to male rats, a larger diversity of studies will be necessary to resolve the quest how lasting ELA-dependent structural changes can contribute to changes in behavior.

References

1. Chen, Y. & Baram, T. Z. Toward Understanding How Early-Life Stress Reprograms Cognitive and Emotional Brain Networks. *Neuropsychopharmacology* 41, 197–206 (2016).
2. Herzberg, M. P. & Gunnar, M. R. Early life stress and brain function: Activity and connectivity associated with processing emotion and reward. *NeuroImage* 209, 116493 (2020).
3. Sandi, C. & Haller, J. Stress and the social brain: behavioural effects and neurobiological mechanisms. Nat. Rev. Neurosci. 16, 290–304 (2015).
4. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nat. Rev. Neurosci.* 17, 652–666 (2016).
5. Teicher, M. H., Gordon, J. B. & Nemeroff, C. B. Recognizing the importance of childhood maltreatment as a critical factor in psychiatric diagnoses, treatment, research, prevention, and education. *Mol. Psychiatry* 1–8 (2021) doi:10.1038/s41380-021-01367-9.
6. Brown, A., Fiori, L. M. & Turecki, G. Bridging Basic and Clinical Research in Early Life Adversity, DNA Methylation, and Major Depressive Disorder. *Front. Genet*. 10, 229 (2019).
7. Hughes, K. et al. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *Lancet Public Health* 2, e356–e366 (2017).
8. Benmhammed, H. et al. Animal Models of Early-Life Adversity. in *Psychiatric Disorders: Methods and Protocols* (ed. Kobeissy, F. H.) 143–161 (Springer, 2019). doi:10.1007/978-1-4939-9554-7_10.
9. Branchi, I. & Cirulli, F. Early experiences: Building up the tools to face the challenges of adult life. *Dev. Psychobiol.* 56, 1661–1674 (2014).
10. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr. Opin. Psychol.* 15, 174–181 (2017).
11. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neurosci. Biobehav. Rev.* 102, 299–307 (2019).
12. Rocha, M., Wang, D., Avila-Quintero, V., Bloch, M. H. & Kaffman, A. Deficits in hippocampal-dependent memory across different rodent models of early life stress: systematic review and meta-analysis. *Transl. Psychiatry* 11, 231 (2021).
13. Wang, D., Levine, J. L. S., Avila-Quintero, V., Bloch, M. & Kaffman, A. Systematic review and meta-analysis: effects of maternal separation on anxiety-like behavior in rodents. *Transl. Psychiatry* 10, 174 (2020).
14. Anacker, C., O'Donnell, K. J. & Meaney, M. J. Early life adversity and the epigenetic programming of hypothalamic-pituitary-adrenal function. *Dialogues Clin. Neurosci.* 16, 321–333 (2014).
15. Kempermann, G., Song, H. & Gage, F. H. Neurogenesis in the Adult Hippocampus. *Cold Spring Harb. Perspect. Biol*. 7, a018812 (2015).
16. Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: A 3 level meta-analysis of rodent studies. *Neurosci. Biobehav. Rev.* 95, 1–16 (2018).
17. Cheung, M. W.-L. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol. Methods* 19, 211–229 (2014).
18. Lissa, C. J. van. Small Sample Meta-Analyses: Exploring heterogeneity using MetaForest. in *Small Sample Size Solutions* (Routledge, 2020).
19. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. *Nat. Neurosci.* 18, 1413–1420 (2015).
20. Bangasser, D. A. & Cuarenta, A. Sex differences in anxiety and depression: circuits and mechanisms. *Nat. Rev. Neurosci.* 1–11 (2021) doi:10.1038/s41583-021-00513-0.
21. Schuler, H., Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. *Effects of early life adversity on immediate early gene expression: systematic review and 3-level meta-analysis of rodent studies.* http://biorxiv.org/lookup/doi/10.1101/2021.06.07.447369 (2021) doi:10.1101/2021.06.07.447369.
22. Levine, S. Regulation of the hypothalamic-pituitary-adrenal axis in the neonatal rat: the role of maternal behavior. *Neurotox. Res.* 4, 557–564 (2002).
23. Marco, E. M. et al. The maternal deprivation animal model revisited. *Neurosci. Biobehav. Rev.* 51, 151–163 (2015).
24. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A Novel Mouse Model for Acute and Long-Lasting Consequences of Early Life Stress. *Endocrinology* 149, 4892–4900 (2008).
25. Champagne, F. A., Francis, D. D., Mar, A. & Meaney, M. J. Variations in maternal care in the rat as a mediating

influence for the effects of environment on development. *Physiol. Behav.* 79, 359–371 (2003).
26. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. *Syst. Rev.* 5, 210 (2016).
27. Hozo, S. P., Djulbegovic, B. & Hozo, I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med. Res. Methodol.* 5, 13 (2005).
28. Rohatgi, Ankit. *Webplotdigitizer*: Version 4.5. (2021).
29. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* 36, 1–48 (2010).
30. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. *J. Neurosci. Methods* 221, 92–102 (2014).
31. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat*. 6, 65–70 (1979).
32. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558 (2002).
33. Hooijmans, C. R. et al. SYRCLE's risk of bias tool for animal studies. *BMC Med. Res. Methodol.* 14, 43 (2014).
34. Sterne, J. A. C. et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343, d4002 (2011).
35. Wickham, H., François, R., Henry, L., Müller, K. & RStudio. *dplyr: A Grammar of Data Manipulation.* (2021).
36. Batalha, V. L. et al. Adenosine A2A receptor blockade reverts hippocampal stress-induced deficits and restores corticosterone circadian oscillation. *Mol. Psychiatry* 18, 320–331 (2013).
37. Oomen, C. A. et al. Severe Early Life Stress Hampers Spatial Learning and Neurogenesis, but Improves Hippocampal Synaptic Plasticity and Emotional Learning under High-Stress Conditions in Adulthood. *J. Neurosci.* 30, 6635–6645 (2010).
38. de Melo, S. R., de David Antoniazzi, C. T., Hossain, S. & Kolb, B. Neonatal Stress Has a Long-Lasting Sex-Dependent Effect on Anxiety-Like Behavior and Neuronal Morphology in the Prefrontal Cortex and Hippocampus. *Dev. Neurosci.* 40, 93–103 (2018).
39. Liu, R. et al. Early postnatal stress suppresses the developmental trajectory of hippocampal pyramidal neurons: the role of CRHR1. *Brain Struct. Funct.* 221, 4525–4536 (2016).
40. Wang, X.-D. et al. Forebrain CRF1 Modulates Early-Life Stress-Programmed Cognitive Deficits. J. Neurosci. 31, 13625–13634 (2011).
41. Bath, K. G. et al. Early life stress leads to developmental and sex selective effects on performance in a novel object placement task. *Neurobiol. Stress* 7, 57–67 (2017).
42. Klein, B. et al. DCX+ neuronal progenitors contribute to new oligodendrocytes during remyelination in the hippocampus. *Sci. Rep.* 10, 20095 (2020).
43. Wojtowicz, J. M. & Kee, N. BrdU assay for neurogenesis in rodents. *Nat. Protoc.* 1, 1399–1405 (2006).
44. Sun, X. & Kaufman, P. D. Ki-67: more than a proliferation marker. *Chromosoma* 127, 175–186 (2018).
45. Ming, G. & Song, H. Adult Neurogenesis in the Mammalian Brain: Significant Answers and Significant Questions. *Neuron* 70, 687–702 (2011).
46. Krugers, H. J. et al. Maternal deprivation and dendritic complexity in the basolateral amygdala. *Neuropharmacology* 62, 534–537 (2012).
47. Parel, S. T. & Peña, C. J. Genome-wide Signatures of Early-Life Stress: Influence of Sex. *Biol. Psychiatry* S0006-3223(20)32119–3 (2020) doi:10.1016/j.biopsych.2020.12.010.
48. Shansky, R. M. & Murphy, A. Z. Considering sex as a biological variable will require a global shift in science culture. *Nat. Neurosci.* 24, 457–464 (2021).
49. Peña, C. J., Nestler, E. J. & Bagot, R. C. Environmental Programming of Susceptibility and Resilience to Stress in Adulthood in Male Mice. *Front. Behav. Neurosci.* 13, 40 (2019).
50. Walker, C.-D. et al. Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress Amst. Neth.* 20, 421–448 (2017).
51. Clément, Y., Calatayud, F. & Belzung, C. Genetic basis of anxiety-like behaviour: a critical review. *Brain Res. Bull.* 57, 57–71 (2002).
52. Daskalakis, N. P. et al. The newborn rat's stress system readily habituates to repeated and prolonged maternal separation, while continuing to respond to stressors in context dependent fashion. *Horm. Behav.* 60, 165–176 (2011).
53. Enthoven, L., de Kloet, E. R. & Oitzl, M. S. Differential development of stress system (re)activity at weaning dependent on time of disruption of maternal care. *Brain Res.* 1217, 62–69 (2008).
54. Bannigan, J. G. The effects of 5-bromodeoxyuridine on fusion of the cranial neural folds in the mouse embryo. *Teratology* 32, 229–239 (1985).

CHAPTER 9

Seven principles of early life adversity in rodents

In preparation

Bonapersona V[1]
Sarabdjitsingh RA[1]
Joëls M[1,2]

1 University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, NL.
2 University Medical Center Groningen, Groningen, NL.

Abstract

Early life adversity (ELA) alters brain development and function, and it is one of the main risk factors for several psychiatric disorders. Rodent models have been used to better understand the underlying biological mechanisms, but results of single studies are rarely reliable due to various sources of biases. Here, we comprehensively review the literature on the effects of ELA on brain and behavior, with a specific focus on the limbic system. We critically appraise the results of 10 meta-analyses investigating the effects of ELA, based on alteration of maternal care during the first postnatal weeks, on several outcomes. From this we delineated seven principles of ELA in rodents. We propose that a community effort is required to keep integrating the accumulating knowledge on the single outcomes. In all, we are at a turning point towards an integrated understanding of the effects of ELA on rodents' brain and behavior.

Early Life Adversity as risk factor for psychopathology

Experience shapes how the developing brain is structured and functions[1]. Adverse experiences can impact brain development[2], with effects visible years after the initial event: one in three adults diagnosed with a psychiatric disorder experienced adversity early in life[3,4]. Early-life adversity (ELA) appears to leave a unique (neuro)biological trace[5]. This notion is based on comparisons of maltreated and non-maltreated individuals with the same diagnosis. For example, differences in grey matter volume between healthy controls and patients with a schizophrenia or bipolar diagnosis were observed only in the patients with a history of moderate-to-high ELA[6]. The neurobiological differences are so profound that it has been suggested that individuals with a history of ELA should have a self-standing DSM diagnosis[5]. A wide-spread idea is that clinical advances can emerge by understanding the connection between ELA and brain circuits' disruption[1]. Better understanding would mean better categorization and better treatment[5]. However, it is unethical to manipulate time and exposure to ELA in humans[7,8], and it is often not feasible to conduct human studies sufficiently powered to examine higher order (genetic and environmental) interactions[7].

Rodents to investigate Early Life Adversity

Many have therefore reverted to rodents to investigate critical aspects of the molecular and cellular mechanisms of ELA that non-invasive human research cannot explore[9]. Rodents and humans are altricial species: they are born undeveloped, and they require care by their parents. This relationship between the primary caregiver and the infant is one of the most important elements for development and most critical environmental factors early in postnatal life[1,10,11]. It can be disrupted in either quantity or quality[12], having an impact on cognitive and emotional processing lasting into adulthood[13]. Cross-species evidence suggests convergence of ELA mechanisms, particularly connected to the stress-system and cortico-limbic network, which are evolutionarily conserved[1]. Rodent models of ELA are therefore suggested to hold face, construct and predictive validity[14]. Undoubtedly, there are inherent problems with using rodents as a model. Although some mechanisms are evolutionarily conserved, a rodent brain is different from a human brain, e.g. neocortical areas are relatively small[15]. Furthermore, some important behaviors (e.g. tasks that require language) cannot be explored in rodents, and in all cases rodent behavior can only be interpreted "from the outside". Lastly, important cultural and social aspects[16] of the human experience cannot be modelled in rodents. Rodents can however be used to model specific aspects of human ELA and what these aspects cause to the brain, while controlling or experimentally varying genes and environment. They can be used for invasive measurements, and have the advantage that the time between ELA and adulthood is much shorter than in humans (2 months rather than 20 years)[14,17].

A comprehensive database of rodent Early Life Adversity

Rodent studies, however, have recently received a wave of methodological criticism[18]: from limited sample sizes used[19], to suboptimal reporting (for example [20]) and the widespread presence of various forms of bias (for example [21,22]). These issues seriously hamper the reproducibility, generalizability and translatability of rodent findings to humans [23,24]. Single, individual rodent studies are unlikely to provide reliable information[19]; yet, this can be achieved by investigating a body of literature with systematic reviews and meta-analyses[25]. While these are generally geared to specific and narrow questions[26], other methods – like evidence maps – have a broader scope: to identify the types of evidence and potential gaps in the literature[27]. These approaches all use systematic methods to minimize bias and to produce more reliable findings to inform decision making[28]. However, neither of these methods addresses the integrated state of a system.

Here, we explore the possibility to integrate these methodologies, to provide a detailed yet broad quantified overview of the stable effects of ELA on rodents' brain and behavior. For this, we critically appraised ten systematic reviews and meta-analyses previously conducted in our own lab (Chapters 4 to 8 [29–32]). Outcomes were investigated at different biological levels: from structural outcomes, such as volume, dendritic morphology, neurogenesis, and potential contributing mediators like brain derived neurotropic factor (BDNF); passing through biochemical indicators of monoamines, GABA and glutamate; immediate early genes; functional data obtained with electrophysiology; and finally behavior. Figure 1 provides core information of how data was collected. Briefly, two systematic searches were conducted (Supplementary note 1) on PubMed and EMBASE or WebOfScience, for behavior and neurobiology separately. We selected publications based on pre-specified inclusion criteria (for full list of criteria, see Supplementary Table 1), identical for all outcomes. Data were extracted for each outcome separately, and processed into a standardized database. The database contains the summary statistics of each comparison between an ELA and a control group, as well as ~100 other variables that categorize different aspects of the experimental design – e.g. details about the population (e.g. species, age); about the ELA model (e.g. timing and duration of separation); about handling of the animals (e.g. cross-fostering, culling, habituation before experiments); about other life experiences (e.g. other chronic stress, origin of the experimental animals); as well as the state of the animals at death (not applicable for behavior). Of note, the animals of the control and ELA group always only differed in the experience (or not) of ELA. If other life events were experienced, e.g. chronic stress during adolescence, these were experienced by both groups. Altogether the database involves close to 300 unique publications that appeared between 1996 and 2020. All data is now collected in a standardized and dynamic database (https://osf.io/eptda/), freely available for further exploration and ready for continuous updating when more data becomes available.

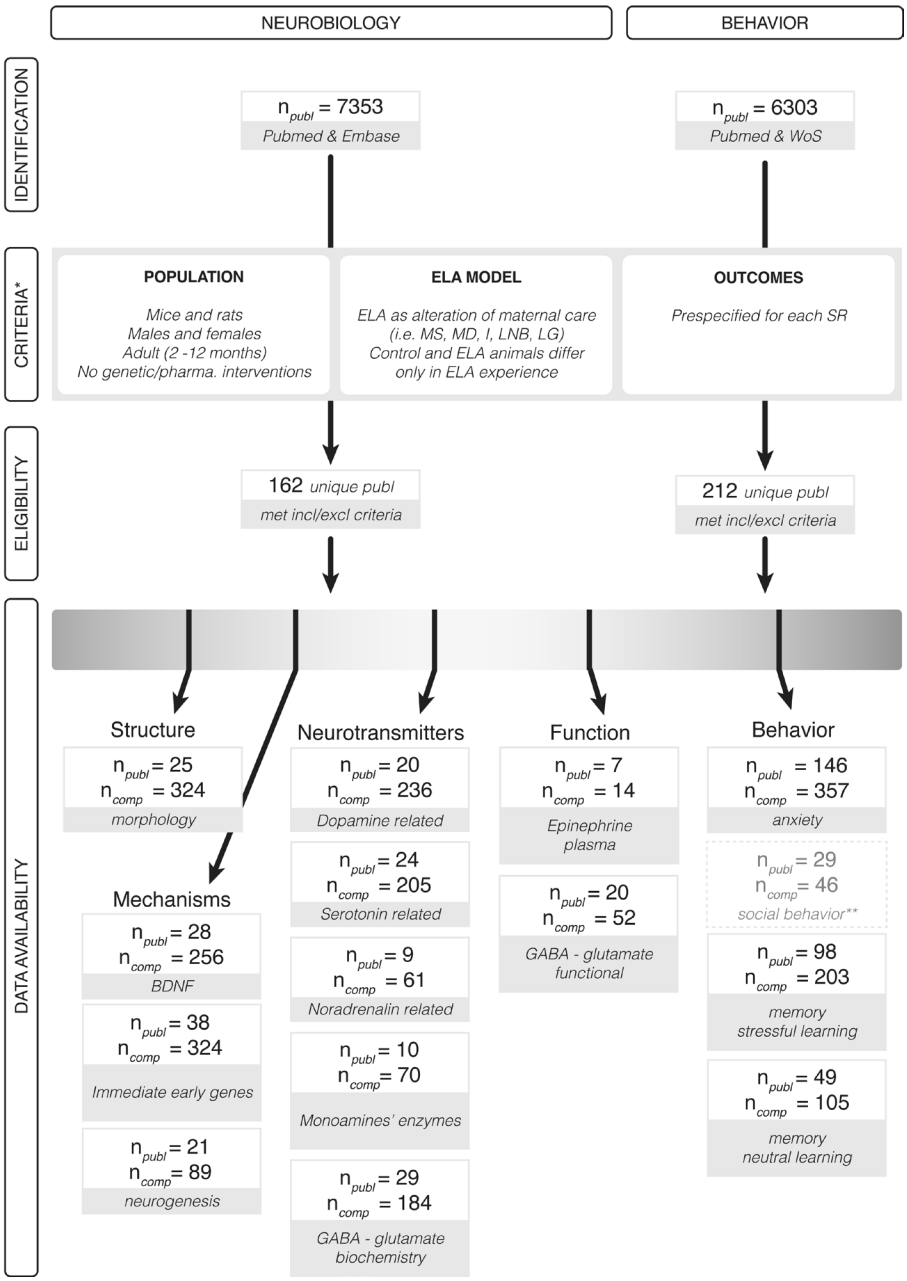


Figure 1. PRISMA overview of available data. Of note, in the two searches we included 162 and 212 publications respectively. However, some of the publications overlapped. The combined dataset includes 298 unique publications. Criteria* = Summary of inclusion criteria. For a full list, see individual publications (Chapters 4-8). Social behavior* = not further discussed in this manuscript, see[30] for more information. WoS = Web of Science (MEDLINE), N_{publ} = number of unique publications, N_{comp} = number of comparisons between a control and an experimental group

The seven principles of rodent Early Life Adversity

Based on our dataset, we identified seven principles of rodent ELA, which are discussed below one by one and summarized at the end in box 1.

Principle #1: The population of ELA rodent studies is fairly homogeneous

Overall, we included 298 studies published between 1996 and 2020. These publications reported 678 experiments (i.e. performed on different sets of animals) and 1949 comparisons between control and ELA groups. We estimate that these experiments reported using ~12843 animals, of which ~77% were males. Most experiments used rats (behavior: 75.9% of exp; neurobiology: 81% of exp, Figure 2A), especially Wistars (behavior: 59.3% of exprats; neurobiology: 46.7% of exprats).

Four models were used to induce ELA, namely: maternal separation, isolation, maternal deprivation, and limited nesting / bedding. Alongside these models involving experimental interference with the mother-pup environment, one model used the natural variation in licking and grooming of the pups by the dam as index for early life environment. We here considered low licking and grooming as a "negative" / ELA-like environment[33].

The ELA model most often used was maternal separation (behavior: 50.5% of exp; neurobiology: 61.4%), with a median of 3h of separation time. For experiments using maternal separation, the protocol in all cases started during the first postnatal week, and usually ended during the second postnatal week (behavior: 65.5% of $exp_{mat\ sep}$; neurobiology: 62% of $exp_{mat\ sep}$). In most experiments, the separation occurred during the light phase (behavior: 88.7% of $exp_{mat\ sep}$; neurobiology: 78.8% of $exp_{mat\ sep}$). Only 20 studies separated the pups from the mother at unpredictable hours, while the others used the same time every day. Concerning the early life environment, cross-fostering was used in 16.8% of expneurobiology and litter size was reported in 64% of expneurobiology (range 5-12 animals per litter). This information is not available for the behavior dataset, but likely comparable.

All animals included in the present dataset were tested in adulthood (inclusion criterion; Figure 2C). The age at the time of testing ranged between 8 and 48 weeks. Age was not specified in 7 publications, although these mentioned that the animals were adults.

Moving to the neurobiological studies, the hippocampus was the brain area most investigated (Figure 2D, 41% of comparisons). This was consistent across all outcomes. Figure 2E reports the frequency of studied outcomes across brain areas. This figure can be considered as an evidence gap map used to identify "gaps" in the literature, i.e. important outcome-by-brain area items that are currently understudied, as well as to evaluate the degree of confidence in the meta-analytic results. For example, Figure 2E shows that norepinephrine, an important neurotransmitter for anxiety-related behavior, has been investigated in only 11 experiments.

In all, this descriptive analysis of the available data points towards the most frequent characteristics, i.e. experiments using maternal separation to investigate hippocampal-specific outcomes in male rats. This population matches 32.3% of the behavior data, and 26% of the neurobiological data. Although there are some variations in protocols, we qualitatively assessed that most protocols were thoroughly standardized. As a consequence, this population is fairly homogeneous.

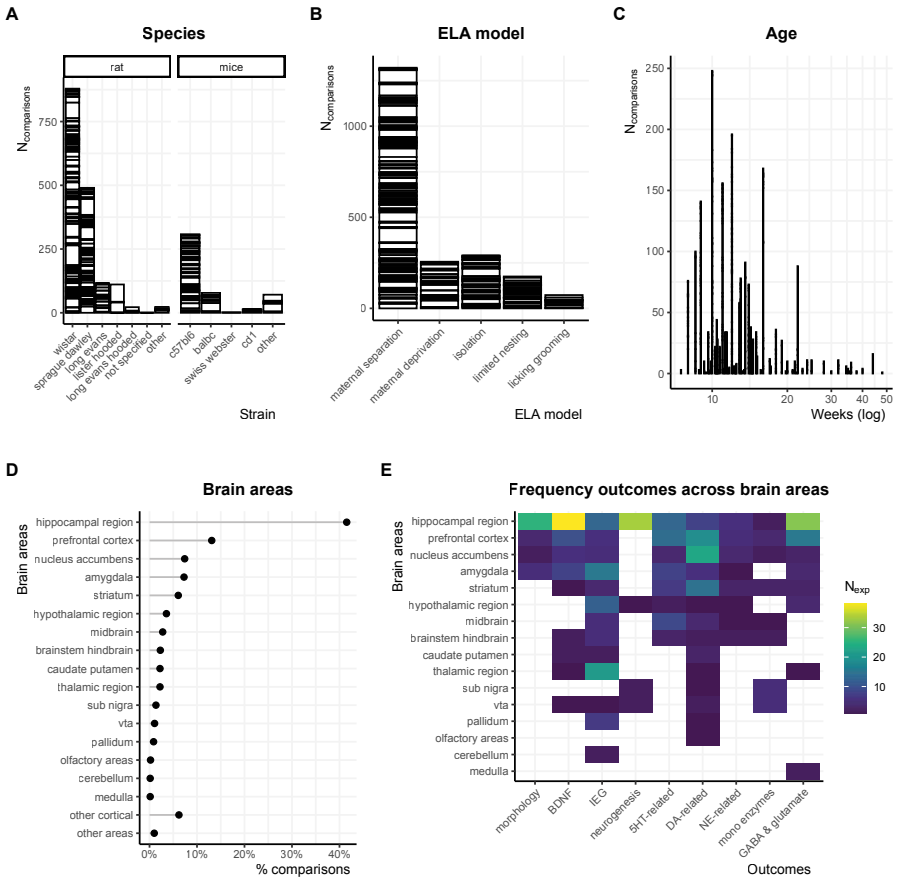


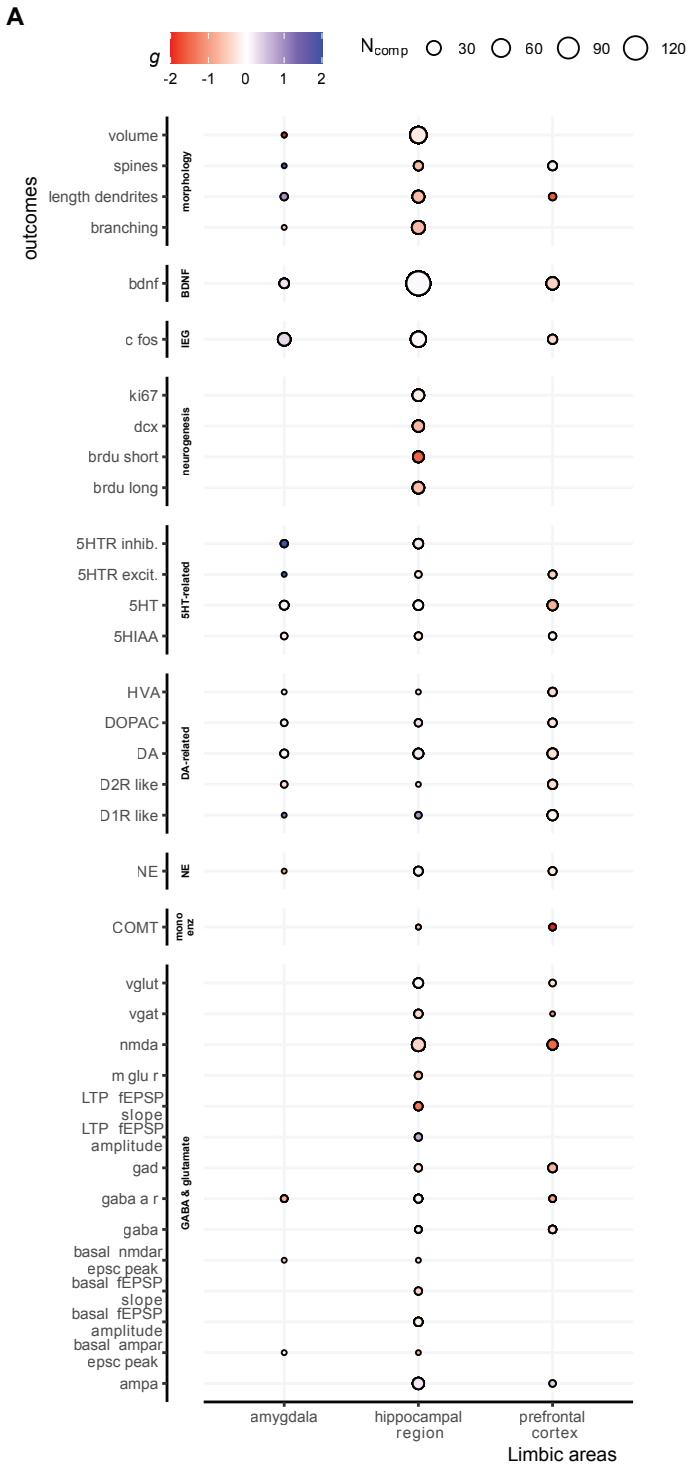
Figure 2. Description of the population. Frequencies of A) species, B) ELA model, C) age, and D) brain areas in our dataset. Individual boxes in the histograms (Fig 2A-C) represent separate publications. E) Evidence gap map. Visual representation of outcomes per brain areas. Color code (legend shown on the right) refers to number of experiments (N_{exp}) that investigated the specific outcome/brain area combination. White spaces suggest "gaps" in the literature, i.e. specific combinations for which no evidence is available, as yet. Glut = glutamate, DA = dopamine, 5HT = serotonin, NE = norepinephrine, mono = monoamines (unspecified), IEG = immediate early genes, sub nigra = substantia nigra, vta = ventral tegmental area.

On the one hand, this is a good life-buoy against the criticism of "comparing apples and pears" in meta-analyses[34]. On the other hand, it raises two issues that deserve further attention. First, a high degree of standardization means a higher probability of committing the "standardization fallacy"[35,36], and to accentuate artefacts[36]. Thus, creating largely homogeneous experiments boosts statistical power, but at the cost of generalizability[18]. Secondly, by investigating only a subset of brain areas with similar ELA models, we miss the opportunity for an integrative view of the effects of ELA on the brain. What are the (holistic) effects across brain areas and neurotransmitter systems? How are these related to critical periods and developmental standpoints? These questions are extremely relevant to improve clinical practice[12], yet they currently remain largely underexplored. This is a serious limitation to our present insights into ELA effects on the rodents' brain.

Principle #2: Acute situation at testing matters

One of the long-standing observations in ELA research is that ELA alters emotion-regulation behaviors, and the brain circuits that support them[1]. In particular, ELA is thought to alter (presumably accelerate[37,38]) the development of the amygdalar-hippocampal-prefrontal circuit, important brain areas of the limbic system. Figure 3A summarizes all meta-analytic results on this circuit for male rodents. Data on females is too sparse to be analyzed in this way, but it is available at https://osf.io/eptda/. The limbic system is responsible for the adult regulation of the activity of the amygdala[37]. In humans it develops in the first 10 years of life, in rodents in the first two postnatal weeks. According to this framework, adult ELA animals are expected to be more anxious, with a bigger amygdala structure, and expanded amygdalar neuronal branches, with increased excitation. Our meta-analytic results (partially) support this view.

Thus, ELA animals show a more anxious phenotype than controls (Figure 3B), as shown in several behavioral tests[30]. This is matched by an increased plasma epinephrine (g[se] = 0.54[0.21], $z = 2.62$, $p = 0.009$), already at baseline (rest/mild arousal) conditions (Figure 3B). Amygdalar excitability and morphology were reported in only a handful of publications, and they are discussed with systematic review in Supplemental Note 2. Another (more common) way to look at cellular activity is measuring the expression of the immediate early gene c-fos[39]. Indeed, c-fos expression is increased after ELA when the animals are at rest / mildly aroused (Figure 3C) throughout the brain[29]. In the amygdala, the effects are further enhanced in animals (both of the control and ELA group) exposed to an acute stressor ($g_{baseline}$[se] = 0.18 [0.19], $z = 1.08$, $p = 0.28$; $g_{acute\ stress}$[se] = 0.64 [0.25], $z = 2.64$, $p = 0.008$, Figure 3B), although the direct comparison between acute situations with a Wald test was not significant ($g_{wald\ test}$[se] = -0.47[0.3], $z = -1.57$, $p = 0.12$, methods available at Supplementary Note 3). Overall, these findings suggest that ELA animals have an "anxiety-prone" phenotype, even at rest or during mild arousal.


A
g
-2
-1
0
1
2
N comp
30
60
90
120
outcomes
volume
spines
length dendrites
branching
morphology
bdnf
BDNF
c fos
IEG
ki67
dcx
brdu short
brdu long
neurogenesis
5HTR inhib.
5HTR excit.
5HT
5HIAA
5HT-related
HVA
DOPAC
DA
D2R like
D1R like
DA-related
NE
NE
COMT
mono ctz
vglut
vgat
nmda
m glu r
LTP fEPSP slope
LTP fEPSP amplitude
gad
gaba a r
gaba
basal nmdar epsc peak
basal fEPSP slope
basal fEPSP amplitude
basal ampar epsc peak
ampa
GABA & glutamate
amygdala
hippocampal region
prefrontal cortex
Limbic areas

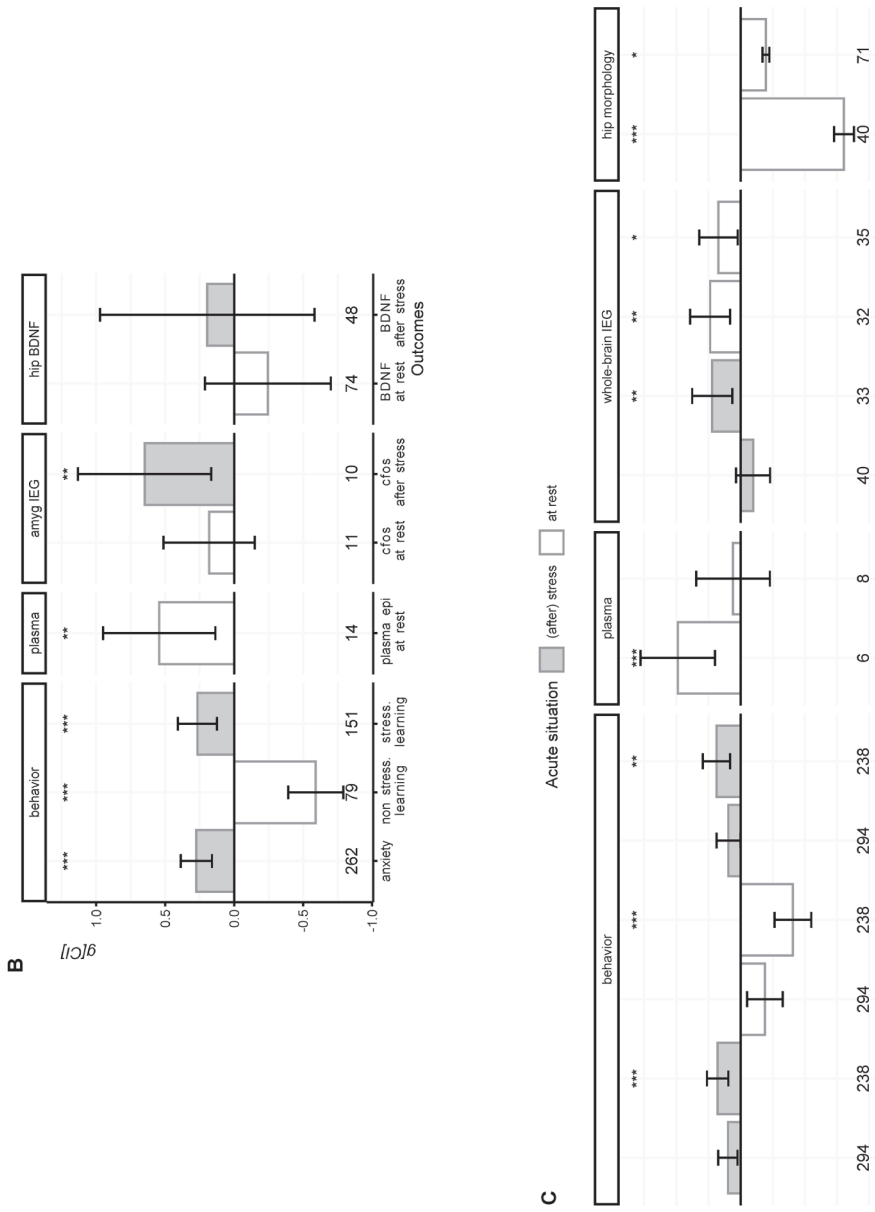


Figure 3. Effects of ELA on behavior and biochemistry. A) Overview of all meta-analytic results in the limbic system, for each domain separately. Dot size represents the amount of available evidence (i.e. the number of comparisons). Blue represents an increase while red a decrease in the ELA group compared to controls, with more intense color indicating a stronger difference in means (g). For visualization purposes, values of g larger than 2, are plotted as 2. 2 indicates an already large effect size[19]. B) ELA animals exhibit an "anxiety prone" phenotype. Focus on the difference between acute rest/mild arousal and stress circumstances. C) Moderating effects of multiple hits (referred to in the figure as '+hits'). excit. = excitatory; hip = hippocampus; inhib. = inhibitory; stress. = stressful; morph = morphology; r. = receptor

Whether this phenotype is appropriate or not for later life circumstances depends on the specific (test) situation in adulthood. For example, ELA animals perform better (i.e. have increased memory) in tasks involving stressful learning such as contextual fear learning, but worse (i.e. have reduced memory) in tasks involving non-stressful learning e.g. object-in-location learning (Figure 3B,[30]). Of note, these tasks critically depend on hippocampal (and prefrontal) functioning. These behavioral results support the hypothesis that ELA alters the development of the limbic system, and points towards alterations at the biochemical and functional level that may be adaptive depending on the acute situation an animal is facing.

To test this hypothesis, we summarize hippocampal and prefrontal outcomes with sufficient evidence (i.e. investigated by at least 3 independent publications, Supplementary Note 4) that could provide a mechanistic overview of the mediating effect of the acute situation (baseline vs stress) on ELA-induced changes. BDNF was the only outcome available both at baseline and in stress circumstances. Although the effects of ELA on BDNF were limited[31], the point estimates of hippocampal BDNF were in opposite directions depending on the acute stress situation (Figure 3B, hippocampus: $g_{stress\text{-}baseline}$[se] = -0.52 [0.13], z = -3.897, p < 0.001). The prefrontal cortex had a similar pattern, but to a much smaller (and non-significant) extent (prefrontal cortex: $g_{stress\text{-}baseline}$ [se] = 0.085 [0.311], z = 0.273, p = 0.785). Of the remaining comparisons that we meta-analyzed (Supplementary Table 2 for a summary of all results), all 24 were measured at rest. Of these, only outcomes related with GABA/glutamate were altered by the effects of ELA; none of the outcomes related to monoamines showed a significant alteration by ELA. At rest, the expression levels of NMDA were decreased in the hippocampus ($g_{hipp\ NMDA}$[se] = -1.46 [0.35], z = -4.107, p < 0.001) as well as in the prefrontal cortex ($g_{pfc\ NMDA}$[se] = -1.666 [0.49], z = -3.406, p < 0.001). Furthermore, LTP — measured as slope of the fEPSP — (g_{LTP}[se] = -1.46 [0.65], z = -2.25, p < 0.024) had a p-value < 0.05 (but not meeting our alpha = 0.01 for false positive rate correction).

All in all, there is a discrepancy in experimental design between functional and biochemical outcomes. At the functional level, the effects of ELA are clearly dependent on the acute stress situation. At the biochemical level, most outcomes are measured at baseline; yet, the positive interaction and the moderating effects of acute situation on BDNF expression brings hope to a potential parallel between biochemistry and function. We conclude that there may be a neurobiological footprint of situation-dependent effects of ELA on behavior, but this clearly requires more in-depth investigations.

Principle #3: Additional negative life experiences ("hits") can enhance the phenotype, but this depends on the outcomes as well as ceiling/floor effects

When investigating the effects of ELA on behavior[30], we observed that additional negative life events ("hits") synergically enhance the effects of ELA on all investigated

behaviors. Examples of additional negative life events were transportation stress in pregnant dams, chronic (unpredictable) stress in adolescence or adulthood, foot-shocks, restraint stress, etc (Supplementary Table 3). These results are in line with the cumulative[40] and three-hits[41] theories of ELA. It is plausible that this remarkable behavioral effect has a neurobiological underpinning, which is likely to involve and integrate multiple systems. In this section, we review the moderator effects of multiple hits on the neuro-substrate. We compare ELA and control animals in two conditions: either both groups experienced additional hits, or they did not. In other words, the difference between ELA and controls was always and only the experience of ELA. The identified effects of "hits" are therefore additive/synergic to those of ELA, rather than main effects.

Anxiety-like behavior was more pronounced in ELA animals compared to controls if both groups experienced additional negative life events (Figure 3C [30]). We therefore reasoned that c-fos expression and plasma epinephrine would also be more pronounced with additional hits. But this was not the case. Throughout the brain, c-fos expression was not moderated by additional hits at rest (Figure 3C [29]). However, after acute stress, the experience of additional hits did enhance the differences between ELA and controls on c-fos expression (Figure 3C [29]). Plasma epinephrine data is available only for the "rest" situation. Here, the effects of ELA were more pronounced when neither group experienced other hits (g_{wald}[se] = -1.1 [0.442], z = -2.646, $p < 0.008$, Figure 3C). This data suggests that the behavioral effects of multiple hits are matched by some of the neurobiological substrates contributing to behavioral outcome, but these likely depend on the acute situation when the animal is tested.

To complete the limbic overview, we re-analyzed the available data on other hippocampal and prefrontal cortical neurobiological outcomes, to investigate the moderating effect of additional negative life experiences. Structurally in the hippocampus, the ELA phenotype was more pronounced when neither control nor ELA groups experienced other hits (Figure 3C [31]). This could also be due to a floor effect, meaning that there may be a limit to the decrease in size of a brain area or the decrease in complexity of a neuron. For neurogenesis[31], BDNF[31] and monoamines-related[32] outcomes, additional hits by themselves were not a significant moderator. It may be that for these outcomes, the interaction between hits and the acute situation is important. However, these complex interactions are yet to be studied. Additional hits turned out to be a significant moderator only on the effect of ELA on prefrontal NMDA level ($Q_M(1) = 24.13$, $p < 0.001$, Figure 3C). This effect is however based on only 3 publications.

Altogether, the influence of multiple life events on ELA-mediated effects depends on the type of outcome. For structural outcomes there may be a floor effect, i.e. structure is not changed beyond the effect of ELA itself. For monoaminergic outcomes and plasma epinephrine, there may be intricate interactions with the acute situation

of testing that are currently not testable. GABA/glutamargic outcomes and c-fos instead may follow the same pattern as behavior. Since not all datasets are sufficiently available for (re)testing, this fourth principle has a tentative character.

Principle #4: ELA amplifies pre-existing differences

ELA does not impact every individual equally[42,43]. Some individuals may be more susceptible than others[44]. For example, the genetic make-up can mediate the effects of ELA on behavior, involving e.g. the dopaminergic[45], endocannabinoid[46], serotonergic[47] and stress systems[48–50]. Pre-existing differences – such as genetic background – can therefore provide the harmonies through which ELA can set the tone of future coping.

The hypothesis that ELA pushes individuals towards more vulnerable or resilient phenotypes can be tested meta-analytically at the population level (Figure 4A). Specifically, we can compare the variation (rather than the mean, as in Principles #2 and #3) of the ELA group to that of the control group. This can be achieved by meta-analyzing the coefficient of variation ratio (CVR, see Supplementary Note 5), which is a measure of difference in variability between the ELA and the control group[51]. Of note, in our data we cannot disentangle if ELA specifically enhances the vulnerable subgroup, the resilient subgroup or both.

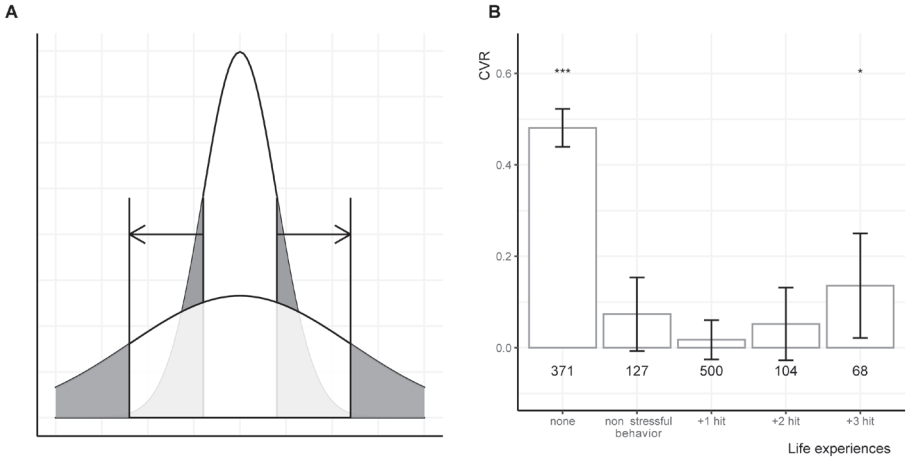


Figure 4. Meta-analysis of variation to evaluate the effects of ELA on the vulnerability/resilience of the population. A) Cartoon image to visualize why ELA could result in changes in variation. The tallest distribution represents a "control" population. The extremes (vulnerable and resilient) of this population are colored in grey. We hypothesize that ELA increases variation, meaning that either one or both arrows are pushed to the extremes. CVR effectively measures the (hypothetical) difference in variation between the shorter (ELA) vs taller (control) normal distributions in the cartoon figure. B) CVR is increased in ELA compared to control animals, especially when both groups did not undergo any other event during their life; i.e., when, besides the early life (ELA or control) experience, they are naïve. *** = $p < 0.001$, * = $p < 0.05$, CVR = Coefficient of Variation Ratio.

Across all studies, the variability of the ELA groups was significantly larger than the control groups' (CVR_{males}(se) = 0.13(0.05), z = 2.83, p = 0.005). This difference was mostly driven by animals that did not experience any other life event besides ELA (or standard rearing conditions), meaning that the animals were otherwise naïve (Figure 4B). On the one hand, this is surprising. We expected ELA differences in CVR (and hence potentially in susceptibility and/or resilience) to be maintained or even exacerbated when facing other life events. On the other hand, these results on CVR can be explained with the allostatic stress theory[52]: if high levels of stress are experienced throughout life, all individuals will eventually reach their maximum allostatic load[53], and ELA animals will no longer have an increased variability compared to those reared under standard conditions and experiencing multiple life events. Given the significant difference between the 'naïve' versus other groups, one could even conclude that ELA may inoculate organisms for later life events[54].

In summary, our results support that ELA can push individuals towards an extreme phenotype compared to the control population, presumably enhancing pre-existing differences in vulnerability and resilience. This underpins, at the population level, the early notion that ELA may promote the extremes of behavior, at the cost of the median[55].

Principles #5: Results are more consistent when closer to structure and function

Variation between individual studies (rather than between experimental groups as in Principle #4) is virtually always present[56]. The variation can have different origins: 1) biologically relevant, due to differences between studies in population, interventions and outcomes; 2) methodological, due to differences in study design or biases; 3) statistical, which should be due to chance alone[57]. In meta-analysis settings, statistical variation is referred to as heterogeneity. A high unexplained heterogeneity means that there are other sources of variation in the data rather than chance (sampling error) alone. This could be due – for example – by yet unexplored biological or methodological factors, or because homogeneity assumptions cannot be met. Meta-analyses of preclinical studies often suffer of a high unexplained heterogeneity[21]. This can be expressed as I^2, a value between 0 and 100%. I^2 describes the percentage of variation between studies due to other factors rather than chance[58]. Of note, the measure I^2 is independent of the number of studies included in the meta-analysis.

In our meta-analyses, we report overall a moderate unexplained heterogeneity across studies (I^2: median [IQR]) = 67.33 [22.975]; mean [SD] = 61.9 [21.64]). The exact values differed largely across outcomes (Table 1), although the population and interventions were highly comparable. As cutoff values for the interpretation of I^2, we use the upper boundary of the range suggested by the Cochrane handbook[57]. The upper boundary was chosen because differences between studies are expected to be larger in preclinical rather than clinical research, on which the Cochrane cutoff is based. I^2

values are interpreted as: 0-40% trivial heterogeneity, 41-60% moderate heterogeneity, 60-90% substantial heterogeneity, 90-100% considerable heterogeneity. Here, we investigated heterogeneity across ELA outcomes, and we discuss the potential origin of its difference. Since the population and interventions in our data are comparable across outcomes, the unexplained heterogeneity likely originates at the level of the measurement.

A statistical explanation is that the deviation of each individual study is so small, that the deviations across studies are not overlapping and the resulting heterogeneity is high. This would occur if there were excessive standardization (see Principle #1). A plausible biological explanation is linked to the biological-time of each outcome, meaning the time (in seconds, minutes, hours) that is required for a biological construct to change. For example, morphology is expected to take a longer time than concentration of neurotransmitters to change from an intervention. Therefore, morphology may be a more stable measurement because it requires a longer time to change. Conversely, electrophysiological measurements may be more precise than concentrations because they are not snapshots, and they are measured in time. Conversely, biochemical measurement could be the "adaptation" strategy of the system: continuously varying to maintain balance and homeostasis. The situation of the animals at measurement may therefore be particularly important. As a consequence, measurement differences may be due not only to measurement error, but to intrinsic dynamic changes of the system.

The meta-analyses on morphology, c-fos expression and electrophysiological outcomes of GABA and glutamate had a trivial-to-moderate heterogeneity (Table 1). This was mainly due to sampling variance, and we deemed it negligible. In other words, we are confident of the consistency of the literature, and we do not expect major additional unknown factors to moderate the effects of ELA on these outcomes. Conversely, BDNF, monoamines, neurogenesis and biochemical indicators of GABA and glutamate instead had a substantial amount of unexplained heterogeneity. For BDNF, the variation was both within and between experiments, e.g. due to the use of different isoforms as experimental outcome. For monoamines, neurogenesis and biochemical indicators of GABA and glutamate, the unexplained I^2 was predominantly due to differences in the experimental design. Of note, we conducted analyses for each outcome separately, choosing the best moderators of the effects for each outcome given the availability of the data. The data here reported is that of the main analyses. However, for each study we thoroughly explored heterogeneity with pre-defined moderators of interest. These analyses are described in more detail elsewhere (Chapter 4 and 6-8 [29–32]). In all, we observe the lowest heterogeneity for outcomes that are the closest to the determination of structure and function.

Table 1. I^2 is elevated specifically for biochemical indicators, rather than morphology and electrophysiology. In other words, the results are more consistent for outcomes that are closest to the determination of structure and function. *Ephys* = electrophysiology, *biochem* = biochemical indicators.

Domain	Unexplained heterogeneity (I^2)	% I^2 Sampling variance	% I^2 Within experiments	% I^2 Between experiments
morphology	28.0	72.95	0.00	27.1
c-fos	44.8	55.01	0.00	45.0
GABA and glutamate ephys	54.2	44.71	16.10	39.2
BDNF	67.3	31.55	40.85	27.6
monoamines	67.6	32.12	28.58	39.3
behavior	72.87	27.13	33.07	39.8
neurogenesis	76.0	23.70	0.98	75.3
GABA and glutamate biochem	94.0	5.43	23.98	70.6

Principles #6: An integrative theory of the effects of ELA on brain and behavior

In the previous principles, we provided an overview of the effects of ELA on brain and behavior, we explored potential moderators, and we interpreted the results towards an integrative view of the effects of ELA. Here, we critically assess previously published theories and evaluate their support based on the current data.

Across the numerous ELA theories (for an excellent overview, see [59]), the effects of ELA on adult behavior are generally explained by two seemingly opposing views[40]: the "cumulative stress" or two-hit (or three-hit [41]) hypothesis, and the "mis-match" hypothesis. In the first, stress exposure during life is seen as "cumulative", meaning that it leads to the build-up of allostatic load[53], thereby increasing the chances of developing a disease [40]. In other words, the more stress, the worse the outcome[60]. On the other hand, the mismatch hypothesis states that ELA triggers an adaptive process, and prepares individuals for a hostile environment later in life[61]. Nederhof & Schmidt previously suggested that these two views could be integrated by considering "programming sensitivity". Specifically, they proposed that the cumulative vs mismatch theories apply to different individuals. The cumulative theory is expected to apply to individuals only marginally programmed by their early environment, while the match-mismatch applies to individuals who experienced strong programming effects [40].

We propose an alternative explanation which incorporate both the cumulative and mis-match theories. Rather than assuming that ELA has biologically-distinct actions in subgroups of individuals, we propose that these theories inadvertently refer to different aspects of the environment. Experiences in line with the cumulative theory are major life events, that have a long-lasting impact on an individual. Conversely, the

mis-match theory describes state- and context- dependent events that are short lived, e.g. whether an individual is under aroused or stressed circumstances when tested. Our data supports that the effects of ELA on behavior as well as several biochemical outcomes are state dependent (Principle #2), and there may be yet unexplored interactions between multiple hits and state (Principle #3). Programming sensitivity can also be incorporated, since the meta-analysis of variation of Principle #4 supports that ELA enhances pre-existing differences. In this view, programming sensitivity could be related to the "genetic predisposition", as described in the three-hit model of vulnerability and resilience[41].

Although we here provide a verbal description of the framework, in future meta-analyses results could be incorporated in statistical models with expert opinion. Currently, the limiting step in this approach is the lack of information about the interaction between the different outcomes of the system. How are the dopamine and serotonin systems related? How does morphology translate into electrophysiology? These questions require a multi-dimensional quantification, starting with an improved reporting of dependency between outcomes. The next step would involve experiments designed based on existing data – and theory – driven frameworks, to enhance our integrative understanding of the effects of ELA on brain and behavior.

Principle #7: Experiments with a fundamental vs translational focus should be planned differently

Every experiment is accompanied by a myriad of choices that may have an impact on the results. Which species to use? Which model? Meta-analyses can help with these choices, as they can be used to investigate the efficacy and stability of elements of the experimental design[62]. In other words, they can aid to identify which species/model/etc maximize the strength of the effects (i.e., efficacy) and are similar among individuals (i.e., stability, as it minimizes interindividual variability). In a meta-analysis setting, these two elements of efficacy and stability can be respectively measured with Hedge's G (g) and CVR. g is a standardized measure of the difference in means; while CVR is a measure of difference in variation. Of note, the variability measure CVR is not the same as the heterogeneity (I^2) described in Principle #5. While CVR is a measure of inter-individual variability across groups, I^2 is a measure of discrepancy between studies. Together, the metrics g and CVR can aid to identify which effects have the potential to be generalizable, i.e. have a high efficacy (large g) and a high stability (low CVR)[62].

Throughout our neurobiological dataset, we measured g and CVR for the same comparisons (Figure 5). For g, we here use the absolute values of the effects to maintain comparability across outcomes. This is likely to lead to an overestimation of the moderator effects in subgroup analyses, but this seems justified since the aim of this analysis was not to estimate effect sizes, but rather to establish a relationship between g and CVR. We focused on the following elements of the experimental design: species, sex of the animals and ELA model.

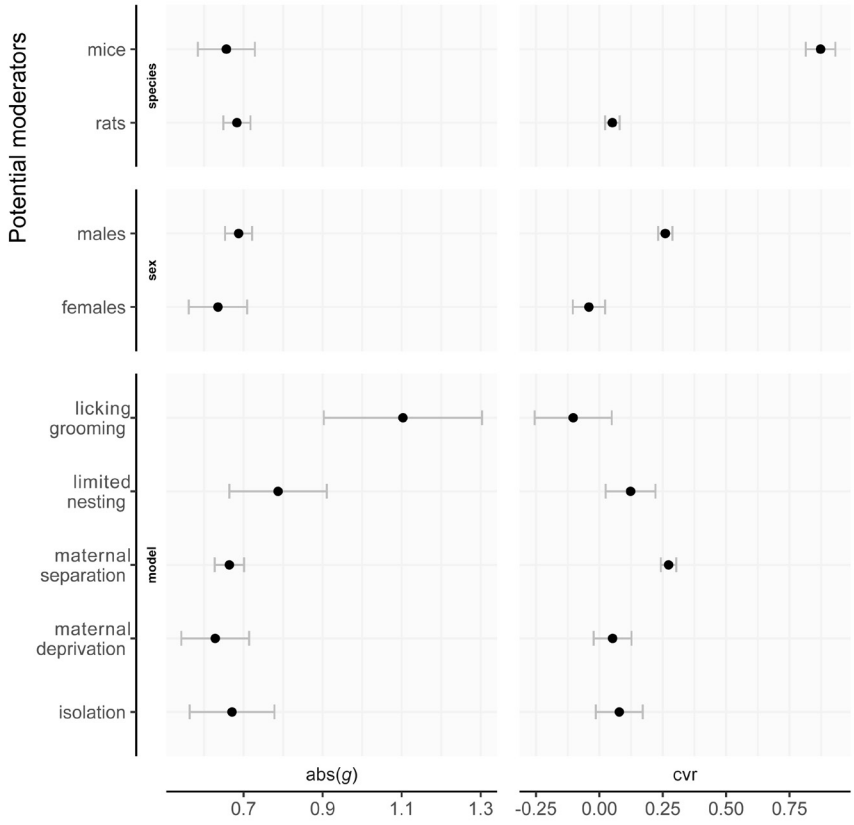


Figure 5. Comparison of g and CVR across multiple potential moderators. These analyses have been conducted on overall measures. The subgroups (e.g. mice and rats for species) do not come from the same publications where "species" was a randomized factor. As a consequence, we cannot exclude study-specific effects since our conclusions depend on the data systematically gathered.

Rats and mice were overall not different in g (g [se] = 0.027 [0.04], z = 0.654, p = 0.513, Figure 5, for methods see Supplementary note 3), meaning that ELA does not consistently give larger differences in means compared to controls in either species. However, mice had a much higher CVR than rats (CVR [se] = -0.65 [0.03], z = -19.56, p < 0.001, Figure 5), suggesting that the effects of ELA across individual mice are more variable than in rats. A biological explanation could be that mice are more state dependent than rats, although the current dataset does not indicate that. A methodological explanation could be that mice are bred in-house more often than rats (in-house breeding: 82% in mice vs 61.2% in rats), which could have consequences for standardization and generalizability[63], and therefore CVR.

Similarly, neither males nor females are systemically more sensitive to the ELA models (g [se] = 0.052 [0.04], z = 1.256, p = 0.209, Figure 5), although outcome-specific sex-differences are plausible[32,42,64,65]. This result was contrary to our expectations[30].

Moreover, males had a much higher CVR than females (CVR [se] = 0.18 [0.3], z = 5.034, p < 0.001, Figure 5), suggesting that the effects of ELA are more stable in females than in males. This seems counter-intuitive, given the expected variation in females due to their menstrual cycle. Yet, the results are in line with several other meta-analyses of variation which showed that males are equally or more variable than females in a variety of interventions (unrelated to ELA)[62,66,67]. At this stage it is unclear, though, whether or not a sex-specific bias may exist, in view of the relatively low number of laboratories investigating female rodents; for instance, labs including females could have subgroup-specific characteristics in their research practice.

Concerning the model used, the naturalistic model based on observations of "licking and grooming" of pups by the dams had the largest efficacy (g [se] = -0.435 [0.1], z = -4.21, p < 0.001, Figure 5) as well as stability (CVR [se] = -0.192 [0.08], z = -2.452, p = 0.014, Figure 5). This closely reproduces subgroup analyses on behavior[30] previously conducted. In other words, extremes of natural variation in maternal care produce stronger and more stable effects that models in which ELA has been induced. Again, some caution is called for, since this model has been successfully applied in only a limited number of laboratories across the world.

Overall, one could argue that – at a meta-analytic level – the most consistent studies were those performed in (female) rats, where licking and grooming was used as ELA model. Following the interpretation of g and CVR by Usui and colleagues, these elements are the ones most likely to be generalizable. However, we here performed overall analyses, with the intent to define trends in the available body of literature. These metrics should be followed-up for any subgroup of interest.

The question remains whether we should strive to standardize these elements as much as possible. We argue that the design of experiments to discover fundamental biological properties should be different than those with a translational aim. Thus, preclinical ELA research has an inherent dual purpose: 1) fundamental, since it improves our basic understanding of brain development, and 2) translational, because it aims to model aspects of human ELA relevant for several disease states. Fundamental research is generally exploratory. Researchers may therefore opt to maximize effect size as much as possible, for example by attempting to minimize variation with standardization, as previously extensively endorsed [68,69] even in the ELA field[8]. By reducing within-study variability, these studies require smaller sample sizes, and are consequently more feasible. This, however, will come at the expense of external validity (see Principle #1), due to the "standardization fallacy"[36]. Conversely, translational research can be exploratory as well as confirmatory, and it generally aims to model aspects of human ELA. Translational research should value external validity over standardization, and thus use a more generalizable population. This can be chosen by investigating g and CVR as in Figure 6, in this case pointing to the use of out-bred strains[63] or systemic heterogenization [62,70,71]. For example, in a fundamental setting, ELA can be applied during

a specific time in development, to investigate e.g. critical periods[7,12,72,73]. In a translational setting, ELA could be investigated with "dimensional" models, which mirror the variety of human ELA. Specifically, a good model should account for equifinality (i.e., different ELA but equal outcome) and multifinality (i.e., equal ELA but different outcome) of ELA[7]. This would improve translational validity, but would hinder the understanding of specific mechanisms[8].

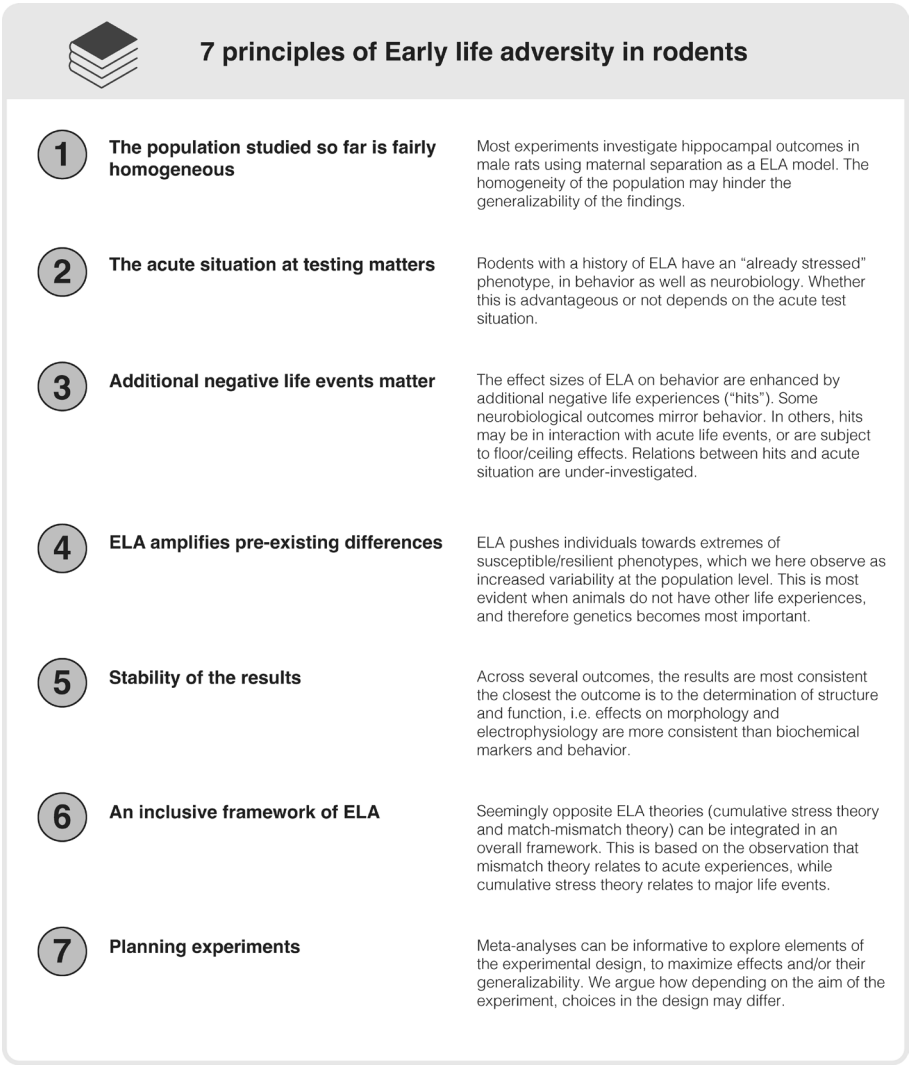
In summary, combined analysis of g and CVR can be used to evaluate robust elements of the experimental design. These can be investigated on a case-to-case basis in our web-portal. Fundamental and translational scientists may opt to use this knowledge differently, depending on the aim of the planned experiment.

Concluding remarks

This review gives a comprehensive overview of rodent studies identified with systematic searches, which appeared between 1996 and 2019 and that describe the effects of early postnatal variation in maternal care – be it natural or caused by interventions – on adult brain and behavior. There are several limitations to consider. First, by restricting the overview to alterations of maternal care in the first postnatal weeks, inevitably we focused on lasting changes in parts of the brain that undergo critical developmental changes in this period, including the amygdala and hippocampus and less so the prefrontal cortex [74]. This may have favored the "emotionality" aspect of ELA, rather than catching the full "dimensional" spectrum. In future, the survey could be extended by studies describing prenatal early life adversity or adverse conditions experienced later in life, e.g. during puberty. Second, only studies describing the effects of ELA on morphology, neurotransmitters / neurotransmission and behavior between 6 weeks of age and 1 year were selected. Excellent single studies indicate that behavioral changes linked to maternal care are already discernable at a much younger age, linked to premature development of the olfactory-amygdalar circuit[11]. Conversely, unique attempts to study behavioral and neuroendocrine properties in very aged animals that underwent 24 h maternal deprivation at postnatal day 3 indicated that favoring the extreme vulnerable and resilient phenotypes is particularly evident in animals >2 years of age[55], in line with Principle #1. These two examples emphasize the need for a more extensive database. Third, based on predefined selection criteria, the influence of genetic background was studied only to a limited extent, i.e. in relation to natural variations; studies using genetic modification were not included in the current database. Finally, the dataset did not address the question to what extent interventions are possible that may prevent, reverse or normalize the effects of ELA (see e.g.[75–77]; and the critical windows in time when interventions are successful.

Despite these limitations, the dataset already allowed to extract seven principles about ELA in rodents, summarized in box 1. Gathering this data required considerable resources and years in the making, but it is now collected in a standardized and

dynamic database (https://osf.io/eptda/), freely available for exploration. We argue that the ELA community should collectively continue along this line, in a joint effort to conceptualize how ELA shapes behavior and neurobiology, through the adaptive value of the individual and of the population. The current overview marks the beginning of a community effort: a step forward towards more robust and generalizable research on ELA, and a more solid base to guide future studies in humans.



Box 1. Summary of the seven principles of ELA. These principles are based on the interpretation of the results from several systematic reviews and meta-analyses.

References

1. Callaghan, B. L., Sullivan, R. M., Howell, B. & Tottenham, N. The international society for developmental psychobiology Sackler symposium: Early adversity and the maturation of emotion circuits—A cross-species analysis. *Developmental Psychobiology* 56, 1635–1650 (2014).
2. Shonkoff, J. P. et al. The Lifelong Effects of Early Childhood Adversity and Toxic Stress. *Pediatrics* 129, e232–e246 (2012).
3. McLaughlin, K. A. et al. Childhood Adversities and First Onset of Psychiatric Disorders in a National Sample of US Adolescents. *Archives of General Psychiatry* 69, 1151–1160 (2012).
4. Struck, N. et al. Childhood maltreatment and adult mental disorders - the prevalence of different types of maltreatment and associations with age of onset and severity of symptoms. *Psychiatry Res* 293, 113398 (2020).
5. Teicher, M. H., Gordon, J. B. & Nemeroff, C. B. Recognizing the importance of childhood maltreatment as a critical factor in psychiatric diagnoses, treatment, research, prevention, and education. *Mol Psychiatry* 1–8 (2021) doi:10.1038/s41380-021-01367-9.
6. Poletti, S. et al. Adverse childhood experiences influence the detrimental effect of bipolar disorder and schizophrenia on cortico-limbic grey matter volumes. *J Affect Disord* 189, 290–297 (2016).
7. Gee, D. G. Early Adversity and Development: Parsing Heterogeneity and Identifying Pathways of Risk and Resilience. *AJP* 178, 998–1013 (2021).
8. Hanson, J. L., Williams, A. V., Bangasser, D. A. & Peña, C. J. Impact of Early Life Stress on Reward Circuit Function and Regulation. *Frontiers in Psychiatry* 12, 1799 (2021).
9. Tost, H., Champagne, F. A. & Meyer-Lindenberg, A. Environmental influence in the brain, human welfare and mental health. *Nat Neurosci* 18, 1421–1431 (2015).
10. Meaney, M. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual review of neuroscience* (2001) doi:10.1146/ANNUREV.NEURO.24.1.1161.
11. Sullivan, R. M., Landers, M., Yeaman, B. & Wilson, D. A. Good memories of bad events in infancy. *Nature* 407, 38–39 (2000).
12. Curley, J. P. & Champagne, F. A. Influence of maternal care on the developing brain: Mechanisms, temporal dynamics and sensitive periods. *Frontiers in Neuroendocrinology* 40, 52–66 (2016).
13. Levis, S. C., Baram, T. Z. & Mahler, S. V. Neurodevelopmental origins of substance use disorders: Evidence from animal models of early-life adversity and addiction. *European Journal of Neuroscience* 1, (2021).
14. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr Opin Psychol* 15, 174–181 (2017).
15. Clark, D. A., Mitra, P. P. & Wang, S. S.-H. Scalable architecture in mammalian brains. Nature 411, 189–193 (2001).
16. Hyde, L. W. et al. An ecological approach to understanding the developing brain: Examples linking poverty, parenting, neighborhoods, and the brain. *Am Psychol* 75, 1245–1259 (2020).
17. van der Veen, R., Bonapersona, V. & Joëls, M. The relevance of a rodent cohort in the Consortium on Individual Development. *Developmental Cognitive Neuroscience* 45, 100846 (2020).
18. Macleod, M. R. Design animal studies better. *Nature* 510, 35–35 (2014).
19. Bonapersona, V., Hoijtink, H., RELACS Consortium, Sarabdjitsingh, R. A. & Joëls, M. Increasing the statistical power of animal experiments with historical control data. *Nat Neurosci* 24, 470–477 (2021).
20. Landis, S. C. et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490, 187–191 (2012).
21. Hooijmans, C. R. et al. Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. *PLOS ONE* 13, e0187271 (2018).
22. Macleod, M. R. et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLOS Biology* 13, e1002273 (2015).
23. Perrin, S. Preclinical research: Make mouse studies work. *Nature* 507, 423–425 (2014).
24. Worp, H. B. van der et al. Can Animal Models of Disease Reliably Inform Human Studies? *PLOS Medicine* 7, e1000245 (2010).
25. Pound, P., Ebrahim, S., Sandercock, P., Bracken, M. B. & Roberts, I. Where is the evidence that animal research benefits humans? *BMJ* 328, 514–517 (2004).
26. Bastian, H., Glasziou, P. & Chalmers, I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 7, e1000326 (2010).
27. Miake-Lye, I. M., Hempel, S., Shanman, R. & Shekelle, P. G. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Reviews* 5, 28 (2016).
28. Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R. & Howells, D. W. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. *J Cereb Blood Flow Metab* 34, 737–742 (2014).
29. Schuler, H., Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. Effects of early life adversity on immediate early gene expression: Systematic review and 3-level meta-analysis of rodent studies. *PLOS ONE* 17, e0253406 (2022).
30. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neuroscience & Biobehavioral Reviews* 102, 299–307 (2019).
31. Joëls, M., Kraaijenvanger, E., Sarabdjitsingh, R. A. & Bonapersona, V. Structural changes after early life adversity

in rodents: a systematic review with meta-analysis. 2022.01.17.476580 (2022) doi:10.1101/2022.01.17.476580.
32. Bonapersona, V., Joëls, M. & van der Veen, R. Effects of early life adversity on biochemical indicators of brain monoamines: a systematic review and meta-analysis of rodent studies. *In preparation* (2022).
33. Liu, Dong et al. Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science* 277, (1997).
34. Hemilä, H. Random-Effects Assumption in Meta-analyses. *JAMA* 322, 81 (2019).
35. Voelkl, B., Würbel, H., Krzywinski, M. & Altman, N. The standardization fallacy. *Nature Methods* 18, 5–7 (2021).
36. Würbel, H. Behaviour and the standardization fallacy. *Nat Genet* 26, 263–263 (2000).
37. Gee, D. G. et al. Early developmental emergence of human amygdala–prefrontal connectivity after maternal deprivation. *PNAS* 110, 15638–15643 (2013).
38. Moriceau, S. & Sullivan, R. M. Maternal presence serves as a switch between learning fear and attraction in infancy. *Nat Neurosci* 9, 1004–1006 (2006).
39. Kovács, K. J. c-Fos as a transcription factor: a stressful (re)view from a functional map. *Neurochem Int* 33, 287–297 (1998).
40. Nederhof, E. & Schmidt, M. V. Mismatch or cumulative stress: toward an integrated hypothesis of programming effects. *Physiol Behav* 106, 691–700 (2012).
41. Daskalakis, N. P., Bagot, R. C., Parker, K. J., Vinkers, C. H. & de Kloet, E. R. The three-hit concept of vulnerability and resilience: toward understanding adaptation to early-life adversity outcome. *Psychoneuroendocrinology* 38, 1858–1873 (2013).
42. Hodes, G. E. & Epperson, C. N. Sex Differences in Vulnerability and Resilience to Stress Across the Life Span. *Biol Psychiatry* 86, 421–432 (2019).
43. Meyer, U. Neurodevelopmental Resilience and Susceptibility to Maternal Immune Activation. *Trends in Neurosciences* 42, 793–806 (2019).
44. Belsky, J., Bakermans-Kranenburg, M. J. & van IJzendoorn, M. H. For Better and For Worse: Differential Susceptibility to Environmental Influences. *Curr Dir Psychol Sci* 16, 300–304 (2007).
45. Zandstra, A. R. E., Ormel, J., Hoekstra, P. J. & Hartman, C. A. Chronic Stressors and Adolescents' Externalizing Problems: Genetic Moderation by Dopamine Receptor D4. The TRAILS Study. *J Abnorm Child Psychol* 46, 73–82 (2018).
46. Lazary, J. et al. Genetic analyses of the endocannabinoid pathway in association with affective phenotypic variants. *Neuroscience Letters* 744, 135600 (2021).
47. Caspi, A., Hariri, A. R., Holmes, A., Uher, R. & Moffitt, T. E. Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. *Am J Psychiatry* 167, 509–527 (2010).
48. van Rossum, E. F. C. et al. Polymorphisms of the glucocorticoid receptor gene and major depression. *Biol Psychiatry* 59, 681–688 (2006).
49. Vinkers, C. H. et al. Mineralocorticoid receptor haplotypes sex-dependently moderate depression susceptibility following childhood maltreatment. *Psychoneuroendocrinology* 54, 90–102 (2015).
50. Wang, Q., Shelton, R. C. & Dwivedi, Y. Interaction between early-life stress and FKBP5 gene variants in major depressive disorder and post-traumatic stress disorder: A systematic review and meta-analysis. *J Affect Disord* 225, 422–428 (2018).
51. Nakagawa, S. et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution* 6, 143–152 (2015).
52. McEwen, B. S. & Stellar, E. Stress and the individual. Mechanisms leading to disease. *Arch Intern Med* 153, 2093–2101 (1993).
53. McEwen, B. S. Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences* 840, 33–44 (1998).
54. Ashokan, A., Sivasubramanian, M. & Mitra, R. Seeding Stress Resilience through Inoculation. *Neural Plast* 2016, 4928081 (2016).
55. Oitzl, M. S., Workel, J. O., Fluttert, M., Frösch, F. & De Kloet, E. R. Maternal deprivation affects behaviour from youth to senescence: amplification of individual differences in spatial learning and memory in senescent Brown Norway rats. *Eur J Neurosci* 12, 3771–3780 (2000).
56. Schroll, J. B., Moustgaard, R. & Gøtzsche, P. C. Dealing with substantial heterogeneity in Cochrane reviews. Cross-sectional study. *BMC Medical Research Methodology* 11, 22 (2011).
57. Higgins, Julian PT et al. Cochrane Handbook for Systematic Reviews of Interventions. in *Cochrane Handbook for Systematic Reviews of Interventions* i–xxi (John Wiley & Sons, Ltd, 2008). doi:10.1002/9780470712184.fmatter.
58. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21, 1539–1558 (2002).
59. Harris, B. N. Stress hypothesis overload: 131 hypotheses exploring the role of stress in tradeoffs, transitions, and health. *General and Comparative Endocrinology* 288, 113355 (2020).
60. Taylor, S. E. Mechanisms linking early life stress to adult health outcomes. *PNAS* 107, 8507–8512 (2010).
61. Schmidt, M. V. Animal models for depression and the mismatch hypothesis of disease. Psychoneuroendocrinology 36, 330–338 (2011).
62. Usui, T., Macleod, M. R., McCann, S. K., Senior, A. M. & Nakagawa, S. Meta-analysis of variation suggests that

embracing variability improves both replicability and generalizability in preclinical research. *PLOS Biology* 19, e3001009 (2021).
63. Tuttle, A. H., Philip, V. M., Chesler, E. J. & Mogil, J. S. Comparing phenotypic variation between inbred and outbred mice. *Nat Methods* 15, 994–996 (2018).
64. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. *Nat Neurosci* 18, 1413–1420 (2015).
65. Bangasser, D. A. & Cuarenta, A. Sex differences in anxiety and depression: circuits and mechanisms. *Nat Rev Neurosci* 1–11 (2021) doi:10.1038/s41583-021-00513-0.
66. Becker, J. B., Prendergast, B. J. & Liang, J. W. Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biology of Sex Differences* 7, 34 (2016).
67. Prendergast, B. J., Onishi, K. G. & Zucker, I. Female mice liberated for inclusion in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews* 40, 1–5 (2014).
68. Festing, Michael F. W. Reduction of animal use: experimental design and quality of experiments. *Lab Anim* 28, 212–221 (1994).
69. van Zutphen, L. F., Baumans, V, & Beynen, A.C. *Principles of Laboratory Animal Science*. (Elsevier, 2001).
70. Richter, S. H. Systematic heterogenization for better reproducibility in animal experimentation. *Lab Anim* 46, 343–349 (2017).
71. Voelkl, B. et al. Reproducibility of animal research in light of biological variation. *Nat Rev Neurosci* 21, 384–393 (2020).
72. Cooke, E. M., Connolly, E. J., Boisvert, D. L. & Hayes, B. E. A Systematic Review of the Biological Correlates and Consequences of Childhood Maltreatment and Adverse Childhood Experiences. *Trauma, Violence, & Abuse* 15248380211021612 (2021) doi:10.1177/15248380211021613.
73. Montalvo-Ortiz, J. L., Gelernter, J., Hudziak, J. & Kaufman, J. RDoC and Translational Perspectives on the Genetics of Trauma-Related Psychiatric Disorders. *Am J Med Genet B Neuropsychiatr Genet* 171, 81–91 (2016).
74. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog Neurobiol* 0, 1–16 (2013).
75. Aisa, B., Tordera, R., Lasheras, B., Del Río, J. & Ramírez, M. J. Cognitive impairment associated to HPA axis hyperactivity after maternal separation in rats. *Psychoneuroendocrinology* 32, 256–266 (2007).
76. Cambon, K. et al. A Synthetic Neural Cell Adhesion Molecule Mimetic Peptide Promotes Synaptogenesis, Enhances Presynaptic Function, and Facilitates Memory Consolidation. *J Neurosci* 24, 4197–4204 (2004).
77. Loi, M. et al. Transient Prepubertal Mifepristone Treatment Normalizes Deficits in Contextual Memory and Neuronal Activity of Adult Male Rats Exposed to Maternal Deprivation. *eNeuro* 4, ENEURO.0253-17.2017 (2017).

Parte C

CHAPTER 10

Increasing the statistical power of animal experiments with historical control data

Nat Neurosci (2021), doi: 10.1038/s41593-020-00792-3

Bonapersona V[1]
Hoijtink H[2]
RELACS[3]
Sarabdjitsingh RA[1]
Joëls M[1,4]

1 University Medical Center Utrecht Brain Center, Utrecht University, The Netherlands.
2 Utrecht University, The Netherlands.
3 Rodent Early Life Adversity Consortium on Stress.
4 University Medical Center Groningen, Groningen, The Netherlands.

Abstract

Low statistical power reduces the reliability of animal research; yet, increasing sample sizes to increase statistical power is problematic for both ethical and practical reasons. We present an alternative solution using Bayesian priors based on historical control data, which capitalizes on the observation that control groups in general are expected to be similar to each other. In a simulation study we show that including data from control groups of previous studies could halve the minimum sample size required to reach the canonical 80% power, or increase power when using the same number of animals. We validated the approach in a dataset based on seven independent rodent studies into the cognitive effects of early-life adversity. We present an open-source tool, *RePAIR*, which can be widely used to apply this approach and increase statistical power, thereby improving the reliability of animal experiments.

Introduction

Before embarking on a new animal study, researchers have to decide how many animals per group are needed to optimize the chance of detecting a real effect rather than a chance finding. When performing a statistical power calculation, power is commonly set a priori at 80% (prospective power), i.e., the expectation is that 80 out of 100 studies investigating a real effect will correctly conclude that the effect exists (true positive), while 20 will not (false negative). As power decreases, the rate of false positive as well as false negative results will increase[1]. Prospective study power therefore directly affects the reliability of the subsequent research findings.

However, a landmark paper by Button et al.[1] estimated, based on 48 meta-analyses of neuroscience studies, that the median power in reality is around 21%, in agreement with previous reports in psychology[3]. Although Button's report was based mainly on studies in humans, a similar discrepancy between prospective and actual power likely exists in animal studies. If so, this would contribute substantially to the reproducibility crisis[4] in animal research[5–8], as single, underpowered studies have a low likelihood of detecting a real effect[1], although they can still be informative when included in meta-analyses[9,10].

To improve reproducibility, previous reports have suggested using systemic heterogenization[7,11], multiple batches[12] or prospective multicenter studies[8,13], alongside changes in research practice and education[8]. These suggestions involve substantial logistics issues and resources; for the foreseeable future, it is likely that the majority of animal experiments will remain single-laboratory. In a single-laboratory setting, an obvious solution to enhancing statistical power would be to increase the number of animals per experiment. For example, for a common effect size Hedge's G = 0.5 (Welch independent samples t-test, α = 0.05), 10 animals/group would correspond to a statistical power of 18%, 30 animals/group to 48%, and 65 animals/group to 81%. Clearly, this is not a feasible solution, not only in terms of the space requirements and financial costs but also in light of the continuing efforts to reduce the number of animals used in research.

How can one ensure that a study has sufficient power without increasing the number of animals per group to unrealistically high levels? An appealing approach would be to recycle data from past experiments, as has been done both in human and animal research[14,15]. In research practice, new studies often build on earlier ones, performed in one's own lab or elsewhere. Here, we focus on the specific example of studies using the same experimental endpoint. The data from previous, similar studies can be incorporated within new experiments by using Bayesian priors, i.e. distributions that describe the mean and variance of an experimental outcome of previous studies. This incorporation can occur already when planning an experiment in the power calculation, or exclusively when analyzing the collected data (although this would require preregistration). Transforming information of previous studies in a

mathematical function is not trivial, and it has been suggested to be one of the most difficult aspects of Bayesian analysis[16]. Priors can be developed by incorporating data from multiple sources (e.g. one's own and others' experiments, or expert knowledge) and through various methodologies (for a review see[16]). Bayesian priors are used in the clinical literature and have already been applied to decrease sample sizes in new experiments (for example[17,18]). Yet, they have been adopted in very few animal studies (for example[19], for a review[14]), which – although remarkable – received limited attention. As a consequence, the powerful message of using historical controls in new experiments has not reached yet the end beneficiary: researchers performing animal experiments.

In this study, we first evaluate the extent of the power problem in animal research, by examining a much larger sample of animal studies than reported before[2]. Next, we show how historical data can be used to limit the number of animals used in a study, by tailoring the Bayesian prior approach to animal experiments. We validate the method and provide an example of how this approach can be applied in daily research practice. We then estimate the impact of the approach on the statistical power of future animal experiments. Lastly, we present RePAIR (Reduction by Prior Animal Informed Research), a user interface optimized for easy use, to facilitate implementation of the methodology.

Results

Many animal experiments are severely underpowered

A common approach to estimate the extent of the power problem in animal research is to calculate statistical power from published literature. Through a systematic search (Supplementary Note 1-2), we identified a large sample of animal studies in the areas of 'neuroscience' and 'metabolism' (n_{study} = 1935) that had previously been included in meta-analyses (n_{ma} = 69). These animal studies had an overall median statistical power of 18% (Figure 1a), which was roughly equal in the two fields (neuroscience: 15%, metabolism: 22%).

Although this approach closely replicated the results of previous reports[2,3], it has major limitations[20]. An alternative approach is to estimate a reasonable prospective power to describe a plausible scenario for new experiments. Since real effect sizes are unknown, we estimated a common range by selecting the medians and quantiles of the distribution identified from published animal studies ($n_{effect\ sizes}$ = 2738). These corresponded to Hedge's G of 0.2, 0.5, 0.9 (Figure 1b), which is almost identical to Cohen's d rule of thumb for small, medium and large effect sizes[21]. Prospective study power was then calculated for this range of effect sizes directly derived from published studies. For large effect sizes, prospective power was sufficient (above 80%) only in 12.5% of studies. This percentage dramatically decreased if smaller effect sizes were considered (Figure 1c).

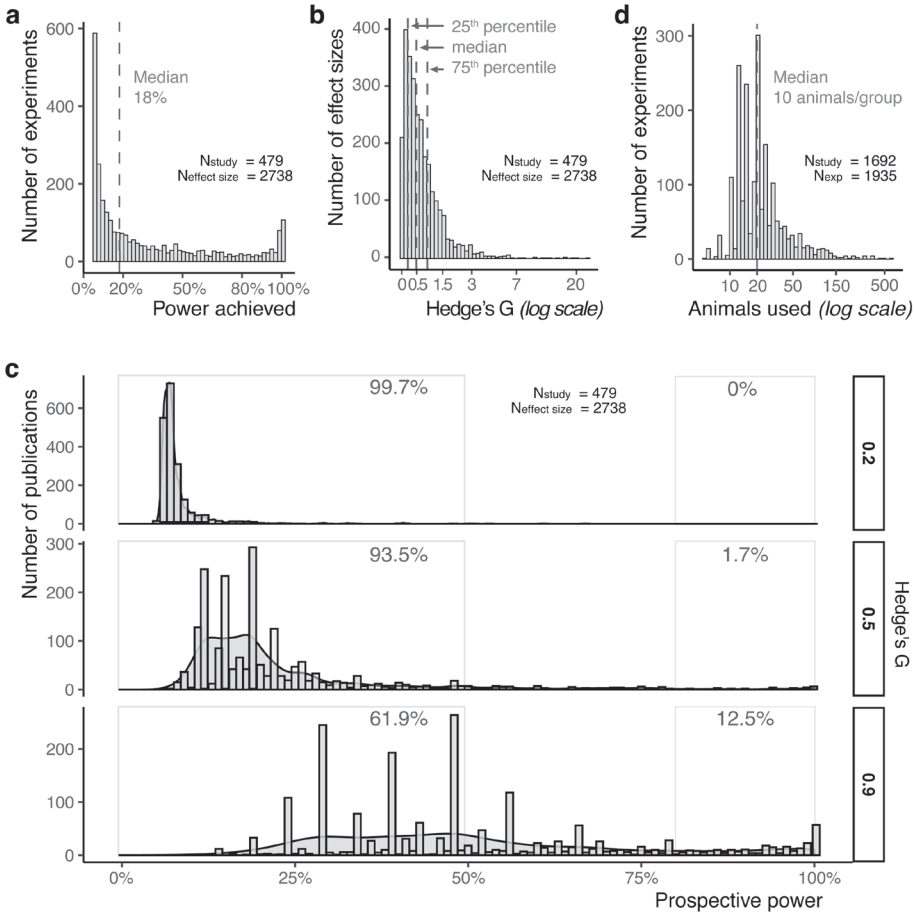


Figure 1. Many animal experiments are severely underpowered. **a)** Power of identified experiments (two-tailed Welch t-test, effect sizes as reported in published papers, 'Data B' in Supplementary Figure 1). Dashed line: median equal to 18%. **b)** Range of common effect sizes in animal literature ('Data B' in Supplementary Figure 1). Dashed lines: percentiles. The related quantities (Hedge's G of 0.2, 0.5, 0.9) were defined as 'small', 'medium' and 'large' effect sizes, respectively. **c)** Prospective power of studies when considering a range of common effect sizes (Figure 1b) and assuming at least one sufficiently powered experiment per publication ('Data B' in Supplementary Figure 1). The highest peaks in the histograms are due to a non-uniform distribution of animals used as shown in Figure 1d. Histogram and density plot of the same data are overlapping. Left box: power ≤50%; right box: power ≥80%. d, Animals per study when considering the two largest independent groups ('Data A'+'Data B' in Supplementary Figure 1). Dashed line: median equal to 20 (~10 animals per group).

Bayesian priors can increase statistical power while limiting sample size

Actual study power is much lower than is commonly assumed (Figure 1c). The most obvious solution would be to increase sample sizes. Currently, a common sample size used is 10 animals per group (Figure 1d). When considering this common sample size and a Welch independent samples t-test (α = 0.05), one would need to assume

an effect size Hedge's G = 1.4 to reach a power of 80%. Such an expected effect size is far larger than what is commonly observed in rodent literature (Figure 1b). If more realistic effect sizes are used, e.g. Hedge's G = 0.2 or 0.9, the required sample size increases to 394 and 21 animals per group, respectively.

An alternative solution is to use data from past experiments in the form of Bayesian priors. We implement this here as a specific application of power priors[22], while adapting an equal-but-discounted[16] approach. Importantly, we applied priors only to the control group and not to the experimental group, as control animals can be more reasonably assumed to belong to the same population (Methods).

We first performed a simulation study to estimate how the use of Bayesian priors influences sample size and power (Figure 2a). The simulation study was based on the formula:

$$n_{con} = n_{exp} - n_{prior} * index$$

where the number of animals of the control group (n_{con}) can be reduced by the number of control animals from prior studies (n_{prior}) multiplied by a weight (*index*, value between 0 and 1) that describes the similarity between control and prior groups. The experimental group (n_{exp}) remains the same. Based on this formula, the number of animals necessary in the control group is effectively diminished (discounted) by the weighed prior. Conversely, if the number of animals remains the same, a further increase in n_{prior} can still be beneficial as power could be enhanced up to its highest boundary, i.e. approaching 100% with large effect sizes (Figure 2a).

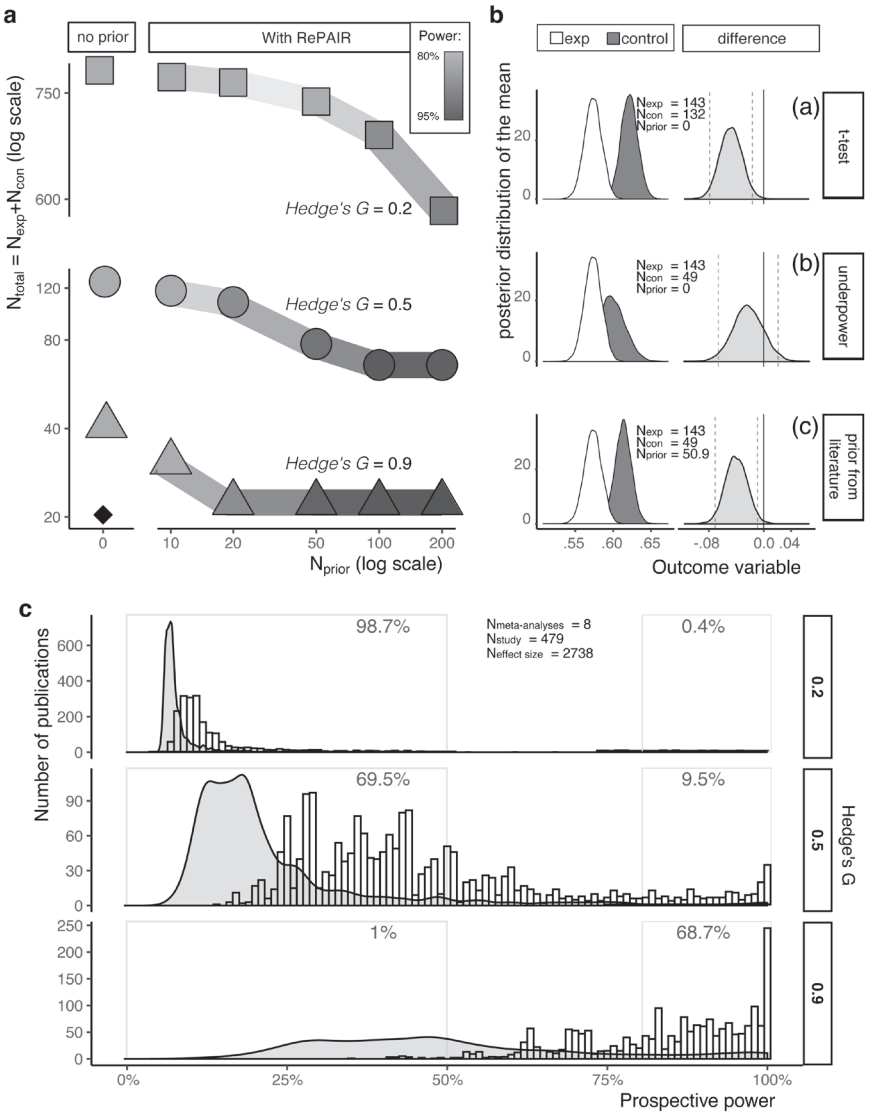


Figure 2. Historical controls can decrease the number of animals required for sufficiently powered research. a) Simulation study on the relationship between prior (index=1), sample size and power. A n_{prior} equal to 0 corresponds to a standard sample size estimation (two-tailed Welch t-test, α=0.05, effect sizes as in Figure1b-c, power=80%). The black diamond indicates the current median sample size. An increase in color intensity signifies an increase in power. As n_{prior} increases, n_{total} decreases until a plateau is reached. Subsequent increases in n_{prior} will result in increased prospective power. **b)** Application of historical controls to the experimental dataset RELACS. Posterior distributions of each group and of their mean difference. The test is significant if 0 (continuous line) is outside the 95% confidence interval (dashed lines) of the means' difference distribution. From the top (Supplementary Table 1), **(a)** analysis without prior provides the same result as a Welch t-test; **(b)** if n_{con} is decreased, the study becomes underpowered; **(c)** but this can be rescued if a prior from (unrelated) published literature is introduced. **c)** Prospective power when using historical controls with an *index*=0.3 (weighted at 30% in the analysis, i.e. n_{prior}=0.3*n_{con} of other studies within the same meta-analysis ("Data A" from Supplementary Figure 1) but maintaining current resources (n_{total} kept the same; n_{con} = n_{total}/3 as recommended rule of thumb) shown as histogram. Grey density plots represent the current prospective power as Figure1c.

Validation in a case study

To test the validity of the proposed method in real-life scenarios, we performed a case study involving experiments assessing the effect of early life adversity (ELA[23]) on spatial learning in adult male mice. The experimental dataset was gathered by aggregating data from single experiments that in principle shared the same design but individually had low power, from several laboratories in the RELACS (Rodent Early Life Adversity Consortium on Stress) consortium. Overall, information of 275 animals (n_{con} = 132, n_{ELA} = 143) was collected, which was larger than required by our prospective power calculation ($n_{con + ELA}$ = 200). Spatial learning was operationalized as discrimination ratio measured in the object-in-location test. In the RELACS dataset, the discrimination ratio was significantly decreased in ELA animals compared to control mice ($t(272.99)=3$, $p=0.003$).

We then mimicked a prospective experiment by reducing the number of control animals of the RELACS dataset to one third of the ELA animals. The new sample size would then be n_{con} = 49 and n_{ELA} = 143. This hypothetical experiment is underpowered since the difference distribution (ELA distribution – control distribution) contains the value 0 in its 95% confidence interval (Figure 2b). Normally, one would argue that the two groups are not different from each other. To 'rescue' the interpretation while still conducting a per se underpowered experiment with 49 control and 143 ELA animals, a Bayesian prior was used. A prior was specified based on relevant yet unrelated (non-ELA) published studies of spatial learning using the object-in-location test. This prior had a cumulative adjusted sample size n_{prior} = 50.9, as measured by the equation described in the previous section. The analysis therefore contained the sample size of ~51 animals for the prior of the control group, 49 control and 143 ELA animals. Although the experiment now hypothesized is still underpowered, the prior rescues the interpretation: the value 0 is outside of the 95% confidence interval of the difference distribution (Figure 2b), and one would conclude that there is evidence that the two groups are different from each other. In other words, this example shows that the same experiment could be conducted with 83 fewer animals (from the 132 control animals of RELACS to the subgroup of 49 animals for the hypothetical experiment) while maintaining a power >80%.

When specifying the prior, every effort was made to reduce subjectivity in the selection of the literature and the definition of the related indices. Yet, other experimenters might have selected different papers with the same task, or assigned different weights. Although it is not possible to exclude this possibility, it is unlikely that it would have had major effects on the results. The distribution of the prior was very similar to the one of the control animals in the RELACS dataset (Figure 3a), which suggests a certain consistency in the measurement values of the experimental endpoint across sources of data.

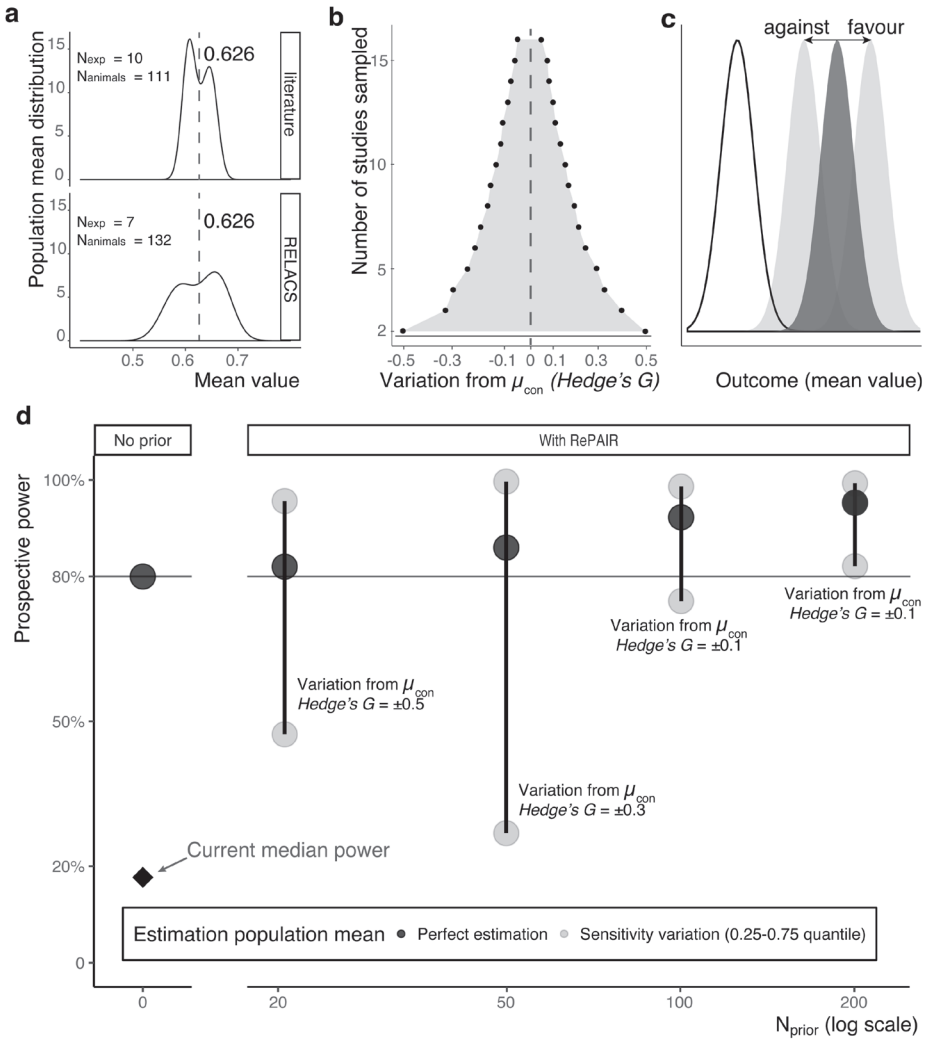


Figure 3. Sensitivity simulation. a) Density distribution of control population means with data from prior literature. Dashed line = mean of the control means. **b)** Range of variation of estimation of populations means (μ_{con}). Relevant deviations were calculated as the 2.5th and 97.5th percentile interval (grey area) of estimated sampled means (10,000 times, 2 to 16 experiments from literature + RELACS combined). Once more than 10 experiments are used, the variation (Hedge's G = 0.1) becomes negligible. **c)** Schematic representation of how variation of estimated population means can be both in favor and against the hypothesis scientifically investigated. White = distribution of the experimental group; grey = distribution of the control group. **d)** Prospective power (Hedge's G = 0.5, equal variances) with historical controls is higher than current practice (black diamond) despite variations in population mean estimation (black line). Since n_{tot} is not consistent due to the increasing of n_{prior}, prospective power can be interpreted only vertically (black lines). Each vertical line displays how prospective power changes if the estimated prior mean is a perfect estimation of the population (dark dot), or deviates from it in favor (top light dot) or against (bottom light dot) the investigated hypothesis. The light dots correspond to rounded valued of the 2.5th and 97.5th interval calculated from (**b).** The exact variation for each percentile interval is written in figure. The progression of the dark dots is an alternative visualization of the increase in color intensity of Figure 2a.

Nonetheless, the issue of subjectivity may arise when considering other experimental endpoints. We evaluated this concern by performing a sensitivity simulation study to mimic variation arising from different selections of literature (Figure 3b-c). Here, we randomly sampled control experiments from an available pool, containing non-ELA literature studies as well as the control studies of the RELACS dataset. This analytical approach to estimate variation has limitations, since researchers would rightfully follow pre-specified criteria to select previous experiments, rather than picking them at random. With the estimated variation, we calculated how random control-study selection would relate to study power (Figure 3d). Overall, the prospective power when using a prior was always larger than the currently estimated 18% (Figure 1a), despite the variations.

Bayesian priors can substantially improve statistical power

Whether Bayesian priors can be applied to new studies depends on the presence of suitable available data from previously performed, similar studies. Although it is difficult to estimate how much suitable data (for a particular experiment) exists in the literature, one could argue that if publications are similar enough to be included in a meta-analysis, they should also be sufficiently similar to be used to calculate a prior.

We recalculated the prospective power displayed in Figure 1c of studies identified by our systematic literature search (Supplementary Figure 1). This time, controls of other studies within the same meta-analysis were used to calculate the prior. New experiments were simulated with the same total number of animals (n_{total}) of the published studies, but distributed differently to the experimental and control group. Since the control group can be aided by the prior, more animals were allocated to the experimental group, according to the rule of thumb $n_{exp}=2* n_{con}$ (Figure 2c).

For Hedge's G = 0.9, application of Bayesian priors increased the percentage of sufficiently powered studies from 12.5% to 69%. These calculations were performed with an *index* of 0.3, which is quite conservative; using an *index* of 1 would yield similar results, with prospective power increasing to 72.5% for large effect sizes.

RePAIR can facilitate implementation

To facilitate the use of Bayesian priors in animal experiments, we created RePAIR (Reduction by Prior Animal Informed Research), an open-source web-based tool (https://osf.io/wvs7m/) that enables anyone designing future experiments to improve the quality of the study design. With a user-friendly interface, one can calculate (multiple) prior parameters from summary statistics of existing data, perform sample size calculations and execute analyses.

RePAIR can also be used to visualize the (potential) heterogeneity between one's own previously acquired control data and control data from other labs; if one's own data differ substantially from those obtained earlier in other laboratories, one could

decide to use only one's own existing control data to calculate the prior, or to not use historical controls at all and instead perform a fully powered experiment.

Sensitivity analyses are essential[16] when priors are specified. To facilitate such analyses, we included in RePAIR the option to perform two types of sensitivity analyses: 1) the leave-one-out sensitivity, to check whether any prior study has substantial influence on the final result, and 2) a sensitivity analysis on the indices by selecting lower or higher indices than those chosen for the analysis. Using the 'leave-one-out' sensitivity, one can assess the impact of each specific experiment on the final analysis. Here, prior parameters are calculated *k* times for each *k* - 1 experiment added: if 3 prior experiments (A, B, C) were added, 3 sensitivity analyses will be conducted (A and B, B and C, C and A). To perform the indices' sensitivity analysis, users have to specify the *index* as a range. The average of the range is used for the main analysis, whereas the lower and higher boundaries of the range are used for the sensitivity analyses. In RePAIR, parameters for sensitivity analyses are automatically calculated when specifying the prior. The resulting file can then be re-uploaded when analyzing data from the new experiment, and sensitivity analyses will be automatically conducted.

Discussion

There is a growing awareness of the reproducibility issue in animal experiments. Study preregistration and the introduction of more rigorous guidelines (e.g. PREPARE for planning of animal experiments and ARRIVE[24] for their reporting) can only partially address this issue. We here describe the (lack of) statistical power in animal studies and explain how the use of Bayesian priors can provide a potential solution. As previously suggested by others (for example[14,17–19]), this statistical method uses historical data to limit the number of animals necessary to perform well-powered research, or to reach higher statistical power with the same number of animals as currently used in experiments. We delineated how to best apply Bayesian priors in the context of animal research, and created RePAIR, a user interface to ease implementation of this approach. This approach can substantially increase prospective power without increasing the total number of animals used. It can be an extremely powerful tool, if correctly implemented and interpreted.

Animal experiments have low statistical power.

The statistical power of animal experiments is much lower than commonly a priori assumed. Although our approach was not conservative, we estimated that at best 12.5% of a large sample of rodent studies were sufficiently powered (i.e. prospective power was larger than 80%). This estimate is a best-case scenario as it is not yet adjusted for any subsequent multiple testing, experimental bias, p-hacking/fishing, selective reporting, etc.

One may wonder why our estimate of sufficiently powered experiments is so low. A technical limitation of our approach is that it considers a range of effect sizes found in literature, and not a minimum effect size of 'biological significance'. Although valuable, the minimum effect size criterion is seldom used in power calculations. We therefore consider our estimate reliable. Besides this technical limitation, several observations can explain why prospective study power is much lower than the commonly assumed 80%. One explanation is that effect sizes are often estimated optimistically in power calculations, since they are based on earlier findings that are liable to (publication) bias[25]. A second explanation is that rodent experiments are frequently exploratory in nature[26], and many scientists opt to use a debatable 'standard' 6 to 10 animals per group. Indeed, the effect size frequently assumed in rodent literature (Hedge's G = 1.4) is much larger than the range of effect sizes that is commonly observed (Hedge's G = [0.2,0.9]). Effect sizes in certain subfields may be more towards the lower (e.g. behavioral phenotyping[9]) or higher (e.g. molecular studies[20]) end of this distribution. Still, this discrepancy between assumed and observed effect sizes contributes to the power problem and reproducibility crisis in animal research in a major way.

Limitations and recommendation for the re-use of historical data

The use of historical control data as here proposed requires the researcher to select experiments and to specify weights via the index. This selection is naturally subjective, and thus can be criticized as introducing bias in an experiment[27]. In the next paragraphs, we discuss how subjectivity might impact an experiment using historical controls, and how these limitations are pragmatically addressed in our methodology. Next, we discuss why using historical controls is a valid approach, despite its subjectivity. Finally, we provide practical recommendations for the re-use of historical control data in new experiments.

When selecting previous experiments, a possible risk is that their cumulative distribution is very different from the one of the new experiment's control group (prior-data conflict)[18]. The prior distribution may then push the control group more towards the experimental group (causing a decrease in power) or further away from it (causing an increase in power); in other words, it can introduce a bias in the posterior distribution, i.e. the distribution obtained from combining prior and new (control) data. The posterior distribution of the control group may then not be a good estimate of the control population, thereby directly impacting (negatively or positively) the power of the study. Previous reports have suggested several ways to mitigate this problem. Some have suggested disregarding the prior altogether, although this would cause a reduction in study power. Others have suggested redistributing the weights of the prior studies based on their relative discrepancies[18,28]. However, we argue that prior-data conflict cannot be adequately addressed in this way. Thus, these solutions are based on the assumption of a correct evaluation of prior-data conflict. This means that a

new experiment was planned with a prior control group, and that the data of the new experiment had already been collected. The evaluation of prior-data conflict then consists of judging whether the prior control and the new control actually belong to the same population. Since the approach presented in this paper is aimed at reducing the number of animals in the new control group as much as possible, the new control group will not be sufficiently large to correctly estimate the new control population and therefore cannot be compared to the prior control population. In other words, we cannot disregard a wealth of previous information based on data from a handful of new animals.

Although we cannot adequately check whether the prior control group is reasonable (i.e. there is no prior-data conflict) after we conduct the new experiment, we can evaluate whether prior control groups are potentially incompatible while we plan the new experiment. Prior controls can be from one's own lab, from others' or a combination. Using information of multiple laboratories can be beneficial. If each laboratory is a single population[7], the overall population can be addressed as a population of populations. As a consequence, results based on information from multiple laboratories should be more generalizable. However, using information of multiple laboratories can also be a major source of variation in the prior distribution, because variation within a laboratory is likely smaller than variations between laboratories[5,13]. An experimenter can check whether one's own prior control data differ largely from prior control data selected from literature, or whether a particular experiment stands out. This evaluation must occur on a case-to-case basis with careful assessment and justification, ideally while planning the experiment. When building a prior, the experimenter can visually compare the distributions of datasets from the selected own or others' previous experiments and assess (for example) whether own prior control data is too different from others', or whether there is an 'odd-one-out' dataset that drives the prior control distribution. The experimenter can then choose to exclude the odd-one-out dataset, or to not use prior control data from other laboratories at all if too different from own prior control data. In both circumstances, the experimenter may nonetheless want to review the potential origin of the differences, for example by comparing experimental design between studies. To facilitate the process of assessing the compatibility of prior control data, the RePAIR app provides a visualization tool; this will aid the experimenter in the process of selecting prior experiments and determining their *index*.

Besides selecting previous experiments subjectively, in our methodology the experimenter also specifies their weight (*index*) subjectively. To avoid subjectivity, one may wonder whether it's necessary to use weights or whether they could be derived from a calculation. The use of weights is in agreement with the common view that past information needs to be somewhat downweighed because experiments are rarely identical[17]. Several methods (for example[17,18,28,29]) have been developed to overcome the subjectivity in defining the weights by analytically deriving them based on the

discrepancy between historical and new data. These methods are appealing and definitely pragmatic in clinical sciences. However, we argue that these methods are not appropriate in animal studies. The argumentation is similar to the one used in the previous paragraphs to deprecate the assessment prior-data conflict in animal experiments: if prior controls are used to reduce sample size as much as possible in the new control group, it cannot be assumed that the new control group – likely based on a small number of animals – will provide a good estimation of the new control population. A correct estimation of the new control population is necessary to evaluate the discrepancy between prior and new control groups. As a consequence, methodologies that analytically derive weights based on this discrepancy cannot be used in the context of animal experiments where the goal is to reduce sample size as much as possible. Therefore, weights are necessary and need to be specified by the experimenter. In our methodology, we use the 'equal-but-discounted' method based on power priors, as suggested by Ibrahim and Chen[22]. Briefly, by setting a certain discount/weight (e.g. *index*=0.5), the sample size is reduced (e.g. from 10 to 5). Scientists themselves (by expert elicitation, an accepted practice in Bayesian statistics[30]) can therefore decide to what extent they value earlier data. Although subjective, even conservative (low) *indices* can be beneficial.

One could argue that the subjective selection of previous experiments and related *indices* is susceptible to gaming and offers yet another 'degree of freedom' when performing analyses. This concern is valid especially for research fields for which little 'past evidence' exists. Until optimal population parameters are known, specification of a prior is subject to variation. At the same time, it is impossible to pre-define how many high-quality studies are necessary for estimating an optimal parameter. We recommend preregistering prior experiments and their indices on suitable platforms like the Open Science Framework (osf.io), preclinicaltrials.eu[31] or the Experimental Design Assistant[32]. During preregistration, scientists should define the prior experiments and related *indices*, and should also describe the rationale behind the choice of experiments and planned sensitivity analyses. Furthermore, scientific societies can facilitate the process of defining reliable priors, for example by establishing expert panels. This could eventually result in an 'atlas' of common control priors in animal research. As the number and quality of experiments increases, more precise estimates of the parameters of the control population can be obtained and, consequently, the subjectivity in selection of experiments and *indices* will reduce.

Despite the above-mentioned limitations, the use of historical controls is desirable and valid. It is desirable because it offers the possibility of increasing statistical power, thereby improving the reliability of animal research. It is valid because it is a translation in statistical terms of assumptions already used in daily research practice. New experiments are usually planned based on information obtained in previous studies. Even though variations between strains and labs clearly exist[7,12,33,34], researchers

have similar expectations about how a control group 'should respond'. Indeed, if this expectation is not met, a researcher would likely not trust the data and conclude that the experiment "did not work" or "needs to be better optimized". In this context, an advantage of rodent studies is that they are relatively well-controlled and often employ 'standard' tests used in many labs. For example, if the plasma concentration of a hormone normally varies between 60 and 100 µg/mL in control animals, an experimenter will rightfully question the validity of control-animal data that show a range between 5 and 10 µg/mL. Translating the above in statistical terms, researchers assume control animals always to belong to the same overall population. This warrants the formal statistical use of priors to supplement control-group data.

The choices involved in building the prior distribution must be considered when interpreting the results, for which sensitivity analysis remains essential[16]. Choosing prior studies and the related *indices* is similar to selecting literature for a new experiment. We recommend considering the quality of the study as well as design variations that likely impacted the results. For example, researchers may select previous experiments obtained in only a specific sex (e.g. females) if the outcome is sex-specific (e.g. ovulation), or both sexes if it is not expected to be sex-specific[35]. Similarly, blinding and randomization may be chosen as inclusion/exclusion criteria, or might be used to define the *index*. The *index* is specified for each study separately. As a rule of thumb, previous reports have attributed a high weight (0.9) to studies that belonged to the same meta-analysis, and lower weight (0.7-0.8) to studies that did not[18]. We suggest a more conservative stand: high weights (0.8-1) could be applied to repeated experiments from the same lab (e.g. different batches), mid weights (0.4-0.8) for experiments that (could) belong to the same meta-analysis, low weights (0.1-0.4) to experiments from other sources. We also recommend specifying a range for the index and conduct sensitivity analyses. RePAIR has in-built features to support each step of the process, from visualization of distribution of prior experiments, to automatic sensitivity analyses.

If sufficient prior information is available, it is theoretically possible to decrease n_{con} to as low as 2 (to still be able to calculate a standard deviation). However, this is not advisable because randomization would be difficult. As a rule of thumb, we recommend that control animals comprise at least one third of the total number of animals in a new experiment. Even though sample size can be no longer reduced, prior information can still be beneficial because it will increase statistical power above 80%.

Finally, if sufficient prior information is not available, priors should not be used; in that case the researcher should perform an appropriately powered experiment, even if this means a sample size of (well) over 20 animals per group is required.

Concluding remarks

The re-use of historical control data in animal experiments can be an extremely

powerful tool to increase statistical power and the reliability of animal studies, if correctly implemented and interpreted. Although here discussed in relation to t-tests, the same approach can be used in more complex experimental designs (e.g. 2x2-ANOVAs) where multiple groups could then be considered as 'controls'. It is a feasible solution to reduce and replace animal use for those research questions where good alternatives to animal testing are not yet available.

Methods

General information

Every effort was made to minimize bias, e.g. data gathering, and analysis was performed blindly, multiple experts were consulted for sensitive information (inclusion/exclusion criteria), and studies' characteristics were prospectively defined. This study was developed after a preliminary analysis of study power and estimation of sample sizes, conducted on a meta-analytic dataset developed previously by our own lab[36]. Part of this data is also used in this publication. Although no ex-ante protocol was preregistered, each component of this study was thoroughly planned in advance unless otherwise stated in each individual section. For data, code and other information about the project, see https://osf.io/wvs7m/ . Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Statistics

To compare control and experimental groups, we used two tailed Welch independent samples t-tests (α = 0.05). We chose a Welch's instead of a student's t-test because it does not assume equal variances between groups. Data distribution was assumed to be normal, but this was not formally tested. Given the small sample sizes of animal experiments, it is likely that also normality tests are underpowered. Bayesian analyses are explained in detail in the following sections.

Evaluation of studies for the systematic review was performed in a random order. Briefly, each study was given a pseudo random number generated in R. This number was then used for the ordering and assessment of publications. For the case-study, presence of randomization was an inclusion criteria. However, we do not have information on how randomization was conducted by the single independent laboratories.

Throughout the study, every effort was made to limit selection and confirmation biases. Inclusion and exclusion criteria for the systematic review were defined before starting the review. The choice of distribution and ranges throughout the analysis (e.g. estimation of effect sizes, sensitivity variation range) was performed once the data was already collected, but prior to any data visualization. For the definition of prior information and the definition of inclusion/exclusion criteria for the case-study, the

researcher (VB) did not have access to the effect sizes, but did have access to meta information of the study (e.g. characteristics of the ELA model).

All analyses were conducted with R (version 4.0.0) in the R studio environment on a macOS Mojave (version 10.14.6). The following R packages were core to this study: 1) tidyverse[37] (version 1.3.0) for general data handling, 2) shiny[38] (version 1.5.0) for RePAIR web-based tool, and 3) MESS[39] (version 0.5.6) for power calculations. The case study power calculation was also confirmed with G*Power[43] (version 3.1.9.2).

Estimation of study power

Since real effect sizes are not known, estimating statistical power of animal research is equivocal. A common approach is to calculate achieved statistical power from meta-analyses identified with a systematic literature search (Supplementary Note 1-2, Supplementary Table 2).

The achieved power is the probability to reject the null-hypothesis (i.e., no difference between the control and experimental group) with the observed sample sizes. Here, this was retrospectively calculated for each set of summary statistics extracted from the systematic literature search ("Data B" from Supplementary Figure 1). Although data may have come from complex experimental designs, we assumed it always belonged to two independent groups (Welch t-test, two tailed, $\alpha = 0.05$, sample size and Hedge's G of "Data B" from Supplementary Figure 1). This retrospective power calculation is a biased estimation of prospective study power, because the larger the p-value observed in a study, the smaller its achieved power[40]. We replicated previous reports[2,3], which used meta-analysis to estimate real effect sizes. This retrospective power calculation was not part of the original study protocol, and was subsequently added.

An alternative approach is to estimate a common prospective study power, thereby partially overcoming the limitations of achieved power calculations. As an experimental design, we assumed two independent groups (Welch t-test, two tailed, $\alpha = 0.05$), while sample sizes were gathered from our systematic search ("Data A" from Supplementary Figure 1). Importantly, only the two largest groups reported in each paper were extracted, assuming that at least the comparison of these two groups were sufficiently powered while all other experiments may have been control experiments. For effect sizes, we aimed to estimate a plausible range - rather than a single value - to mimic scenarios of researchers initiating a new study.

To estimate a plausible range of effect sizes in preclinical literature, we calculated the 25th, 50th and 75th percentiles of Hedge's G's absolute values and defined them as small, medium and large effect sizes, respectively (based on "Data B" from Supplementary Figure 1). Blinded to the results, we chose the 25-75% interval instead of the 95% confidence interval, to avoid extreme values. Extremely low effect sizes may not be biologically relevant and are confounded by null effects, while extremely

high values may lead to interpretation issues and are confounded by over-estimations due to biases. We confirmed (see code at https://osf.io/wvs7m/) that these values are replicable by applying the same methodology to a separate dataset[20,41].

Within this framework, prospective power is the probability to reject the null-hypothesis if the effect size is set equal to a small, medium and large value, respectively. A simple experimental design was assumed (t-test), while sample sizes and effect sizes were estimated from literature. As a consequence, this approach to calculate prospective power portrays a plausible scenario that a new researcher may expect.

Simulation study on the relationship between prior information, sample size and statistical power

The mathematical derivation of the algorithm for prior distributions[42] is described in detail in the Supplementary Note 3. In our study, priors were built based on conjugate distributions, meaning distributions that multiplied by the likelihood function would create a posterior distribution, which summarizes information of previous and current studies with respect to the mean of the control group. The posterior distribution is of the same family as the prior distribution. We therefore chose the prior distribution for the mean in the control and experimental group to be normal, and for the variance to be inverse-χ^2. Although modern computing power is reducing the need for conjugacy[16], we preferred this method for its solid mathematical foundation and the assumption of normality seemed appropriate as it is frequently used in preclinical sciences.

Of note, informative priors (namely priors based on previous experiments) were applied only to the control group. The mean and the variance of the experimental group also have a prior and a posterior distribution. However, the prior distribution of the experimental group is "uninformative", meaning that it will not have impact on the results. Therefore, the posterior distributions that describe mean and variance of the experimental group in our approach depended only on the information of the current experiment.

We performed a simulation study to evaluate to which extent a prior could reduce the number of animals necessary and how this would influence study power. The more informative a prior for the mean in the control group, the more influence it will have on the conclusions of the experiment. Mean and variance of data in the control group were kept identical in all conditions (μ_{con}=0, σ^2_{con}=1); therefore, the influence of the prior was dependent only on its varying sample size n_{prior}. Supplementary Table 3 summarizes all factors varied in the simulation. For each combination of factors, 10,000 datasets were sampled from the corresponding population.

Firstly, we calculated how many animals (n_{total}= n_{con}+ n_{exp}) one would need to perform experiments with the determined characteristics, given a standard sample size calculation (n_{prior} = 0). This was later confirmed by G*Power[43]. The calculation assumed a balanced design, meaning n_{con}=n_{exp}. Secondly, we decreased n_{con} by adding

n_{prior} while keeping n_{exp} the same. Since it would be illogical for n_{con} to become negative when $n_{prior} > n_{con}$, n_{con} is minimally 2, which is the lowest possible sample size to compute a standard deviation. The total number of animals used in then:

$$n_{total} = n_{exp} + n_{con}$$

$$n_{con} = n_{exp} - n_{prior}$$

$$n_{prior} = \sum_{(p=1)}^{p} n_p * index_p$$

where the number of animals of the control group (n_{con}) is diminished by the effective number of prior animals (n_{prior}), meaning the sum of the animals of each experiment used to define the prior (n_p) multiplied by the respective weight ($index_p$). The index is a value between 0 and 1. An *index* of 0.3 means that the information in the prior study at hand will be only be weighed for 30% in the analysis. In the simulation, we set *index*=1 and we assumed that the prior is a perfect estimation of the population, although this issue will be further addressed with a sensitivity simulation study (Section "Case study"). For analyses, researchers may opt to vary this value depending on the degree of similarity of the prior experiments to the current study. For more information about this topic, see "expert elicitation"[30].

Case study

For validation and as an example, we applied Bayesian priors as described in the previous sections to an experimental dataset. Here, the prior for the control group was specified from unrelated literature, while the prior of the experimental group was uninformative.

To this purpose, a well-powered dataset investigating a real and reproducible difference between two groups was required. We defined as "real" and "reproducible" an effect that persists in a high quality, well-powered meta-analysis. These criteria were met by the effects of early life adversity (ELA) on memory after non-stressful learning, as identified by a recent meta-analysis of literature previously conducted by our own lab[9]. From this study, an effect size of Hedge's G = 0.4 was estimated to describe the difference in performance on the object-in-location memory task between controls and animals that experienced ELA with the limited bedding and nesting (LBN) model[23]. Considering a Welch two tailed independent means t-test and an α of 0.05, 200 animals would be required to achieve a power of 80%.

Due to the paucity of power of preclinical studies, it is not surprising that we were unable to identify any study on this experimental outcome using (at least) 200 animals. Even though no single laboratory works with such sample sizes, the required

power could be reached by combining data of multiple laboratories. To this end, we created RELACS (Rodent Early Life Adversity Consortium on Stress), a unique rodent consortium constituted by several laboratories around the globe working on ELA.

We identified relevant authors from a recent systematic search of our lab[9], as well as via our network (Supplementary Note 4). The consortium was prospectively founded and ultimately included 7 independent experiments that met the specified criteria for this particular study. We calculated for each experiment (i.e. an independent set of animals), a measure of discrimination (discrimination ratio) as the ratio between the time spent in the novel location divided by the total exploration time, meaning the sum of the time spent in the novel and the familiar location ($discrimination\ ratio = \frac{time_{novel}}{time_{novel} + time_{familiar}}$).

When analyzed independently, a p-value < 0.05 was reached in only 2 out of 7 experiments, in agreement with the low power of preclinical studies. By combining the 7 experiments, we reached a sample size of 275 animals, distributed as n_{con}=132 and n_{ELA}=143. The effect size Hedge's G = 0.37 calculated in the RELACS dataset was similar to the one estimated from literature (Hedge's G = 0.4). We concluded that this dataset meets the required criteria to validate RePAIR: it describes a reproducible effect as shown by the meta-analysis, and it is sufficiently powered since sample size is larger than the expected 200.

Of note, aggregating data from multiple laboratories in such way would normally be unadvisable as it does not meet the criteria of an individual participant data meta-analysis. However, we used this approach here because our intent was to "mimic" a well-powered experiment, which was otherwise not available in the literature.

To specify a prior from unrelated studies, one of us (VB) selected relevant literature, to mimic planning an experiment with the same characteristics (Supplementary Table 4) as the RELACS dataset, i.e. investigating memory after non-stressful learning with the object-in-location task in adult (age between 9 and 41 weeks, median = 18 weeks) male mice. The researcher was requested to select 8 publications that she would use to set up her study, while focusing on the control and not the experimental group. The selected publications did not belong to the ELA field, and were not used elsewhere in this manuscript. Furthermore, for each study the researcher defined a similarity *index*, a number between 0 and 1 that would express how similar the control group of each literature study was to the experiment that she was planning to perform (1 = identical/equal). Two publications reported the same outcome on two separate groups of animals. Both experiments were considered, albeit with a lower *index*. The process was overseen by a senior researcher (RAS).

As the experimental dataset and prior specification were identified as described above, we had all the elements to validate that the Bayesian approach would reach with fewer animals the same conclusions as current practice. First, we performed a

Welch independent samples t-test (two tailed, $\alpha = 0.05$) on the RELACS dataset to replicate that control and ELA groups differ in discrimination (p-value < 0.05) in the object-in-location task. We then performed the same test, but with fewer animals in the control group and an informative prior. Several tests (Supplementary Table 1) were conducted as control.

Although VB selected the prior blinded to the results of the RELACS dataset, prior specification had some degree of subjectivity, i.e. another researcher may choose different publications on which to base their study. To experimentally quantify relevant variation in article selection, we simulated many different priors by picking at random 10,000 times *k* experiments (*k* equal to 2 through 16) from the 17 identified in total (10 from VB's literature selection + 7 of the RELACS dataset). Variation in article selection for each *k* was calculated as the 2.5th and 97.5th percentiles to avoid extreme values. Changes in Hedge's G between 0.1 and 0.5 could appropriately describe the variation across *k*, and 10 articles are here sufficient for a stable estimation of the population parameters. Of note, the sampling occurred from a finite population, where 17 experiments represent the reference value of the estimated variations. As a consequence, the intervals may be underestimated.

With this experimentally derived estimation of population mean's variation, we conducted a sensitivity simulation study to investigate how prior control population mean's variation affected prospective study power. Of note, this variation can act both in favor or against the hypothesis experimentally investigated, depending on whether the prior control population mean moves towards or away from the population mean of the experimental group. Despite this limitation, we preferred this approach of experimentally deriving variation values over using a canonical variation of Hedge's G = 0.1.

We preferred using number of animals rather than number of experiments in the sensitivity simulation, to keep consistency with the first power simulation study. The relationship between number of sampled experiments and number of animals is not straightforward. For example, one can achieve a n_{prior}=20 with just one experiment, or two (e.g. each of *n*=10), or three (e.g. *n*=9+*n*= 6+*n*=5). To transform variations due to experiments' selection to variations linked to sample sizes, we identified across the *k* * 10,000 sampled estimations of means, animals roughly equivalent to 20, 50, 100, 200 (n_{prior} in our sensitivity simulation). In these subgroups, we calculated the 2.5th and 97.5th percentiles, and visually validated their consistency. These values were used in the sensitivity simulation study to vary prior control population means (between 0 and ±0.5 Hedge's G depending on n_{prior}). All factors of the sensitivity simulation were kept identical to the previous simulation study (Supplementary Table 3).

Estimating how prior control information can impact statistical power with the current total number of animals used

We estimated the increase in prospective power if the Bayesian prior methodology would be used in new animal experiments with the resources currently available. We considered each study identified within each meta-analysis ("Data A" from Supplementary Figure 1) as a new experiment where n_{tot} was kept the same, but animals were redistributed in favor of the experimental group ($n_{exp}=2 * n_{con}$, according to our rule of thumb). The controls of all other studies within the same meta-analysis were then considered as priors. In other words, n_{prior} was calculated from the cumulative n_{con} of all other papers included within the same meta-analysis. This cumulative n_{prior} was then multiplied by the similarity *index*=0.3, meaning that we valued the degree of similarity of the control groups of studies included in the meta-analysis to be 30%. In this circumstance, the value of 0.3 is arbitrary. To evaluate how the similarity index affects power, we also calculated prospective power with a similarity index of 1.

Prospective power was calculated in the case of a two tailed Welch independent means t-test, for the plausible range of effect sizes previously identified (Section "Estimation of study power"), when considering an $\alpha = 0.05$. Since we adopted the same methodology and the same data, the immediate potential impact can be assessed by comparing the prospective power of without information of previously experiments (previously calculated) and with.

Lastly, we created a web user interface, RePAIR (Reduction by Prior Animal Informed Research), to facilitate the implementation of the Bayesian prior methodology to improve statistical power in animal experimentation. The supporting code is also freely available (https://osf.io/wvs7m/).

References

1. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* 2, e124 (2005).
2. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376 (2013).
3. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* 3, (2016).
4. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* 1, 1–9 (2017).
5. Crabbe, J. C., Wahlsten, D. & Dudek, B. C. Genetics of mouse behavior: Interactions with laboratory environment. *Science* (80-.). 284, 1670–1672 (1999).
6. Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: How much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10, 712–713 (2011).
7. Voelkl, B. et al. Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* (2020). doi:10.1038/s41583-020-0313-3
8. Macleod, M. & Mohan, S. Reproducibility and Rigor in Animal-Based Research. *ILAR J.* 60, 17–23 (2019).
9. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neurosci. Biobehav. Rev.* 102, 299–307 (2019).
10. Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. Meta-analysis and the science of research synthesis. *Nature* 555, 175–182 (2018).
11. Richter, S. H., Garner, J. P. & Würbel, H. Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nat. Methods* 6, 257–261 (2009).
12. Karp, N. A. Reproducible preclinical research—Is embracing variability the answer? *PLoS Biol.* 16, 1–5 (2018).
13. Richter, S. H. et al. Effect of population heterogenization on the reproducibility of mouse behavior: A multi-laboratory study. *PLoS One* 6, (2011).
14. Kramer, M. & Font, E. Reducing sample size in experiments with animals: historical controls and related strategies. *Biol. Rev.* 92, 431–445 (2017).
15. Brakenhoff, T., Roes, K. & Nikolakopoulos, S. Bayesian sample size re-estimation using power priors. *Stat. Methods Med. Res.* 096228021877231 (2018). doi:10.1177/0962280218772315
16. Spiegelhalter, D. J., Abrams, K. R. & Myles, J. P. *Bayesian approaches to clinical trials and health-care evaluation.* 1, (2004).
17. Galwey, N. W. Supplementation of a clinical trial by historical control data: is the prospect of dynamic borrowing an illusion? *Stat. Med.* 36, 899–916 (2017).
18. Mutsvari, T., Tytgat, D. & Walley, R. Addressing potential prior-data conflict when using informative priors in proof-of-concept studies. *Pharm. Stat.* 15, 28–36 (2016).
19. Walley, R. et al. Using Bayesian analysis in repeated preclinical in vivo studies for a more effective use of animals. *Pharm. Stat.* 15, 277–285 (2016).
20. Nord, C. L., Valton, V., Wood, J. & Roiser, J. P. Power-up: A Reanalysis of 'Power Failure' in Neuroscience Using Mixture Modeling. *J. Neurosci.* 37, 8051–8061 (2017).
21. Cohen, J. *Statistical power analysis for the behavioral sciences.* (Lawrence Erlbaum Associates, 1977).
22. Ibrahim, J. G. & Chen, M. H. Power prior distributions for regression models. *Stat. Sci.* 15, 46–60 (2000).
23. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A Novel Mouse Model for Acute and Long-Lasting Consequences of Early Life Stress. *Endocrinology* 149, 4892–4900 (2008).
24. Percie du Sert, N. et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biol.* 18, e3000410 (2020).
25. Ioannidis, J. P. A. Why most discovered true associations are inflated. *Epidemiology* 19, 640–648 (2008).
26. Rubin, E. J. & Fortune, S. M. Misunderstanding the goals of animal research. *BMJ* 360, 29321149 (2018).
27. Gelman, A. Objections to Bayesian statistics. *Bayesian Anal.* 3, 445–450 (2008).
28. Viele, K. et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm. Stat.* 13, 41–54 (2014).
29. Neuenschwander, B., Capkun-Niggli, G., Branson, M. & Spiegelhalter, D. J. Summarizing historical information on controls in clinical trials. *Clin. Trials* 7, 5–18 (2010).
30. O'Hagan, A. Expert Knowledge Elicitation: Subjective but Scientific. *Am. Stat.* 73, 69–81 (2019).
31. van der Naald, M., Wenker, S., Doevendans, P. A., Wever, K. E. & Chamuleau, S. A. J. Publication rate in preclinical research: a plea for preregistration. *BMJ Open Sci.* 4, e100051 (2020).
32. Du Sert, N. P. et al. The Experimental Design Assistant. *Nat. Methods* 14, 1024–1025 (2017).
33. Crabbe, J. C. & Phillips, T. J. Mother nature meets mother nurture. *Nature neuroscience* 6, 440–442 (2003).
34. Kafkafi, N. et al. Addressing reproducibility in single-laboratory phenotyping experiments. *Nat. Methods* 14, 462–464 (2017).
35. Shansky, R. M. Are hormones a "female problem" for animal research? *Science* (80-.). 364, 825–826 (2019).
36. Bonapersona, V., Joëls, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: a 3 level meta-analysis of rodent studies. *Neurosci. Biobehav. Rev.* 95, 1–16 (2018).
37. Wickham, H. et al. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686 (2019).
38. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for *R. R Packag. version 1.5.0* (2020).

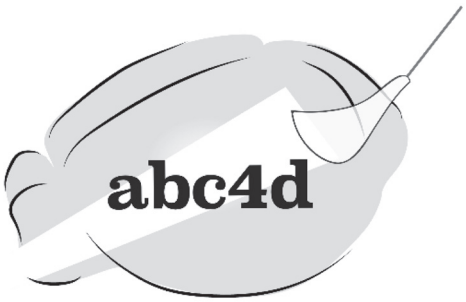
39. Ekstrøm, C. T. MESS: Miscellaneous Esoteric Statistical Scripts. *R package version 0.5.6* (2019).
40. Lenth, R. V. Some practical guidelines for effective sample size determination. *Am. Stat.* 55, 187–193 (2001).
41. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365 (2013).
42. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian data analysis.* (Chapman & Hall, 1995).
43. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191 (2007).

Chapter 11

Software developed

Software is revolutionizing research in biology, has an impact on research productivity[1] and is a critical part of open and reproducible workflow. Software has also been an integral part of my PhD: I relied on other's Open Software for my analyses, and I created Open Software for others to use (and improve). Although the Open Software supporting this thesis has been partially introduced in the respective chapters, I here summarize its main features.

abc4d



abc4d is an R package for data preprocessing and analysis of whole-brain microscopy data over time. Specifically, the package contains functions to 1) automatically (un) blind the dataset to limit bias while coding the analysis, 2) clean the data from unspecific protein binding and damaged brain areas, 3) deal with missing values and batch effects, 4) normalize and standardize the data depending on the analysis of choice, as well as 5) several analyses. This software is interoperable with several others for image processing, most notably *Clearmap*[2] and *CellFinder*[3]. In other words, this software was built as a continuation of these pre-existing ones, to cover a gap in the literature. This R package has been developed as part of Chapter 2.

RePAIR

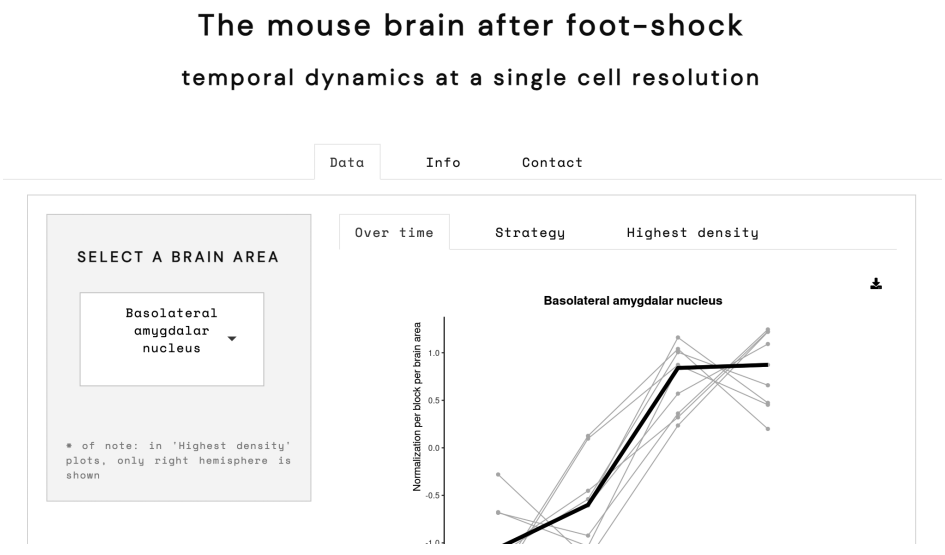
RePAIR is an interactive statistical software that uses previously obtained information to decrease the number of animals and perform well-powered research. This software is built within the Bayesian statistical framework, and it adopts power priors with the equal-but-discounted approach. The software can be used to 1) define prior parameters, 2) calculate sample size, and 3) analyze results. As a bonus, the software can also be used to visualize how one's own experiment compares to those of the literature. Currently, this needs to be inputted by the individual

researchers. In a future version, we envision that *RePAIR* can be directly connected with a database, where researchers can select the evidence to use. The user-interface was specifically designed and tested to be as intuitive and easy as possible for applied researchers. We also created video tutorials to facilitate its implementation. This interactive statistical software has been developed as part of Chapter 10.

Interactive data visualizations

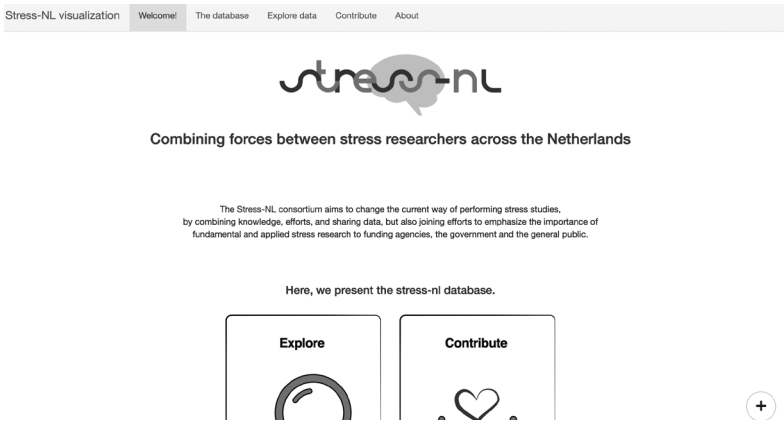
We used interactive data visualizations to increase the transparency of research data and to facilitate their exploration by researchers. Via user-interfaces, researchers can select the (sub)data of interest, visualize it graphically and perform basic exploratory analyses. Practical examples are provided below for each type of data.

Experimental data: The mouse brain after foot-shock.



Researchers can select the brain area of interest, according to the categorization of the Allen Brain Atlas[4]. The app returns a graphical representation of 1) the normalized c-fos+ cell count for each experimental animal (in figure, grey dot) in each experimental block (in figure, grey line), 2) whether the brain area becomes active by increasing the number or the intensity of c-fos+ cells (in figure, "Strategy" tab), and 3) a 3D representation of highest density sub-parts of the brain area (in figure, "Highest density" tab). The underlying data as well as source code is also openly available. This interactive visualization supports Chapter 2.

Consortia data: stress-NL database.



Within this interactive visualization, researchers can select the population of interest (e.g. healthy females between 20 and 30 years old with). Across all studies of the consortium, the tool outputs not only the available data and meta-data, but also visualizations and summary statistics (i.e. mean, median and deviations) of one particular variable, i.e. cortisol concentration after acute stress. This interactive visualization was not included in the final stress-NL database project, and it is no longer available online. Yet, it highlights an important possibility. While human data is protected by privacy and cannot (in most cases) be openly shared, user interfaces can be created to make the user interact and explore the data, while not directly seeing it. Specifically, this could be an important tool for meta-researchers, who could perform meta-analyses on more detailed populations. This interactive visualization was created as part of Chapter 4.

Meta-analysis data: MaBapp and MaDEapp.

These user-interfaces enable researchers to explore meta-analysis data. Here, you can select a specific population of interest (e.g. male mice that performed an object in location task). The tool outputs 1) all publications available in the database with the characteristics of interest, 2) forest, 3) funnel and 4) cumulative plots of the meta-analytic data. This can be used to plan future experiments, and for sample size calculation. These interactive visualizations were created as part of Chapter 5 and 6.

References

1. Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.-Y. & Kitano, H. Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* 12, 821–832 (2011).
2. Renier, N. et al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. *Cell* 165, 1789–1802 (2016).
3. Tyson, A. L. et al. A deep learning algorithm for 3D cell detection in whole mouse brain image datasets. *PLOS Computational Biology* 17, e1009074 (2021).
4. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953. e20 (2020)

CHAPTER 12

General discussion

Discussion

There are few experiences as engulfing as the subjective experience of stress. It affects multiple biological systems, with temporal and spatial specificity. It depends on genetic predisposition, early- and late- life events, and it leads to different responses depending on its type (e.g. psychological vs physiological) and frequency (e.g. acute vs chronic). This complex system has historically been studied by isolating each specific feature in an experiment. While this approach has been extremely successful, understanding stress as a system, rather than a collection of organs and hormones, can bring us closed to comprehend how it works in nature, in health and for the future improvement of disease. In this thesis, we examined stress as a biological system rather than isolating a specific feature in a single experiment. Specifically, we integrated information from multiple sources (i.e., literature, consortia, atlases, and newly generated data) to increase our understanding of the effects of acute stress and chronic stress experienced in early life on brain and behavior. Our approach was fully grounded in Open Science practices of collaboration, data and code sharing, as well as in software development.

After a summary of the main findings of this thesis, this chapter proceeds with a general discussion of two overarching themes, i.e. Open Science and data re-use. We specifically focused on how this thesis incorporates these approaches, their limitations and what our findings could mean for future research.

Summary and contextualization of main findings

This thesis is subdivided into three sections (Part A to C), each addressing one of the aims of this thesis:

A. To integrate information related to the healthy acute stress response (in rodents and humans);
B. To integrate information related to chronic stress experienced early in life (in rodents);
C. To develop methodologies for information integration.

The aim of Part A was to integrate information related to the healthy acute stress response. We conducted two studies focusing on the rodent brain (Chapter 2) and on salivary cortisol concentration in humans (Chapter 3). While the studies have been conducted in two different species, they both investigate acute stress as a dynamic process that occurs in time.

Chapter 2 investigated the effects of a specific type of acute stress (i.e., foot-shock) on the whole brain, rather than on prespecified brain areas (i.e., regions of interest). Acute stress leads to the sequential activation of functional brain networks, as demonstrated in several human studies using functional magnetic resonance imaging (fMRI, for reviews[1,2]). We replicated this key aspect of the acutely stressed

human brain, but now for the first time in mice. Rather than fMRI, we used whole-brain immunohistochemistry for its superior spatial resolution, to which we added a pseudo-time metric. In our experiment, 96% of all brain areas investigated had an increased expression of c-fos (marker of cellular activity). Hypothalamic areas stood out as being the most active, as well as the first to be activated, followed by amygdalar, prefrontal, hippocampal and finally thalamic nuclei. Importantly, we could move beyond the spatial resolution of circuits at which scale we replicated the human findings, and we could zoom in to single cells. This allowed us to identify shifts within – additionally to between – brain areas over time after stress, which we illustrated for the case of the basolateral amygdala. Moreover, while some brain areas showed an increase in the number of c-fos+ cells, others dramatically increased the c-fos intensity in just a subset of cells, reminiscent of engrams. This "strategy" changed after foot-shock in half of the brain areas. Throughout the project, we conducted our analyses while keeping in mind reproducibility as a key principle, and we facilitated their use by others by developing an R package. Most of our analyses tested the robustness of our (exploratory) findings across the different batches, besides evaluating their "statistical significance". This study had to be conducted in separate batches to enable feasibility in handling the material. Besides using a block design (i.e. an animal of each experimental group is randomized within a block) to maintain the robustness of the experiment, we saw the different batches as pseudo-replicates. We aimed to be as transparent as possible in the reporting of our methodology: the supplementary information contains detailed lists of all analyses considered to answer each question, and why a specific approach was ultimately chosen. We consider this one of the core values of Open Science: walking the reader through the process of scientific thinking, rather than providing merely the result as an ultimate answer. In all, Chapter 2 meets this thesis's aim by 1) integrating information across multiple brain areas and spatial resolutions over time, and 2) leading to the development of an R package to facilitate future analyses of whole-brain data.

The aim of Chapter 3 was to facilitate future integration of information related to the acute stress response (in a laboratory setting) of humans. In this chapter, we introduced the stress-NL database. This database was born from a collaborative initiative that involves 12 Dutch research groups, which worked together to achieve an accurate inventory of (neuro)biological, physiological and behavioral data from laboratory-based human studies that used acute stress tests. We provide example analyses of the usability of the data, specifically focusing on cortisol concentration after stress. In the first proof-of-concept analysis, we describe the effect size difference between males and females in cortisol concentration after acute stress, measured as area under the curve with respect to increase (AUC_I). In 18 out of 23 studies that met the criteria for this analysis, males had a consistently larger AUC_I than females, although effect sizes varied greatly (from d = 0.1 to d = 1.1). In the second analysis,

we again compared males and females, but this time females were distinguished between using and not using oral contraceptives. The results suggest that the use of oral contraceptives dampens the cortisol peak value in females. To our surprise, only 4 experiments in the stress-NL database met the criteria for this analysis, i.e. reported sufficient participants not using oral contraceptives. The qualitative description of the database already shows the potential as well as the challenges of the approach. The database contains individual participant information of over 5500 participants (41% female), of age between 6 and 99 years. The type of acute stress paradigm, the outcomes, and the time of measurement were only partially overlapping across the different experiments. The heterogeneity of experiments within the stress-NL database reflects the heterogeneity of the literature. While the database allows in principle for re-, meta-, and proof-of-principle analyses, the next challenge is analytical. Previous research has highlighted that data aggregation may not always be appropriate for neuroendocrine data[3,4]. Future analyses should therefore be based on state-of-the art individual participant data analysis or Bayesian evidence synthesis. The promise of the flexibility of the database comes together with the increased statistical expertise. In all, Chapter 3 describes the novel stress-NL database, which will facilitate future integration of information related to the human acute stress response.

Conclusion of Part A

The studies of Part A illustrate that:

1) acute stress exposure changes cellular activity in many areas of the rodent brain, with patterns that align with earlier studies on the human brain;
2) combining existing datasets on acute stress in humans allows to discern patterns in the neuroendocrine response, an approach that might also be useful in animal research.

Independently of whether integration of information is applied to animal or human data, it is most efficacious when aimed at investigating the stability and reproducibility of effects.

The aim of Part B (Chapter 4 to 9) was to integrate information related to chronic stress experienced early in life by rodents. Adversities early in life can have long lasting consequences on brain development, and are one of the main risk factors for several mental health disorders. Here, we provide a quantitative description of the effects of early life adversity (ELA) on behavior and neurobiology. The approach was of

systematic review and meta-analysis of the literature. We focused on long-lasting changes of ELA, measured in adulthood. While each chapter of Part B focuses on a different group of outcomes, we maintained consistency in the approach: search string, inclusion/exclusion criteria and analyses are comparable across chapters. The only exception is Chapter 5, which includes prenatal as well as postnatal ELA models, as will be explained in more detail below.

In Chapter 4, we established a causal link between ELA and changes in adult behavior in mice and rats. Human epidemiological studies robustly associate the experience of (various forms of) early adversity with various mental health disorders, most notably anxiety, depression, suicide attempts and substance abuse[5]. However, these studies often rely on prospective or retrospective observational data. Due to confounding elements and reverse causality inherent to observational data, causal inference on the effects of ELA on adult behavioral changes is yet to be established[6] in humans. Many therefore reverted to animal studies[7], where genetic influences can be controlled, and ELA can be experimentally induced[8]. Yet, the ELA rodent literature is extremely heterogeneous: systematically reviewing the literature was essential to disentangle whether the emerging paradoxical effects of ELA on behavior were due to real biological differences (e.g. were robust across behavioral tests) or whether differences between studies could be attributed to methodological factors (e.g. type of ELA model, species, etc.). We concluded that ELA, here defined as alteration of maternal care, increases anxiety, improves memory after stressful learning, reduces memory after non-stressful learning and impairs social behavior, especially in males. The effects were further enhanced by other negative life experiences ("hits"). In all, Chapter 4 provides robust evidence on the causal effects of ELA on behavior, by integrating the information of 212 independent publications.

Chapter 5 to 8 investigate different aspects of the neurobiological effects caused by ELA in rodents. In Chapter 5, we investigated the effects of prenatal and postnatal ELA on the dopaminergic system. Converging evidence suggests that the dopaminergic system is involved in mediating the influence of ELA on vulnerability to psychopathology[9]: from its developmental period to its strong association with several mental disorders[10]. In this chapter, we analyzed 90 rodent publications to evaluate whether ELA indeed causes long-lasting changes to the dopaminergic system. The final dataset included 41 publications investigating prenatal ELA, and 49 publications investigating postnatal ELA. These were analyzed in two separate analyses since they impact distinct critical periods. To our surprise, the effects were rather limited. The striatum was the brain area most impacted, especially in the prenatal dataset. We expanded upon these findings in Chapter 6, to which we included an analysis of the brain's expression of other monoamines, i.e. serotonin and noradrenaline. Here, we used stricter inclusion criteria than Chapter 5 (but identical to Chapters 4, 7 and 8), with the intention to decrease the heterogeneity of the population, thereby

potentially uncovering effects hidden in our previous study. Ultimately, we included 47 publications investigating the effects of ELA, in this chapter restricted to alterations of maternal care, on biochemical indicators of monoaminergic systems. Our quantitative analysis focused on males, due to the limited number of experiments performed in female rodents. Even with the more homogeneous population, we confirmed our two previous findings of Chapter 5: 1) the meta-analytic changes induced by ELA on monoaminergic systems are limited, and 2) the effects of ELA on the dopaminergic system are most evident in the striatum. Furthermore, our results suggested that the prefrontal cortex is the brain area most sensitive to ELA effects on the serotonergic system. While the effects on dopamine were most evident at the level of receptors, the effects on serotonin were most evident at the level of concentration of serotonin and its metabolite 5HIAA. Lastly, noradrenaline was investigated by a remarkably low number of publications, highlighting an evident gap in the literature. Chapter 6 therefore reproduces the findings on the effects of ELA on the dopaminergic system, and extends them to other monoamines.

In Chapter 7, we investigated another aspect of cellular activity. Rather than neurotransmitters (Chapter 5 and 6), here we investigated the effects of ELA on the brain's expression of immediate early genes. Immediate early genes (IEG) increase their expression upon calcium influx, and are often used as markers of cellular activity[11]. Although the downstream products of IEG are diverse (e.g. transcription factors, postsynaptic proteins, secretory factors), their functions are surprisingly homogeneous. IEGs are generally related to cellular processes, such as dendrite and spine development, synapse formation and elimination, and regulation of excitatory/inhibitory balance. Here, we synthetized data of 39 publications. Specifically, we quantitatively meta-analyzed data in males, where the IEG c-fos was measured; a qualitative analysis of other IEG and of females is available at the level of a systematic review. At rest, ELA increased c-fos expression, irrespective of other life events. After an acute stressor (i.e. both in the ELA and control groups), ELA increased c-fos expression only if (control and ELA) animals had experienced other negative life experiences. These results suggest that ELA creates an "already stressed" phenotype, comparable to ELA-naïve, acutely stressed animals.

Lastly, Chapter 8 investigated the effects of ELA on structural plasticity, specifically on morphology, neurogenesis and BDNF expression as a potential contributing mediator. We synthetized the results of 64 publications, the majority of which investigated outcomes in males and in the hippocampus. Concerning morphology, we meta-analytically concluded that ELA decreased the volume, as well as the number and length of dendrites in the male hippocampus. The effects were consistent with those described by other (adult) chronic stress paradigms, with apical dendrites being affected more than basal. Furthermore, ELA decreased the expression of several neurogenesis markers, while the results on BDNF expression were not significant. This

should however not be considered as evidence of absence of an effect, especially because the unexplained heterogeneity was moderate.

Overall, Chapters 4 to 8 quantify stable effects of ELA on rodents' brain and behavior. Our approach is of systematic review and meta-analysis of the literature, and our conclusions are based on an entire body of literature. These chapters therefore meet this thesis' requirements because they integrate past information on the effects of chronic stress on brain and behavior. Perhaps more remarkably, these studies review the same population. Effectively, this means that we can provide a general overview of the effects of ELA. This "general" overview is limited to the outcomes gathered by us, but it is more comprehensive than any other review in the ELA field. We therefore reasoned that one of the great advantages of these studies was their integration, a meta integration. This concept ultimately materialized in Chapter 9. In Chapter 9, we experiment with a new way of integrating information: we provide a detailed yet broad overview of ELA. This chapter is a commentary, but it is supported by quantitative meta-analytical statements. With this approach, we extracted 7 principles of ELA in rodents, cumulatively based on nearly 700 experiments comparing control and ELA animals. These principles can be considered a discussion of Chapter 4 to 8, and can be read in detail in Chapter 9. Two points particularly stood out. First, our meta-analyses suggest that there may be interactions not only between early- and late-life events, but also with the acute state the animal is in. Although this conclusion may seem obvious, it is rarely considered when planning experiments. Furthermore, this observation can help unify theories of ELA (specifically, cumulative and match-mismatch theories) which are generally considered in direct opposition. Second, ELA was found to increase variation, favoring the vulnerable and/or resilient phenotype, at the expense of the 'average' phenotype. This was especially evident when animals did not experience any other life events, and therefore the genetic makeup becomes more important. For an in-depth discussion of these topics, see Principle #4 and #6 in Chapter 9. In all, Chapter 9 integrates meta-information about the effects of ELA on brain and behavior. Its conclusions increase our (theoretical) understanding of how ELA works, and provide valuable recommendations for future experiments.

Although the methodologies used in Part B follow the gold standard of the preclinical meta-analysis field, we gave our own methodological twist. We were the first to apply MetaForest[12] to rodent data. This is an exploratory approach to identify the most important moderators in the dataset (e.g. type of model used, species...). This data-driven technique adapts random forests (a machine learning algorithm) for meta-analysis by means of bootstrap sampling. MetaForest ranks moderators based on their influence on the effect size, and it can be a great advantage for rodents' studies, where the potential moderators are likely to be multiple, interacting, and with non-linear effects. Secondly, we were among the first to apply to animal literature a 3-level fixed effect model rather than the more common random effects. This model accounts

for the violation of the assumption of independency when the data is collected from the same animals, thereby improving the robustness of the conclusions drawn. Lastly, we developed interactive data visualizations so that any researcher could perform his/her own meta-analysis by selecting the population and outcomes of interest. Within the user-friendly interfaces developed, a wide variety of features can be selected, such as ELA models and their components (e.g. type, timing, predictability), age, sex, etc. With the information provided, the app returns relevant publications, forest, funnel and cumulative plots. These analyses are not confirmatory, but they are valuable because they can directly impact research practice. Specifically, they can be useful 1) to explore the literature, 2) to define new hypotheses, 3) to evaluate publication bias and replicability of the findings, and 4) to estimate realistic effect sizes on which to ground future research.

Conclusion Part B

The effects of ELA on behavior and neurobiology depend on a complex interaction between early-, late- life events, as well as the acute state of the animal. By using systematic reviews and meta-analyses, we can integrate information of the literature to identify robust conclusions. This led to the extraction of 7 principles of ELA, which not only give insight in what is currently known but can also guide future research.

The aim of Part C (Chapters 10 and 11) was to develop methodologies for information integration. The work of the previous chapters included in this thesis sparked two observations: 1) in rodents, control groups are in many aspects comparable to each other, and 2) making conclusions based on multiple sources is a scientist's daily job. The chapters included in this section transform these concepts into practical tools for scientists.

In Chapter 10, we show how historical control data can be integrated in new experiments to improve their statistical power. We estimated that animal experiments are generally severely (prospectively) underpowered, a conclusion based on 479 publications in the metabolism and neuroscience fields. This was true across a range of possible effect sizes likely to be found in literature. At best, 12.5% of animal experiments were found to be sufficiently powered, meaning that their prospective power was larger than 80%. Based on this observation, performing properly powered experiments would require an increase of sample size in the dozens. Clearly, this would not be a feasible solution, not only in terms of the space requirements and financial

costs but also due to the continuing efforts to reduce the number of animals used in research. We therefore proposed in Chapter 10 an alternative solution. Information of previous experiments can be integrated in new experiments using Bayesian priors, directly impacting the statistical power of animal experiments. Importantly, this is applied only to the control and not to the experimental group, since control animals can be assumed to belong to the same population (see Chapter 10 for an in-depth discussion on this important assumption). We validated this approach in a case study on the effects of ELA on spatial learning in adult male mice. The experimental data was gathered by aggregating data from single experiments that in principle shared the same design, but individually had low power. We founded a consortium (RELACS, Rodent Early Life Adversity Consortium on Stress) of 10 laboratories, which collectively used 275 animals to answer this specific question. To facilitate the implementation of the method, we developed *RePAIR*, an open-source web-based tool to apply Bayesian priors to new experiments. *RePAIR* can be used to 1) calculate prior parameters from the summary statistics of existing data, 2) perform sample sizes calculation and 3) execute analyses. It can also be used to visualize the (potential) heterogeneity between one's own previous acquired control data and control data of other laboratories. The strength of this approach is that it is intuitive, and it is a translation in statistical terms of assumptions already used in daily research practice. New experiments are usually planned based on information obtained in previous studies. Even though variations between strains or laboratories clearly exist, animal researchers often have similar expectations about how a control group "should respond". If this expectation is not met, a researcher would likely "not trust the data" and conclude that the experiment "did not work" or "needs to be better optimized". In all, Chapter 10 introduces an intuitive method for information integration, easily used by researchers conducting animal studies.

In Chapter 11, we summarized the software developed in the previous chapters. We developed three different types of software: 1) an R package (abc4d, Chapter 2), 2) a web-based statistical software (*RePAIR*, Chapter 10), and 3) various interactive data visualizations (Chapters 2 to 5). The R package can be used to preprocess (e.g. handling of missing values, batch effect correction, normalization and standardization) and analyze data from whole-brain immunohistochemistry experiments, from the networks to the single cell level, and with the possibility to include a time dynamic. *RePAIR* can be used to incorporate data from previous experiments into new ones, and walks the scientist through every step: prior specification, power calculation, and data analysis. Lastly, the interactive data visualizations allow anyone to directly interact with the data published within a research article, rather than with an excel file. Here, anyone can select features of interest, and the apps deliver visualizations of the data. This was created for the whole-brain data discussed in Chapter 2, for the stress-NL consortia data described in Chapter 3, and for several meta-analyses. In all, Chapter 11 meets this

thesis' aim because it summarizes the software developed in this thesis that can be used to integrate various forms of data.

Conclusion Part C

Information of previous control animal groups can be used to improve the statistical power of animal experiments. Open Software can be developed to aid scientists to integrate information, and as a knowledge utilization strategy for the smoother communication of scientific results.

In conclusion, in this thesis we integrated information related to the acute and chronic (i.e. ELA) stress response. This was achieved 1) by performing new experiments in rodents, 2) by founding (RELACS) and expanding the collaboration within (stress-NL) consortia, 3) by extensively and systematically reviewing the literature, and lastly 4) by providing data and scripts freely available online. It is a first step towards a comprehensive, systems-view of *stress*.

In the following sections, we focused on two main overarching themes within this thesis: Open Science and data re-use. For an in-depth discussion of the biological themes within the thesis, please see Chapter 9.

An Open Science approach to research

An important part of this thesis was founded in the Open Science approach. Open Science is an umbrella term to define practices aimed at increasing transparency, accessibility, and reproducibility of scientific research. Although the concept of Open Science in its modern connotation was already introduced in 1985[13], Open Science as a movement started to kick in with a series of events in the early 2000s, linked especially to Open Software[14,15]: e.g. the rising use of internet, the institution of creative common licenses[16], and the launch of cloud computing as part of Amazon Web Services. While these events are not directly linked to academia, they enabled and showed first-hand the power of sharing data, information and technology. During the same years, Plos Biology was launched as the first Open Access journal; Ioannidis published "Why most published research findings are false"[17]; and Open Source initiatives started to emerge, such as the chemistry challenge of producing the drug praziquantel as a single enantiomer[18]. These events respectively added to the historical context the elements of accessibility of scientific findings, problems of reproducibility, and importance of collaboration: the recipe of Open Science was ready. Open Science stands for knowledge that is transparent, accessible, shared and community

developed (reviewed in[19]). This was the environment in which I received my education, and I am proudly in the first generation of researchers for whom there is no "pre" to open science.

The values of open science can be used directly to improve the research method (Figure 1A[20,21]). When I started my research career in 2015, Open Science was a "learning by doing"; nowadays, several resources can aid researchers to integrate elements of Open Science within their research practice (e.g. Open Scholarship Knowledge Base[22] and R4E[23]). The Open Science toolkit includes – but is not limited to – 1) preregistration, 2) sharing of data and research tools, 3) preprints and open access publishing, 4) open peer review and 5) transparent research evaluation and reward. In the following paragraphs, I will discuss how I applied this toolkit on the work of this thesis, and what were its tangible advantages. I will focus on the items 1 and 2, since they can be applied directly by individuals, even at the PhD-level.

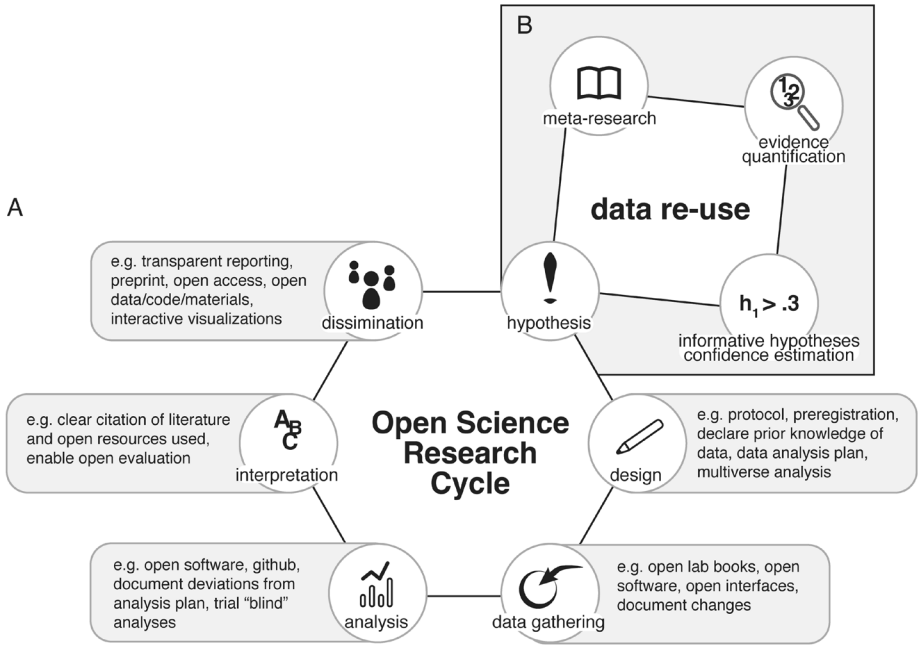


Figure 1. Open Science and the research cycle. A) Schematic overview of the research cycle with recommendation of research science practices to implement at each stage (grey boxes). The list is not exhaustive, and it has been partially adapted from[20,21,24]. B) Schematic overview of how in the future data re-use can help refining hypotheses, thereby reducing research waste. Briefly, literature and pre-existing data can be systematically addressed, text-mined, etc. Their evidence can be quantified, for example with meta-analyses or Bayesian evidence synthesis. In turn, this can be used to estimate the confidence (and therefore the suitability) in informative hypotheses. Formal models can also be built to generate plausible data. This *in silico* approach can then be used to refine our hypotheses and re-design our experiments.

Preregistration (item 1) describes the act to specify in advance (parts of) a scientific study and submit it to a registry before it is conducted[25,26]. The main goal of preregistration is to distinguish between exploratory (i.e. hypothesis generating) and confirmatory (i.e. hypothesis confirming) research. By respecting this distinction, scientists can directly recognize the impact of hindsight and confirmation bias[27,28]. More recently, it has been suggested that exploratory studies (e.g. fundamental preclinical studies) can also benefit of preregistration, in the form of registered reports[29]. This would consist of a log, where researchers would add (positive or negative) data step-by-step, as they proceed with the study. This log would capture the entire trajectory of a study, and justify the selection or omission of data in a final manuscript[29]. Others have suggested that preregistration makes a study neither better nor worse[30,31], yet reviewers could be inclined to (arbitrarily) value preregistered analyses higher[32] (reviewed in [33]). While preregistration can help in preventing bad research practices, it does not automatically incentivize good research practice[31].

While I tried to address all these important issues in my research, I purposely decided to not preregister any of my studies. While some have argued that Science is "show me" and not "trust me"[34], I believe that trust is an intrinsic part of the scientific process, which should be fostered rather than controlled. Besides publication bias, preregistration of exploratory research (such as the one mainly conducted in this thesis) is aimed to antagonize selective reporting, undisclosed analytical flexibility and insufficient study power[29]. My solution to these issues has been a thorough and detailed approach to transparent reporting. Wherever appropriate (e.g. meta-analyses, Chapters 4-8 and 10), protocols were prepared prior the beginning of the studies, and made available (but not 'preregistered') on their open science framework page. Of note, outlines of studies conducted in the Netherlands are generally publicly available due to grant applications and/or proposals for the Animal Ethics Committee. Additionally, we took care in meticulously reporting in our method sections any deviations from such protocols, as well as whether each question was confirmatory or exploratory. Importantly, we thoroughly specified why certain decisions were made and which methodologies were considered, even when developing analyses' tools ourselves (Chapter 2). Methods sections can therefore walk the reader through the process of scientific reasoning itself, showing how science rather than the procedure was performed. Concerning publication bias, all our studies have been posted on preprint services (i.e., item 3 of the Open Science toolkit) prior to their submission to academic journals. This is by no means a solution to the complex problem of publication bias, but rather a practical operation actionable by individual researchers and supportive in accelerating the progress of science. To me, the solution for improving research practice remains education, rather than preregistration.

Transparent reporting directly links to the second item of the Open Science toolkit, i.e. *sharing of data and research tools*. In line with transparent reporting, all studies

published in this thesis are associated to an open science framework page, where all data, code and materials (e.g. protocols) are available. An exception is Chapter 3, whose data is available via the stress-NL database. Our scripts are shared with version control and can be used to reproduce analyses from the raw files until the final manuscripts' figures. Furthermore, our latest projects were public since their commencement, rather than just upon submission.

Sharing data and tools exceeds transparent reporting: sharing enables re-use, directly impacting the value of research. While many have criticized data sharing due to the logistic, economic and technical burden alongside the challenge of appropriately attributing credit (for example,[35–37]), others have argued that the added value of data sharing overweighs the costs[38]. Specifically, data sharing enables re-mixing and combining, re-analysis with new methods, hypothesis generation, meta-analysis and bias minimization[38]. This thesis would not exist if others would not have shared their data and work. Chapter 2 would not have been possible without the freely available Allen Brain Atlas. Chapter 3 and 10 would have been impossible without the collaboration of several laboratories, respectively organized in the stress-NL and RELACS consortia. Chapter 4 to 9 would have been impossible without data from publications, open software such as *WebPlotDigitalizer* and R packages such as metafor[39] and metaForest[12]. While we were the main researchers on these projects, we relied on a strong scientific community. For this reason, we aimed to pay our debt forward: we shared our data and codes, and developed software for others to use (Chapter 11). Specifically, while the development of software has been essential to answer our research questions, it might facilitate knowledge utilization in future research. Besides the analyses developed (*abc4d* and *RePAIR*, in Chapter 2 and 10 respectively) that can be directly re-used and re-purposed, our interactive data visualizations foster transparency and serve a general purpose of communicating our own research. This initiative can facilitate scientists to use and explore the data, without the burden of downloading and processing it. As the number of available publications increases by the hour, scientists require ways to summarize and re-use information that go beyond reading publications one-by-one. Sharing of data and tools (as developed and implemented in this thesis) can therefore foster collaboration and knowledge dissemination among researchers. It can decrease research waste and improve the design of new experiments. This is particularly valuable in animal research: we support re-use of historical animal data as an animal-free innovation, while realizing that exploratory studies that propel science forward may remain necessary. Although this aspect of knowledge dissemination targets scientists directly, it has a wider impact for society since it aims to improve research practices and the way in which scientists make conclusions.

Open Science has and will revolutionize our approach to conduct and report scientific research. I tentatively drafted how it could also impact the research cycle

itself (Figure 1B). Specifically, I view that in the future open data can be systematically re-used to appropriately integrate information from multiple sources, to quantify the collective knowledge on (opposing) theories, and ultimately to estimate the confidence in our deriving (formal) hypotheses. *RePAIR* (Chapter 10) is a first step in this direction. The *pros*, *cons* and *hows* of this possible future are further addressed in the next section. Any time is good to incorporate Open Science within one's practice. For example, the next step I will take to improve my scientific practice is in conducting multiverse analyses. Rather than conducting just one analysis, multiverse analysis[40] consists in preprocessing and conducting all "plausible analyses" which correspond to a large set of "reasonable scenarios". The advantage of multiverse analysis is to identify results that are stable across analysis' choices. In all, a cardinal question that emerged from this thesis has been "do I need to perform an experiment to answer this question?". Perhaps meta-research, collaboration, strong logic, and intuition can bring us much further along the way of discovery.

Data re-use: approach and limitations

Research data is generally crafted for specific research purposes[41], and it is therefore liable to local and historical artefacts[42,43]. Once the data is removed from its context, there is an inherent information loss[44]. Besides the issues inherently related to research practice described in the previous section, data re-use has an additional caveat: prior knowledge of the data. Prior knowledge of the data increases the risk of bias, meaning that it could push the researchers in pursuing a specific research idea[24]. This would be the case not only if scientists directly pre-test a hypothesis, but also if they test related variables, as it is often the case in e.g. cohort and longitudinal studies[24]. Furthermore, data re-use necessarily leads to increased analytical flexibility, since its core value is that the same dataset is analyzed multiple times. Analytical flexibility describes the multiple choices a scientist needs to take to preprocess and analyze the data[45]. These choices can lead to contrasting conclusions, as showcased in the neuroimaging field[46]. Indeed, the rate of false positive findings increases with the number of analyses conducted[47]. In other words, the probability of finding important associations in the data decreases over time, a concept referred to as dataset decay[47]. These concerns need to be critically addressed when re-using research data: while meta-data and ontologies can facilitate data integration[48] and changing alpha value can delay dataset decay[47], whether data integration is appropriate needs to be evaluated on a case-by-case basis. Robust and transparent methodology and analysis are therefore essential.

Re-using (others') data also means being dependent on it, specifically on its availability and quality. Besides the possible influences of publication bias, meta-analyses are dependent on the studies that have been performed in the literature. For example, a recurrent limitation of our analyses was that female data was consistently

insufficient for a meta-analytical quantification. Although we tried to evaluate potential sex-differences (see Chapters 4 to 9), it remains to be established whether the sex-differences observed are due to lab-specific effects (discussed in Chapter 9), meaning that laboratories experimenting with females have sub-group characteristics that distinguish them from other laboratories. Another example is the behavioral test "object-in-context task", which has been used in Chapter 4 as an indicator memory after non-stressful learning. It is generally assumed that rodents have a preference for novel over familiar objects when in non-stressful circumstances[49]. Accordingly in the object-in-context task, a rodent is considered to have a better memory if it spends more time exploring the object-context combination that is novel over the one that is familiar (i.e., from the sample phase). Recently, this interpretation has been challenged: the preference of the novel or familiar object may underlie a behavioral strategy, rather than merely providing a measure for memory[50]. In Chapter 4, we concluded that ELA decreases memory after non-stressful learning. This result was robust, but the analyses were based on the original interpretation of the object-in-context task. Meta-analyses are dependent on the availability of the data: if an entire field uses a faulty interpretation or paradigm, the conclusions of meta-analyses can be reproducible, but scientifically wrong. Availability aside, the quality of the meta- or re- analysis is also dependent on the quality of the included studies. This concept is often referred to as *garbage-in, garbage-out*. Meta-analyses try to address this issue 1) by weighting the effects e.g. based on their precision (i.e. inverse of variance), 2) by assessing the risk of bias and 3) by systematically reviewing a large body of literature. In this thesis, we used SYRCLE's tool to assess the risk of bias in animal studies[51]. Of the hundreds of papers reviewed in this thesis, none reported on all items. Therefore, the current risk of bias tool for rodent studies often led to inconclusive judgements concerning the possible "garbage" in the data. Notably, reporting has markedly improved in recent years presumably due to the adoption of guidelines[52–54], which will facilitate bias assessment in the future. In all, as the quality and quantity of the primary literature increases, the strength of meta- / re-analyses' conclusions will improve.

In this thesis, we re-use data in 4 contexts: 1) atlases (Chapter 2), 2) consortia (Chapter 3 and 10), 3) literature (Chapter 4 to 9), and 4) historical controls (Chapter 10). In the following paragraphs, we address the contexts of 1 to 3, with a focus on their limitations. The re-use of data as historical controls (context 4) has been thoroughly discussed in Chapter 10, where we direct the reader interested in further reading.

Atlases. Concerning atlases, in Chapter 2 we use the Allen Brain Reference Atlas (ABA)[55] to align and annotate cells that expressed the protein c-fos, and we used c-fos expression values reported by the in-situ hybridization ABA experiments for one of our control analyses. Although atlases enable different researchers to work in a common space, we are assuming that there are negligible study-specific differences between our and the ABA studies. Estimating the extent of this assumption is specifically

challenging because of the limited information available on the animals used for the atlas. The ABA reference atlas in its current (after 2017, not previous) version is composed by averaging 1675 brains from C57bl/6 mice[55], grouped housed and older than P54 (i.e., adult). Of these, about a third is females. While the authors state that there were no sex differences[55], others have reported differences in volume of brain areas across sexes in mice[56]. Furthermore, strain-specific effects have also been observed in genes' expression levels[57]. As a consequence, the impact of using the ABA atlas while experimenting on a different population (e.g. sex, strain, age) remains to be established. In Chapter 2, we used male mice of the same strain and age, therefore the use of the ABA reference atlas seemed appropriate.

Consortia. The consortia data we collected was at the individual participant level, both for Chapter 3 and 10. Individual participant data (IPD) has several advantages, such as using a consistent unit and method of analysis, handling missing values uniformly, and greater statistical power to examine interactions with participant-level covariates. However, analyzing IPD data as if it were from a single experiment would ignore differences in study-specific characteristics, e.g. protocols[58]. As a consequence, there would be no insight in the heterogeneity of results between studies and one may inadvertently encounter ecological bias. Previous analyses have reported that endocrine data may be specifically difficult to aggregate[4], outside of a strict IPD meta-analysis context. In Chapter 3, we provide example analysis on the effects of acute stress on cortisol concentrations, using the stress-NL database. While we do not perform a formal analysis, our proof-of-concepts were in line with 2-stages IPD meta-analyses. Specifically, we first evaluated differences between males and females within each study (first stage), and we then compared the estimates across studies (second stage), thereby addressing potential study-specific effects. Conversely, in Chapter 10, we used data from the RELACS consortium to obtain a dataset sufficiently large to validate our *RePAIR* statistical method. Here, we simply aggregated data as if it were a single experiment. By doing so, we assumed that all participating laboratories were studying the same underlying effects. In our specific application, this assumption seemed appropriate because 1) the studies were extremely similar to each other, and 2) the aim of the consortium was to gather a sufficiently large dataset rather than estimating an effect size or evaluating moderating mechanisms.

Literature. The systematic reviews and meta-analyses of literature reported in this thesis adhered to the state-of-the-art methodology in the field. While systematic reviews are routine in clinical research (e.g. Cochrane), they became common in animal research only recently (for a review, see[59]). Specifically, systematic reviews can be used in a translational context to assess both the external and internal validity of prespecified research findings[60]. Systematic reviews apply a structured approach to identify and assess all data relevant to a specific research question, with the intent to minimize bias[60]. They can be followed by meta-analyses, a statistical process that

quantitatively compares and summarizes the separate studies. In clinical practice the main goal of meta-analysis is to estimate an effect size; conversely, in animal studies meta-analysis is most relevant for the exploration of heterogeneity[60], i.e. for the identification of those aspects of the experimental design that enhance/diminish the effect of the intervention investigated. These advantages can be translated in research questions such as: "In rodents, does X alter Y? And does Z moderate the effect of X on Y?". For example, in my thesis: "In rodents, does ELA alter structural plasticity (Chapter 8)? Does the experience of other negative life events moderate this effect?". These questions (and systematic reviews in general) have the characteristic of addressing only one specific outcome. While it is possible to perform meta-analyses of multiple outcomes when the correlations between outcomes are known, between-outcomes correlations are often not reported in preclinical literature. In my opinion, there is a discrepancy between the research questions a fundamental scientist is interested in, and the research question that can actually be asked with current meta-analyses techniques. For example, a fundamental scientist could ask "Which part of the (e.g.) dopaminergic system is most sensitive to the effects of ELA? And which parts change as an adaptation mechanism?" rather than "does X impact Y". In this thesis, we addressed this issue in Chapter 9 by qualitatively integrating the ELA effects of the brain and behavior, where we identified 7 emerging principles of ELA in rodents. In the future, I view the opportunity of developing a methodology to achieve the same quantitatively, rather than qualitatively. While overcoming the lack of between-outcomes' correlations is challenging, this could be achieved by combining elements of meta-analysis with elements of Bayesian evidence synthesis and of network theory. For example, Bayesian evidence synthesis could be used to integrate data from multiple sources (e.g. expert opinion) for those variables for which meta-analytic data is not (yet) available. Network theory could be fostered to identify which nodes (i.e. outcomes) are most important, given the meta-analyses and Bayesian evidence synthesis data. Such methodologies would be game changing, as they would foster the power of old experiments also for fundamental research purposes. The natural next step of these "networks" of evidence on (e.g.) ELA would be to be sustained by the entire community, e.g. in the form of living reviews. Scientific societies could play a role in moderating these discussions, and in leading the change to a cumulative and sustainable science. Our collective knowledge could then be centralized, updated, and at everyone's disposal.

Conclusion

Understanding the stress response as a system, rather than a collection of organs and hormones, brings us closer to comprehend how it actually works in nature, in health and for future treatment or prevention of disease. This thesis made an attempt to integrate information in the stress field: information that was already available with new information; human data with animal data; data from systems to cells; data on behaviour with data on potential neurobiological mechanisms. We used an open science approach to achieve this integration. While many of the details remain unknown, the work of this thesis has highlighted that collaboration, transparency, reproducibility, and (meta)analysis can support us while we re-learn how to research.

References

1. Hermans, E. J., Henckens, M. J. A. G., Joëls, M. & Fernández, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences* 37, 304–314 (2014).
2. Berretz, G., Packheiser, J., Kumsta, R., Wolf, O. T. & Ocklenburg, S. The brain under stress-A systematic review and activation likelihood estimation meta-analysis of changes in BOLD signal associated with acute stress exposure. *Neurosci Biobehav Rev* 124, 89–99 (2021).
3. Miller, R. & Plessow, F. Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations. *Psychoneuroendocrinology* 38, 941–946 (2013).
4. Miller, R. et al. The CIRCORT database: Reference ranges and seasonal changes in diurnal salivary cortisol derived from a meta-dataset comprised of 15 field studies. *Psychoneuroendocrinology* 73, 16–23 (2016).
5. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nat Rev Neurosci* 17, 652–666 (2016).
6. Gage, S. H., Munafò, M. R. & Davey Smith, G. Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research. *Annu. Rev. Psychol.* 67, 567–585 (2016).
7. Tost, H., Champagne, F. A. & Meyer-Lindenberg, A. Environmental influence in the brain, human welfare and mental health. *Nat Neurosci* 18, 1421–1431 (2015).
8. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr Opin Psychol* 15, 174–181 (2017).
9. Gatzke-Kopp, L. M. The canary in the coalmine: The sensitivity of mesolimbic dopamine to environmental adversity during development. *Neuroscience & Biobehavioral Reviews* 35, 794–803 (2011).
10. Money, K. & Stanwood, G. Developmental origins of brain disorders: roles for dopamine. *Frontiers in Cellular Neuroscience* 7, (2013).
11. Okuno, H. Regulation and function of immediate-early genes in the brain: Beyond neuronal activity markers. *Neuroscience Research* 69, 175–186 (2011).
12. Lissa, C. J. van. Small Sample Meta-Analyses: Exploring heterogeneity using MetaForest. in *Small Sample Size Solutions* (Routledge, 2020).
13. Chubin, D. E. Open Science and Closed Science: Tradeoffs in a Democracy. Science, *Technology, & Human Values* 10, 73–80 (1985).
14. Voytek, B. Social Media, Open Science, and Data Science Are Inextricably Linked. Neuron 96, 1219–1222 (2017).
15. Tennant, J. et al. A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship. (2020) doi:10.31235/osf.io/2kxq8.
16. History - Creative Commons. https://web.archive.org/web/20111007165253/http://creativecommons.org/about/history (2011).
17. Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLOS Medicine* 2, e124 (2005).
18. Woelfle, M., Olliaro, P. & Todd, M. H. Open science is a research accelerator. *Nature Chem* 3, 745–748 (2011).
19. Vicente-Saez, R. & Martinez-Fuentes, C. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research* 88, 428–436 (2018).
20. Szkuta, K. & Osimo, D. Exploring the implications of science 2.0 dynamics for R&I policies. *Technical Report* (2012) doi:10.13140/RG.2.2.35128.03842.
21. Munafò, M. R. et al. A manifesto for reproducible science. *Nat Hum Behav* 1, 1–9 (2017).
22. Center for Open Science. Open Scholarship Knowledge Base. https://www.cos.io/communities/open-scholarship-knowledge-base.
23. Auer, S. et al. A community-led initiative for training in reproducible research. *Elife* 10, e64719 (2021).
24. Baldwin, J. R., Pingault, J.-B., Schoeler, T., Sallis, H. M. & Munafò, M. R. Protecting against researcher bias in secondary data analysis: challenges and potential solutions. *Eur J Epidemiol* (2022) doi:10.1007/s10654-021-00839-0.
25. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. & Mellor, D. T. The preregistration revolution. *PNAS* 115, 2600–2606 (2018).
26. Peirce, C. S. A theory of probable inference. in *Studies in logic by members of the Johns Hopkins University* 126–181 (Little, Brown and Co, 1883). doi:10.1037/12811-007.
27. Nuzzo, R. How scientists fool themselves – and how they can stop. *Nature* 526, 182–185 (2015).
28. Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J. & Kievit, R. A. An Agenda for Purely Confirmatory Research. *Perspect Psychol Sci* 7, 632–638 (2012).
29. Dirnagl, U. Preregistration of exploratory research: Learning from the golden age of discovery. *PLOS Biology* 18, e3000690 (2020).
30. Lakens, D. The value of preregistration for psychological science: A conceptual analysis. JSage 62, 221–230 (2019).
31. Morey, R. You must tug that thread: why treating preregistration as a gold standard might incentivize poor behavior. *Psychonomic Society Featured Content* https://featuredcontent.psychonomic.org/you-must-tug-that-thread-why-treating-preregistration-as-a-gold-standard-might-incentivize-poor-behavior/ (2019).
32. Oberauer, K. Preregistration of a forking path – What does it add to the garden of evidence? *Psychonomic Society Featured Content* https://featuredcontent.psychonomic.org/preregistration-of-a-forking-path-what-does-it-add-to-the-garden-of-evidence/ (2019).

33. Toth, A. A. *et al. Study* preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology* No Pagination Specified-No Pagination Specified (2020) doi:10.1007/s10869-020-09695-3.
34. Stark, P. B. Before reproducibility must come preproducibility. *Nature* 557, 613–613 (2018).
35. Tenopir, C. et al. Data sharing by scientists: practices and perceptions. *PLoS One* 6, e21101 (2011).
36. Berman, F. & Cerf, V. Science priorities. Who will pay for public access to research data? *Science* 341, 616–617 (2013).
37. Sterling, T. D. & Weinkam, J. J. Sharing scientific data. *Commun. ACM* 33, 112–119 (1990).
38. Voytek, B. The Virtuous Cycle of a Data Ecosystem. *PLOS Computational Biology* 12, e1005037 (2016).
39. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 36, 1–48 (2010).
40. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci* 11, 702–712 (2016).
41. Pasquetto, I. V., Borgman, C. L. & Wofford, M. F. Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review* 1, (2019).
42. 'Raw Data' Is an *Oxymoron*. (MIT Press, 2013).
43. Star, S. L. & Griesemer, J. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 387–420 (1989).
44. Leonelli, S. *Data-Centric Biology: A Philosophical Study.* (University of Chicago Press, 2016).
45. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychol Sci* 22, 1359–1366 (2011).
46. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582, 84–88 (2020).
47. Thompson, W. H., Wright, J., Bissett, P. G. & Poldrack, R. A. Dataset decay and the problem of sequential analyses on open datasets. *eLife* 9, e53498 (2020).
48. Jones, M. B., Schildhauer, M. P., Reichman, O. J. & Bowers, S. The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. *Annu. Rev. Ecol. Evol. Syst.* 37, 519–544 (2006).
49. Ennaceur, A. One-trial object recognition in rats and mice: methodological and theoretical issues. *Behav Brain Res* 215, 244–254 (2010).
50. Sep, M. S. C., Vellinga, M., Sarabdjitsingh, R. A. & Joëls, M. The rodent object-in-context task: A systematic review and meta-analysis of important variables. *PLOS ONE* 16, e0249102 (2021).
51. Hooijmans, C. R. et al. SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology* 14, 43 (2014).
52. NPQIP Collaborative group. Did a change in Nature journals' editorial policy for life sciences research improve reporting? *BMJ Open Sci* 3, e000035 (2019).
53. Sert, N. P. du et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biology* 18, e3000410 (2020).
54. Ramirez, F. D. et al. Methodological Rigor in Preclinical Cardiovascular Studies: Targets to Enhance Reproducibility and Promote Research Translation. *Circ Res* 120, 1916–1926 (2017).
55. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953. e20 (2020).
56. Qiu, L. R. et al. Mouse MRI shows brain areas relatively larger in males emerge before those larger in females. *Nat Commun* 9, 2615 (2018).
57. Morris, J. A. et al. Divergent and nonuniform gene expression patterns in mouse brain. *PNAS* 107, 19049–19054 (2010).
58. Tierney, J. F. et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLOS Medicine* 12, e1001855 (2015).
59. Luijk, J. van et al. Systematic Reviews of Animal Studies; Missing Link in Translational Research? *PLOS ONE* 9, e89981 (2014).
60. Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R. & Howells, D. W. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. *J Cereb Blood Flow Metab* 34, 737–742 (2014).

Appendix

Nederlandse samenvatting

Er zijn maar weinig ervaringen die zoveel impact hebben als de subjectieve ervaring van stress. Stress beïnvloedt meerdere biologische systemen en is zowel in de tijd als in de ruimte specifiek. Het hangt af van genetische aanleg, vroege en late levensgebeurtenissen, en het leidt tot verschillende reacties afhankelijk van het type (b.v. psychologisch vs. fysiologisch) en de frequentie (b.v. acuut vs. chronisch). Dit complexe systeem werd in het verleden bestudeerd door elk specifiek kenmerk in een experiment te isoleren. Hoewel deze benadering zeer succesvol is geweest, kan het begrijpen van stress als een systeem, in plaats van een verzameling organen en hormonen, ons dichter bij het begrijpen van hoe het werkt in de natuur, in gezondheid en voor de toekomstige verbetering van ziekte brengen. In dit proefschrift hebben we stress onderzocht als een biologisch systeem in plaats van een specifieke eigenschap te isoleren in een enkel experiment. In het bijzonder integreerden we informatie uit meerdere bronnen (d.w.z. literatuur, consortia, atlassen, en nieuw gegenereerde gegevens) om ons begrip te vergroten van de effecten van acute en chronische stress ervaren in het vroege leven op hersenen en gedrag. Onze aanpak was volledig gebaseerd op de Open Science praktijken van samenwerking, het delen van gegevens en code, evenals in softwareontwikkeling.

Na een samenvatting van de belangrijkste bevindingen van dit proefschrift, gaat dit hoofdstuk verder met een algemene bespreking van twee overkoepelende thema's, namelijk Open Science en het hergebruik van gegevens. We hebben ons specifiek gericht op de manier waarop deze benaderingen in dit proefschrift zijn verwerkt, hun beperkingen en wat onze bevindingen zouden kunnen betekenen voor toekomstig onderzoek.

Samenvatting en contextualiseren van de voornaamste bevindingen

Dit proefschrift is onderverdeeld in drie delen (deel A tot en met C), die elk betrekking hebben op een van de doelstellingen van dit proefschrift:

A. Het integreren van informatie met betrekking tot de gezonde acute stressrespons (bij knaagdieren en mensen);
B. Het integreren van informatie met betrekking tot chronische stress die vroeg in het leven wordt ervaren (bij knaagdieren);
C. Het ontwikkelen van methodologieën voor informatie-integratie.

Het doel van **Peel A** was het integreren van informatie met betrekking tot de gezonde acute stressrespons. We hebben twee studies uitgevoerd die zich richten op de hersenen van knaagdieren (Hoofdstuk 2) en op de cortisolconcentratie in speeksel bij mensen (Hoofdstuk 3). Hoewel de studies zijn uitgevoerd in twee verschillende diersoorten, onderzoeken ze beide acute stress als een dynamisch proces dat in de tijd plaatsvindt.

Hoofdstuk 2 onderzoekt de effecten van een specifiek type acute stress (i.e., voetschok) op de gehele hersenen, in plaats van op vooraf gespecificeerde hersengebieden (i.e., regio's waar de onderzoeker in geïnteresseerd is). Acute stress leidt tot de opeenvolgende activatie van functionele hersennetwerken, zoals aangetoond in verschillende humane studies met behulp van functionele magnetische resonantie imaging (fMRI, voor reviews[1,2]). Wij hebben dit belangrijke aspect van het acuut gestresste menselijke brein nagebootst, maar nu voor het eerst in muizen. In plaats van fMRI gebruikten we immunohistochemie van de hele hersenen vanwege de superieure ruimtelijke resolutie, waaraan we een pseudo-tijd metriek toevoegden. In ons experiment is in 96% van alle onderzochte hersengebieden een verhoogde expressie van c-fos (marker van cellulaire activiteit) te zien. De hypothalamische gebieden bleken het meest actief en werden ook als eerste geactiveerd, gevolgd door de amygdala-, prefrontale, hippocampus- en ten slotte thalamische kernen. Belangrijk is dat we verder konden gaan dan de ruimtelijke resolutie van circuits op welke schaal we de menselijke bevindingen repliceerden, en we konden inzoomen op afzonderlijke cellen. Hierdoor konden we verschuivingen identificeren binnen - en ook tussen - hersengebieden in de tijd na stress, wat we illustreerden voor het geval van de basolaterale amygdala. Bovendien, terwijl sommige hersengebieden een toename van het aantal c-fos+ cellen vertoonden, verhoogden andere dramatisch de c-fos intensiteit in slechts een subset van cellen, wat doet denken aan engrammen. Deze "strategie" veranderde na de voetschok in de helft van de hersengebieden. Gedurende het hele project hebben we onze analyses uitgevoerd met in gedachten reproduceerbaarheid als een belangrijk principe, en we hebben het gebruik ervan door anderen vergemakkelijkt door een R-package te ontwikkelen. De meeste van onze analyses testten de robuustheid van onze (exploratieve) bevindingen over de verschillende experimenten heen, naast het evalueren van hun "statistische significantie". Deze studie moest in afzonderlijke experimenten worden uitgevoerd om de haalbaarheid van de behandeling van het materiaal mogelijk te maken. Naast het gebruik van een blokdesign (i.e. een dier van elke experimentele groep is gerandomiseerd binnen een blok) om de robuustheid van het experiment te behouden, zagen wij de verschillende experimenten als pseudo-replicaten. Wij hebben ernaar gestreefd zo transparant mogelijk te zijn in de rapportage van onze methodologie: de aanvullende informatie bevat gedetailleerde lijsten van alle analyses die zijn overwogen om elke vraag te beantwoorden, en waarom uiteindelijk voor een specifieke aanpak is gekozen. Wij beschouwen dit als een van de kernwaarden van Open Science: de lezer door het proces van het wetenschappelijk denken leiden, in plaats van alleen het resultaat als een ultiem antwoord te geven. Al met al voldoet Hoofdstuk 2 aan de doelstelling van dit proefschrift door 1) informatie te integreren over meerdere hersengebieden en ruimtelijke resoluties in de tijd, en 2) te leiden tot de ontwikkeling van een R-pakket om toekomstige analyses van hele-hersen data te vergemakkelijken.

Het doel van Hoofdstuk 3 was om toekomstige integratie van informatie met betrekking tot de acute stressrespons (in een gecontroleerde laboratorium omgeving) van mensen te vergemakkelijken. In dit hoofdstuk hebben we de stress-NL database geïntroduceerd. Deze database is ontstaan uit een samenwerkingsinitiatief van 12 Nederlandse onderzoeksgroepen, die hebben samengewerkt om te komen tot een nauwkeurige inventarisatie van (neuro)biologische, fysiologische en gedragsgegevens uit laboratorium-gebaseerde humane studies die gebruik maakten van acute stress testen. Wij geven voorbeeldanalyses van de bruikbaarheid van de gegevens, specifiek gericht op de cortisolconcentratie na stress. In de eerste proof-of-concept analyse beschrijven we het verschil in effectgrootte tussen mannen en vrouwen in de cortisolconcentratie na acute stress, gemeten als oppervlakte onder de curve met betrekking tot toename (AUCi). In 18 van de 23 studies die aan de criteria voor deze analyse voldeden, hadden mannen een consequent grotere AUCi dan vrouwen, hoewel de effectgroottes sterk varieerden (van d = 0,1 tot d = 1,1). In de tweede analyse vergeleken we opnieuw mannen en vrouwen, maar deze keer werd onderscheid gemaakt tussen vrouwen die orale anticonceptie gebruiken en vrouwen die geen orale anticonceptie gebruiken. De resultaten suggereren dat het gebruik van orale anticonceptiemiddelen de cortisolpiekwaarde bij vrouwen dempt. Tot onze verrassing voldeden slechts 4 experimenten in de stress-NL database aan de criteria voor deze analyse, i.e., dat er voldoende deelnemers waren die geen orale anticonceptiva gebruikten. De kwalitatieve beschrijving van de databank toont het potentieel, maar ook de uitdagingen van de aanpak. De database bevat individuele deelnemersinformatie van meer dan 5500 deelnemers (41% vrouw), in de leeftijd van 6 tot 99 jaar. Het type acuut stressparadigma, de uitkomsten en het tijdstip van meting waren slechts gedeeltelijk overlappend over de verschillende experimenten. De heterogeniteit van de experimenten binnen de stress-NL databank weerspiegelt de heterogeniteit van de literatuur. Hoewel de database in principe re-, meta-, en proof-of-principle analyses mogelijk maakt, is de volgende uitdaging analytisch. Eerder onderzoek heeft uitgewezen dat aggregatie van gegevens niet altijd geschikt is voor neuro-endocriene gegevens [3,4]. Toekomstige analyses moeten daarom worden gebaseerd op de modernste gegevensanalyse voor individuele deelnemers of op 'Bayesian evidence synthesis' van bewijsmateriaal. De belofte van de flexibiliteit van de database gaat samen met de toegenomen statistische expertise. Alles bij elkaar genomen, beschrijft Hoofdstuk 3 de nieuwe stress-NL database, die toekomstige integratie van informatie met betrekking tot de menselijke acute stress respons zal vergemakkelijken.

Conclusie van Deel A

De studies van Deel A illustreren dat:

1) acute blootstelling aan stress de cellulaire activiteit in veel gebieden van de

knaagdierhersenen verandert, met patronen die overeenkomen met eerdere studies naar de menselijke hersenen;

2) het combineren van bestaande datasets over acute stress bij mensen het mogelijk maakt om patronen in de neuro-endocriene respons te onderscheiden, een aanpak die ook nuttig zou kunnen zijn bij dieronderzoek.

Ongeacht of de integratie van informatie wordt toegepast op dierlijke of menselijke gegevens, de integratie is het meest doeltreffendst wanneer zij gericht is op het onderzoeken van de stabiliteit en de reproduceerbaarheid van effecten.

Het doel van **Deel B** (Hoofdstuk 4 tot 9) was om informatie te integreren die verband houdt met chronische stress die knaagdieren vroeg in hun leven ondervinden. Tegenslagen op jonge leeftijd kunnen langdurige gevolgen hebben voor de ontwikkeling van de hersenen, en zijn een van de belangrijkste risicofactoren voor verschillende psychische stoornissen. Hier geven we een kwantitatieve beschrijving van de effecten van vroeg in het leven ervaren tegenspoed (ELA) op gedrag en neurobiologie. De aanpak bestond uit het systematisch samenvatten van de literatuur (systematic review) en het meta-analyseren van de literatuur. We richtten ons op langdurige veranderingen van ELA, gemeten in de volwassenheid. Hoewel elk hoofdstuk van Deel B zich richt op een andere groep uitkomsten, hebben we de aanpak consistent gehouden: zoekcriteria, in- en exclusiecriteria en analyses zijn vergelijkbaar tussen de hoofdstukken. De enige uitzondering is Hoofdstuk 5, dat zowel prenatale als postnatale ELA-modellen omvat, zoals hieronder nader zal worden toegelicht.

In Hoofdstuk 4 hebben we een oorzakelijk verband aangetoond tussen ELA en veranderingen in het gedrag van volwassenen bij muizen en ratten. Epidemiologische studies bij mensen brengen het ervaren van (verschillende vormen van) vroege tegenspoed in verband met verschillende psychische stoornissen, met name angst, depressie, suïcidepogingen en middelenmisbruik[5]. Deze studies zijn echter vaak gebaseerd op prospectieve of retrospectieve observationele gegevens. Als gevolg van verstorende elementen en omgekeerde causaliteit die inherent zijn aan observationele gegevens, moeten causale gevolgtrekkingen over de effecten van ELA op gedragsveranderingen bij volwassenen nog worden vastgesteld[6] bij mensen. Daarom hebben velen hun toevlucht genomen tot dierstudies[7] , waarbij genetische invloeden kunnen worden gecontroleerd en ELA experimenteel kan worden geïnduceerd[8]. De literatuur over ELA bij knaagdieren is echter uiterst heterogeen: een systematisch onderzoek van de literatuur was van essentieel belang om te bepalen of de paradoxale effecten van ELA op het gedrag te wijten waren aan echte biologische verschillen (bv. robuust in de verschillende gedragstests) of dat de verschillen tussen de studies te wijten waren aan methodologische factoren (bv. type ELA-model, diersoort, enz.). We concludeerden dat ELA, hier gedefinieerd als verandering van moederlijke zorg, angst

verhoogt, het geheugen verbetert na stressvol leren, het geheugen vermindert na niet-stressvol leren en sociaal gedrag belemmert, vooral bij mannetjes. De effecten werden verder versterkt door andere negatieve levenservaringen ("hits"). Alles bij elkaar levert Hoofdstuk 4 robuust bewijs voor de causale effecten van ELA op gedrag, door de informatie van 212 onafhankelijke publicaties te integreren.

In de Hoofdstukken 5 tot en met 8 worden verschillende aspecten van de neurobiologische effecten van ELA bij knaagdieren onderzocht. In Hoofdstuk 5 onderzoeken we de effecten van prenatale en postnatale ELA op het dopaminerge systeem. Convergerende aanwijzingen suggereren dat het dopaminerge systeem betrokken is bij het mediëren van de invloed van ELA op kwetsbaarheid voor psychopathologie[9]: van de ontwikkelingsperiode tot de sterke associatie met verschillende psychische stoornissen[10]. In dit hoofdstuk hebben we 90 knaagdierpublicaties geanalyseerd om na te gaan of ELA inderdaad langdurige veranderingen in het dopaminerge systeem veroorzaakt. De uiteindelijke dataset bevatte 41 publicaties die prenatale ELA onderzochten, en 49 publicaties die postnatale ELA onderzochten. Deze werden geanalyseerd in twee afzonderlijke analyses, omdat ze betrekking hebben op verschillende kritische perioden. Tot onze verrassing waren de effecten vrij beperkt. Het striatum was het hersengebied dat het meest beïnvloed werd, vooral in de prenatale dataset. Wij hebben deze bevindingen verder uitgewerkt in Hoofdstuk 6, waarin wij een analyse hebben opgenomen van de expressie in de hersenen van andere monoamines, i.e., serotonine en noradrenaline. Hier hebben we strengere inclusiecriteria gehanteerd dan in Hoofdstuk 5 (maar identiek aan de Hoofdstukken 4, 7 en 8), met de bedoeling de heterogeniteit van de populatie te verminderen en zo mogelijk effecten aan het licht te brengen die in onze eerdere studie verborgen bleven. Uiteindelijk gebruiken we 47 publicaties die de effecten van ELA onderzochten, in dit hoofdstuk beperkt tot veranderingen in de zorg voor de moeder, op biochemische indicatoren van monoaminerge systemen. Onze kwantitatieve analyse richtte zich op mannetjes, vanwege het beperkte aantal experimenten uitgevoerd bij vrouwelijke knaagdieren. Zelfs met de meer homogene populatie, bevestigden we onze twee eerdere bevindingen uit Hoofdstuk *5*: 1) de meta-analytische veranderingen geïnduceerd door ELA op monoaminerge systemen zijn beperkt, en 2) de effecten van ELA op het dopaminerge systeem zijn het duidelijkst in het striatum. Bovendien suggereerden onze resultaten dat de prefrontale cortex het hersengebied is dat het meest gevoelig is voor ELA effecten op het serotonerge systeem. Terwijl de effecten op dopamine het duidelijkst waren op het niveau van de receptoren, waren de effecten op serotonine het duidelijkst op het niveau van de concentratie van serotonine en zijn metaboliet 5HIAA. Tenslotte werd noradrenaline onderzocht met een opmerkelijk laag aantal publicaties, dit brengt duidelijke gat in de literatuur aan het licht. In Hoofdstuk 6 reproduceren we de bevindingen over de effecten van ELA op het dopaminerge systeem, en breidt deze uit tot andere monoamines.

In Hoofdstuk 7 onderzochten we een ander aspect van cellulaire activiteit. In plaats van neurotransmitters (Hoofdstuk 5 en 6), onderzochten we hier de effecten van ELA op de expressie in de hersenen van "immediate early genes". Immediate early genes (IEG) verhogen hun expressie bij instroom van calcium, en worden vaak gebruikt als markers van cellulaire activiteit. Hoewel de downstream producten van IEG divers zijn (b.v. transcriptiefactoren, postsynaptische eiwitten, secretiefactoren), zijn hun functies verrassend homogeen. IEGs zijn over het algemeen gerelateerd aan cellulaire processen, zoals dendriet en wervelkolom ontwikkeling, synaps vorming en eliminatie, en regulering van excitatoir/inhibitoir evenwicht. Hier hebben we een synthese gemaakt van gegevens uit 39 publicaties. In het bijzonder hebben we een kwantitatieve meta-analyse uitgevoerd met de gegevens van mannen, waarbij het IEG c-fos werd gemeten; een kwalitatieve analyse van andere IEG en van vrouwen is beschikbaar op het niveau van een systematische review. In rust verhoogde ELA de c-fos expressie, onafhankelijk van andere levensgebeurtenissen. Na een acute stressor (dus zowel in de ELA als in de controlegroepen) verhoogde ELA de c-fos expressie alleen als (controle en ELA) dieren andere negatieve levenservaringen hadden meegemaakt. Deze resultaten suggereren dat ELA een "reeds gestresst" fenotype creëert, vergelijkbaar met ELA-naïeve, acuut gestresste dieren.

Tenslotte onderzocht Hoofdstuk 8 de effecten van ELA op structurele plasticiteit, specifiek op morfologie, neurogenese en BDNF expressie als een potentiële mediatoren. We hebben de resultaten van 64 publicaties samengevat, waarvan de meerderheid de resultaten bij mannetjes en in de hippocampus onderzocht. Wat betreft morfologie concludeerden wij meta-analytisch dat ELA zowel het volume, als het aantal en de lengte van dendrieten in de mannelijke hippocampus verminderde. De effecten waren consistent met die beschreven door andere (volwassen) chronische stress paradigma's, waarbij apicale dendrieten meer werden aangetast dan basale. Bovendien verminderde ELA de expressie van verschillende neurogenese markers, terwijl de resultaten op BDNF expressie niet significant waren. Dit mag echter niet worden beschouwd als bewijs voor de afwezigheid van een effect, vooral omdat de onverklaarde heterogeniteit matig was.

In het algemeen kwantificeren de Hoofdstukken 4 tot en met 8 stabiele effecten van ELA op de hersenen en het gedrag van knaagdieren. Wij benaderen de literatuur door middel van een systematische review en meta-analyse, en onze conclusies zijn gebaseerd op de gehele literatuur. Deze hoofdstukken voldoen dus aan de eisen van dit proefschrift, omdat ze informatie uit het verleden over de effecten van chronische stress op hersenen en gedrag integreren. Misschien nog opmerkelijker is dat deze studies dezelfde populatie bestuderen. Effectief betekent dit dat we een algemeen overzicht kunnen geven van de effecten van ELA. Dit "algemene" overzicht is beperkt tot de door ons verzamelde uitkomsten, maar het is uitgebreider dan enig ander overzicht in het ELA veld. Wij redeneerden daarom dat een van de grote voordelen

van deze studies hun integratie was, een meta-integratie. Dit concept krijgt uiteindelijk gestalte in Hoofdstuk 9. In Hoofdstuk 9 experimenteren we met een nieuwe manier om informatie te integreren: we geven een gedetailleerd en toch breed overzicht van ELA. Dit hoofdstuk is een commentaar, maar het wordt ondersteund door kwantitatieve meta-analytische uitspraken. Met deze aanpak hebben we 7 principes van ELA bij knaagdieren geëxtraheerd, cumulatief gebaseerd op bijna 700 experimenten waarin controle- en ELA-dieren worden vergeleken. Deze principes kunnen worden beschouwd als een discussie van de Hoofdstukken 4 tot en met 8, en kunnen in detail worden gelezen in Hoofdstuk 9. Twee punten vielen in het bijzonder op. Ten eerste suggereren onze meta-analyses dat er niet alleen interacties kunnen zijn tussen vroege en late levensgebeurtenissen, maar ook met de acute toestand waarin het dier zich bevindt. Hoewel deze conclusie voor de hand lijkt te liggen, wordt zij zelden in overweging genomen bij het plannen van experimenten. Bovendien kan deze observatie helpen bij het verenigen van theorieën over ELA (specifiek, cumulatieve en match-mismatch theorieën) die over het algemeen als lijnrecht tegenover elkaar staan. Ten tweede bleek ELA de variatie te vergroten, ten gunste van het kwetsbare en/of veerkrachtige fenotype, ten koste van het 'gemiddelde' fenotype. Dit was vooral duidelijk wanneer de dieren geen andere levensgebeurtenissen meemaakten, en de genetische opmaak dus belangrijker wordt. Voor een diepgaande bespreking van deze onderwerpen, zie Principe #4 en #6 in Hoofdstuk 9. Al met al integreert Hoofdstuk 9 meta-informatie over de effecten van ELA op hersenen en gedrag. De conclusies vergroten ons (theoretisch) begrip van hoe ELA werkt, en geven waardevolle aanbevelingen voor toekomstige experimenten.

Hoewel de in **Deel B** gebruikte methodologieën de gouden standaard van de preklinische meta-analyse volgen, hebben wij er onze eigen methodologische draai aan gegeven. Wij waren de eersten die MetaForest toepasten op knaagdier data. Dit is een verkennende aanpak om de belangrijkste moderatoren in de dataset te identificeren (bv. type van gebruikt model, soort...). Deze data-gedreven techniek past random forests (een machine-learning algoritme) aan voor meta-analyse door middel van bootstrap sampling. MetaForest rangschikt moderatoren op basis van hun invloed op de effectgrootte, en het kan een groot voordeel zijn voor knaagdierstudies, waar de potentiële moderatoren waarschijnlijk meervoudig zijn, interageren, en niet-lineaire effecten hebben. Ten tweede waren wij een van de eersten die in de dierenliteratuur een model met vaste effecten op 3 niveaus toepasten in plaats van de meer gebruikelijke modellen met willekeurige effecten. Dit model houdt rekening met de schending van de veronderstelling van onafhankelijkheid wanneer de gegevens van dezelfde dieren worden verzameld, en verbetert zo de robuustheid van de getrokken conclusies. Tenslotte hebben wij interactieve visualisaties van de gegevens ontwikkeld, zodat iedere onderzoeker zijn eigen meta-analyse kan uitvoeren door de populatie en

de uitkomsten van belang te selecteren. Binnen de ontwikkelde gebruiksvriendelijke interfaces kan een grote verscheidenheid aan kenmerken worden geselecteerd, zoals ELA-modellen en hun componenten (bv. type, timing, voorspelbaarheid), leeftijd, geslacht, enz. Met de verstrekte informatie geeft de app relevante publicaties, forest, funnel en cumulatieve plots. Deze analyses zijn niet bevestigend, maar ze zijn waardevol omdat ze de onderzoekspraktijk direct kunnen beïnvloeden. Ze kunnen met name nuttig zijn 1) om de literatuur te verkennen, 2) om nieuwe hypothesen te definiëren, 3) om publicatiebias en de repliceerbaarheid van de bevindingen te evalueren, en 4) om realistische effectgroottes te schatten waarop toekomstig onderzoek kan worden gebaseerd.

Conclusie Deel B

De effecten van ELA op gedrag en neurobiologie zijn afhankelijk van een complexe interactie tussen vroege en late levensgebeurtenissen, alsmede de acute toestand van het dier. Door gebruik te maken van systematische reviews en meta-analyses, kunnen we informatie uit de literatuur integreren om tot robuuste conclusies te komen. Dit heeft geleid tot de extractie van 7 principes van ELA, die niet alleen inzicht geven in wat momenteel bekend is, maar ook richting kunnen geven aan toekomstig onderzoek.

Het doel van **Deel C** (Hoofdstukken 10 en 11) was het ontwikkelen van methodologieën voor informatie-integratie. Het werk van de voorgaande hoofdstukken in dit proefschrift leidde tot twee observaties: 1) bij knaagdieren zijn controlegroepen in veel opzichten vergelijkbaar met elkaar, en 2) conclusies trekken op basis van meerdere bronnen is het dagelijks werk van een wetenschapper. De hoofdstukken in dit hoofdstuk zetten deze concepten om in praktische hulpmiddelen voor wetenschappers.

In Hoofdstuk 10 laten we zien hoe historische controlegegevens in nieuwe experimenten kunnen worden geïntegreerd om de statistische kracht ervan te vergroten. We schatten dat dierexperimenten over het algemeen ernstig (prospectief) underpowered zijn, een conclusie op basis van 479 publicaties op het gebied van metabolisme en neurowetenschappen. Dit gold voor een hele reeks mogelijke effectgroottes die in de literatuur kunnen worden gevonden. In het beste geval bleek 12,5% van de dierexperimenten een toereikende power te hebben, wat betekent dat hun prospectieve power groter was dan 80%. Op basis van deze observatie zou het uitvoeren van voldoende bekrachtigde experimenten een toename van de steekproefgrootte met tientallen vereisen. Het is duidelijk dat dit geen haalbare oplossing zou zijn, niet alleen vanwege de benodigde ruimte en de financiële kosten, maar ook vanwege de voortdurende inspanningen om het aantal in onderzoek gebruikte dieren te verminderen. Daarom hebben wij in Hoofdstuk 10 een alternatieve oplossing voorgesteld. Informatie van eerdere experimenten kan met behulp van Bayesiaanse priors in nieuwe experimenten worden geïntegreerd, wat rechtstreeks van

invloed is op de statistische kracht van dierexperimenten. Belangrijk is dat dit alleen wordt toegepast op de controlegroep en niet op de experimentele groep, aangezien aangenomen kan worden dat controledieren tot dezelfde populatie behoren (zie Hoofdstuk 10 voor een diepgaande discussie over deze belangrijke aanname). We hebben deze benadering gevalideerd in een case studie naar de effecten van ELA op ruimtelijk leren bij volwassen mannelijke muizen. De experimentele gegevens werden verzameld door het samenvoegen van gegevens van afzonderlijke experimenten die in principe hetzelfde ontwerp hadden, maar individueel een lage power. Wij richtten een consortium op (RELACS, Rodent Early Life Adversity Consortium on Stress) van 10 laboratoria, die gezamenlijk 275 dieren gebruikten om deze specifieke vraag te beantwoorden. Om de implementatie van de methode te vergemakkelijken, ontwikkelden we RePAIR, een open-source web-based tool om Bayesiaanse priors toe te passen op nieuwe experimenten. RePAIR kan gebruikt worden om 1) prior parameters te berekenen uit de samenvattende statistieken van bestaande data, 2) steekproefgroottes te berekenen en 3) analyses uit te voeren. Het kan ook worden gebruikt om de (potentiële) heterogeniteit tussen de eigen eerder verkregen controlegegevens en controlegegevens van andere laboratoria te visualiseren. De kracht van deze aanpak is dat hij intuïtief is, en dat het een vertaling in statistische termen is van veronderstellingen die reeds in de dagelijkse onderzoekspraktijk worden gebruikt. Nieuwe experimenten worden meestal gepland op basis van de informatie die in eerdere studies is verkregen. Ook al bestaan er duidelijk verschillen tussen stammen of laboratoria, toch hebben dieronderzoekers vaak dezelfde verwachtingen over hoe een controlegroep "zou moeten reageren". Als niet aan deze verwachting wordt voldaan, zal een onderzoeker de gegevens waarschijnlijk "niet vertrouwen" en concluderen dat het experiment "niet heeft gewerkt" of "beter moet worden geoptimaliseerd". Al met al introduceert Hoofdstuk 10 een intuïtieve methode voor informatie-integratie, die gemakkelijk kan worden gebruikt door onderzoekers die dierstudies uitvoeren.

In Hoofdstuk 11 vatten we de software samen die in de voorgaande hoofdstukken is ontwikkeld. We hebben drie verschillende soorten software ontwikkeld: 1) een R pakket (abc4d, Hoofdstuk 2), 2) een web-gebaseerde statistische software (RePAIR, Hoofdstuk 10), en 3) verschillende interactieve data visualisaties (Hoofdstukken 2 tot 5). Het R pakket kan worden gebruikt voor het voorbewerken (b.v. behandeling van ontbrekende waarden, batch effect correctie, normalisatie en standaardisatie) en analyseren van gegevens van experimenten gebaseerd op immunohistochemie van het gehele brein, van het netwerk tot het enkele cel niveau, en met de mogelijkheid om een tijdsdynamiek op te nemen. RePAIR kan worden gebruikt om gegevens van eerdere experimenten te integreren in nieuwe experimenten, en begeleidt de wetenschapper door elke stap: voorafgaande specificatie, berekening van de power, en data-analyse. Ten slotte maken de interactieve datavisualisaties het voor iedereen mogelijk om

rechtstreeks de gegevens die in een onderzoeksartikel zijn gepubliceerd te visualiseren, in plaats van met een excelbestand. Hier kan iedereen kenmerken selecteren die van belang zijn, en de apps leveren visualisaties van de gegevens. Dit is gemaakt voor de hele-hersen data besproken in Hoofdstuk 2, voor de stress-NL consortia data beschreven in Hoofdstuk 3, en voor verschillende meta-analyses. Al met al voldoet Hoofdstuk 11 aan het doel van dit proefschrift omdat het een samenvatting geeft van de in dit proefschrift ontwikkelde software die gebruikt kan worden om verschillende vormen van data te integreren.

Conclusie Deel C

Informatie van eerdere controle diergroepen kan worden gebruikt om de statistische kracht van dierproeven te verbeteren. Open software kan worden ontwikkeld om wetenschappers te helpen informatie te integreren, en als een kennisbenuttingsstrategie voor een vlottere communicatie van wetenschappelijke resultaten.

Concluderend, in dit proefschrift hebben we informatie met betrekking tot de acute en chronische (d.w.z. ELA) stressrespons geïntegreerd. Dit werd bereikt 1) door het uitvoeren van nieuwe experimenten in knaagdieren, 2) door het oprichten (RELACS) en uitbreiden van de samenwerking binnen (stress-NL) consortia, 3) door het uitvoerig en systematisch reviewen van de literatuur, en tenslotte 4) door het vrij online beschikbaar stellen van data en scripts. Het is een eerste stap in de richting van een alomvattend, systeem-overzicht van stress.

Layman summary

Stress is like a digital image: it is made of small squares with a meaning themselves, but much more interesting together. By performing experiments, we are increasingly good at defining the colors and details of each square. Our next challenge in stress research is therefore to connect the squares and let the picture emerge.

In this thesis, we give a head start on connecting the squares. We studied stress that is short-lived (i.e. acute) in adulthood, and stress that is long-term (i.e. chronic) in early life. Importantly, we studied how to connect the squares, and we created methods to do it.

By connecting the squares in research on acute stress, we showed that mice and humans use similar parts of the brain, and in a similar order after acute stress.

By connecting the squares of chronic stress in early-life, we found 7 principles of what early-life adversity does to the brain and behavior. Among these principles, two particularly stood out. Firstly, stress depends not only on (early-to-adult) life experiences, but also on the acute situation at testing. Secondly, early-life adversity makes animals either more resilient or vulnerable, at the expenses of the "average". These 7 principles give an insight on what is currently known, and can also guide future research.

By studying how to connect the squares, we created a computer program that can re-use information from old animal experiments. This can be used to reduce the use of animals in studies for which there are no animal-free alternatives yet.

We connected the squares in our research questions, but also in the way we do Science. We founded RELACS and participated in stress-NL. These are two associations of scientists where research data is collected, shared and re-used. Furthermore, we limited the number of animal experiments by thoroughly reading, summarizing and analyzing the literature of several decades of previous research. Lastly, we have made all our data and analyzes available online.

Researching stress as a picture rather than a collection of squares brings us closer to understand how stress actually works in nature, in health and for the future treatment or prevention of disease. This thesis is a first step towards connecting the squares: old and new experiments, human and animal experiments, experiments on behavior with experiments on cells. While many details remain to uncover, we show that working together, being transparent, and connecting squares can be canes to support us while we re-learn how research.

Riassunto generale

Lo stress è come un'immagine digitale: è fatto di piccoli quadrati con un significato proprio, ma molto più interessanti insieme. Quando facciamo esperimenti, siamo molto bravi a definire i colori e le forme di ciascun quadrato. La nostra prossima sfida nella ricercar sullo stress è quindi quella di collegare i quadrati e far emergere l'immagine.

In questa tesi, abbiamo iniziato a collegare i quadrati. Abbiamo studiato lo stress di breve durata (cioè acuto) nell'età adulta e lo stress a lungo termine (cioè cronico) nei primi anni di vita. Inoltre, abbiamo studiato come collegare i quadrati e abbiamo sviluppato dei metodi per farlo.

Collegando i quadrati nella ricerca sullo stress acuto, abbiamo dimostrato che i topi e gli esseri umani utilizzano le stesse parti del cervello e nello stesso ordine dopo lo stress acuto.

Collegando i quadrati dello stress cronico nella prima infanzia, abbiamo trovato 7 principi di ciò che le avversità della prima infanzia fanno al cervello e al comportamento. Tra questi principi, ne spiccano due. Innanzitutto, lo stress dipende dalle esperienze di vita (dai primi anni all'età adulta), ma anche dalla situazione presente in cui una persona (o animale) si trova. In secondo luogo, le avversità della prima infanzia rendono gli animali o più resilienti o più vulnerabili, a scapito della "media". I nostri 7 principi forniscono una visione di ciò che è attualmente noto e possono anche guidare la ricerca futura.

Studiando come collegare i quadrati, abbiamo creato un programma informatico per riutilizzare le informazioni provenienti da vecchi esperimenti sugli animali. Questo può essere usato per ridurre l'uso di animali in studi per i quali non esistono ancora alternative senza animali.

Abbiamo collegato i quadrati non solo nelle nostre domande di ricerca, ma anche nel modo in cui facciamo Scienza. Abbiamo fondato RELACS e partecipato a stress-NL. Queste sono due associazioni di scienziati in cui i dati di ricerca vengono raccolti, condivisi e riutilizzati. Inoltre, abbiamo limitato il numero di esperimenti sugli animali leggendo, riassumendo e analizzando a fondo la letteratura di diversi decenni di ricerche sullo stress. Infine, abbiamo messo i dati e le analisi online a disposizione di tutti.

Studiare lo stress come un'immagine piuttosto che come un insieme di quadrati ci avvicina a capire come funziona effettivamente in natura, nella salute e per il futuro trattamento o prevenzione delle malattie. Questa tesi è un primo passo per collegare i quadrati: esperimenti vecchi e nuovi, esperimenti sull'uomo e sugli animali, esperimenti sul comportamento ed esperimenti sulle cellule. Sebbene rimangano ancora molti dettagli da scoprire, questa tesi dimostra che lavorare insieme, essere trasparenti e collegare i quadrati sono come stampelle per supportarci mentre re-impariamo a ricercare.

Lekensamenvatting

Stress is als een digitaal beeld: het is opgebouwd uit kleine vierkantjes met hun eigen betekenis, maar samen veel interessanter. Als we experimenten doen, zijn we heel goed in het bepalen van de kleuren en vormen van elk vierkant. Onze volgende uitdaging is om de vierkantjes te verbinden en het resulterende beeld naar buiten te brengen.

In dit proefschrift zijn we begonnen met het verbinden van de vierkanten. Wij onderzochten kortdurende (d.w.z. acute) stress op volwassen leeftijd en langdurende (d.w.z. chronische) stress in het vroege leven. We bestudeerd hoe we de vierkanten kunnen verbinden en hebben we methoden ontwikkeld om dat te doen.

Door de vierkanten in het acute stress-onderzoek met elkaar te verbinden, toonden we aan dat muizen en mensen na acute stress dezelfde delen van de hersenen gebruiken en in dezelfde volgorde.

Door de verbanden te leggen tussen chronische stress in de vroege kindertijd, vonden we 7 principes van wat tegenspoed in de vroege kindertijd doet met de hersenen en het gedrag. Twee van deze beginselen vallen op. Ten eerste is stress afhankelijk van levenservaringen (van de vroege jaren tot de volwassenheid), maar ook van de huidige situatie waarin een mens (of dier) zich bevindt. Ten tweede maakt tegenspoed in de vroege jaren dieren ofwel veerkrachtiger ofwel kwetsbaarder, ten nadele van het "gemiddelde". Onze 7 beginselen geven inzicht in wat momenteel bekend is en kunnen ook richting geven aan toekomstig onderzoek.

Door te bestuderen hoe de vierkantjes met elkaar verbonden kunnen worden, hebben we een computerprogramma gemaakt dat informatie uit oude dierproeven kan hergebruiken. Dit kan worden gebruikt om het gebruik van dieren te verminderen bij studies waarvoor nog geen proefdiervrije alternatieven bestaan.

Het concept van de verbindende vierkanten is toegepast op onze onderzoeksvragen, maar ook op onze manier van wetenschap bedrijven. Wij hebben RELACS opgericht en deelgenomen aan Stress-NL. Dit zijn twee verenigingen van wetenschappers waar gegevens worden verzameld, gedeeld en hergebruikt. Bovendien hebben wij het aantal dierproeven beperkt door de literatuur van verscheidene decennia stressonderzoek te lezen, samen te vatten en grondig te analyseren. Ten slotte hebben we de gegevens en analyses voor iedereen online toegankelijk gemaakt.

Door stress te bestuderen als een beeld in plaats van een reeks vierkanten komen we dichter bij het begrijpen van hoe het werkelijk werkt in de natuur, in de gezondheid en het verbeteren van de toekomstige behandeling of preventie van ziekte. Deze dissertatie is een eerste stap naar het verbinden van de vierkanten: oude en nieuwe experimenten, menselijke en dierlijke experimenten, gedragsexperimenten en cel experimenten. Hoewel nog veel details moeten worden ontdekt, toont deze dissertatie aan dat samenwerken, transparant zijn en de vierkanten verbinden, hulpstukken zijn om ons opnieuw te leren hoe we onderzoek moeten doen.

Curriculum vitae

RESEARCH EXPERIENCE

since April 2022	Utrecht University (NL) — Postdoc Methodology and Statistics; Visiting researcher at QUEST (BIH, Berlin, DE)
2017 - Jan 2022	UMC Utrecht (NL) — PhD candidate Neuroscience and Statistics
2017	Univ. of Edinburgh (GB) — Research internship endocrinology
2015 - 2016	Utrecht University (NL) — Research internship animal models

SOFTWARE DEVELOPED

2020 - 2021	Developer R package abc4d *https://valeriabonapersona.github.io/abc4d/*
2019 - 2020	Developer interactive statistical software RePAIR *https://utrecht-university.shinyapps.io/repair/*

OTHER RELEVANT EXPERIENCE

2019 - 2021	Database developer and maintainer (stress-nl)
2018 - 2021	Consortium founder and coordinator (RELACS, Rodent Early Life Adversity Consortium on Stress)
2017 - 2021	Teachich: supervision 9 students for interships/theses, lecturer in 13 courses, supervision 6 practicals/workshops. Curently working towards a teaching qualification for higher education (BKO).
2009 - 2021	Student representative (PhD: 2017- present; Master: 2015-2017; High school: 2009-2012)

EDUCATION

2017 - 2022	Postgraduate courses (selection): Differential equations (MOOC, 2020); Advanced OMICS algorhythms and data analysis (CS&D, 2020); Individual participant data meta-analyses (Julius Center, 2020); Neurobiology of stress and resilience (NSAS, 2018); Teaching Neuroscience (UU, 2020/2021), Reproducible code (2019), Philosophy of Science (2018)
2015 - 2017	MSc Neuroscience and Cognition (cum laude)
2012 - 2015	Bachelor of Science (honors, cum laude)
2006 - 2016	Diploma Superiore in Piano Performance (9.5/10)
2007 - 2012	Liceo Scientifico (97/100)

ACHIEVEMENTS

2022 Top Paper Prize at Dutch Neuroscience Meeting (NL)
2021 Oral presentation prize, BCRM research day (NL)
2020 Travel grant, S4: small sample size solutions (NL), *cancelled*
Travel grant, European College of Neuropsychpharm. Workshop (FR), *cancelled*
2019 2nd prize at FIGON: PhD competition in pharmachology (NL)
Best oral presentation, Translational Neuroscience day, Utrecht University (NL)
Best oral presentation, ONWAR PhD symposium (NL)
Travel grant EBBS, Pragure (CZ)
2018 Best poster presentation, Dutch Neuroscience Meeting (NL)
2017 Travel grants for master internship (cumul. ~4500e)

KNOWELDGE UTILIZATION AND PUBLIC ENGAGEMENT

2020-present RePAIR is being implemented in 3 academic labs (NL, BE and USA) and by 2 companies (NL & USA)
Talk for lay audience, part of "Proberen lukt altijd", Tivoli Vredenburg, Utrecht (NL)
My work has been covered twice in the lay journal Spectrum News
2017-present Developer 4x interactive visualizations of research data
e.g.: *https://utrecht-university.shinyapps.io/brain_after_footshock/*

PUBLICATIONS

Bonapersona, V., Schuler, H., ..., Joëls, M. & Sarabdjitsingh, R.A. (2022). The mouse brain after foot-shock: temporal dynamics at a single cell resolution. PNAS, doi: 10.1073/pnas.2114002119

Bonapersona, V., Born, F.J., ..., Joëls, M., Vinkers, C.H. (2022) The STRESS-NL database: a resource for human acute stress studies across the Netherlands. Psychoneuroendocrinology in press.

Schuler, H.^, **Bonapersona, V.**^, Joëls, M. & Sarabdjitsingh, R. A. (2022). Effects of early life adversity on immediate early gene expression: systematic review and 3-level meta-analysis of rodent studies. PlosOne in press. ^ contributed equally

Bonapersona, V., RELACS consortium, Hoijtink, H., Sarabdjitsingh, R. A. & Joëls, M. (2021). Increasing the statistical power of animal experiments with historical control data. Nature Neuroscience, 24(4), 470-477

van der Veen, R., **Bonapersona, V.**, Joëls, M. (2020). The relevance of a rodent cohort in the Consortium on Individual Development. Developmental cognitive neuroscience, 45, 100846.

Bonapersona, V., Kentrop, J., ..., Joels, M. & Sarabdjitsingh, R. A. (2019). The behavioral phenotype of early life adversity: a 3-level meta-analysis of rodent studies. Neuroscience & Biobehavioral Reviews, 102, 299-307.

Bonapersona, V., ..., Joels, M. & Sarabdjitsingh, R. A. (2019). Sex-Dependent modulation of acute stress reactivity after early life stress in mice: relevance of mineralocorticoid receptor expression. Frontiers in behavioral neuroscience, 13, 181.

Bonapersona, V., Joels, M. & Sarabdjitsingh, R. A. (2018). Effects of early life stress on biochemical indicators of the dopaminergic system: a 3 level meta-analysis of rodent studies. Neuroscience & Biobehavioral Reviews, 95, 1-16.

Joëls, M., Kraaijenvanger, E., Sarabdjitsingh, R.A., **Bonapersona, V.** (*submitted*) Structural changes after early life adversity in rodents: a systematic review with meta-analysis.

Bonapersona, V., Sarabdjitsingh, R.A. & Joëls, M. (*submitted*) 7 principles of early life adversity in rodents. Submitted.

Bonapersona, V., Joëls, M., van der Veen, R. *(in preparation)* Effects of early life adversity on monoamines' expression in rodents: a systematic review and meta-analysis.

ABSTRACTS AND PRESENTATIONS

I wrote 14 conference abstracts, 10 of which as first author. Additionally, I was invited as speaker at 16 conferences/symposia.

Acknowledgements

This work was possible only with the love, compassion, support, wisdom (or lack thereof) and wit of many of you. If you are reading this, chance is that this thesis wouldn't be here without you. I am grateful that you have accompanied me, through all my doubts and my emotions, in all the different phases and shapes that I have taken throughout my PhD. The past 5 years have been the best of my life.

Dear **Marian**, what a ride it has been! Early in my PhD, you said "*Making a scientist is like making a child: you need time, love and money*". It is a complete mystery to me how you have managed to do this for each one of us. I have always felt extremely supported in doing my best science, and also to discover *what* science to do and *how* to do it. You have taught me the importance of boundaries: sometimes they keep us safe, other times they exist only to be overcome. Thank you for teaching me to dream without limits, while practically getting everything done. With your nurturing, I could discover, develop, and apply my talents. What a luxury having you as my supervisor! Your mentorship really moulded me: I gained self-confidence, ability of reasoning and wit. Besides the science, thank you for allowing us to see the woman behind the success. Thank you for telling me stories of your youth, of your family, and of your struggles. I have enjoyed dearly our conversations on planes, at restaurants, in cars and on zoom. In my heart, I am your last PhD student – Rob cannot say anything about it.

Dear **Herbert**, thank you for teaching me that a job doesn't define us, and that – rather – we can mould everything in our life to fit us. I've joked often about "Herbert's way of doing things", but it taught me how to set priorities and to think independently. Throughout my PhD, I gathered a "Herbert mantra": pointers to follow if you want to succeed in statistics. This wisdom is here written with a permanent DOI to be passed on and on to future generations. *"1) If you don't understand it, don't trust it. 2) Probably someone already thought about it. 3) If someone already thought about it, then there's an R package for it."* Lastly, you are right: informative hypotheses are the best! I hope one day Science will realize it too.

Dear **Angela**, thank you for seeing my talent since the very beginning, and to teach me how to explore and use it. I have always felt extremely safe with you, and I value a lot your opinion and feedback. Thank you for showing me that people can change what they want and what their priorities are – which is not only possible, but also wise. Thank you for having been there through children and accidents, my ups and downs, and when changing job. I really admire your unique sensitivity and ability to observe (and to cook, which I wish I could have tasted more). I am very grateful of the support (in all its forms) that you have given me throughout the years.

Joëls trio! You then became a quartet, then a quintet. Our science has been music for my ears. **Jiska, Jelle, Milou, Katerina, Ruth, Henk, Rixt, Angela, Eline.** Thank you for the cakes, the walks on the beach, for the sharing of ideas about life and science, and for playing along with any game I could think of during the stress meetings. Thank you all for being true friends besides colleagues. **Jiska**, thank you for introducing me to the Joëls' fan club as a big sister: the "food drawer", the "understanding Marian 101", the "if you don't plan it to go well, it also cannot go bad" and the "your red-light district". Sharing the office with you has been really fun and enjoyable – maybe not the most productive. Thank you for surviving reading 6000 articles, and for being there with your kindness and smile when it felt like my life was falling apart. You will be hell of a mum. **Jelle** (aka "the true scientist"), taking a break to steal your coffee has always been an extra reason to come to the office. Thank you for including me in the monthly calendar day, for the constant optimistic attitude towards life, and for sharing with me the love for mountains and depressive music. Thank you for introducing me with Jiska to the "conference life" – and what an introduction. **Milou**, you have let me see and appreciate how many things I am, and I hope you can also see how many things you are. Your intelligence spreads on so many different levels, and I hope to be there when you'll feel like its fully expressed. Thank you for showing me that one can always ask, and that we do not really need to fit into a box. Thank you for all the times we exchanged ideas, for all the chats with all the wines, for all the art and all the science. **Eline** (aka "excel master"), thank you for having taken up such a challenge with helping me. From friend to colleague to friend again! Thank you for all the time and Tijm, and for our now 7 years long friendship. **Ruth**, thank you for all that you have taught me in the lab. You were always present and open to help. Thank you for not losing your patience on me. Even if our roads have already departed from some time, the only thing you really missed was the escalation of my chocolate addiction. **Rixt**, thank you for always putting at heart an interest in my work. Our relationship has changed during the years – thank you for upgrading me from student to peer. I am really impressed of your "let's see how it goes" and "let's do the best we can" attitude towards life and science. I hope that the future will bring you loads of luck and satisfaction, and that your naturalistic and thought-through science will be appreciated more and more.

The Science would have not been as good (nor as fun) without a strong and supportive community – **Translational Neuroscience, CID consortium, and collaborators (RELACS, stress-NL)** from everywhere else. A special thanks to **Rob, Joeri, Janssen, Silvie, Eva, Kitty** for all the fun at conferences, and **Aniko** for having such a wonderful group. Thank you, **Mathias, Frank, Kevin, Caspar, Carmen, Lotte** and **Tracy** for (in)directly inspiring me throughout the years. You have left a mark in my education. Thank you **Geert** for the support throughout Master and PhD. Thank you

for all the work that you do to keep us sane and provide us with excellent education: we really appreciate it. Thank you, **Ramona, Jessy, Paul, Karlijn** and **Lieke** for the PhD-related laughs and tears. I have beautiful memories of each of you impressed on my heart. Thank you **Youri** and **Roger Koot** for all your help with my "favourite" project. It would have never happened without you. No exaggeration.

Heike, Lieke, Lara, Ally, Elisa, Frieda: you have been such wonderful students. I am super proud of all of you, of how you have grown and where you are at. You rock! Thank you **Heike** for always doing your best, and challenging me to do my best. It has been really fun to brainstorm and learn new things together. Thank you **Frieda** for wanting me as a supervisor even given the circumstances, for confiding in me and in my opinion. This role that you requested of me made me grow so much, and I'm extremely grateful.

FRIENDS! Your time to shine has come. You kept me sane throughout the years: well done. All of you are part of my extended family, and we should all hold hands and dance in a circle at my defence. Just kidding. **Ramona**, my paranymph, my sister from a different mum, mouse catcher, housemate, bed mate (luckily neither of us snores), sauna companion, etc. Thank you for all the nights spent chatting on the floor in your room, and all the days chatting on the floor in your office. To all our new adventures, hopefully without broken knees and broken heads. **Fabi** and **Lewis**, you have saved me when you took me into your house, life and heart. No words can describe how grateful I am to you both. Thank you for creating such a loving environment in the house – our little nest with **Nacho**. **Fabi**, thank you for accompanying me in my own process without never judging, and for allowing me to be your friend. It is remarkable how accepting of others and of others' difficulties you are. I love you very much. **Dirk Jan** (not Dirk-Jan), you have been my "school" companion since UCU... you have always been there! How much we have grown. Thank you for sharing so many interests with me: with science and music our conversations are never dull. From new years' parties to ice-creams and horses. Thank you for being my paranymph. I could not imagine anyone else accompanying me through this last phase, to the closure of this Dutch chapter. **Lucie** and **Paula**, thank you for all the calls, and for always be there to share the joys and sorrows of life with me. I miss you both, and I can't wait for us to be re-united. No matter where in the world we are, we are together. **Marielle, Gerard, Marguerite** and **Serge**, thank you for being my "surrogate parents" whenever I needed. It has been lovely! **Klara** and **Emma**, thank you for all the drinks and ldjskfhbgsd about PhD. Our meetings always feel like these static moments in our ever-changing life. How many plot twists we had! How wrong we were about what the future would have brought us! Thank you for sharing so many laughs together. Klara, you are one of the most remarkable women I know. You have always been such an example. Emma, I am

so extremely proud of you and how you have changed your life around. **Deli**, thank you for listening to my excitement about science and yoga. Thank you for hosting me in your house and in your heart for so long. No matter where we will be in the world, you can count on me. **Miel**, thank you for giving me a ~~second~~ third chance. I will always look forward to our walks in the woods. Thank you, **Mette, Jesse, Thomas, Tijn, Bob, Damien, Natasha, Rikkie, Deli** for all the partying. You reminded me that life should just be enjoyed with a pinch of craziness. I didn't really have "teenage years", and our group really allowed me to have that experience. **Jesse**, thank you for all the talks, dinners, love and understanding (and for De Banka). Our friendship is sealed in my heart – outside time. Thank you for creating **Arbor**, and letting me be its aunt. I love you both dearly. **Oguzhan**, thank you for having chosen and invested in our friendship. I grow after each of our conversations, and you continue to surprise me. Thank you for challenging time and time again my thoughts and my beliefs. **Franka** (aka *"bioinformatics' bible"*) thank you for all the times we have been girly or sciency together. Because we can have it all – a beautiful skin and an overactive brain. I am looking forward continuing to do exactly the same until when we are old, wrinkly, and we will have finally managed to talk about everything we want. **Rachel**, thank you for keeping me always with my feet glued to the floor. **Reshma**, thank you for sharing the love and passion of yoga with me, and for teaching me to not take anything (including myself) too seriously. Grazie **Gigi** e **Dema** perche' nonostante gli anni passati siete ancora al mio fianco. La vostra presenza e supporto e' molto importante per me, e voi siete gli unici che mi possono prendere in giro per cose successe 15 anni fa.

Henk, Anja, Marlies, Jasper, Nolan, Arlo, Lisanne, Frank & Gijs: thank you for having let me into your life and family.

Grazie **mamma** e **papá** per avermi insegnato che non ci sono limiti e per avermi dato la possibilitá di sognare e fare a mio piacimento. Grazie per avermi insegnato che la testa e' sempre a servizio di un cuore aperto, e che il piu' grande successo e' poter condividere la vita con le persone che amo. Grazie mammina anche per aver fatto tutta la impaginazione della tesi. Che pazienza! **Sorellina**, grazie per accompagnarmi, con la tua presenza, il tuo amore e la tua saggezza. Siamo opposte ma identiche, due facce della stessa medaglia, camminando le vie di questa – e molte altre – vite insieme. Grazie anche per aver portato **Gaia** a questo mondo, la piú grande maestra di tutti noi. Grazie **Diego** per renderle entrambi felici. **Nonni**, grazie per esserci sempre – anche oggi, anche se non e' comodo e vicino a casa. Che ottantenni! Spero che oggi possiate scoprire quanto sia speciale che il primo membro della nostra famiglia abbia ottenuto un PhD. **Nonna tosta group**: la nostra famiglia mi piace ogni anno di piú. Vi amo tanto.

Thank you **Eelco** for all that cannot be expressed with words, and for liking trains as much as I do.

Supplementary Information for Chapter 2

The mouse brain after foot-shock in 4D: temporal dynamics at a single cell resolution.

Contents

Supplementary Notes

Supplementary Note 1: c-fos

c-fos is a proto-oncogene of the Fos family, which acts as a transcription factor upon heterodimerization with a member of the Jun family [1,2]. With the exception of a few constitutively active brain areas, c-fos is not expressed under baseline, i.e. non-stressed, circumstances[3] but transiently induced by mild-to-severe acute stimuli, with activity-dependent intensity[4].

Supplementary Methods

Animals and husbandry. 8 to 10 weeks old male C57Bl/6JOlaHsd mice were purchased from Envigo (Harlan, France) in 3 separate batches. The animals were habituated to the animal facility for at least two weeks, then tested at the age of 10 to 14 weeks. Until the experimental day, the animals were housed in groups of five in type II L cages (365x207x140mm, Tecniplast®) on a 12h dark/light cycle (light phase: 9.00AM–9.00PM), 22±2 °C, humidity at ± 64%, with *ad libitum* access to standard chow (Special Diet Services, UK, sdsdiets.com) and tap water. Experimental cages were placed on racks without a specific order and left undisturbed in the same experimental room, except for weekly cleaning by animal caretakers unfamiliar with the study design. A copy of the work-protocol (license: 527/16/644/01/06, 527/18/4806/01/01) as well as a step-by-step protocol can be found on the Open Science Framework page of the project (osf.io/8muvw). All animal procedures were approved by the Animal Ethical

Committee at Utrecht University (license: AVD1150020184806), the Netherlands. Every effort was taken to minimize animal suffering in accordance with the FELASA guidelines and the Dutch regulation for housing and care of laboratory animals. For the main analysis, a total of 9 male mice per experimental group (n_{block} = 9, $n_{time\ point}$ = 4, $n_{animals}$ = 36) were used.

Experimental design. The primary aim of this study was to develop a methodology to be able to analyze brain-wide activation over time, specifically of the stress system. The experiments were designed to address batch effects, missing values and normalization. An overview of the experimental procedure, data cleaning, preprocessing, and analysis can be found in Figure 1.

The main experiment used a uniform block design, where each block corresponded to a separate cage of animals. Each cage contained an animal of each experimental group (being 'time after foot-shock' and control; $n_{time\ point}$ = 4). The animals were killed at three different time points after receiving a foot-shock ($n_{time\ point\ foot\text{-}shock}$ = 3). A control group underwent the same identical procedure but did not receive the foot-shock. This group is considered time point 0 min (t_0). Of note, this control group did experience a novel environment, and its activation should therefore be considered not as 'baseline' but as mildly stressed. The time points refer to the moment of transcardial perfusion, and were chosen to comprehensively model all phases of the stress response. They cover the initiation (t = 30 min, t_{30}), maintenance (t = 90 min, t_{90}) and termination (t = 180 min, t_{180}) of the HPA axis response, while considering the required time to synthesize c-fos [4] as well as the delay of 30 minutes of increase of corticosteroid in brain tissue compared to blood [5]. Randomization of the experimental groups occurred within the cage, and samples of the same block were processed together both pre- and post- mortem. In other words, each experimental group was represented in every four animals (block), but the order of experimental groups differed across blocks. This type of design is important for batch effects' correction.

The following control experiments were performed. A 15 min time point ($n_{animals}$ = 3) to validate that indeed 30 min is the earliest time point to detect an increase in c-fos protein expression. A 300 min time point ($n_{animals}$ = 2) to validate that c-fos expression eventually decreases. A no primary antibody staining ($n_{animals}$ = 3) to estimate the extent of unspecific secondary antibody binding. A home-cage control group ($n_{animals}$ = 3) was used to validate that the c-fos expression of the t_0 group was due to the experimental procedure (i.e., placement into a novel cage). To minimize the sample size of animals used, the animals of these control experiments were processed together with those of the main experiments, so that the comparison groups could be reused. As a consequence, these control experiments are only qualitatively assessed, and no formal statistical analysis is performed to avoid unnecessary multiple testing. Data of the control experiments is presented paired to the t_0 group of the same batch.

Initially, we aimed to include males as well as females (see work protocol at osf.io/8muvw) in the experiment. However, the percentage of animals lost due to unforeseen circumstances (due to the antibody) was much larger than anticipated (30% instead of 10%). That is, the secondary antibody was less stable than expected; as a consequence, the quality of the scans was often insufficient for a brain-wide quantification. We therefore simplified the experimental design and used only 9 male mice per experimental group ($n_{block} = 9$, $n_{time\ point} = 4$, $n_{animals} = 36$). These were processed in three different batches of 3, 4, and 2 blocks, respectively. We chose males instead of females because more male samples had sufficient quality after the first two batches. This qualitative assessment took place before c-fos quantification. The sample size of 9 animals per group was in line with or larger than manuscripts previously published using whole-brain microscopy [6–8]. Furthermore, it was sufficiently large to identify, study and mitigate batch effects. No experimental animal was excluded from the study for reasons other than insufficient staining quality. For a full description of the missing animals, see Supplementary Table 2.

Foot-shock induction. At any given experimental day, a cage was brought to the experimental room. Each animal of the cage was placed in a separate foot-shock box (floor area: 250x300mm) at the same time. Since the animals were not earmarked, selection bias was limited by randomizing the order of the shock boxes when placing the animals. In this way, we aimed to limit a "shock box" specific effect. After 60 seconds, the experimental animals received a single foot-shock (0.8 mA, 2 sec). The t_0 animal of each block was always placed in a "sham shock box". This box was identical to the others, but it did not give a shock. After another 30 seconds, all animals were removed from the shock boxes. The t_0 animals were euthanized immediately, whereas animals of the foot-shock groups were single-housed in new cages waiting for transcardial perfusion according to their specified time point. The new cages were enriched with *ad libitum* chow and water, as well as bedding material of the home cage to limit arousal due to a novel environment.

To acquire meaningful, yet above threshold c-fos expression, the optimal t_0 condition (home-cage vs novel environment) and foot-shock intensity (0.4 mA vs 0.8 mA, assessed at t = 90 min) were established in a pilot study. The results of the pilot study were only qualitatively assessed.

The investigator (RD) performing the foot-shock procedure and the perfusion was not blinded to condition, since she needed to confirm that the foot-shock was successfully applied and that animals were perfused at the correct time point.

Perfusion and tissue preparation. Euthanasia was performed with an intraperitoneal injection of 0.1 mL pentobarbital (Euthanimal 20%, 200 mg/mL) ~10 minutes prior to transcardial perfusion. The animals were perfused with ice-cold 1x PBS until blood clearance, followed by perfusion with ice-cold 4% PFA/1x PBS. Brains were extracted

from the skull of perfused animals and stored in 4% PFA/1x PBS overnight for post-fixation.

Brains were cleared and stained following the iDisco+ protocol [7]. Briefly, 24h post-fixation, samples were washed with 1x PBS. The olfactory bulbs and cerebellum were trimmed. A methanol/H_2O gradient was applied to dehydrate the tissue, followed by decolorization with 5% hydrogen-peroxide. After rehydration with a methanol/H_2O gradient, brains were permeabilized and remained in a blocking buffer for two days. Samples were then incubated in the primary antibody for seven days (rabbit anti c-fos, Synaptic Systems, cat. 226003, 1/1000 in 92% PTwH / 3% Donkey Serum / 5% DMSO). Following washing with 1x PBS / 0.2%Tween-20 / 0.1%Heparin (PTwH), brains were incubated for another seven days in the secondary antibody (Donkey anti-Rabbit IgG (H+L) Alexa 647™, Thermo Fisher Scientific, cat. A31573, 1/1000 in 97% PTwH / 3% Donkey Serum). Lastly, samples were washed with PTwH, dehydrated with a methanol/H_2O gradient, incubated in 66% DCM / 33% Methanol, washed with 100% DCM, and then cleared and stored in the clearing agent dibenzyl ether (DBE), protected from the light. The total time required to complete the protocol is 25 days. The investigators processing the tissue samples were blind to the experimental groups.

Imaging with Light-Sheet microscopy. Starting three days after clearing, samples were imaged with a light-sheet microscope (UltraMicroscopeII, LaVision BioTec), equipped with an MVPLAPO 2x/0.5NA objective (Olympus), an MVX-10 Zoom Body (Olympus) and a 10mm working distance dipping cap. The images were recorded with an sCMOS camera (Neo 5.5 sCMOS, Andor Technology Ltd; image size: 2560x2160 pixels; pixel size: 6.5x6.5μm^2) using the software InspectorPro (v.5.0285.0; LaVision BioTec). The samples were scanned in horizontal slices (step size: 3μm; effective magnification: $zoom_{body\ objective}$ + dipping lens 2x0.63x=1.26x; sheet width: 60%) with two-sided illumination using the built-in blending algorithm. Two image stacks per sample were taken consecutively, without moving the sample in between recordings. This is essential for the later correct annotation of c-fos positive (c-fos+) cells to brain areas. To record the fluorescence of c-fos+ cells, we used a Coherent OBIS 647-120 LX laser (emission filter: 676/29). The images were recorded at 70% laser power, sheet NA of 0.076 (results in a 10μm thick sheet), and an exposure time of 100 msec, as well as with horizontal focus to reduce z-slice distortion (steps: 20). To highlight the intrinsic fluorescence of the tissue for registration of the sample to a template, we imaged with a Coherent OBIS 488-50 LX Laser (filter: 525/50nm). The images were recorded at 80% laser power, sheet NA at 0.109 (results in a 7μm thick sheet), and exposure time of 100 msec, without horizontal focus.

The investigators (VB, HS) conducting imaging and image processing were blind to condition. Samples were imaged in ascending order, with sample numbers being randomly assigned during perfusion.

Image processing: cell detection. c-fos+ cells were detected from the 647nm image stack using Imaris (v9.2.0, Bitplane). A spots object was created and parameterized to detect cells with estimated xy-diameter of 10μm and an estimated z-diameter of 35μm (to avoid overcounting cells due to z-plane distortion). Thereafter, detected cells were filtered by quality (>18) and required to cross a threshold of minimum intensity (>225). The quality filter verifies that the shape of the detected cells aligns to the spots object within a threshold specified automatically by the algorithm. These parameters were optimized on pilot imagined brains to achieve optimal signal-to-noise ratio while avoiding ceiling effects of the cell intensity parameter. The xyz-position of each cell was exported for later annotation to a reference atlas.

Image processing: alignment and annotation. Each cleared brain was registered to the Allen Mouse Brain 25 μm reference Atlas [9] by using ClearMap [7] as an interface to the open-source software Elastix v5.0 [10]. Registration was performed using the autofluorescent images (488nm image stack). As per our setting, the images were first rotated, sheared and scaled with affline transformation, then translated onto the reference atlas with b-spline transformation.
The transformation matrix as calculated by Elastix was applied to the xyz coordinates of the Imaris detected cells using Transformix. In this way, detected cells were transformed into a template space, and each c-fos+ cell was assigned to a brain region. This was possible because each sample remained in the same position for both image stacks.
To estimate the error in the approach, we drew in ImageJ (Fiji) artificial spots ($n_{spots} = 72$) in both hemispheres ($n_{animals} = 12$) in three different locations (i.e., within dentate gyrus, mammillothalamic tract, amygdalar capsule). To estimate the error of the alignment, we calculated the distance between the expected position of the artificial spots and the position resulting from the alignment procedure.
The output for further data analysis is the xyz-position of every cell with their corresponding area code. The code was then translated to the respective brain area according to the brain region table provided by Renier and colleagues [7], which follows the hierarchical organization of the Allen Brain Atlas (ABA).
At this point, images have been transformed in machine-readable numbers. Other tools can be used until here. To continue with the following steps, one is only required to have files with xyz coordinates for each cell, together with their annotated brain area.

Quality control and data pre-processing. Quality control, data pre-processing and analysis were planned on a subset of the data ($n_{blocks} = 3$), and then later extended to the full dataset. The experimenter coding the analyses (VB) was blinded to the experimental condition.
To mitigate technical noise, a series of quality control steps were performed on the

xyz annotated coordinates of c-fos+ cells. We removed (false positive) cells from brain areas in which no counts were expected, either because these areas contain no brain tissue (background, ventricular system) or because they were trimmed from the sample (olfactory bulb, cerebellum, hindbrain). Next, we grouped the highest resolution areas of the ABA in line with the ABA hierarchical organization. The aim was to preserve as much spatial specificity as allowed by alignment inaccuracies in areas likely to be stress or c-fos sensitive, while minimizing the total number of brain areas to ease interpretation and to avoid unnecessary subsequent multiple testing. Accordingly, the categorization considered the region-specific distribution of glucocorticoid receptors as well as the region-specific c-fos expression after acute stress [11,12]. The hierarchical relationship of ABA areas is not complete, meaning not all larger brain areas can be fully subdivided into smaller brain areas. These "left over" spaces were removed from the analysis since they were deemed not interpretable.
An illumination artifact was present in all samples on the outside borders of the brain and the ventricles, presumably due to unspecific antibody binding. Initially, we aimed to use the no primary antibody control group to correct for this artifact; however, this was not possible as the number of c-fos+ cells in the no primary antibody control group was minimal. As an alternative solution, a mask of 75μm thickness was modeled along the inside border of the brain and the ventricles of the aligned samples (Supplementary Figure 3 a), and cells that fell within the mask coordinates were removed from further analysis. The size of the mask (25 through 175μm) was piloted in 3 samples. Ultimately, 89 brain areas were included in the analysis (Supplementary Table 1).
Lastly, we removed xyz coordinates with extremely high c-fos intensity. We qualitatively assessed histograms of the maximum intensity of c-fos+ xyz coordinates per brain area, and compared them across samples to identify potential unspecific binding of the protein ("spots"). The potential candidates had 2- to 10-fold higher intensity than others within the same brain area. These were checked against the raw scans and removed if they did not appear as "cells" during a qualitative evaluation.

Outlying values. The selection of parameters for cell identification, the removal of the illumination artefact, the managing of areas with small volumes, and the removal of mis-labelled spots are procedures to limit as much as possible the presence of outliers. Despite these efforts, residual biological / technical outliers could be expected, either at the single cell level (e.g. unspecific antibody binding) or at the brain area level (e.g. disproportionate activation). Due to the limited sample size and the batch effects, the identification of outlying values was not trivial. We therefore chose to not use any rule (e.g. 3SD away from the mean) or statistical test to detect / exclude / replace outliers. Rather, we assumed that they may occur uniformly across samples, thereby giving rise to increased variation. To mitigate their effects, we used medians and quantiles rather than means to summarize the data.

Missing values. The main source of missing value was the loss of animals due to insufficient staining quality (see 'Experimental Design' in Methods).

A second source of missing values was due to damaged brain areas during the experimental procedure. c-fos+ cells were counted per brain area across the whole brain. Damaged areas were manually detected in the 488nm image stack independently by at least two of three researchers (VB, HS, RD). The researchers were blinded to the experimental condition, and discrepancies were resolved with discussion. c-fos+ cells of damaged areas were removed, and then imputed by mirroring the xyz cells' coordinates of the same brain area of the opposite hemisphere. Although this approach inherently assumes no differences between hemispheres, we preferred it to a multiple imputation approach because it did not require batch effects' mitigation and it could be performed at a single cell rather than at brain area level.

A third source of missing values was linked to cell detection. The cell detection algorithm requires the definition of a minimum c-fos+ intensity. This parameter was optimized in pilot experiments, and it was kept identical throughout all experimental brains. In principle this is not a problem, since by rigorous standardization it is possible to mitigate batch effects and obtain comparable relative statistics. However, when a brain area has no active cells at t_0, it needs to be further evaluated to conduct a proper standardization. In our experiment, two brain areas (FRP and AHN) had no c-fos+ cells in one t_0 sample. Since this occurred only in one sample, we considered these brain areas as missing, not as zeros for analyses that required standardization with ratios to baseline.

Preprocessing for region-based statistics. Additional pre-processing is required for region-based analyses. In Figure 1, we summarize which pre-processing steps were required for which type of analysis.

In region-based analyses, the total number of c-fos+ cells (i.e. absolute counts) was calculated per brain area. However, cell counts are by definition not normally distributed; rather, they follow a Poisson or (negative) binomial distribution. We therefore applied a Box-Cox transformation per block (i.e. each set for four different timepoints), so that our data would resemble a normal distribution.

Different brain areas have different sizes; therefore, absolute counts of c-fos+ cells are not indicative of how active a certain brain area is. In analysis where different brain areas are compared, absolute counts need to be normalized to the size of the brain area. We therefore calculated the number of c-fos+ cells per thousand of the total cells in each brain area, by adapting the atlas by Erö and colleagues [13]. We used the total cell count estimation rather than that of only neurons because several publications have reported c-fos+ glia and astrocytes (for a review, see [14]), and it is in agreement with the presence of c-fos+ cells in the fiber tracts of our own data.

The number of c-fos+ cells differed across batches, although the relationship across

time points was consistent within batches. Therefore, a normalization step was required to make the data more comparable across batches. Z-score normalization was performed per block, i.e. a unit of one sample per experimental group. With z-score normalization, the data is scaled with a mean of 0 and a standard deviation of 1, according to the formula $(\chi - \mu)/\sigma$, where χ is the observed value, μ is the mean of the sample, and σ corresponds to the standard deviation of the sample.

Region-based analyses: active brain regions. With the exception of the single-cell strategy analysis, the analyses were planned on a subset of data (n=3), and later extended to the full dataset. The experimenter coding the analyses (VB) was blinded to the experimental groups.

We tested which brain areas had a significant increase from baseline in c-fos+ cell count per thousand of total cells ($n_{cfos+ / tot}$). The dependent variable was scaled and normalized as explained in the data pre-processing section. We performed pairwise comparisons (Welch t-test, one-sided, alpha = 0.05) for each time point (t_{30}, t_{90}, t_{180}) against t_0. P-values were adjusted with the Benjamini-Hochberg (BH) procedure. For visualization only, we transformed the p-values with a $-\log_{10}$ transformation, and we grouped the brain areas according to the ABA embryological origin.

Next, we tested which brain areas were most active. The analysis was independently performed per block; therefore, no other batch-effect correction step was taken. For each block, we calculated the top 5% of $n_{cfos+ / tot}$ independently of time point, and thereby identified per block the most active brain areas. Next, we counted how often a specific brain area was categorized as most active. We considered a brain area to be consistent across samples if it was present in at least 5 out of 9 blocks in a particular time point. If we consider the process to be random under the null hypothesis, the probability of a brain area to be present in 5 out of 9 blocks would be 0.1%. This probability was estimated with a simulation study. We simulated 1000 independent experiments, and each experiment consisted of 4 time points with 9 independent iterations (i.e., $n_{blocks} = 9$). For each iteration, we selected 18 brain areas, meaning the 5% of 90 ($n_{brain\ areas}$) * 4 ($n_{time\ points}$). Each brain area had an equal probability of being selected (i.e., uniform distribution), and a brain area could be picked multiple times within each iteration (i.e., block), up to 4 ($n_{time\ point} = 4$). Then, we calculated across 1000 experiments the probability of a brain area to be in the top 5% of the distribution. This simulation gave information about how likely it is that the representation in the top 5% was chance.

Since c-fos is not uniformly distributed across the brain, we performed a simulation study to assess whether the pattern obtained was due to the baseline spatial distribution of c-fos. We downloaded via the ABA's API the mRNA c-fos expression levels of 3 experiments that passed the ABA quality check (id: 80342219; 79912554; 68442895). We calculated the mean and standard deviation for each of the brain

areas available ($n_{brain\ areas}$ = 8). Since the resolution available for c-fos expression is lower than the resolution in our dataset, we assumed that the c-fos expression available corresponded to the location and scale parameters of a normal distribution defined by all the sub-areas. In other words, for each sub-area we sampled values from a normal distribution with location and scale parameters equal to those derived from the ABA atlas. This was performed for 9 independent samples ($n_{samples}$ = 4) for each time point ($n_{time\ point}$ = 4). Since the selection may not be linked only to baseline c-fos expression, but also to a natural increase due to the foot-shock, we multiplied the baseline expression levels with the overall increase in c-fos across time points in our experiment. For each brain area, we calculated the median of the ratio between each foot-shock time point (t_{30}, t_{90}, t_{180}) and t_0. This ratio was then multiplied by the estimated c-fos expression values. Of note, the ratio at t_0 was always 1, meaning that the expression levels were estimated only from the ABA. We performed this simulation 1000 times, thereby simulating 1000 independent experiments. For each simulated experiment, we considered 9 blocks and 4 time points, as in our actual experiment. Then, for each block in each experiment, we selected the brain areas whose expression was in the top 5% of the distribution. Across the 1000 experiments, we then calculated the mean and standard deviation.

Region-based analyses: order of activation. Since brain areas displayed a temporal dynamic pattern, we aimed to order the brain areas based on their c-fos+ expression. Ordering brain areas based on the time of their activation is not trivial, especially since in 3D microscopy time is discrete ($n_{time\ point}$ = 4) rather than continuous (as, for example, in fMRI). Additionally, c-fos protein is not transient, but it peaks ~90 min and decays ~180 min after a stimulus. This dynamic may even be different depending on the brain area [4]. With these challenges in mind, we aimed to analytically create a pseudo-time to increase the temporal resolution, which would in turn allow to order brain areas.
Among the approaches considered (Supplementary Table 3), we ultimately ordered areas based on the estimated time of mid-activation across blocks. c-fos+ cell counts were Z-transformed. Then, for each brain area we calculated the median across blocks of each time point. We interpolated a linear model between each two consecutive timepoints: this line is the "continuum" of pseudo-time. To order the brain areas, we then considered at which pseudo-time point, c-fos activation reached its mid-activation level. Since the data was Z transformed, this corresponded to reaching the value 0. To limit the sensitivity of the pseudo-time, we binned the pseudo-time variable in bouts of 10 minutes. Each brain area was therefore grouped to the closest bout (binning). This approach has the advantage to create a criterion on which to categorize brain areas, but it does not consider the range (error) among which it could happen. The approach is ideal for areas that have one point of activation; it is biased for brain areas with a biphasic activation (e.g., at the beginning and at the end of our time curve).

For interpretation and visualization purposes, we classified the brain areas (Supplementary Table 4) according to functional networks relevant to stress. We followed Henckens and colleagues' [15] results, to which the an amygdalar group was added.

Combining voxel based and single cell analysis: sub-brain areas. We questioned whether c-fos+ cells are uniform within a brain area, or whether there are locations in which c-fos+ cells are most dense, i.e., are in closer proximity to each other. For this, we selected a brain area important for the stress response, i.e. the basolateral amygdala (BLA). All other brain areas can be visualized on the *abc4d* app.

For each sample separately, we estimated the probability density of the BLA at each cell, by using a kernel density estimation with Gaussian function. Kernel densities are routinely used to smooth data from a finite sample to make inferences about a population. Here, they were used to estimate the cell density within a brain area. Next, we filtered the cells with highest density per sample.

The BLA was divided into voxels of 30μm per side. Considering that there could be an alignment error of (maximum) ~23μm (Supplementary Figure 2b), we considered 30μm the minimum, interpretable size. To look for consistency across samples, we calculated how many samples had at least one cell in each voxel, and considered 3 the minimum for consistency. In each xyz direction, we calculated per time point the median and interquartile of the voxels' position. Of note, due to the batch effects, calculating number of cells (or other measures of activation) across samples would have had little value, and would need to be standardized. We therefore opted for this more straightforward approach.

To determine whether our observations were due to a chance process, we randomly attributed each sample to a time point, and perform the exact same analysis.

Combining region-based and single cell analysis: strategy. We hypothesized that brain areas can show activation with different strategies. With the "count strategy", a brain area increases the *number* of c-fos+ cells with a low c-fos expression (intensity); with an "intensity strategy", c-fos+ cells increase in *intensity* rather than number.

To test this hypothesis, we analyzed t_0 samples, where we calculated the n_{cfos+} of each brain area, as well as the mean intensity of the cells in that area (intensity refers to the maximum intensity as reported by Imaris). Here, the mean rather than the median was intentionally used to be able to observe the increase in intensity due to a subgroup of cells within a brain area. For this analysis, we did not perform a batch effects correction; rather we took advantage of the differences across blocks. In our cell detection methodology ('Image processing: cell detection' in Supplementary Methods), cells are identified as c-fos+ depending on intensity. This relationship should always be the same, irrespective of batch effects or time points. To quantify the

relationship analytically, we therefore interpolated a linear model between the raw c-fos+ cell count and median intensity, of all brain areas of all samples.

The linear model was used as a discriminant criterion to classify whether a brain area had a strategy more towards intensity (above the regression line) or towards count (below the regression line). This categorization was performed for each brain area of each t_0 sample ($n_{animal} = 9$) independently. We then calculated the probability of a brain area to belong to a certain categorization. The resulting variable was continuous between the values of 0 (i.e., all samples had an intensity strategy) and 1 (i.e., all samples had a count strategy).

If the categorization of brain areas would be a random process, the probability of brain areas to belong to a certain categorization would be normally distributed around $\mu = 0.5$ under the null hypothesis (i.e. brain areas do not have a strategy). To validate that the null hypothesis would indeed follow a normal distribution, we performed a simulation study. In this study, we used the exact same analysis, but the values for intensity and c-fos+ cell count were drawn independently from a Poisson distribution $P(\lambda$). The value for lambda λ was selected by qualitatively comparing the distribution of intensity and count in the data with computer generated Poisson distributions with different lambdas. However, the interpretation would not change if different values of lambda would be selected.

Next, we questioned whether brain areas may change strategy after stress, relatively to t_0. Our experiment was not powered sufficiently to answer this particular research question, and therefore results should be interpreted as exploratory. From the categorization probability, we selected those brain areas that were consistent across samples, i.e. that had a specific categorization in at least 6 out of 9 samples. Of these, we selected those with a consistent change (increase or decrease) in count and / or intensity in at least one of the foot-shock time points (t_{30}, t_{90}, t_{180}). We calculated the pairwise difference between t_0 and each foot-shock time point for count as well as intensity. From this, we calculated the rate of change (count over intensity) per block for each time point. To compare data across brain areas, we converted the rates across samples to standardized mean differences (Hedge's g). We then classified brain areas as having changed strategy after stress if the effect size was below 0.5 or higher than 2, meaning that the relative increase in activity must have doubled towards intensity or towards count after stress compared to baseline.

Software. We developed the R package abc4d ("Analysis Brain Cellular activation in 4 Dimensions") to ease the implementation of data pre-processing and analysis. Furthermore, we developed a web tool (https://vbonapersona.shinyapps.io/brain_after_footshock/) to interactively visualize the effects of acute stress on the brain area of choice.

All analyses were conducted with R (version 4.0.0) in the R studio environment on a macOS Mojave (version 10.14.6). The following R packages were core to this study: 1) *tidyverse* (version 1.3.0) for general data handling and visualization; *shiny* (v 1.6.0) for the generation of the web interface; *ComplexHeatmap* (v 2.4.3) for heatmap visualization.

Supplementary Figures

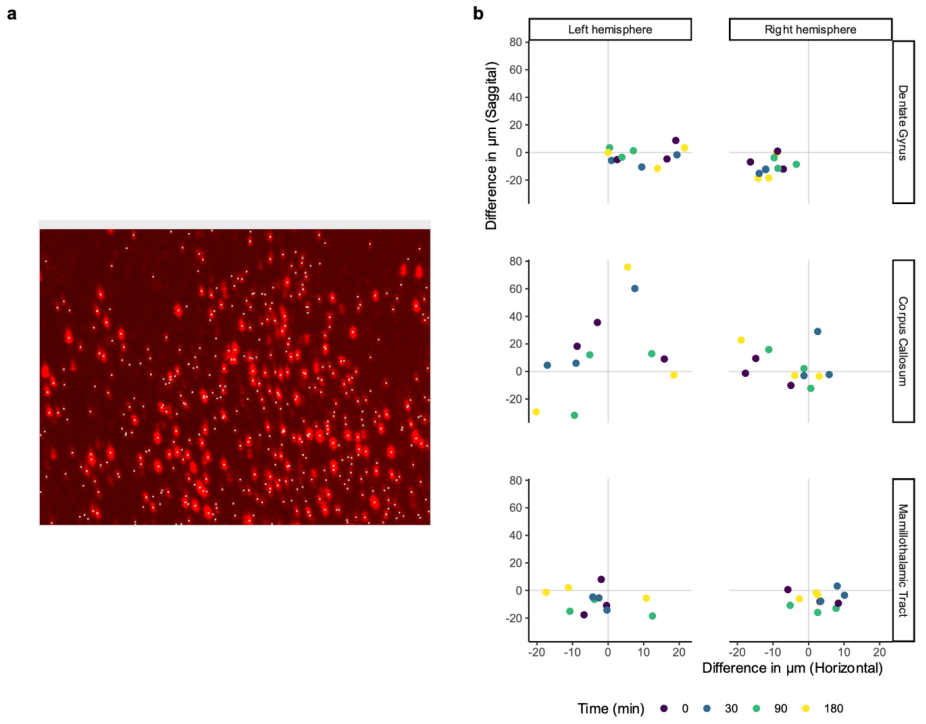


Figure S1. Cell detection and alignment validation. a) Representative example of staining of c-fos+ cells (bright red). White squares represent objects identified by the Imaris algorithm as cells. b) Validation of alignment. Error of the alignment represented as distance between real and aligned objects along the horizontal and sagittal axis. Error was calculated in three separate brain areas (horizontal facets) for n = 3 samples per time point.

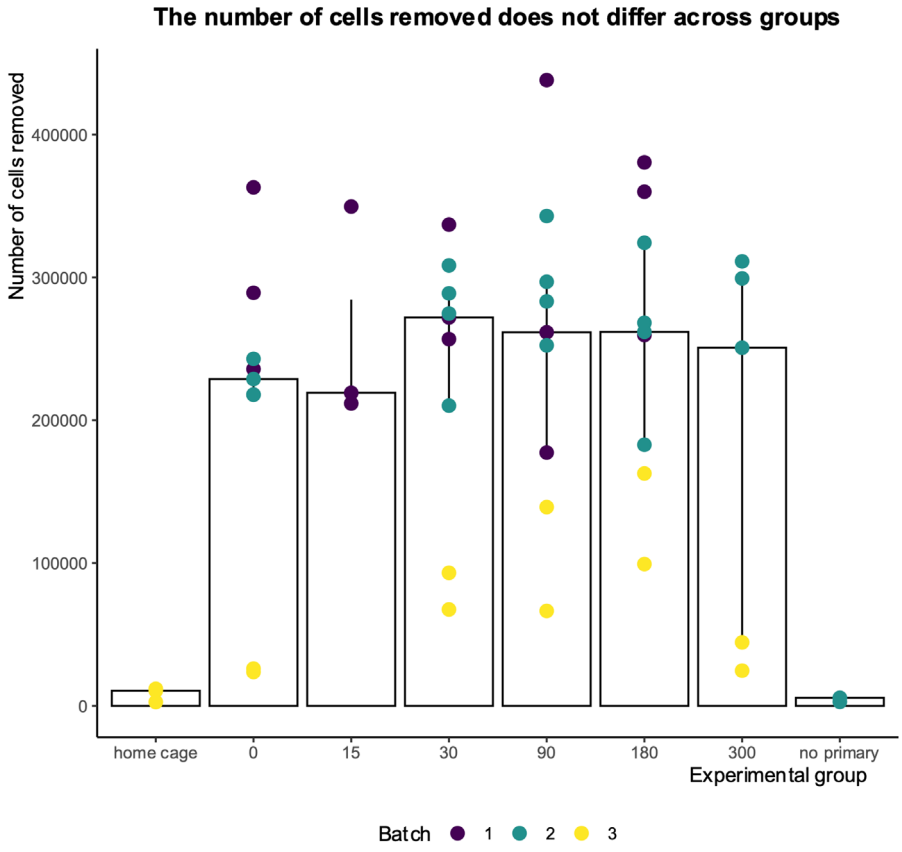


Figure S2. Data cleaning. Number of 'cells' removed during the data cleaning procedure across all groups. Each dot corresponds to one sample. Data presented as median and IQR. Of note, t_{15} and t_{300} were only investigated in control experiments, and not across all batches (Supplementary Figure 3). Sample sizes (n): $n_{home\ cage} = 3$; $n_{0} = 9$; $n_{15} = 3$; $n_{30} = 0$; $n_{90} = 9$; $n_{180} = 9$; $n_{300} = 6$; $n_{no\ primary} = 3$.

S

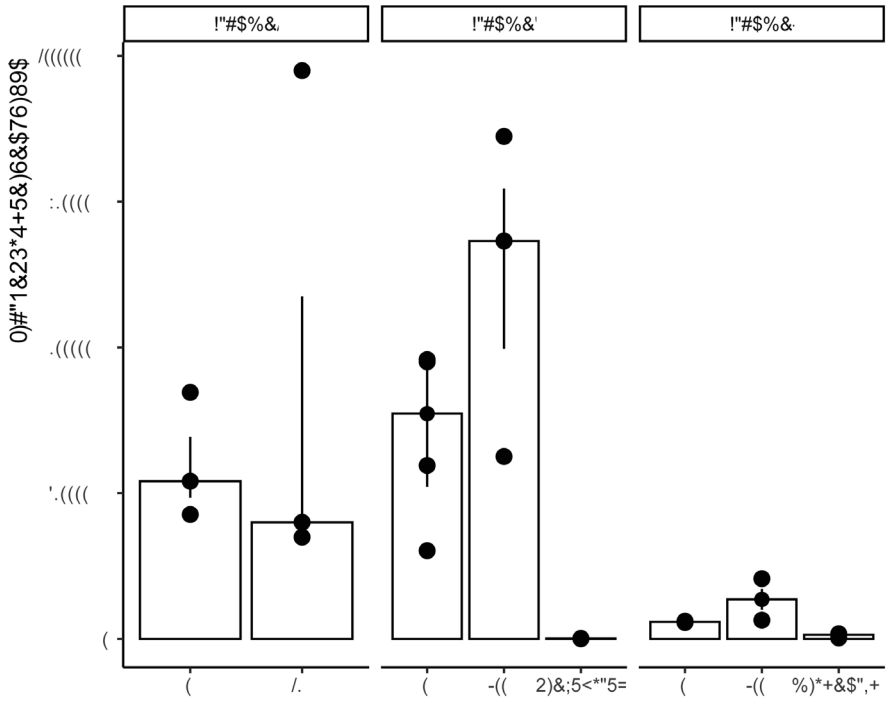


Figure S3. Total number of c-fos+ cells of control experiments. 15 min after foot-shock is insufficient to detect an increase in c-fos expression. At 300 minutes, c-fos+ cell count is comparable to t_0. Home cage group has lower c-fos+ cell count than the respective t_0 group. No primary antibody group has nearly no counts. Each dot represents a sample, with the bar indicating the median, and the errorbar the interquartiles (IQR). The control experiments are represented separately with t_0 groups of the same batch. Sample sizes (n): n_0 = 9 (3 + 4 + 2 in each batch); n_{15} = 3; n_{300} = 6; $n_{\text{no primary}}$ = 3; $n_{\text{home cage}}$ = 3.

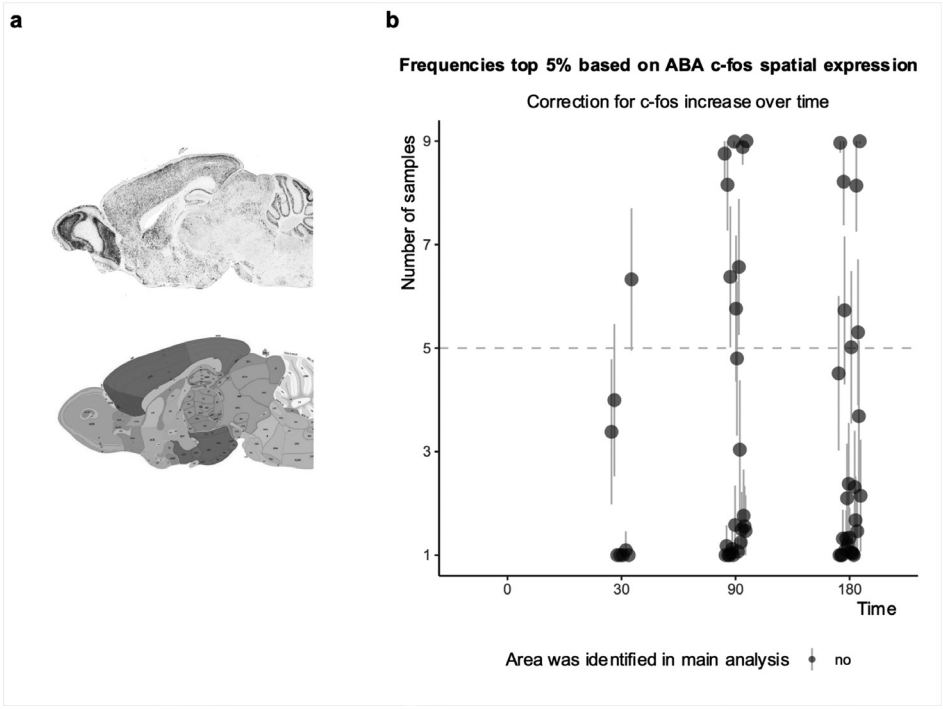


Figure S4. Simulation of most active brain areas based on c-fos ABA mRNA expression and increase of c-fos+ cells over time. a) Regional distribution of c-fos as displayed by the Allen Brain Atlas (ABA). Red: hypothalamic areas. b) Results of the simulation study. An *in silico* dataset was created by using a sampling approach. c-fos mRNA expression values were downloaded from all the experiments available at the ABA API ($n_{experiment} = 3$). These values were used to sample weights to mimic what one would expect if the data were only due to c-fos expression and increase of c-fos over time. The procedure was repeated 1000 times. Each dot corresponds to a brain area that was present in the top 5% of the c-fos+ cell counts (per thousand of total) in one block. The vertical line of each dot represents the 95% confidence interval across the 1000 simulated experiments. None of the brain areas that were consistent in more than 5 samples was present in the actual experimental data.

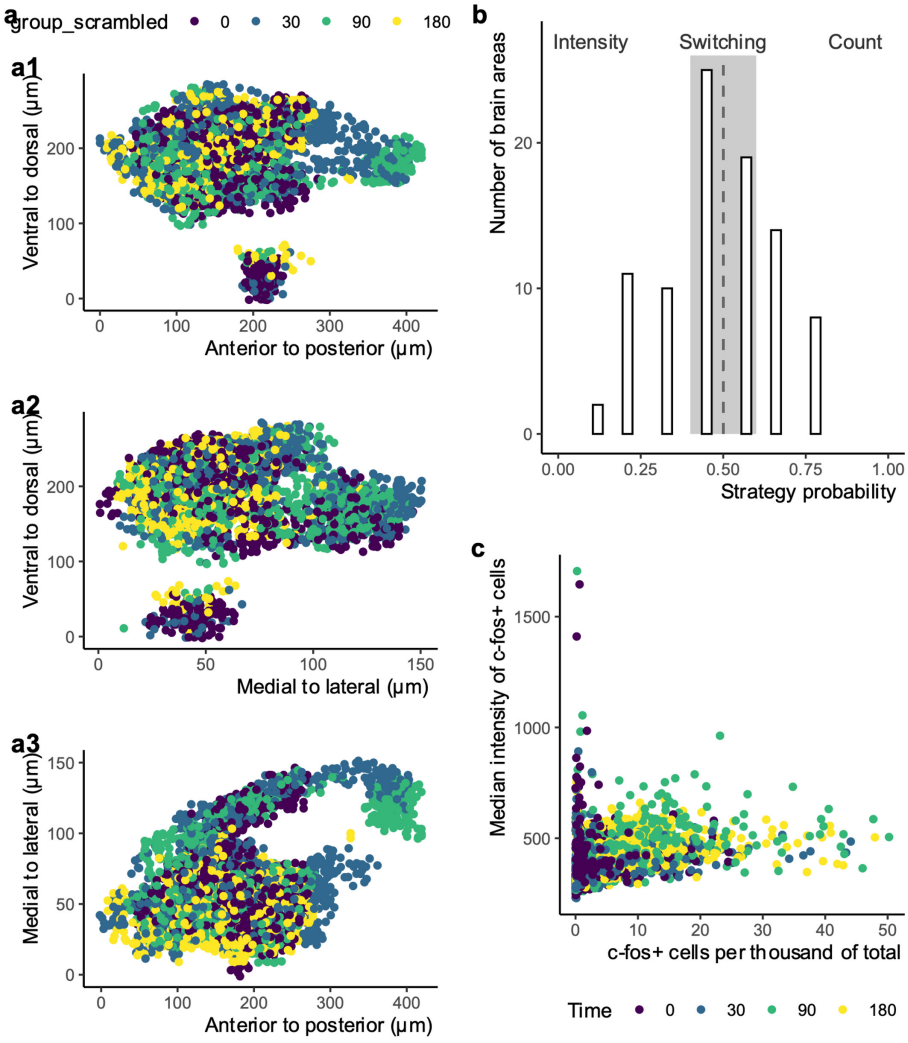


Figure S5. a) c-fos+ cells in highly dense areas across scrambled time point. a1, a2, a3 refer to the different 2D views of the xyz coordinates. Time points were randomly allocated for each sample, so that each block (i.e. a unit of 4 time points) had still one sample per time point. b) Strategy probability of brain areas according to our hypothesis, i.e. the relationship between count and intensity is due to the technical set-up. c) Relationship between c-fos+ count and median intensity. Contrary to expectations, count of c-fos+ cells and mean intensity are not correlated to each other.

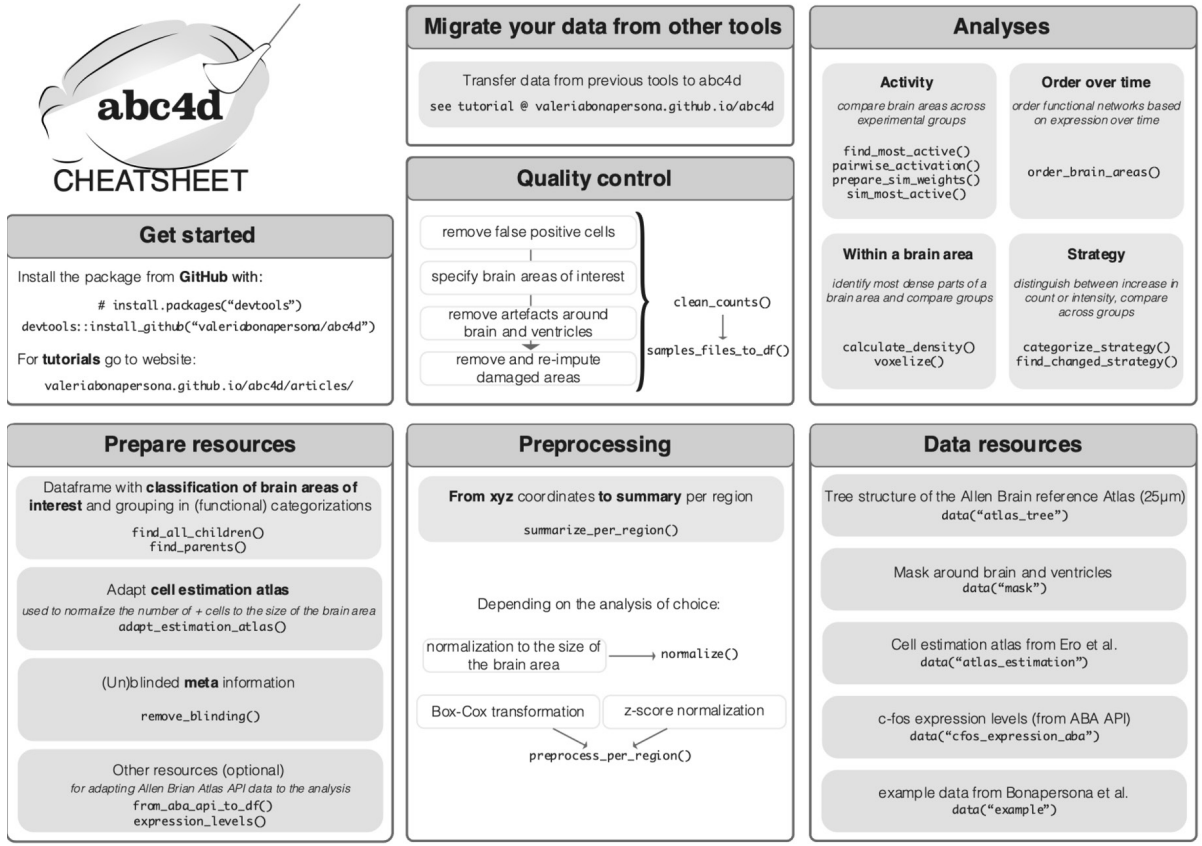


Figure S6. Cheat-sheet of *abc4d* package.

Supplementary Tables

Table S1. List of brain areas included in the analysis. The categorization follows the structure and acronym of the Allen Brain Reference atlas (25μm). ID is the code used by the ABA. Of note, we excluded 6 brain areas because we deemed that their size was too small for a confident alignment, namely: subfornical organ, vascular organ of the lamina terminalis, bed nucleus of the anterior commissure, bed nucleus of the accessory olfactory tract, fasciola chierea, induseum griseum.

ID	Name brain area	Acronym
23	Anterior amygdalar area	AAA
31	Anterior cingulate area	ACA
56	Nucleus accumbens	ACB
88	Anterior hypothalamic nucleus	AHN
95	Agranular insular area	AI
223	Arcuate hypothalamic nucleus	ARH
239	Anterior group of the dorsal thalamus	ATN
247	Auditory areas	AUD
295	Basolateral amygdalar nucleus	BLA
319	Basomedial amygdalar nucleus	BMA
351	Bed nuclei of the stria terminalis	BST
382	Field CA1	CA1
423	Field CA2	CA2
463	Field CA3	CA3
776	corpus callosum	cc
536	Central amygdalar nucleus	CEA
583	Claustrum	CLA
631	Cortical amygdalar area	COA
672	Caudoputamen	CP
784	corticospinal tract	cst
726	Dentate gyrus	DG
830	Dorsomedial nucleus of the hypothalamus	DMH
856	Thalamus polymodal association cortex related	DORpm
864	Thalamus sensory-motor cortex related	DORsm
814	Dorsal peduncular area	DP
895	Ectorhinal area	ECT
909	Entorhinal area	ENT
942	Endopiriform nucleus	EP
958	Epithalamus	EPI
1000	extrapyramidal fiber systems	eps
184	Frontal pole cerebral cortex	FRP
998	Fundus of striatum	FS
1057	Gustatory areas	GU
1105	Intercalated amygdalar nucleus	IA
44	Infralimbic area	ILA
51	Intralaminar nuclei of the dorsal thalamus	ILM
59	Intermediodorsal nucleus of the thalamus	IMD
131	Lateral amygdalar nucleus	LA
138	Lateral group of the dorsal thalamus	LAT
896	thalamus related	lfbst
194	Lateral hypothalamic area	LHA

ID	Name brain area	Acronym
226	Lateral preoptic area	LPO
275	Lateral septal complex	LSX
290	Hypothalamic lateral zone	LZ
290	Hypothalamic lateral zone	LZ
323	Midbrain motor related	MBmot
331	Mammillary body	MBO
339	Midbrain sensory related	MBsen
348	Midbrain behavioral state related	MBsta
362	Mediodorsal nucleus of thalamus	MD
403	Medial amygdalar nucleus	MEA
991	medial forebrain bundle system	mfbs
500	Somatomotor areas	MO
515	Medial preoptic nucleus	MPN
904	Medial septal complex	MSC
619	Nucleus of the lateral olfactory tract	NLOT
698	Olfactory areas	OLF
714	Orbital area	ORB
754	Olfactory tubercle	OT
780	Posterior amygdalar nucleus	PA
788	Piriform-amygdalar area	PAA
818	Pallidum dorsal region	PALd
826	Pallidum medial region	PALm
835	Pallidum ventral region	PALv
843	Parasubiculum	PAR
922	Perirhinal area	PERI
946	Posterior hypothalamic nucleus	PH
972	Prelimbic area	PL
1037	Postsubiculum	POST
1084	Presubiculum	PRE
1109	Parastrial nucleus	PS
63	Paraventricular hypothalamic nucleus descending division	PVHd
141	Periventricular region	PVR
149	Paraventricular nucleus of the thalamus	PVT
157	Periventricular zone	PVZ
165	Midbrain raphe nuclei	RAmb
254	Retrosplenial area	RSP
262	Reticular nucleus of the thalamus	RT
453	Somatosensory areas	SS
502	Subiculum	SUB
541	Temporal association areas	TEa
877	tectospinal pathway	tsp
589	Taenia tecta	TT
614	Tuberal nucleus	TU
629	Ventral anterior-lateral complex of the thalamus	VAL
669	Visual areas	VIS
677	Visceral area	VISC
685	Ventral medial nucleus of the thalamus	VM
693	Ventromedial hypothalamic nucleus	VMH
709	Ventral posterior complex of the thalamus	VP

Table S2. Missing values. a) List of missing animals with reasons. b) Damaged brain areas. These were removed from the analysis and re-imputed.

a) Missing animals

Batch	Animals missing	Reason
2	11 male animals, 40 female animals	Staining faded.
3	3 male animals	Staining faded (2 animals), scanning mistake (1 animal). The 3 animals belonged to two separate blocks. To not exclude both blocks completely, we merged the remaining sample into one block by selecting the best quality stainings.

b) Damaged areas

Sample ID	Damaged areas
13	AAA left, CP left, OT right, AAA right, CP right, ENTl right
14	CP left, AAA left, VISC right, Alp right
15	MOp right, SSp right
16	PIR left, ENTl left, PL right, MOs right
18	CTX left
19	MOp left, AId left, GU left, AIv left
21	SSs left, CP left, PAR left, HPF left, AAA right
22	RSPd left, VISp left, PRE left, AAA right
23	RSPd left, RSPagl left, VISpm left, RSPv left
24	SI left, FS left, CP left, AAA right, CP right
25	PTLp left
26	PAG left, ICe left, SCs left
27	SSs left, PAA left, PIR left, COApl left, COAa left
34	AAA right, CP right, SSs right, MEAav right, PERI right, ECT right
35	AAA right, CP right, CEAm right, AUDd right, VISl right, TEa right, PRE right, ec right, dhc right, PRE right, SUBv right
36	PERI left , Alp right, CP right, AAA right
38	PERI left , ENTl left
39	VISpl left , VISp left , ec left , dhc left , PRE left , MOs , MOp , TEa , ECT , ENTm , PAR , ec , dhc
40	VISC left, Alp left, CP left, ECT right, ec right, dhc right, PAR right, ENTl right
41	OT left, PIR left, CP left, OT right, PIR right, AAA right, FS right, CP right, Alp right
42	MOs left , RSPv left , RSPd left , NA left , MOs right, TEa right, ECT right, PERI right, SUBv-sp right, ENT right, PAR right
43	OT left , SI left , FS left , CP left , PTLp right, TEa right
44	AId left, MOp left, ORBl left, CP left, OT left, FS left, CP left, VISpl left, ec left, dhc left, POST left, OT right, FS right, CP right
45	OT left , FS left , CP left , ECT left , OT right, FS right, CP right
46	OT left, FS left, CP left, CP right, OT right, FS right, AUDd right, AUDpo right, ECT right
47	TTd left , AON left , AAA left , CP left , IA left , ENTl left , ECT left , CP right, AAA right, IA right, ENTl right, ECT right
48	OT left, SI left, ACB left, CP left, Alp left, OT right, SI right, FS right, CP right, ECT right, ENTl right
49	VISpl left, ECT left, RSPd right, RSPv right

Sample ID	Damaged areas
50	AId left, AON left, PIR left, AIv left, GU left, AAA right, CP right
51	AAA left, CP left, VISp left, VISal left, VISam left, VISp left, VISl left, POST left, PRE left, AAA right, CP right, AUDd right, PTLp right, ENT right, PAR right
51	SSs left, CP left, ECT left, TEa left, ENTm right
106	SI right, FS right, SUBv-sp right
107	ACB left, OT left, PIR left, CP left, SI left
108	SSp left , GU left , OT right, SI right, ACB right, CP right, POST right, ec right, VISpm right, VISp right, RSPv right, RSPd right
109	MOs right
112	ECT right, PERI right, ENTl right
113	VISpm left, VISp left, ec left, dhc left, POST left, MOs right, MOp right, PRE right, HPF right, SUBv-sp right, alv right, ec right
115	OT left
116	SUBv-sp right
117	PIR right, PAR right, ENTm right, ENTl right
118	SSs left , FS left , act left , CP left
119	SSp right
120	CP left
121	CP right
122	TEa right, SUBv-sp right, PAR right

Table S3. List of analytical approaches considered for ordering brain areas based on c-fos activation.

Approach	Brief explanation	Not pursued because:
Clustering	Clustering to reduce dimensions, then order the cluster groups. The pseudo-time would then have resolution equal to the number of clusters. Ordering could be achieved by comparing to a simulated model of possible clusters out of theory (example: cluster with only initial activation at t^{30}; cluster with activation at t^{90}; clustering with activation at t^{90} as well as t^{180}).	All brain areas were activated; therefore, very minimal difference would appear between clusters. Furthermore, creating "expected" cluster models is not trivial.
Derivatives	Identify the steepest derivative between each two consecutive time points. This can be performed per sample (probabilistic approach) or on the median across samples. It might be able to identify multiple activations (e.g. if first and third derivatives are steeper than the second).	Too many rules (e.g. only one derivative is the steepest, two derivatives are the steepest...), therefore it has the same problem as creating the "expected cluster" model as described above. Furthermore, very pseudo-time resolution (n = 3).
Peaks	Fit a loess curve for each sample and identify the maxima. Each maxima is considered a peak. Advantage that it can identify multiple activations for a single brain area	Since all brain areas were so activated, many peaks would appear within the same range (poor pseudo-time resolution). As a consequence, too much importance would be given to the type of curve used to fit the data.

Table S4. Functional categorization of brain areas. Brain areas were classified in functional groups relevant to the stress response, by adapting [15]. *cx = cortex.*

Functional categorization	Brain area	Acronym
Amygdala	Anterior amygdalar area	AAA
Amygdala	Basolateral amygdalar nucleus	BLA
Amygdala	Basomedial amygdalar nucleus	BMA
Amygdala	Central amygdalar nucleus	CEA
Amygdala	Cortical amygdalar area	COA
Amygdala	Intercalated amygdalar nucleus	IA
Amygdala	Lateral amygdalar nucleus	LA
Amygdala	Medial amygdalar nucleus	MEA
Amygdala	Posterior amygdalar nucleus	PA
Amygdala	Piriform-amygdalar area	PAA
Hippocampus	Field CA1	CA1
Hippocampus	Field CA2	CA2
Hippocampus	Field CA3	CA3
Hippocampus	Dentate gyrus	DG
Hippocampus	Entorhinal area	ENT
Hippocampus	Parasubiculum	PAR
Hippocampus	Postsubiculum	POST
Hippocampus	Presubiculum	PRE
Hippocampus	Subiculum	SUB
Hypothalamus	Anterior hypothalamic nucleus	AHN
Hypothalamus	Arcuate hypothalamic nucleus	ARH
Hypothalamus	Dorsomedial nucleus of the hypothalamus	DMH
Hypothalamus	Lateral hypothalamic area	LHA
Hypothalamus	Lateral preoptic area	LPO
Hypothalamus	Hypothalamic lateral zone	LZ
Hypothalamus	Hypothalamic lateral zone	LZ
Hypothalamus	Mammillary body	MBO
Hypothalamus	Medial preoptic nucleus	MPN
Hypothalamus	Posterior hypothalamic nucleus	PH
Hypothalamus	Parastrial nucleus	PS
Hypothalamus	Paraventricular hypothalamic nucleus descending division	PVHd
Hypothalamus	Periventricular region	PVR
Hypothalamus	Periventricular zone	PVZ
Hypothalamus	Tuberal nucleus	TU
Hypothalamus	Ventromedial hypothalamic nucleus	VMH
Motor cx	Somatomotor areas	MO
Prefrontal cx	Anterior cingulate area	ACA
Prefrontal cx	Orbital area	ORB
Prefrontal cx	Prelimbic area	PL
Prefrontal cx	Taenia tecta	TT
Primary somatosensory cx	Somatosensory areas	SS
Thalamus	Anterior group of the dorsal thalamus	ATN

Functional categorization	Brain area	Acronym
Thalamus	Thalamus polymodal association cortex related	DORpm
Thalamus	Thalamus sensory-motor cortex related	DORsm
Thalamus	Epithalamus	EPI
Thalamus	Intralaminar nuclei of the dorsal thalamus	ILM
Thalamus	Intermediodorsal nucleus of the thalamus	IMD
Thalamus	Lateral group of the dorsal thalamus	LAT
Thalamus	Mediodorsal nucleus of thalamus	MD
Thalamus	Paraventricular nucleus of the thalamus	PVT
Thalamus	Reticular nucleus of the thalamus	RT
Thalamus	Ventral anterior-lateral complex of the thalamus	VAL
Thalamus	Ventral medial nucleus of the thalamus	VM
Thalamus	Ventral posterior complex of the thalamus	VP
Visual cx	Visual areas	VIS

Supplementary references

1. Hughes, P. & Dragunow, M. Induction of immediate-early genes and the control of neurotransmitter-regulated gene expression within the nervous system. 47, 133–178 (1995).
2. Kouzarides, T. & Ziff, E. The role of the leucine zipper in the fos–jun interaction. *Nature* 336, 646–651 (1988).
3. Luckman, S., Dyball, R. & Leng, G. Induction of c-fos expression in hypothalamic magnocellular neurons requires synaptic activation and not simply increased spike activity. *J Neurosci* 14, 4825–4830 (1994).
4. Kovács, K. J. Invited review c-Fos as a transcription factor: a stressful (re)view from a functional map. *Neurochemistry International* 33, 287–297 (1998).
5. Qian, X. et al. A Rapid Release of Corticosteroid-Binding Globulin from the Liver Restrains the Glucocorticoid Hormone Response to Acute Stress. *Endocrinology* 152, 3738–3748 (2011).
6. Kimbrough, A. et al. Brain-wide functional architecture remodeling by alcohol dependence and abstinence. *Proc Natl Acad Sci* USA 117, 2149–2159 (2020).
7. Renier, N. et al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. *Cell* 165, 1789–1802 (2016).
8. Wheeler, A. L. et al. Identification of a Functional Connectome for Long-Term Fear Memory in Mice. *PLOS Computational Biology* 9, e1002853 (2013).
9. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953. e20 (2020).
10. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging* 29, 196–205 (2010).
11. Cullinan, W. E., Herman, J. P., Battaglia, D. F., Akil, H. & Watson, S. J. Pattern and time course of immediate early gene expression in rat brain following acute stress. *Neuroscience* 64, 477–505 (1995).
12. Morimoto, M., Morita, N., Ozawa, H., Yokoyama, K. & Kawata, M. Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat brain: an immunohistochemical and in situ hybridization study. *Neuroscience Research* 26, 235–269 (1996).
13. Erö, C., Gewaltig, M.-O., Keller, D. & Markram, H. A Cell Atlas for the Mouse Brain. Front. *Neuroinform*. 12, (2018).
14. McReynolds, J. R., Christianson, J. P., Blacktop, J. M. & Mantsch, J. R. What does the Fos say? Using Fos-based approaches to understand the contribution of stress to substance use disorders. *Neurobiology of Stress* 9, 271-285 (2018).
15. Henckens, M. J. A. G. et al. Stress-induced alterations in large-scale functional networks of the rodent brain. *NeuroImage* 105, 312–322 (2015).

Supplementary Information for Chapter 3

The STRESS-NL database: a resource for human acute stress studies across the Netherlands.

Contents

Supplementary tables

Table S1. Table summarizing the information contained in the meta-data.

Variable	Description	Type
Sample size	Sample size used in the study, and available in the database	Numeric
Acute stress test	Type of acute stress test used in the experimental procedure	Categorical
Time of the day	Time of the day during which the experiment was conducted	Categorical
Cortisol	Whether cortisol was measured	Boollean
Alpha amylase	Whether alpha amylase was measured	Boollean
Subjective stress	Whether subjective stress was measured	Boollean
Questionnaires	List of questionnaires conducted	Descriptive string
Cognitive tasks	List of cognitive tasks conducted	Descriptive string
Physiological tests	List of physiological tests (e.g. heart rate)	Descriptive string
Brain activity	List of measures of brain activity (e.g. EEG, MRI, fMRI) taken	Descriptive string
Genetics	Whether genetics information (e.g. GWAS, SNPs) were measured	Bollean
Transcriptomics, epigenomics, proteomics, metabolomics	Whether any of these were measured	Bollean
Immune system	List of immune system measures collected	Descriptive string
Biobank	Whether biobank tissues are available	Bollean
Other	List of other measures that do not fall in the previous categories	Descriptive string

Table S2. Summary of the general characteristics of the Stress-nl studies. Of note, the summary statistics provided refer to the numbers included in the database and they may differ to what published in the original studies. Order *Fig. 4 = Order of the experiments as in Figure 4. Fem = females*

Order Fig. 4	Author (year)	Doi	Acute stress test	Sample size	% Fem.	Age	Health status	N cortisol timepoint	N cortisol samples
1	Giesbrecht (2007)	10.1097/01.nmd.0000253822.60618.60	TSST	125	48.8%	19.8[2.6]	healthy	4	500
2	Smeets (2011)	10.1016/j.psyneuen.2010.08.001	SECPT	152	55.3%	19.9[1.84]	healthy	5	456
3	de Brouwer (2014)	10.1111/bjd.12697	TSST	89	46.1%	58.6[10.34]	healthy, rheumatism, psoriasis	6	516
4	Oei (2012)	10.1093/scan/nsr024	TSST	34	0%	24.2[3.41]	healthy	5	170
5	Quaedflieg (2013a)	10.1111/psyp.12058	MAST	105	19%		healthy	8	518
6	Hartman (2013)	10.1016/j.biopsycho.2013.05.009	adapted TSST	211	39.3%	16[0]	preadolescent DSM-IV diagnosis	5	1038
7	Smeets (2007)	10.1016/j.biopsycho.2007.07.001	TSST	52	75%	23.1[3.81]	healthy	2	104
8	Houtepen (2015)	10.1016/j.euroneuro.2014.10.005	adapted TSST	127	54.3%	45.5[14.18]	healthy, borderline disorder 1 and siblings	7	873
10	Nelemans (2017)	10.1016/j.biopsycho.2017.03.003	PST	344	44.2%	17.2[0.52]	SAD	7	2362
11	Smeets (2006a)	10.1016/j.biopsycho.2005.09.004	TSST	150	50%	19.8[2.52]	healthy	4	600
12	van Campen (2015)	10.1093/brain/awv157	adapted TSST	93	61.3%	11.5[3.16]	healthy, epilepsy	9	837
13	Roelofs	unpublished	MAST	92	46.7%	17.2[0.15]	healthy	6	551
14	Zhang (2019)	10.1016/j.neuroimage.2019.01.063	SCEPT, MA	335	22%		healthy	5	
16	Smeets (2010)	10.1016/j.biopsycho.2010.02.015	TSST	68	50%	22.3[3.05]	healthy	5	340
17	Smeets (2006b)	10.1016/j.ijpsycho.2005.11.007	TSST	60	50%	19.6[1.82]	healthy	4	236
18	Smeets (2012)	10.3758/s13423-011-0180-z	TSST	80	0%	22.5[4.52]	healthy	7	553
19	Bouma (2009)	10.1016/j.psyneuen.2009.01.003	adapted TSST	715	50.9%	15.6[0.66]	healthy	5	3492
20	Quaedflieg (2013b)	10.1016/j.psyneuen.2013.09.002	MAST	34	50%	21.4[3.43]	healthy	5	170
21	Oei (2014)	10.1016/j.psyneuen.2013.10.005	TSST	40	0%	25.8[17.28]	healthy	4	157
22	Quaedflieg (2015)	10.1016/j.biopsycho.2014.11.014	MAST	69	56.5%	20.8[2.73]	healthy	7	474
23	Smeets (2009a)	10.1016/j.psyneuen.2009.03.001	TSST	48	0%	20.7[3.29]	healthy	5	240
23	Tollenaar (2009)	10.1080/09658210802665845	TSST	110	0%	21.5[3.09]	healthy	7	532

Order Fig. 4	Author (year)	Doi	Acute stress test	Sample size	% Fem.	Age	Health status	N cortisol timepoint	N cortisol samples
24	Smeets (2012)	10.1016/j.psyneuen.2012.04.012	cold pressure test, SECPT, P SECPT, MAST, TSST	160	0%	21.6[2.93]	healthy	7	1000
25	Bakvis (2009)	10.1016/j.yebeh.2009.09.006	cold pressure test	39	76.9%	33.2[12.07]	healthy, psychogenic seizures	9	350
26	de Rooji (2006)	10.1016/j.psyneuen.2006.09.007	social stressor	694	52%	58.2[0.96]	not available	7	
28	Evers	unpublished	M-PASAT	60	0%	22.8[3.34]	healthy	4	714
29	Jansen (2016)	unpublished	TSST	130	0%	20.5[1.67]	healthy	4	485
30	Bakvis (2010)	10.1111/j.1528-1167.2009.02394.x	TSST	39	84.6%	24.8[6.47]	Healthy, psychogenic seizures	11	428
31	Roelofs (2009)	10.1016/j.biopsych.2008.08.022	TSST	18	50%	24[0]	social phobia	10	180
31	Fernandez	10.1016/j.biopsych.2014.07.034	aversive movie	240	0%	21.9[2.62]	healthy	3	718
32	Smeets (2006c)	10.1037/0735-7044.120.6.1204	TSST	40	0%	19.2[1.36]	healthy	7	278
33	Oei (2018)	10.1016/j.biopsycho.2018.02.018	TSST	116	48.3%	66.4[6.13]	healthy, alcohol addiction, past diagnosis of PTSD or MDD or burnout or anxiety disorder	4	453
34	Evers	unpublished	TSST	48	100%	21.9[2.31]	healthy	6	286
35	Cornelisse (2011)	10.1038/npp.2011.162	TSST	140	38.6%	21.1[1.17]	healthy	6	377
36	Oei (2006)	10.1080/10253890600965773	TSST	20	0%	22[3.93]	healthy	5	67
37	Vinkers (2013)	10.1016/j.psyneuen.2012.12.012	adapted TSST	159	50.3%	23.5[0]	healthy	8	1257
38	de Brower (2011)	10.1371/journal.pone.0027432	TSST	74	58.1%	58.8[10.77]	rheumatism	6	432
39	Smeets (2009b)	10.1016/j.yhbeh.2009.01.011	TSST	64	50%	25.9[4.33]	healthy	4	251
40	Hermans (2011)	10.1126/science.1209603	aversive movie	160	33.8%	22.5[3.55]	healthy	2	320
41	Tollenaar (2020)	10.1016/j.heliyon.2020.e04488	TSST	124	58%%	24.1[3.51]	healthy	3	371
42	Tollenaar (2008)	10.1016/j.actpsy.2007.10.007							

Supplementary Information for Chapter 4

The behavioral phenotype of early life adversity: a 3-level meta-analysis of rodent studies.

Contents

Supplementary Methods

Definitions

The table below reports an explanation/definition of (technical) terms used throughout the manuscript.

Terms	Definitions and assumptions
Behavioral test	Experimental method to measure behavior in a standardized manner
Outcome variables	Outcomes reported for each behavioral test. A complete list of included behavioral tests and related variables can be found in S1.3.
Individual comparison	Each effect size measured between a control and an experimental group with a history of ELA. Often, multiple outcome variables were measured and reported for each behavioral test performed. In an attempt to limit hierarchy of the data, we rated a priori how well a variable described the behavioral domain that the test aimed to operationalize. If multiple variables were reported, we selected only the one with the highest rating. It follows that each behavioral test is represented in the dataset by only one individual comparison. Rating of the variables for each behavior test can be found in S1.3.
Experiment	Ensemble of individual comparisons (each representing a different behavioral test) from the same groups of animals. Individual comparisons within the same experiment were considered dependent on each other. Any publication can report multiple experiments. Individual comparisons from different experiments within the same publication are considered independent of each other as they derive from different animals. If a publication did not mention that different cohorts of animals were used, we assumed that all behavioral tests and related outcomes were performed in the same animals and therefore belonged to the same experiment.
Nest	Unit of aggregation in the multi-level model. Here, it corresponds to the "experiment" level as individual comparisons within the same experiment derive from the same animals and are therefore dependent on each other.

Search string

PubMed

("early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "early life adversity*" OR "early stress"[tiab] OR "neonatal stress"[tiab] OR "postnatal stress"[tiab] OR "perinatal stress"[tiab] OR "neonatally stressed"[tiab] OR "early adverse experience"[tiab] OR "perinatally stressed"[tiab] OR "early adverse experiences"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "maternal separation"[tiab] OR "maternal deprivation"[tiab] OR "maternal care"[tiab] OR "isolation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR licking and grooming"[tiab] OR "licking-grooming"[tiab] OR "licking/grooming"[tiab])

AND

("murine"[tiab] OR "rodentia"[tiab] OR "rodent"[Tiab] OR "rodents"[Tiab] OR "rodentia"[tiab] OR mus[Tiab] OR murinae[Tiab] OR muridae[Tiab] OR "mice"[MeSH Terms] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[MeSH Terms] OR "rat"[tiab] OR "rats"[tiab])

AND

("Behavior, Animal"[Mesh] OR "behaviour"[tiab] OR "behavior"[tiab] OR "behaviours"[tiab] OR "behaviors"[tiab] OR "behav*"[tiab] OR "behavioural test"[tiab] OR "behavioural tests"[tiab] OR "behavioral test"[tiab] OR "behavioral tests"[tiab] OR "test, behavioral"[tiab] OR "test, behavioural"[tiab] OR "tests, behavioral"[tiab] OR "tests,behavioural"[tiab] OR "anxiety"[tiab] OR "fear"[tiab] OR "anxiety/fear"[tiab] OR "anxiety-fear"[tiab] OR "emotional learning"[tiab] OR "non-stressful learning"[tiab] OR "non stressful learning"[tiab] OR "social behaviour"[tiab] OR "social behavior"[tiab] OR "sexual behaviour"[tiab] OR "sexual behavior"[tiab] OR "radial arm"[tiab] OR "T maze"[tiab] OR "Ymaze"[tiab] OR "what where which task"[tiab] OR "what-where-which task"[tiab] OR "object in location"[tiab] OR "object in context"[tiab] OR "object recognition"[tiab] OR "object discrimination"[tiab] OR "barnes maze"[tiab] OR "holeboard"[tiab] OR "circular maze"[tiab] OR "Morris water maze"[tiab] OR "spontaneous alteration task"[tiab] OR "maze learning"[tiab] OR "active avoidance"[tiab] OR "spring test"[tiab] OR "inhibitory avoidance"[tiab] OR "passive avoidance"[tiab] OR "defensive withdrawal"[tiab] OR "fear conditioning"[tiab] OR "cat box"[tiab] OR "elevated plus maze"[tiab] OR "EPM"[tiab] OR "cross maze"[tiab] OR "open field"[tiab] OR "concentric square field test"[tiab] OR "place preference"[tiab] OR "place avoidance"[tiab] OR "light/dark test"[tiab] OR "light dark test"[tiab] OR "light-dark test"[tiab] OR "light/dark box"[tiab] OR "light dark box"[tiab] OR "light-dark box"[tiab] OR "object exploration"[tiab] OR "square field test"[tiab] OR "shuttle box"[tiab] OR "social interaction"[tiab] OR "three chambers"[tiab] OR "3 chambers"[tiab] OR "three

chamber"[tiab] OR "3 chamber"[tiab] OR "1 chamber"[tiab] OR "one chamber"[tiab] OR "emotional witness stress"[tiab] OR "social play"[tiab] OR "social approach test"[tiab] OR "social encounter test"[tiab] OR "social interaction test"[tiab] OR "social preference test"[tiab] OR "social learning"[tiab] OR "social preference"[tiab] OR "social hierarchy"[tiab] OR "dominance"[tiab] OR "tube test"[tiab] OR "resident test" [tiab] OR "intruder test"[tiab] OR "resident intruder test"[tiab] OR "resident/intruder test"[tiab] OR "resident-intruder test"[tiab] OR "competitive behaviour"[tiab] OR "competitive behaviour"[tiab] OR "play fighting behaviour"[tiab] OR "play fighting behaviour"[tiab] OR "play-fighting behaviour"[tiab] OR "play-fighting behavior"[tiab] OR "play/fighting behaviour"[tiab] OR "play/fighting behavior"[tiab])

WebOfScience

"early life stress" OR "ELS" OR "early life adversity" OR "early life adversities" OR "early life adversity*" OR "early stress" OR "neonatal stress" OR "postnatal stress" OR "perinatal stress" OR "neonatally stressed" OR "early adverse experience" OR "perinatally stressed" OR "early adverse experiences" OR "postnatal manipulation" OR "postnatal manipulations" OR "perinatal manipulation" OR "perinatal manipulations" OR "maternal separation" OR "maternal deprivation" OR "maternal care" OR "isolation" OR "limited bedding" OR "limited nesting" OR "limited material" OR "licking and grooming" OR "licking-grooming" OR "licking/grooming"

AND

"murine" OR "rodentia" OR "rodent" OR "rodents" OR "rodentia" OR mus OR murinae OR muridae OR "mice" OR "mouse" OR "rat" OR "rats"

AND

"behaviour" OR "behavior" OR "behaviours" OR "behaviors" OR "behav*" OR "behavioural test" OR "behavioural tests" OR "behavioral test" OR "behavioral tests" OR "test, behavioral" OR "test, behavioural" OR "tests, behavioral" OR "tests,behavioural" OR "anxiety" OR "fear" OR "anxiety/fear" OR "anxiety-fear" OR "emotional learning" OR "non-stressful learning" OR "non stressful learning" OR "social behaviour" OR "social behavior" OR "sexual behaviour" OR "sexual behavior" OR "radial arm" OR "T maze" OR "Ymaze" OR "what where which task" OR "what-where-which task" OR "object in location" OR "object in context" OR "object recognition" OR "object discrimination" OR "barnes maze" OR "holeboard" OR "circular maze" OR "Morris water maze" OR "spontaneous alteration task" OR "maze learning" OR "active avoidance" OR "spring test" OR "inhibitory avoidance" OR "passive avoidance" OR "defensive withdrawal" OR "fear conditioning" OR "elevated plus maze" OR "EPM" OR "cross maze" OR "open field" OR "concentric square field test" OR "place preference" OR "place avoidance" OR "light/dark test" OR "light dark test" OR "light-dark test" OR "light/dark box" OR "light dark box" OR

"light-dark box" OR "object exploration" OR "square field test" OR "shuttle box" OR "social interaction" OR "three chambers" OR "3 chambers" OR "three chamber" OR "3 chamber" OR "1 chamber" OR "one chamber" OR "emotional witness stress" OR "social play" OR "social approach test" OR "social encounter test" OR "social interaction test" OR "social preference test" OR "social learning" OR "social preference" OR "social hierarchy" OR "dominance" OR "tube test" OR "resident test" OR "intruder test" OR "resident intruder test" OR "resident/intruder test" OR "resident-intruder test" OR "competitive behaviour" OR "competitive behaviour" OR "play fighting behaviour" OR "play fighting behaviour" OR "play-fighting behaviour" OR "play-fighting behavior" OR "play/fighting behaviour" OR "play/fighting behavior"

Classification of behavioral tests in behavioral domains

Prior the beginning of the study, four experts (JK, RvdV, MJ & Ruth Damsteegt) were consulted for the selection of tests and related outcomes, as well as for their classification in behavioral domains. Overall, we aimed to extract the variable for each test that best represented the described domain. However, often the most reported outcome for a certain test is not necessarily the best at representing the categorized domain. For example, in the anxiety-like test "elevated plus maze", the most reported outcome is "time spent in open arms". Arguably "time spent in the closed arms" is a more direct measure of anxiety-like behavior: the more anxious the animal, the more the time spent in closed arms. In such circumstances, if a paper reported both outcomes, we extracted the most common (in this case, time spent in the open arms), in the intent to avoid unnecessary heterogeneity. The experts agreed on which variables for each test best expressed the categorized behavioral domain (e.g. anxiety-like, memory), and ranked them based on their importance.

Table legend: Importance = ranking for variable selection (for details see above); multiplication effect size for model = effect sizes were multiplied whenever necessary by -1 so that an increase in Hedge's G would indicate an increase in anxiety-like behavior, improved memory after stressful learning, impaired memory after neutral learning and decreased social behavior; † = Tests included in the systematic review, but not in the confirmatory analysis. Specific inclusion/exclusion criteria for these tests are specified in the footnote below.

Behavioral test	Outcomes reported	Importance	Multiplication effect size for model	Comments
Anxiety-like				
Defensive withdrawal	Time spent in the center	1	-1	
	Time spent in the tube	2	1	
	Latency to exit the tube	3	1	
Elevated zero maze	Time spent in open arms	1	-1	
	Time spent in closed arms	2	1	
	entriesOpen	3	-1	

Behavioral test	Outcomes reported	Importance	Multiplication effect size for model	Comments
Elevated plus maze	Time spent in open arms	1	-1	
EPM	Amount entries in open arms	2	-1	
Fear conditioning	Amount of time spent freezing	1	1	After first footshock (no retention)
Forced swim test	Time spent immobile	1	1	during first exposure
	Time spent struggling	2	-1	during first exposure
Light/Dark Box	Time spent in the dark compartment	1	1	
	Time spent in the light compartment	2	-1	
	Latency to enter the dark compartment	3	-1	animal starts test in light compartment
Novelty-induced reduction of feeding/ drinking	Latency to feed	1	1	
	Time spent in the center	2	-1	
	Time spent feeding	3	-1	
Open field *	Time spent in the center	1	-1	
	Time spent in the periphery	2	1	
	Distance moved in the center	3	-1	
	Latency to enter the center	4	1	animal starts test in the periphery
	Amount entries in the center	5	-1	
Tail suspension test	Amount of time spent immobile	1	1	
Memory after stressful learning				
Fear Conditioning	Amount of time spent freezing	1	1	At re-exposure (retention time scored in separate variable). Fear can also be "social"
Forced swim test	Time spent immobile	1	1	At re-exposure (retention time scored in separate variable)
	Latency to immobility	2	-1	At re-exposure (retention time scored in separate variable)
	Distance moved	3	-1	At re-exposure (retention time scored in separate variable)
	Frequency immobility scored	4	1	At re-exposure (retention time scored in separate variable)
Morris water Maze (water temperature <24°C)*	Time spent in target quadrant	1	1	
	Distance swum in target quadrant	2	1	
	Latency to find platform	3	-1	If probe trial not present*
Shuttle box	Amount of avoidance responses	1	1	
	Latency to avoid	2	-1	
	Latency to enter avoidance compartment	2	-1	Same as latency to avoid

Behavioral test	Outcomes reported	Importance	Multiplication effect size for model	Comments
Memory after neutral learning				
Morris water Maze (water temperature >26°C)*	Time spent in target quadrant	1	1	
	Distance swum in target quadrant	2	1	
	Latency to find platform	3	-1	If probe trial not present*
Object in context	Discrimination index	1	-1	
Object in location	Discrimination index	1	-1	
	Time spent with novel object	2	-1	
Object recognition	Discrimination intex	1	-1	
	Time spent with novel object	2	-1	
	Ratio time spent novel / familiar object	3	-1	
Social recognition	Discrimination index	1	-1	novel vs familiar animal
	Time spent with novel animal	2	-1	novel vs familiar animal
Temporal order task	Discrimination index	1	-1	
T maze	Time spent in novel arm	1	-1	
Y maze	Time spent in novel arm	1	-1	
Social behavior				
Resident intruder test	Time spent in aggressive behavior	1	1	
	Tiime spent attacking	2	1	
	Latency to first aggression	3	-1	
	Latency to first attack	4	-1	
	Total amount of aggressive behavior	5	1	
	Time spent in offensive posture	6	1	
Social interaction	Time spent in social interaction	1	-1	
	Amount of social interaction	2	-1	
Social open field	Time spent in social proximity	1	-1	
Social play	Time spent in social interaction	1	-1	
Social preference	Preference index	1	-1	animal vs inanimate object
	Time spent in social interaction	2	-1	
	Time spent in social tube	3	-1	
Other †				
8-arms radial maze	Errors in working memory			
T maze	Alternation			
Y maze	Alternation			
Step down inhibitory avoidance	Latency to step down			at retest
Morris Water Maze *				If water temperature between 24°C and 26°C

* = Inclusion/exclusion criteria for specific tests:

- Open field is included only if test length <15 min. If test length >15min, the test is considered a measure of locomotor activity and not anxiety-like behavior.
- In the open field, "amount of crossings" is considered a measure of locomotor activity and not anxiety-like behavior. Tests reporting this as the only measure are not included.
- The Morris Water Maze test is considered stressful if water temperature <24°C, non-stressful if water temperature >26°C. Water "at room temperature" was considered as 24°C. If water temperature was between 24°C and 26°C, it was considered not classifiable in either the non-stressful or stressful domain. Nonetheless, these comparisons were included in the exploratory part.
- Working memory is excluded from meta-analysis due to controversial domain categorization in memory after stressful / non-stressful learning. Nonetheless, it is included in the systematic review as it provides information about memory retention and repetition.
- Step down-inhibitory avoidance is excluded from the meta-analysis because it is questionable whether for an animal remaining on a platform for a long time would be better or worse than an inescapable footshock. To the experts, it was therefore controversial to define the directionality of the effect. Nonetheless, we include this test in the systematic review. Furthermore, we ran a sensitivity analysis by including step down-inhibitory avoidance as part of the memory after stressful learning domain and verified that the interpretation did not change.

Definition of multiple hits

Prior to the beginning of the study, we defined elements that would constitute "multiple hits"[24]. Although this would ideally be a continuous variable (e.g. severity), we solely categorize its presence/absence due to the complexity and subjectivity of the classification. Animals were considered in the "multiple hits" group if they had one of the following (in addition to ELA):

Considered multiple hits	Not considered multiple hits
Stressful behavioral test performed previously (e.g. FST, fear conditioning)	Intragastric saline
Footshocks	Saline injections
Chronic (mild) unpredictable stress	Vaginal smears
Chronic constant light	Daily handling by experimenter
Chronic restraint	
Chronic individual housing	
Vaginal balloon distention	
Cannula implantation, mock surgeries, blood sampling, isofluorane anaesthesia	
Dams transported pregnant	
Stress prone strain (BALB/C, wistar Kyoto, DBA)	

****Note:** manipulated genetic background were excluded from the meta-analysis (S1.5) and therefore could not be included in the definition of vulnerability, despite it being an important factor

Inclusion/exclusion criteria

Study selection was performed independently by two researchers (VB and JK), who were blinded to the studies' results. The inclusion and exclusion criteria were specified prior to the beginning of the study.

Criteria	Comments
Inclusion	
Peer reviewed original publications in English	
Mice and rats	
ELA starts before P14	ELA model can extend after P14
ELA as alteration of maternal care[85] separation of the pup from the mother (maternal separation[86] / deprivation[85]) separation of the pup from mother and siblings (isolation) limited bedding and nesting[68] licking and grooming[87]	We define as 'separation' those models in which the mother was repeatedly separated from the pups (e.g. 3 hours a day for 2 weeks). We define as deprivation those models in which the mother was separated once from the pups for a prolonged time (e.g. 1 time 24 hours, or 2 times 12 hours). In other words, the categorization in maternal separation/deprivation depends on the model used and not the naming used in the papers. The separation/deprivation/isolation models are adaptations of Levine's original model. These adaptations are based on the observation that dams often leave the nest to forage for 15-30 min periods[88]. For this reason, we consider "adverse" and therefore include only those studies in which the duration of separation/deprivation/isolation time was >1h.
Behavioral testing during adult age	Older than 8 weeks but younger than 1 year
Behavioral tests among the listed a priori	See 1.3
Exclusion	
Specific pathogen free animals	
Ovariectomized females	
Sex not specified	Publication is included if sex is retrieved after contacting the authors
Males and females pooled	Publication is included if summary statistics of males and females separately are received after contacting the authors
Fasting before behavioral test (unless part of the test itself)	
Handling, gentling and communal nesting as ELA models	
Maternal separation with early weaning[89]	Early weaning is defined as separation of the pups from the mother at P17. If early weaning is only in the experimental group, the experiment is excluded. If early weaning occurred in both control and experimental group, the study is included and early weaning is considered a factor that could increase vulnerability
Handling as control group	
Genetic manipulations	
Animals bred for high/low anxiety-like behavior or novelty response or sensitivity/resilience to depression	
Animals separated in high/low performance	
Administration of any drug or alcohol via any route	e.g. Drug injections before testing, methamphetamine conditioned place preference tests
Any manipulation to previous generations	
Other *	Inclusion/exclusion criteria specific to certain tests. See footnote of S1.3.

Choosing the outcomes to include in the meta-analysis is often a non-straightforward task. Despite our intent to be as comprehensive as possible in the definition of inclusion/exclusion criteria prior to the beginning of the study, the list was not exhaustive. We therefore added inclusion/exclusion criteria during data collection:

- Unless the test required training, the test is included only during the first exposure
- If the data is presented in multiple time bins, we extracted the mean value reported most similar to other papers of the same category. In particular, we selected:
    - The first time bin for anxiety-like behavior
    - The last time bin for learning during the Morris Water Maze (intent to be as similar as possible to the probe test)
- Memory extinction and reversal learning were excluded

Details extraction of statistical information

Effect size was preferably calculated from mean, standard deviation (SD) and amount of animals (n) for each group (control and experimental). Points of note:

- If only the standard error of the mean (SEM) was reported, SD was calculated as SEM*$\sqrt{n}$. If the number of animals was reported as a range (e.g. 6-8 animals per group), we used the mean of this number (e.g. 7 animals per group).
- If median and interquartile range (IQR) were reported instead of mean and SD, we assumed normality and considered median = mean and the IQR as SEM/6[90]. We confirmed that this assumption did not alter our interpretation of the results by conducting a sensitivity analysis in which studies reporting medians were excluded (n_{comp} = 5, n_{exp} = 4, n_{stud} = 4).
- If total n was provided, n was equally split across groups. If n was not mentioned, could not be calculated from the degrees of freedom nor could be retrieved from the authors, we considered n to be equal to the n average of all other comparisons combined.
- If a single control group was used to compare experimental groups in which ELA was induced with different models (e.g. maternal deprivation at P4 vs maternal deprivation and P9[91]), the sample size of the control group was equally divided as control for each experimental group (e.g. n=10 overall in control group becomes n=5 for control of maternal deprivation at P4 and n=5 for control of maternal deprivation at P9)[30].

In our intent to classify as extensively as possible the methodological heterogeneity between different studies, we categorize >40 variables. However, not every publication reported on each of these, giving rise to missing information. Due to model estimation requirements with MetaForest25, it is necessary to estimate missing values. Following

standard practice, we imputed the median of the variable of interest in case of continuous variables, and the most common category in case of categorical variables.

Sensitivity analysis & Analysis of influential cases

We conducted the following sensitivity analysis:

- Specified prior to the analysis:
  - Outlying and influential cases
    - We identified outlying and influential cases according to Viechtbauer & Cheung's definition[35]. We qualitatively investigated the identified outlying and influential cases, but we could not identify any specific pattern of characteristics. The identified comparisons were removed and we evaluated the consistency of the results as sensitivity analysis
    - We identified potentially outlying cases also according to Tabachnick and Fidell's definition[92]. We conducted a sensitivity analysis by removing them from the analysis and verifying results' consistency.
  - Blinded and randomized studies
    - According to the standards of meta-analysis, we should include only studies which were blinded as well as randomized. However, only a few comparisons had these characteristics. For this reason, we chose to perform the main analysis on the full dataset.
    - To check for the influence of blinding and randomization on the effects sizes estimated in the main analysis, we performed a sensitivity analysis by including a "blinded and randomized" variable as a moderator in our model. We confirmed that this moderator was not significant (**males**: Q(1) = 0.316, p = 0.574; **females**: Q(1) = 3.263, p = 0.07).
  - Tests which were only reported by at least 4 publications (including second hit)
  - Risk of potential bias
    - We evaluated whether increase in potential bias corresponded to an increase in effect sizes.
- Specified after conduction of the analysis:
  - Effects of converting medians (IQR) to means(SD)
  - Publications reporting medians (IQR) were excluded, and we confirmed the consistency of our results

Publication bias assessment details

Publication bias was assessed with several methods. Although this may seem redundant, this approach was selected to balance out the pros and cons of each method

Test	Pros	Cons
Qualitative investigation of funnel plot	Estimated values derive from the built 3-level mixed effect model	Qualitative and not quantitative
Egger's regression followed by test for funnel plot asymmetry	Frequently used, quantitative	Does not consider the 3-level design
Begg's test	Frequently used, quantitative	Does not consider the 3-level design
File drawer analysis with fail and safe test	Addresses file drawer problem (only significant results are published). It provides an estimate of how many studies are necessary to nullify the effect found	Does not consider the 3-level design nor the moderators of the effect
Trim and fill	Aims to both identify and correct funnel plot asymmetry. It provides an estimate of the number of missing studies	Does not consider the 3-level design nor the moderators of the effect. Furthermore, it is known to perform poorly when substantial heterogeneity is present.

Tuning parameters metaforest

MetaForest's tuning parameters were selected from a 10-fold cross-validation according to the author's instructions[45]. We tested which type of weights (random-effects, fixed-effects or unweighted) provided the best model fit, with how many moderators available at each split (2, 4 or 6), and what was the most appropriate minimum size of the node (2, 4, or 6) to allow for splitting. Root-mean-square error (RMSE) was used to select the optimal model using the smallest value. This led to the selection of the following parameters: uniform weighting metaForest, 4 candidate moderators available at each split, 2 as minimum node size. The estimated residual heterogeneity of the model was $\tau^2 = 0.46$. We investigated the marginal bivariate relationship of each moderator by averaging its effect size over the values of all other moderators. The resulting partial dependency graphs can be obtained with the R script accompanying the text.

Supplementary results

MISSING VALUE DETAILS

We were not able to retrieve information from the following publications:

- 14 manuscripts published before 2008:
  - 55,93–101
- Full text of 8 publications was not found (authors contacted):
  - 102–109
  - It cannot be evaluated whether they were suitable for inclusion
- Authors from 9 manuscripts were contacted but no answer was received:
  - 99,110–116

COMPARISONS EXCLUDED FROM META-ANALYSIS

Below we provide details on comparisons excluded from the meta-analysis due to controversial domain categorization (S1.3). These comparisons were nonetheless analyzed at a systematic review level and are present in the published dataset (https://osf.io/ra947/).

Table legend: n_{comp}= amount of comparisons, n_{exp} = amount of experiments from which the comparisons were retrieved, n_{stud} = amount of studies from which the comparisons were retrieved.

Test	n_{comp}	n_{exp}	n_{stud}	Comments
Morris water maze	4	4	4	Water temperature between 24 and 26°C
8 arm radial maze	2	2	1	Working memory
T maze	1	1	1	Working memory
Y maze	9	7	3	Working memory
Step down avoidance	6	6	6	

DESCRIPTIVE INFORMATION ON STUDY CHARACTERISTICS

ELA models Several ELA models are used in the literature to disrupt maternal care. Primarily, these can be distinguished according to the **type** (=which paradigm) and the **timing** (=which postnatal day) of the model. Furthermore, there are specific characteristics within each model that can be altered. These are:

- For separation/deprivation/isolation: the animals are placed in a new cage or remain in the homecage
- For separation/deprivation/isolation: the duration of separation/deprivation/isolation can differ in length
- For separation/deprivation/isolation: the control group can be untouched, animal facility reared, handled <5min, or can derive from a split-litter design
- For separation/isolation: the protocol can be either predictable (every day at roughly the same time) or unpredictable

- For separation/isolation: the protocol occurs during the light/dark phase of the cycle

When considering **timing** as critical periods (model in first-and-second or first-to-third postnatal week): we identified 41 different protocols (322 theoretically possible). In particular, 14 variations of ELA models made up 85.8% of comparisons. This suggests that although there are variations in the protocols, the models are fairly consistent across the literature.

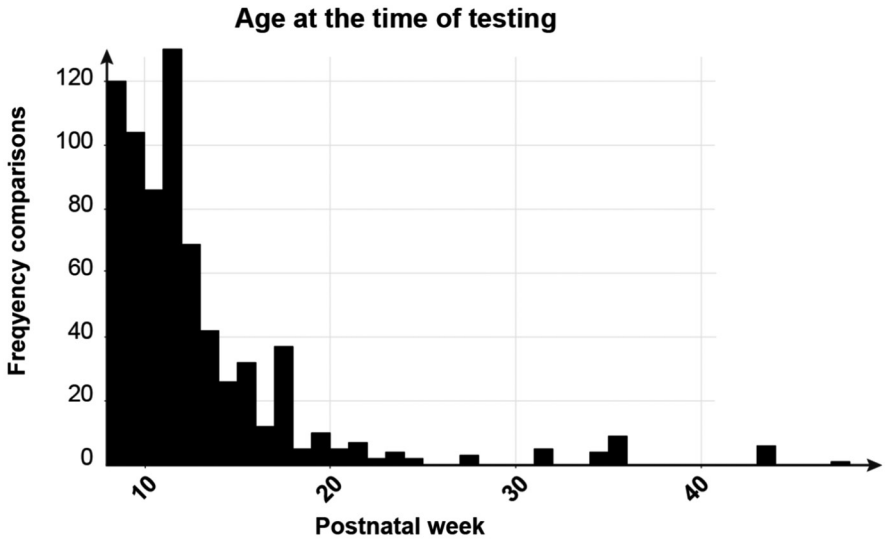
The table below provides descriptive information on the quantity of comparisons (n_{comp}) and of relative experiments (n_{exp}) across ELA models.

ELS Model	n_{comp}	nexp
Maternal separation	347	203
Maternal deprivation	82	49
Isolation	186	104
Limited nesting and bedding	76	34
Licking and grooming	20	12

Species and strains The table below describes the amount of experiments (n_{exp}) for each strain used.

Species	Strain	n_{exp}
Mice	BalbC	11
	C57Bl/6	65
	CD1	7
	DBA	4
	NMRI	3
	Other	6
	swissWebster	1
Rats	Lister Hooded	3
	Long Evans	19
	Long Evans Hooded	9
	SpragueDawley	78
	Wistar	181
	Wistar Kyoto	2
	Other	3
	Not specified	10

Age The histogram below displays the distribution of age of the animals at the time of testing expressed as postnatal week. We included animals tested for behavior older than 8 weeks of age, but younger than 1 year (S1.5). Although it has been reported that the effects of ELA on behavior (memory in particular) may become more evident in older animals, the amount of comparisons of this age group in our study was not sufficient to further explore this hypothesis.



Domains and tests The tables below display the distribution of comparisons (ncomp), experiments (nexp) and studies (nstud) across A) sex and domains, B) behavioral tests.

A)	Males			Females		
Domain	n_{comp}	n_{exp}	n_{stud}	n_{comp}	n_{exp}	n_{stud}
Anxiety-like behavior	262	198	135	95	69	50
Memory after stressful learning	151	136	88	52	45	27
Memory after non-stressful learning	79	56	45	26	19	17
Social behavior	40	36	29	6	6	5

B)	males			females		
Test	ncomp	nexp	nstud	ncomp	nexp	nstud
Defensive Withdrawal	7	7	4	1	1	1
Elevated Zero Maze	12	11	7	1	1	1
Elevated Plus Maze	105	105	77	38	38	29
Fear conditioning (anxiety-like)	17	17	12	7	7	7
Forced Swim Test (anxiety-like)	29	29	19	5	5	4
Light/Dark box	19	19	11	9	9	5

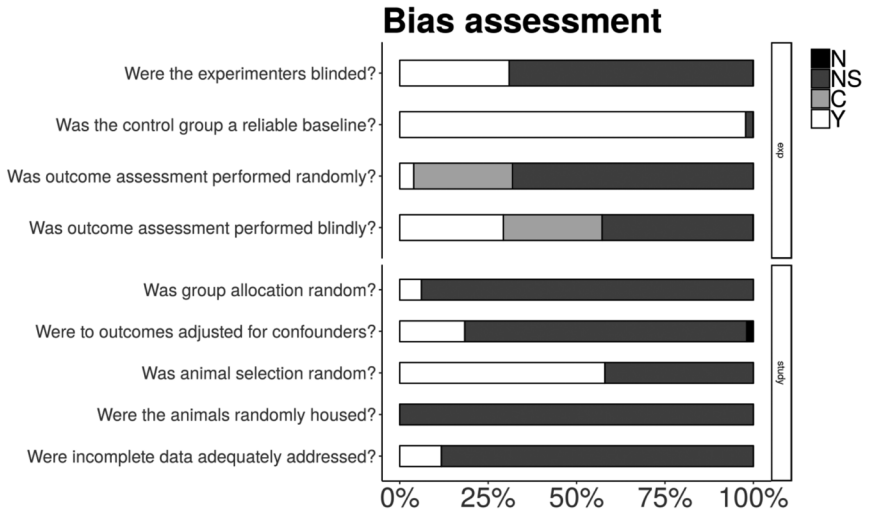
Novelty induced-suppression of feeding and drinking	6	6	5	1	1	1
Open field	64	64	46	32	32	19
Tail suspension test	3	3	3	1	1	1
Fear conditioning (stressful learning)	45	37	28	18	13	12
Forced swim test (stressful learning)	52	52	37	17	17	10
Morris water maze (stressful learning)	36	31	25	6	6	5
Shuttle box	17	17	6	11	11	5
Social Fear Conditioning	1	1	1			
Morris water maze (neutral learning)	2	2	2			
Object in Context	2	2	2	1	1	1
Object in Location	12	10	9	5	5	5
Object Recognition	39	35	31	15	15	14
Social Recognition	15	13	9	3	3	2
Temporal Order Task	4	4	4			
Y Maze (neutral memory)	6	6	5	2	2	2
Resident intruder test	10	8	7			
Social Interaction	23	23	18	6	6	5
Social Open Field	2	2	1			
Social Play	1	1	1			
Social Preference	4	4	3			
Morris Water Maze (excl)	4	4	4			
Radial Maze 8 Arm	1	1	1	1	1	1
Step Down Avoidance	6	6	6			
T Maze	1	1	1			
Y Maze (excl)	7	7	3	2	2	1

Risk of bias assessment

Risk of bias assessment was performed according to SYRCLE guidelines, and by distinguishing between bias at an experiment- or study- level. No publication reported information on all SYRCLE potential bias items. Overall, "not specified" was the most common score (54.8%). In 44.8% of the cases, measures to prevent bias were reported. This includes computerized approaches. 41 studies yielding a total of 145 comparisons reported being blinded as well as randomized. For sensitivity analysis, amount of potential bias was operationalized by summing the risk of bias of each item according to the definition: "yes" = 0, "unclear" = 0.5, "no" = 1. Computerized approaches were

considered as "0" bias. This produced a continuous variable between 0 (no risk bias) and 10 (maximum risk of bias).

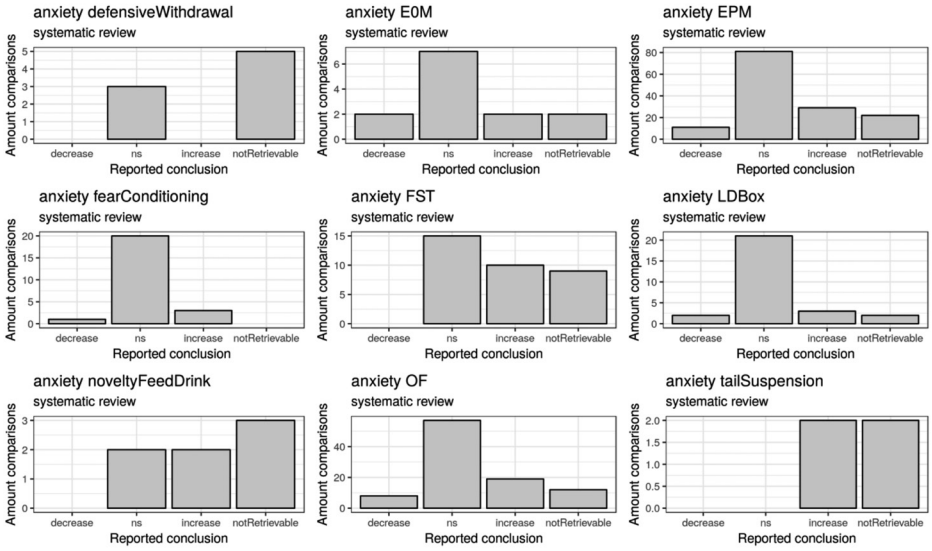
Figure legend: N = Bias was not prevented; NS = it was not specified whether measure to prevent bias were applied; C = computerized approach; Y = measures to prevent bias were used



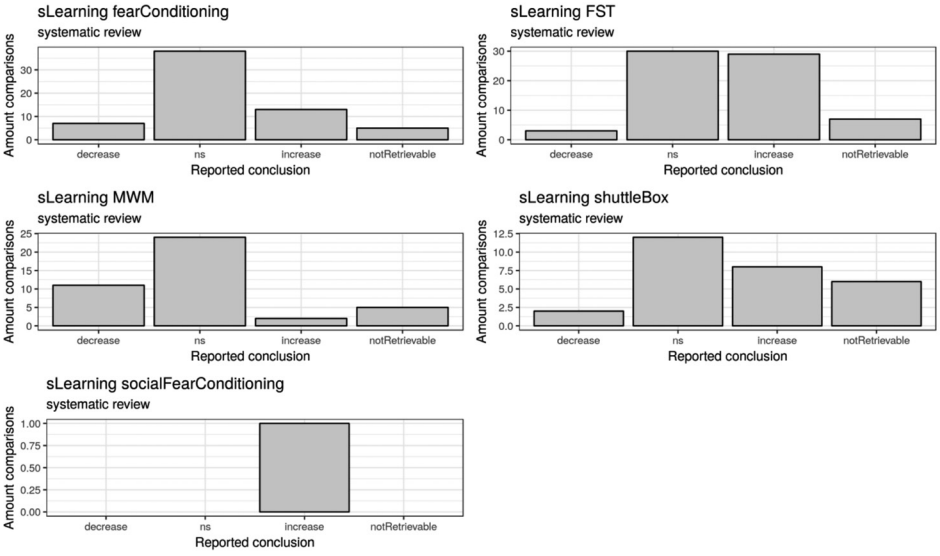
Results at a systematic review level

On a systematic review level, we evaluated the directionality of the effects of ELS on each behavioral test used. These are expressed as decrease, ns = not significant, increase, and notApplicable = it could not be deduced directly from the data reported. "Increase" should be interpreted as an enhancement of the behavior reported (more anxious, more memory, more social behavior). For example, "increase" in the elevated plus maze signifies that the animals were more anxious. This could mean that they spent less time in or entered fewer times into the open arms. The figures below represent data for each behavior test at a systematic review level in A) anxiety-like behavior, B) memory after stressful learning, C) memory after non-stressful learning, D) social behavior, and E) tests not included in the meta-analysis.

A) Anxiety-like behavior

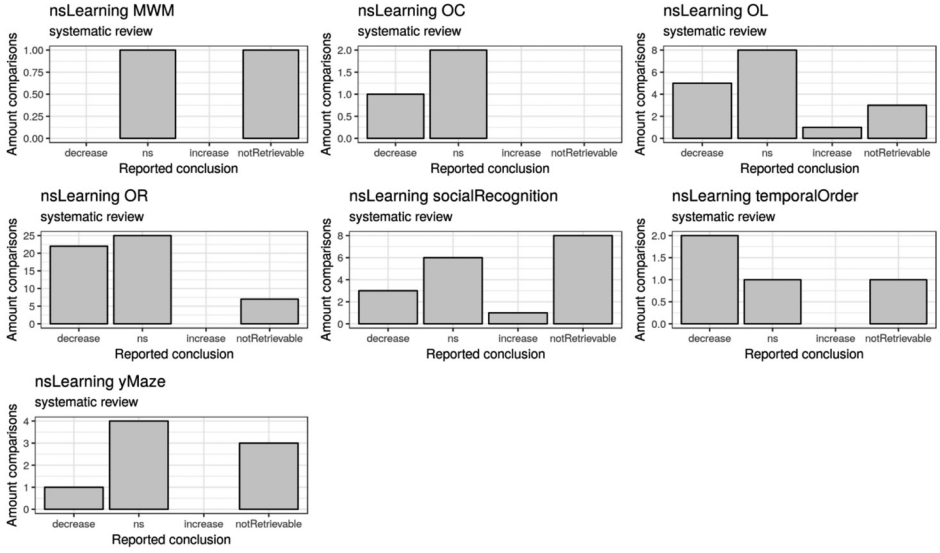


B) Memory after stressful learning

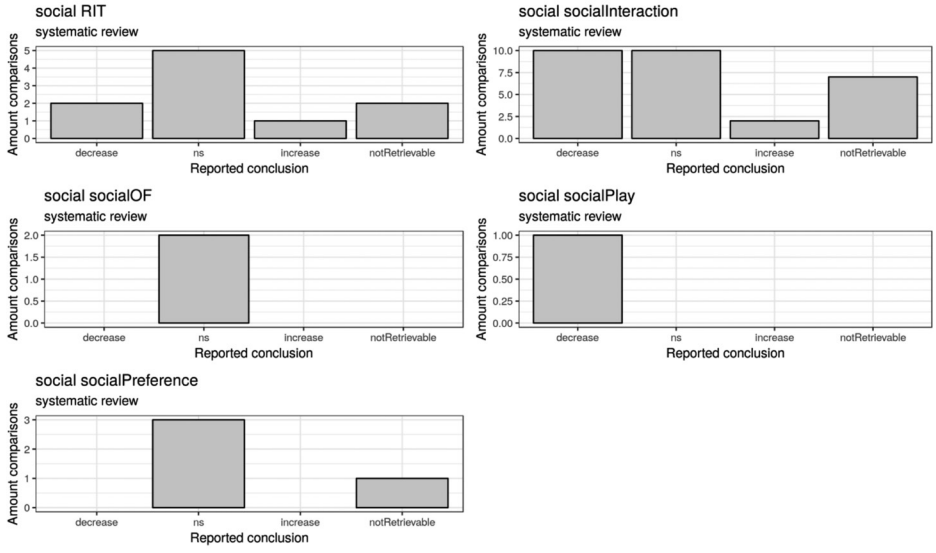


S

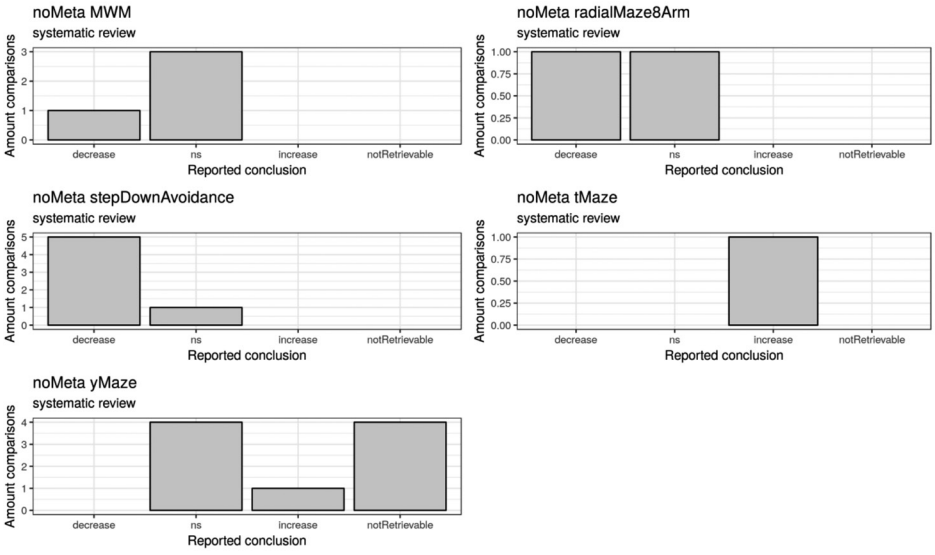
C) Memory after non-stressful learning



D) Social behavior



E) Not included in the meta-analysis



Statistics main results: males

The table below summarizes the results of the hypotheses-testing analysis in males.

Table legend: ci.lb = lower boundary confidence interval; ci.ub = upper boundary confidence interval; effectsize = estimated Hedge's G; se = standard error of the estimated Hedge's G; z-value = z-value of the test; p-value = uncorrected p-value of the test; p-value_bonf = corrected p-value for family-wise comparison with Bonferroni; _Corr = refers to effectsize, ci.lb and ci.ub which have been flipped to ease interpretation (Method Section); sLearning = memory after stressful learning; nsLearning = memory after non-stessful learning; domainHit = statistics of experiments with multiple hits vs experiments without; domainNo = only experiments without multiple hits; domainYes = only experiments with multiple hits.

	test	ci.lb	ci.ub	effectsize	se	Zvalue	Pvalue	Pvalue_bonf	effectsize_Corr	ci.lb_Corr	ci.ub_Corr
Domains: main effects											
	anxiety	0.1649	0.3911	0.278	0.0577	4.8194	0	0	0.278	0.1649	0.3911
	sLearning	0.1409	0.4255	0.2832	0.0726	3.8997	0.0001	0.0004	0.2832	0.1409	0.4255
	nsLearning	0.3953	0.7923	0.5938	0.1013	5.8604	0	0	-0.5938	-0.7923	-0.3953
	social	0.3477	0.8801	0.6139	0.1358	4.5211	0	0	-0.6139	-0.8801	-0.3477
Increased vulnerability: main effect											
	hit	0.018	0.4264	0.2222	0.1042	2.1315	0.0331	0.0331	0.2222	0.018	0.4264
Effect increased vulnerability for each domain											
	anxietyHit	-0.0619	0.3811	0.1596	0.113	1.4123	0.1579	0.6316	0.1596	-0.0619	0.3811
	sLearningHit	-0.095	0.4676	0.1863	0.1435	1.2989	0.194	0.776	0.1863	-0.095	0.4676
	nsLearningHit	0.0395	0.8309	0.4352	0.2019	2.1556	0.0311	0.1244	-0.4352	-0.8309	-0.0395
	socialHit	-0.4228	0.638	0.1076	0.2706	0.3975	0.691	1	-0.1076	-0.638	0.4228
Posthocs											
	anxietyNo	0.0477	0.3487	0.1982	0.0768	2.5808	0.0099	0.0792	0.1982	0.0477	0.3487
	sLearningNo	0.0031	0.3771	0.1901	0.0954	1.9921	0.0464	0.3712	0.1901	0.0031	0.3771
	nsLearningNo	0.1	0.6524	0.3762	0.1409	2.6699	0.0076	0.0608	-0.3762	-0.6524	-0.1
	socialNo	0.2084	0.912	0.5602	0.1795	3.1208	0.0018	0.0144	-0.5602	-0.912	-0.2084
	anxietyYes	0.1922	0.5234	0.3578	0.0845	4.2349	0	0	0.3578	0.1922	0.5234
	sLearningYes	0.1641	0.5887	0.3764	0.1083	3.4743	0.0005	0.004	0.3764	0.1641	0.5887
	nsLearningYes	0.5271	1.0959	0.8115	0.1451	5.5914	0	0	-0.8115	-1.0959	-0.5271
	socialYes	0.2694	1.066	0.6677	0.2032	3.2867	0.001	0.008	-0.6677	-1.066	-0.2694

Statistics main results: females

In the female dataset, we were unable to confirm our hypothesis in any of the domains investigated. In particular, in females with a history of ELA, we could not confirm changes in anxiety-like behavior (*HedgesG* [95%CI] = .101 [-.035,.236], z = 1.459, p =.59), memory after stressful learning (*HedgesG* [95%CI] = .192 [.014, .37], z = 2.11, p =.14), memory after non-stressful learning (*HedgesG* [95%CI] = -.284 [-.532, -.0355], z = 2.24, p = .1), or in social behavior (*HedgesG* [95%CI]=.011 [-.405,.428], z = -.053, p = .957).

Table legend: ci.lb = lower boundary confidence interval; ci.ub = upper boundary confidence interval; effectsize = estimated Hedge's G; se = standard error of the estimated Hedge's G; z-value = z-value of the test; p-value = uncorrected p-value of the test; p-value_bonf = corrected p-value for family-wise comparison with Bonferroni; _Corr = refers to effectsize, ci.lb and ci.ub which have been flipped to ease interpretation (Method Section); sLearning = memory after stressful learning; nsLearning = memory after non-stessful learning; domainHit = statistics of experiments with multiple hits vs experiments without; domainNo = only experiments without multiple hits; domainYes = only experiments with multiple hits.

	test	ci.lb	ci.ub	effectsize	se	Zvalue	Pvalue	Pvalue_bonf	effectsize_Corr	ci.lb_Corr	ci.ub_Corr
Domains: main effects											
	anxiety	-0.0347	0.2361	0.1007	0.0691	1.4587	0.1447	0.5788	0.1007	-0.0347	0.2361
	sLearning	0.0138	0.3698	0.1918	0.0908	2.1123	0.0347	0.1388	0.1918	0.0138	0.3698
	nsLearning	0.0355	0.5321	0.2838	0.1267	2.2403	0.0251	0.1004	-0.2838	-0.5321	-0.0355
	social	-0.4276	0.405	-0.0113	0.2124	-0.0534	0.9574	1	0.0113	-0.405	0.4276
Increased vulnerability: main effect											
	hit	-0.0033	0.5965	0.2966	0.153	1.9387	0.0525	0.0525	0.2966	-0.0033	0.5965
Effect increased vulnerability for each domain											
	anxietyHit	-0.1384	0.3826	0.1221	0.1329	0.9188	0.3582	1	0.1221	-0.1384	0.3826
	sLearningHit	-0.001	0.6752	0.3371	0.1725	1.954	0.0507	0.2028	0.3371	-0.001	0.6752
	nsLearningHit	0.0693	1.0611	0.5652	0.253	2.2338	0.0255	0.102	-0.5652	-1.0611	-0.0693
	socialHit	-0.6703	0.9941	0.1619	0.4246	0.3814	0.7029	1	-0.1619	-0.9941	0.6703
Posthocs											
	anxietyNo	-0.1324	0.2118	0.0397	0.0878	0.4518	0.6514	1	0.0397	-0.1324	0.2118
	sLearningNo	-0.1816	0.228	0.0232	0.1045	0.2222	0.8242	1	0.0232	-0.1816	0.228
	nsLearningNo	-0.2997	0.3021	0.0012	0.1535	0.0076	0.9939	1	-0.0012	-0.3021	0.2997
	socialNo	-0.5882	0.4036	-0.0923	0.253	-0.3649	0.7152	1	0.0923	-0.4036	0.5882
	anxietyYes	-0.0407	0.3643	0.1618	0.1033	1.5667	0.1172	0.9376	0.1618	-0.0407	0.3643
	sLearningYes	0.0802	0.6404	0.3603	0.1429	2.5204	0.0117	0.0936	0.3603	0.0802	0.6404
	nsLearningYes	0.1719	0.9609	0.5664	0.2013	2.8133	0.0049	0.0392	-0.5664	-0.9609	-0.1719
	socialYes	-0.599	0.7382	0.0696	0.3411	0.2041	0.8383	1	-0.0696	-0.7382	0.599

Publication bias results

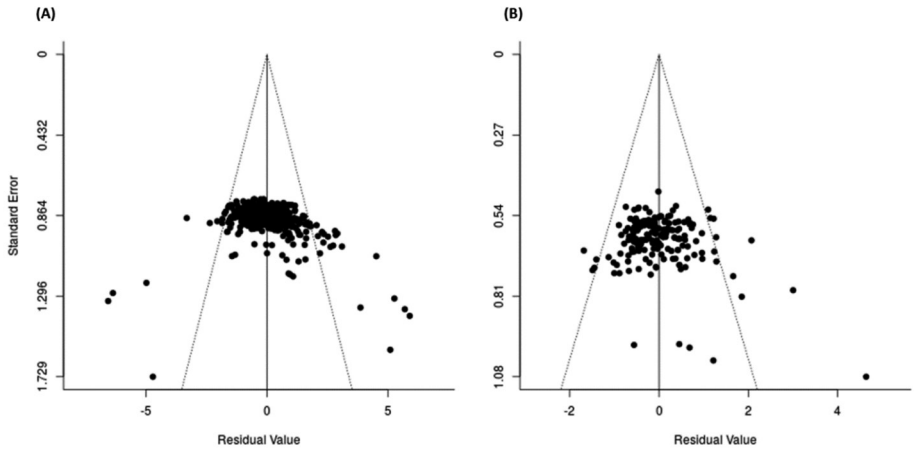
Details on the tests used to evaluate the influence of publication bias are described in S1.8.

Males Publication bias is evident from qualitative evaluation of funnel plot asymmetry (Figure A in S2.8), Egger's regression (z = 7.501, p < .001) and Begg's test (z = 7.3961, p < .0001). However, fail-safe file drawer analysis revealed that >7600000 unpublished, filed, or un-retrieved comparisons averaging null results would be required to bring the average unweighted effect size to non-significance. Similarly, trim-and-fill analysis (based on random effects meta-analysis) estimates that 0 studies are missing. These results suggest that although there is evidence of publication bias, the model seems not influenced by it as the interpretation of the results would not change.

Females There is evidence of publication bias both from the funnel plot asymme-

try (Figure B in S2.8), Egger's regression (based on random-effects meta-analysis, z = 2.329, p = 0.020) and Begg's test (z = 2.424, p = 0.015). However, fail-safe file drawer analysis was not performed as the overall meta-analysis was not significant (Q(8) = 14.384, p = .07). Trim-and-fill analysis (based on random effects meta-analysis) estimates that 0 studies are missing.

Figure S2.8. Funnel plots for publication bias evaluation in the A) males' and B) females' datasets.



Sensitivity analyses

A summary of all sensitivity analyses performed can be found in supplementary material. Any researcher interested in analyzing sensitivity analyses in more detail is referred to the publicly available dataset and analysis script used (https://osf.io/ra947/).

Study of distribution of variance

In **males**, the within- (σw^2 = 0.296) and the between-variance component (σb^2 = 0.245) differed significantly from 0 (p<0.000), which indicates that the variation in effect sizes is accounted for by differences within as well as between experiments. Conversely, in **females**, the between-experiment variance component (σw^2 = 0.000, p = 1.0) was negligible, while the within-variance component (σb^2 = 0.1838) differed significantly from 0 (p<0.000). This indicates that the variation in effect sizes is accounted mainly by differences between experiments.

Directed exploration: metaforest partial dependence plots

Partial dependence plots are visualizations in which effect size is predicted as a function of the average over all other predictor variables. All partial dependence plots can be visualized by running the provided R-script (https://osf.io/ra947/).

Instructions:

1) Open the R project provided
2) Install any R package that might be missing.
3) Prepare environment by running the code section "Environment Preparation"
4) Prepare dataset for MetaForest analysis by running the code section "MetaForest: dataset preparation"
5) Perform MetaForest analysis by running the code section "MetaForest: tuning" (depending on your computer, it may take 30min-1h)
6) Save partial dependence plots in the Rproject folder by running the code section "MetaForest: plots".

Supplementary references

1. Kessler, R. C. et al. Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. The British journal of psychiatry : the journal of mental science 197, 378–85 (2010).
2. Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. Nature Reviews Neuroscience 17, 652–666 (2016).
3. Masson, M., East-Richard, C. & Cellard, C. A meta-analysis on the impact of psychiatric disorders and maltreatment on cognition. Neuropsychology 30, 143–156 (2016).
4. Nelson, C. A. et al. Cognitive Recovery in Socially Deprived Young Children: The Bucharest Early Intervention Project. Science 318, 1937–1940 (2007).
5. Vargas, T. et al. Childhood Trauma and Neurocognition in Adults With Psychotic Disorders: A Systematic Review and Meta-analysis. Schizophrenia Bulletin 1–14 (2018) doi:10.1093/schbul/sby150.
6. Millan, M. J. et al. Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. Nature Reviews Drug Discovery 11, 141–168 (2012).
7. Monfils, M. H. & Holmes, E. A. Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. The Lancet Psychiatry 0366, (2018).
8. Liberzon, I. & Abelson, J. L. Context processing and the neurobiology of post-traumatic stress disorder. Neuron 92, 14–30 (2016).
9. Green, M. F., Horan, W. P. & Lee, J. Social cognition in schizophrenia. Nature Reviews Neuroscience 16, 620–631 (2015).
10. Ammerman, R. T., Cassisi, J. E., Hersen, M. & Van Hasselt, V. B. Consequences of physical abuse and neglect in children. Clinical Psychology Review 6, 291–310 (1986).
11. Knop, J., Joëls, M. & van der Veen, R. The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. Current Opinion in Psychology 15, 174–181 (2017).
12. Bowlby, J. Maternal care and mental health. World Health Organization Monograph Series 179 (1951).
13. Harlow, H. & Harlow, M. The affectional systems. in Behavior of nonhuman primates (eds. Schrier, A., Harlow, H. & Stollnitz, F.) 287–334 (Academic Press, 1965).
14. Meaney, M. J. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. Annual Review of Neuroscience 24, 1161–1192 (2001).
15. Levine, S. Developmental determinants of sensitivity and resistance to stress. Psychoneuroendocrinology 30, 939–946 (2005).
16. Hofer, M. A. Hidden Regulatory Processes in Early Social Relationships. in Perspectives in Ethology: social behavior (eds. Bateson, P. & Klopfer, P.) 135–201 (Springer, 1978).
17. Benetti, F. et al. Early postnatal maternal deprivation in rats induces memory deficits in adult life that can be reversed by donepezil and galantamine. International Journal of Developmental Neuroscience 27, 59–64 (2009).
18. Champagne, D. L. et al. Maternal Care and Hippocampal Plasticity: Evidence for Experience-Dependent Structural Plasticity, Altered Synaptic Functioning, and Differential Responsiveness to Glucocorticoids and Stress. Journal of Neuroscience 28, 6037–6045 (2008).
19. Kanatsou, S. et al. Overexpression of Mineralocorticoid Receptors in the Mouse Forebrain Partly Alleviates the Effects of Chronic Early Life Stress on Spatial Memory, Neurogenesis and Synaptic Function in the Dentate Gyrus. Frontiers in Cellular Neuroscience 11, 1–13 (2017).
20. Ivy, A. S. et al. Hippocampal Dysfunction and Cognitive Impairments Provoked by Chronic Early-Life Stress Involve Excessive Activation of CRH Receptors. Journal of Neuroscience 30, 13005–13015 (2010).
21. Bredy, T. W., Humpartzoomian, R. A., Cain, D. P. & Meaney, M. J. Partial reversal of the effect of maternal care on cognitive function through environmental enrichment. Neuroscience 118, 571–576 (2003).

22. Mello, P. B., Benetti, F., Cammarota, M. & Izquierdo, I. Physical exercise can reverse the deficit in fear memory induced by maternal deprivation. Neurobiology of Learning and Memory 92, 364–369 (2009).
23. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience 14, 365–376 (2013).
24. Daskalakis, N. P. et al. The three-hit concept of vulnerability and resilience: Toward understanding adaptation to early-life adversity outcome. Psychoneuroendocrinology 38, 1858–1873 (2013).
25. van Lissa, C. J. MetaForest: Exploring heterogeneity in meta-analysis using random forests. PsyArXiv (2018) doi:10.31234/osf.io/myg6s.
26. De Vries, R. B. M. et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. Evidence-based Preclinical Medicine 1, 1–9 (2015).
27. Leenaars, M. et al. A step-by-step guide to systematically identify all relevant animal studies. Laboratory Animals 46, 24–31 (2012).
28. Moher, D. et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine 6, (2009).
29. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software 36, 1–48 (2010).
30. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. Journal of Neuroscience Methods 221, 92–102 (2014).
31. Bonapersona, V., Joels, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: A 3 level meta-analysis of rodent studies. Neuroscience and biobehavioral reviews 95, 1–16 (2018).
32. Aarts, E., Verhage, M., Veenvliet, J. V, Dolan, C. V & van der Sluis, S. A solution to dependency: using multilevel analysis to accommodate nested data. Nature Neuroscience 17, 491–496 (2014).
33. Cheung, M. W. L. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. Psychological Methods 19, 211–229 (2014).
34. Vesterinen, H. M. et al. Meta-analysis of data from animal studies: A practical guide. Journal of Neuroscience Methods 221, 92–102 (2014).
35. Viechtbauer, W. & Cheung, M. W.-L. Outlier and influence diagnostics for meta-analysis. Research Synthesis Methods 1, 112–125 (2010).
36. Rosenthal, R. The file drawer problem and tolerance for null results. Psychological Bulletin 86, 638–641 (1979).
37. Duval, S. & Tweedie, R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56, 455–463 (2000).
38. Hooijmans, C. R. et al. SYRCLE ' s risk of bias tool for animal studies. BMC Medical Research Methodology 14, 1–9 (2014).
39. Assink, M. & Wibbelink, C. J. M. Fitting three-level meta-analytic models in R: A step-by-step tutorial. The Quantitative Methods for Psychology 12, 154–174 (2016).
40. Hastie, T., Friedman, J. & Tibshirani, R. The elements of statistical learning: Data Mining, Inference and Prediction. (Springer Series in Statistics, 2009).
41. Walker, C.-D. et al. Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. Stress 0, 1–28 (2017).
42. Loi, M. et al. Effects of early-life stress on cognitive function and hippocampal structure in female rodents. Neuroscience 342, 101–119 (2017).
43. Karp, N. A. Reproducible preclinical research—Is embracing variability the answer? PLoS Biology 16, 1–5 (2018).
44. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2015).
45. van Lissa, C. J. Package ' metaforest '. Cran (2018).
46. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for R. (2017).
47. Wickham, H., François, R., Henry, L., Müller, K. & RSudio. Package ' dplyr ' Version 0.7.6. (2018).
48. Walker, A. K. et al. Neonatal lipopolysaccharide and adult stress exposure predisposes rats to anxiety-like behaviour and blunted corticosterone responses: Implications for the double-hit hypothesis. Psychoneuroendocrinology 34, 1515–1525 (2009).
49. Pechtel, P. & Pizzagalli, D. A. Effects of early life stress on cognitive and affective function: An integrated review of human literature. Psychopharmacology 214, 55–70 (2011).
50. Suor, J. H., Sturge-Apple, M. L., Davies, P. T., Cicchetti, D. & Manning, L. G. Tracing Differential Pathways of Risk: Associations Among Family Adversity, Cortisol, and Cognitive Functioning in Childhood. Child Development 86, 1142–1158 (2015).
51. Ioannidis, J. P. A. Why most published research findings are false. PLoS medicine 2, e124 (2005).
52. Hooijmans, C. R. & Ritskes-Hoitinga, M. Progress in using systematic reviews of animal studies to improve translational research. PLoS Medicine 10, 1–4 (2013).
53. Rosenthal, R. Meta-analytic procedures for social research. (Sage, 1991).
54. Pryce, C. R., Bettschen, D., Nanz-Bahr, N. I. & Feldon, J. Comparison of the effects of early handling and early deprivation on conditioned stimulus, context, and spatial learning and memory in adult rats. Behavioral Neuroscience 117, 883–893 (2003).

55. Penke, Z. et al. Postnatal maternal deprivation produces long-lasting modifications of the stress response, feeding and stress-related behaviour in the rat. The European journal of neuroscience 14, 747–755 (2001).
56. Veenema, A. H., Bredewold, R. & Neumann, I. D. Opposite effects of maternal separation on intermale and maternal aggression in C57BL/6 mice: link to hypothalamic vasopressin and oxytocin immunoreactivity. Psychoneuroendocrinology 32, 437–450 (2007).
57. Antonic, A. et al. Stem Cell Transplantation in Traumatic Spinal Cord Injury: A Systematic Review and Meta-Analysis of Animal Studies. PLoS Biology 11, (2013).
58. Egan, K. J., Sena, E. S. & Vesterinen, H. M. Making the most of animal data - Improving the prospect of success in pragmatic trials in the neurosciences. Trials 12, no pagination (2011).
59. Cooper, H., Hedges, L. V & Valentine, J. C. The handbook of research synthesis and meta-analysis 2nd edition. (Russel Sage Foundation, 2009).
60. Hensch, T. K. Critical period plasticity in local cortical circuits. Nature Reviews Neuroscience 6, 877–888 (2005).
61. Levine, S. & Lewis, G. W. Critical period for effects of infantile experience on maturation of stress response. Science 129, 42–43 (1959).
62. Ellenbroek, B. A. & Cools, A. R. Early maternal deprivation and prepulse inhibition: The role of the postdeprivation environment. Pharmacology Biochemistry and Behavior 73, 177–184 (2002).
63. Witek-Janusek, L. Pituitary-adrenal response to bacterial endotoxin in developing rats. American Journal Physiology 255, E525–E530 (1988).
64. Van Oers, H. J. j., De Kloet, E. R. & Levine, S. Early vs. late maternal deprivation differentially alters the endocrine and hypothalamic responses to stress. Developmental Brain Research 111, 245–252 (1998).
65. Peña, C. J. et al. Early life stress confers lifelong stress scsceptibility in mice via ventral tegmental area OTX2. Science 1, 1185–1188 (2017).
66. Singh-Taylor, A. et al. NRSF-dependent epigenetic mechanisms contribute to programming of stress-sensitive neurons by neonatal experience, promoting resilience. Molecular Psychiatry 23, 648–657 (2018).
67. Suchecki, D. Maternal regulation of the infant's hypothalamic-pituitary-adrenal axis stress response: Seymour 'Gig' Levine's legacy to neuroendocrinology. Journal of Neuroendocrinology 30, 1–17 (2018).
68. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. Endocrinology 149, 4892–4900 (2008).
69. Jennions, M. D. & Møller, A. P. Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. Proceedings of the Royal Society B: Biological Sciences 269, 43–48 (2002).
70. Wigger, A. & Neumann, I. D. Periodic maternal deprivation induces gender-dependent alterations in behavioral and neuroendocrine responses to emotional stress in adult rats. Physiology and Behavior 66, 293–302 (1999).
71. Oitzl, M., Workel, J., Fluttert, M., Frösch, F. & Ronde Kloet, E. Maternal deprivation affects behaviour from youth to senescence: amplication of individual differences in spatial learning and memory in senescent Brown Norway rats. European Journal of Neuroscience 12, 3771–3780 (2000).
72. Millstein, R. A. & Holmes, A. Effects of repeated maternal separation on anxiety- and depression-related phenotypes in different mouse strains. Neuroscience and Biobehavioral Reviews 31, 3–17 (2007).
73. Teicher, M. H. & Samson, J. A. Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. American Journal of Psychiatry 170, 1114–1133 (2013).
74. Baram, T. Z. et al. Fragmentation and unpredictability of early-life experience in mental disorders. American Journal of Psychiatry 169, 907–915 (2012).
75. Whipple, E. E. & Webster-Stratton, C. The role of parental stress in physically abusive families. Child Abuse & Neglect 15, 279–291 (1991).
76. Cohen, N. Language impairment and psychopathology in infants, children, and adolescnets. (Sage Publications, 2001).
77. Miller, A. B. et al. Dimensions of deprivation and threat, psychopathology, and potential mediators: A multi-year longitudinal analysis. Journal of Abnormal Psychology 127, 160–170 (2018).
78. Bale, T. L. et al. Early life programming and neurodevelopmental disorders. Biological Psychiatry 68, 314–319 (2010).
79. Riecher-Rössler, A. Sex and gender differences in mental disorders. The Lancet Psychiatry 4, 8–9 (2017).
80. Plotsky, P., Bradley, C. & Anand, K. Behavioral and neuroendocrine consequences of neonatal stress. Pain Research and Clinical Management 10, 77–100 (2000).
81. Bale, T. L. & Epperson, C. N. Sex differences and stress across the lifespan. Nature neuroscience 18, 1413–20 (2015).
82. Oomen, C. A. et al. Opposite effects of early maternal deprivation on neurogenesis in male versus female rats. PLoS ONE 4, (2009).
83. van Hasselt, F. N., Boudewijns, Z. S. R. M., Van Der Knaap, N. J. F., Krugers, H. J. & Joëls, M. Maternal Care Received by Individual Pups Correlates with Adult CA1 Dendritic Morphology and Synaptic Plasticity in a Sex-Dependent Manner. Journal of Neuroendocrinology 24, 331–340 (2012).
84. McCarthy, M. M. Multifaceted origins of sex differences in the brain. Philosophical Transactions of the Royal Society B: Biological Sciences 371, (2016).
85. Levine, S. Regulation of the hypothalamic-pituitary-adrenal axis in the neonatal rat: the role of maternal behavior. Neurotoxicity research 4, 557–564 (2002).
86. Sanchez, M. M., Ladd, C. O. & Plotsky, P. M. Early adverse experience as a developmental risk factor for later psychopathology: Eveidence from rodent and promate models. Development and Psychopathology 13, 419–449

(2001).
87. Champagne, F. A., Francis, D. D., Mar, A. & Meaney, M. J. Variations in maternal care in the rat as a mediating influence for the effects of environment on development. Physiology & behavior 79, 359–371 (2003).
88. Leon, M., Croskerry, P. G. & Smith, G. K. Thermal control of mother-young contact in rats. Physiology & Behavior 21, 793–811 (1978).
89. Carlyle, B. C. et al. Maternal separation with early weaning: A rodent model providing novel insights into neglect associated developmental deficits. Development and Psychopathology 24, 1401–1416 (2012).
90. Hozo, S. P., Djulbegovic, B. & Hozo, I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology 5, 1–10 (2005).
91. Lehmann, J., Logeay, C. & Feldon, J. Long-term effects of a single 24-hour maternal separation on three different latent inhibition paradigms. Psychobiology 28, 411–419 (2000).
92. Tabachnick, B. G. & Fidell, L. S. Using multivariate statistics. (Pearson Education,2013, 2013).
93. Lehmann, J., Pryce, C. R., Bettschen, D. & Feldon, J. The maternal separation paradigm and adult emotionality and cognition in male and female Wistar rats. Pharmacology Biochemistry and Behavior 64, 705–715 (1999).
94. Lehmann, J., Stöhr, T. & Feldon, J. Long-term effects of prenatal stress experience and postnatal maternal separation on emotionality and attentional processes. Behavioural Brain Research 107, 133–144 (2000).
95. Liu, D., Diorio, J., Day, J. C., Francis, D. D. & Meaney, M. J. Maternal care, hippocampal synaptogenesis and cognitive development in rats. Nature Neuroscience 3, 799–806 (2000).
96. Daniels, W. M. U., Pietersen, C. Y., Carstens, M. E. & Stein, D. J. Maternal separation in rats leads to anxiety-like behavior and a blunted ACTH response and altered neurotransmitter levels in response to a subsequent stressor. Metabolic brain disease 19, 3–14 (2004).
97. Macrí, S., Mason, G. J. & Würbel, H. Dissociation in the effects of neonatal maternal separations on maternal care and the offspring's HPA and fear responses in rats. European Journal of Neuroscience 20, 1017–1024 (2004).
98. Garner, B., Wood, S. J., Pantelis, C. & van den Buuse, M. Early maternal deprivation reduces prepulse inhibition and impairs spatial learning ability in adulthood: No further effect of post-pubertal chronic corticosterone treatment. Behavioural Brain Research 176, 323–332 (2007).
99. Knuth, E. D. & Etgen, A. M. Long-term behavioral consequences of brief, repeated neonatal isolation. Brain Research 1128, 139–147 (2007).
100. Rüedi-Bettschen, D. et al. Early deprivation leads to altered behavioural, autonomic and endocrine responses to environmental challenge in adult Fischer rats. European Journal of Neuroscience 24, 2879–2893 (2006).
101. Guijarro, J. Z. et al. Effects of brief and long maternal separations on the HPA axis activity and the performance of rats on context and tone fear conditioning. Behavioural Brain Research 184, 101–108 (2007).
102. Hofer, M. A. The role of nutrition in the physiological and behavioral effects of early maternal separation on infant rats. Psychosomatic medicine 35, 350–9 (1973).
103. Plaut, S. M. Maternal deprivation in mice and rats. Laboratory animal science 22, 594 (1972).
104. Hofer, M. A. Toward a developmental basis for disease predisposition: the effects of early maternal separation on brain, behavior, and cardiovascular system. Research publications - Association for Research in Nervous and Mental Disease 59, 209–28 (1981).
105. Hofer, M. A. Effects of reserpine and amphetamine on the development of hyperactivity in maternally deprived rat pups. Psychosomatic medicine 42, 513–20 (1980).
106. Kehoe, P., Shoemaker, W. J., Triano, L., Hoffman, J. & Arons, C. Repeated isolation in the neonatal rat produces alterations in behavior and ventral striatal dopamine release in the juvenile after amphetamine challenge. Behav Neurosci 110, 1435–1444 (1996).
107. Tönjes, R., Hecht, K., Brautzsch, M., Lucius, R. & Dörner, G. Behavioural changes in adult rats produced by early postnatal maternal deprivation and treatment with choline chloride. Experimental and clinical endocrinology 88, 151–7 (1986).
108. McVey Neufeld, K.-A. et al. Neurobehavioural effects of Lactobacillus rhamnosus GG alone and in combination with prebiotics polydextrose and galactooligosaccharide in male rats exposed to early-life stress. Nutritional neuroscience 1–10 (2017) doi:10.1080/1028415X.2017.1397875.
109. Hofer, M. A. Early relationships as regulators of infant physiology and behavior. Acta Paediatrica 83, 9–18 (1994).
110. Lai, M. C., Yang, S. N. & Huang, L. T. Neonatal Isolation Enhances Anxiety-like Behavior Following Early-life Seizure in Rats. Pediatrics and Neonatology 49, 19–25 (2008).
111. Wang, L., Jiao, J. & Dulawa, S. C. Infant maternal separation impairs adult cognitive performance in BALB/cJ mice. Psychopharmacology 216, 207–218 (2011).
112. A.L., V. et al. Neurobiological effects of neonatal maternal separation and post-weaning environmental enrichment. Behavioural Brain Research 240, 110–118 (2013).
113. Sachs, B. D. et al. The effects of brain serotonin deficiency on behavioural disinhibition and anxiety-like behaviour following mild early life stress. International Journal of Neuropsychopharmacology 16, 2081–2094 (2013).
114. Wang, Q., Li, M., Du, W., Shao, F. & Wang, W. The different effects of maternal separation on spatial learning and reversal learning in rats. Behavioural Brain Research 280, 16–23 (2015).
115. Franks, B., Champagne, F. A. & Curley, J. P. Postnatal maternal care predicts divergent weaning strategies and the development of social behavior. Developmental psychobiology 57, 809–817 (2015).
116. Fabricius, K., Wörtwein, G. & Pakkenberg, B. The impact of maternal separation on adult mouse behaviour and on the total neuron number in the mouse hippocampus. Brain structure & function 212, 403–16 (2008).

Supplementary Information for Chapter 5

Effects of early life adversity on biochemical indicators of the dopaminergic system: a 3-level meta-analysis of rodent studies.

Contents

S1 Appendix | Search string

Pubmed

("dopamine"[MeSH Terms] OR "dopamine"[tiab] OR "dopaminergic"[tiab] OR "dopamin*"[tiab] OR "DRD1"[tiab] OR "DRD2"[tiab] OR "DRD3"[tiab] OR "DRD4"[tiab] OR "DRD5"[tiab] OR "tyrosine hydroxylase"[tiab] OR "3,4-dihydroxyphenethylamine"[tiab] OR "L-DOPA"[tiab] OR "D1-like"[tiab] OR "D2-like"[tiab] OR "DAT"[tiab]

OR

"dopamine" OR "dopaminergic" OR "dopamin*" OR "DRD1" OR "DRD2" OR "DRD3" OR "DRD4" OR "DRD5" OR "tyrosine hydroxylase" OR "3,4-dihydroxyphenethylamine" OR "L-DOPA" OR "D1-like" OR "D2-like")

AND

("early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "early stress"[tiab] OR "prenatally stressed"[tiab] OR "neonatal stress"[tiab] OR "postnatal stress"[tiab] OR "prenatal stress"[tiab] OR "perinatal stress"[tiab] OR "neonatally stressed"[tiab] OR "prenatal adversity"[tiab] OR "prenatal adversities"[tiab] OR "perinatal adversity"[tiab] OR "perinatal adversities"[tiab] OR "gentling"[tiab] OR "early handling"[tiab] OR "neonatal handling" OR "early

stress"[tiab] OR "early adverse experience"[tiab] OR "perinatally stressed"[tiab] OR "early adverse experiences"[tiab] OR "maternal separation"[tiab] OR "maternal deprivation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR "paternal care"[tiab] OR "maternal care"[tiab] OR "paternal separation"[tiab] OR "paternal deprivation"[tiab]
OR "licking and grooming"[tiab] OR "licking-grooming"[tiab] OR "licking/grooming"[tiab] OR "communal housing"[tiab] OR "communal nesting"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "prenatal manipulation"[tiab] OR "prenatal manipulations"[tiab] OR "perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "prenatal LPS"[tiab] OR "prenatal lipopolysaccharide"[tiab] OR "perinatal LPS"[tiab] OR "perinatal lipopolysaccharide"[tiab] OR "postnatal LPS"[tiab] OR "postnatal lipopolysaccharide"[tiab] OR "perinatal restraint"[tiab] OR "prenatal restraint"[tiab] OR "early life glucocorticoid exposure"[tiab] OR "early-life glucocorticoid exposure"[tiab] OR "postnatal glucocorticoid exposure"[tiab] OR "neonatal glucocorticoid exposure"[tiab] OR "prenatal glucocorticoid exposure"[tiab] OR "perinatal glucocorticoid exposure"[tiab] OR "maternal immune activation"[tiab] OR "MIA"[tiab] OR "prenatal poly"[tiab] OR "prenatal flu"[tiab] OR "perinatal poly"[tiab] OR "perinatal flu"[tiab] OR "prenatal chronic unpredictable stress"[tiab] OR "perinatal chronic unpredictable stress"[tiab] OR "postnatal chronic unpredictable stress"[tiab] OR "neonatal chronic unpredictable stress"[tiab] OR "prenatal chronic variable stress"[tiab] OR "perinatal chronic variable stress"[tiab] OR "postnatal chronic variable stress"[tiab] OR "neonatal chronic variable stress"[tiab]

OR

"early life stress" OR "ELS" OR "early life adversity" OR "early life adversities" OR "early stress" OR "prenatally stressed" OR "neonatal stress" OR "postnatal stress" OR "prenatal stress" OR "perinatal stress" OR "neonatally stressed" OR "prenatal adversity" OR "prenatal adversities" OR "perinatal adversity" OR "perinatal adversities" OR "gentling" OR "early handling" OR "neonatal handling" OR "early stress" OR "early adverse experience" OR "perinatally stressed" OR "early adverse experiences" OR "maternal separation" OR "maternal deprivation" OR "limited bedding" OR "limited nesting" OR "limited material" OR "paternal care" OR "maternal care" OR "paternal separation" OR "paternal deprivation" OR "licking and grooming" OR "licking-grooming" OR "licking/grooming" OR "communal housing" OR "communal nesting" OR "postnatal manipulation" OR "postnatal manipulations" OR "prenatal manipulation" OR "prenatal manipulations" OR "perinatal manipulation" OR "perinatal manipulations" OR "prenatal LPS" OR "prenatal lipopolysaccharide" OR "perinatal LPS" OR "perinatal lipopolysaccharide" OR "postnatal LPS" OR "postnatal lipopolysaccharide" OR "perinatal restraint" OR "prenatal restraint" OR "early life glucocorticoid exposure" OR "early-life glucocorticoid exposure" OR

"postnatal glucocorticoid exposure" OR "neonatal glucocorticoid exposure" OR "prenatal glucocorticoid exposure" OR "perinatal glucocorticoid exposure" OR "maternal immune activation" OR "MIA" OR "prenatal poly" OR "prenatal flu" OR "perinatal poly" OR "perinatal flu" OR "prenatal chronic unpredictable stress" OR "perinatal chronic unpredictable stress" OR "postnatal chronic unpredictable stress" OR "neonatal chronic unpredictable stress" OR "prenatal chronic variable stress" OR "perinatal chronic variable stress" OR "postnatal chronic variable stress" OR "neonatal chronic variable stress")

AND

(mus[Tiab] OR murine[Tiab] OR wood mouse[tiab] OR murine[Tiab] OR muridae[Tiab] OR cotton rat[tiab] OR cotton rats[tiab] OR hamster[tiab] OR hamsters[tiab] OR cricetinae[tiab] OR rodentia[Tiab] OR rodent[Tiab] OR rodents[Tiab] OR "rodentia"[MeSH Terms] OR "rodentia"[tiab] OR "mice"[MeSH Terms] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[MeSH Terms] OR "rat"[tiab] OR "rats"[tiab] OR "rodent"[tiab] OR "dengus"[tiab]

OR

mus OR murine OR wood mouse OR murinae OR muridae OR cotton rat OR cotton rats OR hamster OR hamsters OR cricetinae OR rodentia OR rodent OR rodents OR "rodentia" OR "rodentia" OR "mice" OR "mouse" OR "rat" OR "rats" OR "rodent" OR "dengus")

Web of Science

"dopamine" OR "dopaminergic" OR "dopamin*" OR "DRD1" OR "DRD2" OR "DRD3" OR "DRD4" OR "DRD5" OR "tyrosine hydroxylase" OR "3 4 dihydroxyphenethylamine" OR "L-DOPA" OR "D1-like" OR "D2-like"

AND

"early life stress" OR "ELS" OR "early life adversity" OR "early life adversities" OR "early stress" OR "prenatally stressed" OR "neonatal stress" OR "postnatal stress" OR "prenatal stress" OR "perinatal stress" OR "neonatally stressed" OR "prenatal adversity" OR "prenatal adversities" OR "perinatal adversity" OR "perinatal adversities" OR "gentling" OR "early handling" OR "neonatal handling" OR "early stress" OR "early adverse experience" OR "perinatally stressed" OR "early adverse experiences" OR "maternal separation" OR "maternal deprivation" OR "limited bedding" OR "limited nesting" OR "limited material" OR "paternal care" OR "maternal care" OR "paternal separation" OR "paternal deprivation" OR "licking and grooming" OR "licking-grooming" OR "licking/grooming" OR "communal housing" OR "communal nesting" OR "postnatal manipulation" OR "postnatal manipulations" OR "prenatal manipulation" OR "prenatal manipulations" OR "perinatal manipulation" OR "perinatal manipulations" OR "prenatal LPS" OR "prenatal lipopolysaccharide" OR "perinatal LPS" OR "perinatal lipopolysaccharide" OR "postnatal LPS" OR "postnatal lipopolysaccharide" OR "perinatal restraint" OR "prenatal restraint" OR "early life glucocorticoid exposure" OR "early-life glucocorticoid exposure" OR "postnatal glucocorticoid exposure" OR "neonatal glucocorticoid exposure" OR "prenatal glucocorticoid exposure" OR "perinatal glucocorticoid exposure" OR "maternal immune activation" OR "MIA" OR "prenatal poly" OR "prenatal flu" OR "perinatal poly" OR "perinatal flu" OR "prenatal chronic unpredictable stress" OR "perinatal chronic unpredictable stress" OR "postnatal chronic unpredictable stress" OR "neonatal chronic unpredictable stress" OR "prenatal chronic variable stress" OR "perinatal chronic variable stress" OR "postnatal chronic variable stress" OR "neonatal chronic variable stress"

AND

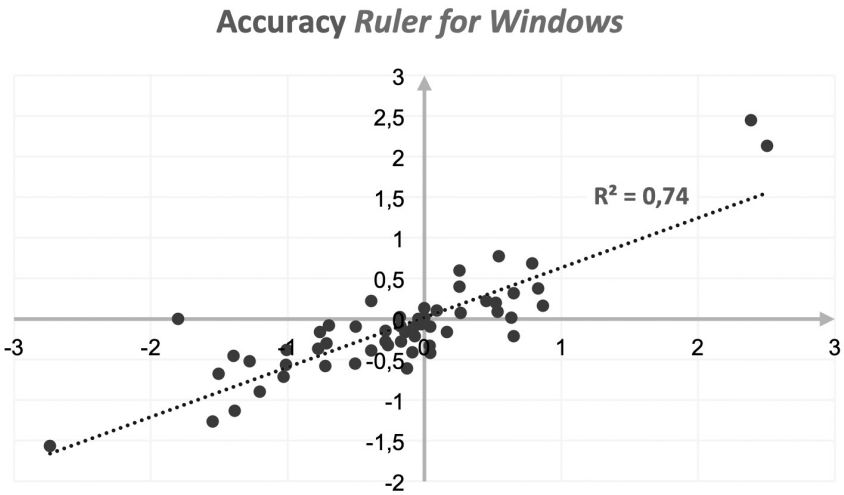
mus OR murine OR wood mouse OR murinae OR muridae OR cotton rat OR cotton rats OR hamster OR hamsters OR cricetinae OR rodentia OR rodent OR rodents OR "rodentia" OR "rodentia" OR "mice" OR "mouse" OR "rat" OR "rats" OR "rodent" OR "dengus"

S2 | Methodology appendix

Table S2 - 1 | Inclusion criteria *specified prior to the beginning of the study. ELS models, species and outcome measures were refined from the original search string due to the limited amount of observations available for such categories. Although licking and grooming as ELS model had enough comparisons, this was excluded from the analysis as all publications originated from the same lab, and it would have therefore not been possible to distinguish between a model- and a lab- dependent effect. LPS = lipopolysaccharide; DA = dopamine; DAT = dopamine transporter; DOPAC = 3,4-Dihydroxyphenylacetic acid; ELS = early life stress; 3-MT = 3-Methoxytyramine; HVA = Homovanillic acid; Th = tyrosine hydroxylase; VTA = ventral tegmental area*

Criteria	Comments
Original primary publications	Unpublished data, reviews and commentaries were excluded
English language	
Studies conducted in mice and rats	
ELS model one of the following: Prenatal stress Injection LPS Injection Poly I:C Restraint Postnatal stress Maternal deprivation/separation (dam removed from the litter) Isolation (pup removed from the litter) Handling (separation <15min/day)	
Experimental output measures: DA DA metabolites (DOPAC, HVA, 3MT) and turnover (DOPAC/DA, HVA/DA, 3MT/DA) Dopamine receptors DAT Th as dopamine precursor	
Outcome assessed in non-stressed conditions, without pharmacological induction.	
Adult animals (>P40) tested	
The brain areas investigated fell into the following categories: Hypothalamic area: Hypothalamus/preoptic/paraventricular nucleus Limbic area hippocampus and amygdala Cortical area (pre)frontal cortex/infralimbic system Striatal area Striatum/caudate putamen/nucleus accumbens VTA area VTA/substantia nigra/midbrain	We selected brain areas involved in part of the dopamine system and involved in the stress response. If a study investigated more than one sub-area of the ones mentioned (e.g. left and right hemisphere, core and shell, etc.), these were considered as separate data points.

Figure S2 - 1 | Accuracy ruler for windows. *For 55 comparisons, data was available both numerically and in graphs. We derived numerical information from the graphs with Ruler for Windows to be able to test its accuracy. The plot below displays the correlation between the calculated values (x axis) and the values given numerically (y axis).*

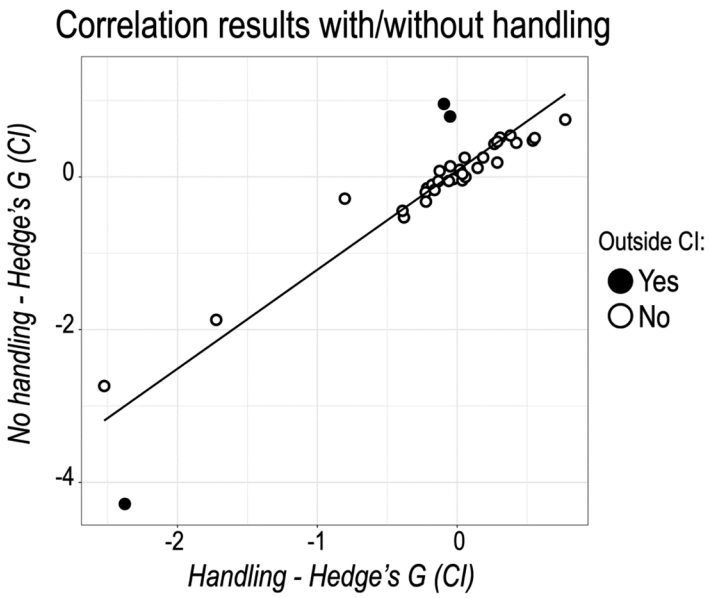


S

Table S2 - 2 | Tool for assessing risk of bias. ** = items from Cochrane's risk of bias tool; Γ = items added to SYRCLE's risk of bias tool. Adapted from (Hooijmans et al., 2014).*

Type of Bias	Item	Domain	Description of domain	Risk of bias question
Selection	1*	**Sequence generation**	Describe the methods used, if any, to generate the allocation sequence in sufficient detail to allow an assessment whether it should produce comparable groups.	Was the group allocation sequence adequately generated and applied?
	2	**Baseline characteristics**	Describe all the possible prognostic factors or animal characteristics, if any, that are compared in order to judge whether or not intervention and control groups were similar at the start of the experiment.	Were the groups similar at baseline or were they adjusted for confounders in the analysis?
	3*	**Allocation concealment**	Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen before or during enrolment.	Did animal selection account for confounders?
Performance	4	**Random housing**	Describe all measures used, if any, to house the animals randomly within the animal room.	Were the animals randomly housed during the experiment?
	5	**Blinding**	Describe all measures used, if any, to blind trial caregivers and researchers from knowing which intervention each animal received. Provide any information relating to whether the intended blinding was effective.	Were the caregivers and/ or investigators blinded from knowledge which intervention each animal received during the experiment?
	6 Γ	**Quality of control**	Describe whether or not the control group was appropriate.	Did the control group offer a reliable baseline?
Detection	7	**Random outcome assessment**	Describe whether or not animals were selected at random for outcome assessment, and which methods to select the animals, if any, were used.	Were animals selected at random for outcome assessment?
	8	**Blinding**	Describe all measures used, if any, to blind outcome assessors from knowing which intervention each animal received. Provide any information relating to whether the intended blinding was effective.	Was the outcome assessor blinded?
Attrition	9*	**Incomplete outcome data**	Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomized animals), reasons for attrition or exclusions, and any re-inclusions in analyses for the review.	Were incomplete outcome data adequately addressed?

Figure S2 - 2 | Impact of handling as an ELS model. *Correlation of the estimated effect sizes for all analyses when comparing the model with and the model without handling. The coherence of the results were evaluated in two ways. First, we correlated the results and verified their consistency (correlation .97). Secondly, for each model we checked whether the estimated effect sizes were contained within the confidence intervals of the other model. This was confirmed in all estimations with ≥3 comparisons. Dots in black were outside confidence intervals of the estimate of the other model. However, each of these had only 2 comparisons in the model with handling and 1 comparison in the model without.*



S3 | Results appendix

Table S3 - 1 | Characteristics across studies. The tables below specify how many comparisons and papers were present in the prenatal as well as postnatal dataset for each potential moderator. As a rule of thumb, 4 comparisons from 3 papers were considered necessary for a meaningful analysis. At times, the grand total sum of papers is lower than the sum of papers of the individual subgroups. For example, the grand total of papers following postnatal ELS is 49. However, the sum of papers of females + males + pooled/unclear is 15+43+3 = 61. This means that 12 papers used females as well as males. *Method = method of assessment; post = postnatal ELS, pre = prenatal ELS.*

	post		pre	
Sex	**comparisons**	**Papers**	**comparisons**	**Papers**
Females	65	15	150	10
Males	401	43	348	35
Pooled/unclear	8	3	37	4
Grand Total	474	49	535	41

	post		pre	
Species-strain	**comparisons**	**Papers**	**comparisons**	**Papers**
Mouse	70	7	172	10
Balb/c			12	1
C57Bl/6J	28	3	132	5
CD1	3	1		
CFW	2	1		
ICR			8	1
Swiss	37	2	2	1
unclear			18	2
Rat	404	42	363	31
Holtzman	8	1		
Lister-Hooded	24	1		
Long Evans	62	5	154	5
SHR	1	1		
Sprague-Dawley	55	16	110	10
Wistar	230	16	99	16
Wistar/Hann	22	1		
WKY	1	1		
unclear	1	1		
Grand Total	474	49	535	41

	post		pre	
Method	**comparisons**	**Papers**	**comparisons**	**Papers**
function	8	3	14	4
protein	429	40	501	36
RNA	37	11	20	5
Grand Total	474	49	535	41

Table S3 – 2 | List of potential moderators specified prior the beginning of the study.

	Potential moderator	Levels	Comments
Biological moderators			
	Outcome measure	DA	
		DOPAC	
		HVA	
		3-MT	
		DR1	
		DR2	
		DR3	not present in prenatal dataset
		DAT	
		Th	
	Brain area	striatal area	
		hypothalamic area	
		limbic area	
		cortical area	
		VTA area	
	sub-brain area		subgroup analysis
		striatum caudate nucleus accumbens	striatal area
		hypothalamus preoptic area	hypothalamic area
		hippocampus amygdala	limbic area
		cingulate cortex (medial)prefrontal cortex	cortical area
		VTA substancia nigra	VTA area
	Sex	Males	
		Females	
		Pooled or unclear	
	Species	Mice	
		Rats	
	Age as continuous variable		
	Level of technique	RNA	
		protein	
		function	
Technical moderators			
	ELS model used	injection LPS	
		injection PolyI:C	prenatal: subgroup analysis
		restraint	
		MD/MS	
		handling	postnatal: subgroup analysis
		injection LPS	
	Amount of potential bias		sensitivity analysis

Table S3-3 | Heterogeneity of moderators: prenatal dataset. The table below displays the test of moderators for each dataset as well as the explained heterogeneity by each potential moderator. Test of moderators was considered significant when the p-value was <.10. Percentage of explained heterogeneity was calculated as percentage improvement in heterogeneity score. *Het = heterogeneity score, df1 = first degree of freedom for moderator test, df2 = second degree of freedom for moderator test, F = f score for the moderator test*

		Test of moderator				
Moderator	Het	df1	df2	F	p-value	% explained het
Outcome	913.745	7	371	3.956	**0.000**	4.3
Brain Area	896.765	4	374	6.144	**0.000**	6.1
Sex	949.804	2	376	0.423	0.655	0.5
Species	952.817	1	377	0.100	0.752	0.2
Age	942.565	1	377	0.197	0.657	1.3
Method of assessment	950.143	2	376	.098	.907	0.5

Table S3 - 4 |Summary effects 3-level model of the prenatal ELS dataset *Hyp = hypothalamic area, limbic = limbic area, striatum = striatal area, VTA = VTA area, Hedges = Hedges G, ci.lb = confidence interval lower boundary, ci.ub = confidence interval upper boundary, com = number of comparisons, exp = number of experiments, papers = number of papers, yes = inclusion in main analysis, no = excluded from main analysis as not enough comparisons are present*

What	Hedges	Pval	ci.lb	ci.ub	Com	Exp	Papers	mainAnalysis
hypDA	0.382	0.080	-0.046	0.809	9	7	6	yes
hypDOPAC	0.314	0.178	-0.143	0.771	7	5	4	yes
hypHVA	0.436	0.158	-0.170	1.043	4	4	3	yes
limbicDA	-0.102	0.586	-0.470	0.266	12	7	5	yes
limbicDOPAC	-0.139	0.476	-0.521	0.244	10	6	4	yes
PFCD1R	-0.335	0.356	-1.047	0.378	5	4	3	yes
PFCD2R	-0.488	0.135	-1.128	0.152	8	4	4	yes
PFCDA	-0.252	0.106	-0.558	0.054	22	18	12	yes
PFCDOPAC	-0.133	0.410	-0.452	0.185	18	12	8	yes
PFCHVA	0.051	0.790	-0.325	0.427	11	9	6	yes
striatumD1R	**0.386**	**0.067**	-0.028	0.799	20	9	7	yes
striatumD2R	0.138	0.522	-0.286	0.563	23	7	7	yes
striatumDA	0.083	0.469	-0.143	0.309	54	35	23	yes
striatumDOPAC	**0.323**	**0.018**	0.056	0.590	32	19	13	yes
striatumHVA	0.199	0.126	-0.056	0.454	37	23	14	yes
striatumTh	**-1.164**	**0.000**	-1.744	-0.584	11	7	5	yes
VTADA	-0.020	0.917	-0.394	0.354	14	6	5	yes
VTADOPAC	-0.020	0.919	-0.414	0.374	10	5	4	yes
VTAHVA	-0.047	0.859	-0.564	0.470	5	5	4	yes
VTATh	-0.131	0.435	-0.462	0.199	32	18	11	yes
hypD1R	0.035	0.951	-1.076	1.146	2	2	1	no
hypD2R	-0.374	0.516	-1.504	0.757	2	2	1	no
hypTh	-3.995	0.000	-5.652	-2.339	2	2	1	no
limbicD1R	0.388	0.292	-0.335	1.111	4	2	1	no
limbicD2R	0.870	0.180	-0.404	2.144	2	2	2	no
limbicHVA	-0.127	0.572	-0.567	0.314	8	3	2	no
limbicMT	-0.277	0.252	-0.752	0.197	6	2	1	no
PFCMT	-0.525	0.314	-1.549	0.498	1	1	1	no
PFCTh	-4.775	0.000	-7.209	-2.342	1	1	1	no
striatumDAT	-1.562	0.074	-3.278	0.154	1	1	1	no
striatumMT	0.464	0.098	-0.086	1.014	4	2	1	no
VTAD2R	-14.885	0.000	-22.219	-7.551	1	1	1	no
VTADAT	-2.696	0.006	-4.626	-0.767	1	1	1	no

Figure S3 - 1 | Forest and funnel plot striatal Th in prenatal dataset. *(A) Forest plot. (B) Funnel plot.*

(A)

(B)

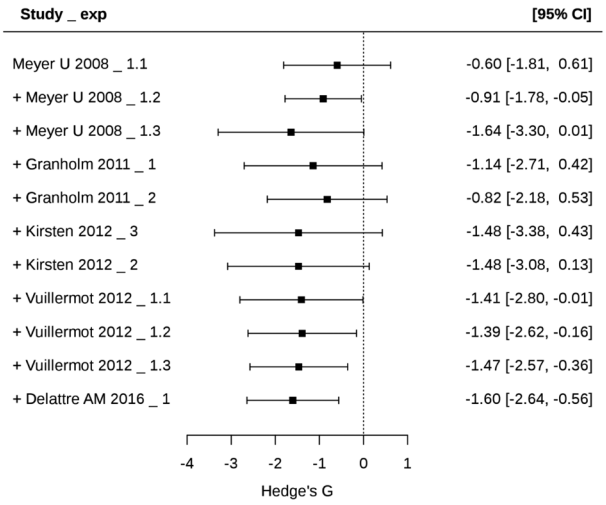

Study _ exp
[95% CI]
Meyer U 2008 _ 1.1
-0.60 [-1.81, 0.61]
+ Meyer U 2008 _ 1.2
-0.91 [-1.78, -0.05]
+ Meyer U 2008 _ 1.3
-1.64 [-3.30, 0.01]
+ Granholm 2011 _ 1
-1.14 [-2.71, 0.42]
+ Granholm 2011 _ 2
-0.82 [-2.18, 0.53]
+ Kirsten 2012 _ 3
-1.48 [-3.38, 0.43]
+ Kirsten 2012 _ 2
-1.48 [-3.08, 0.13]
+ Vuillermot 2012 _ 1.1
-1.41 [-2.80, -0.01]
+ Vuillermot 2012 _ 1.2
-1.39 [-2.62, -0.16]
+ Vuillermot 2012 _ 1.3
-1.47 [-2.57, -0.36]
+ Delattre AM 2016 _ 1
-1.60 [-2.64, -0.56]
-4
-3
-2
-1
0
1
Hedge's G

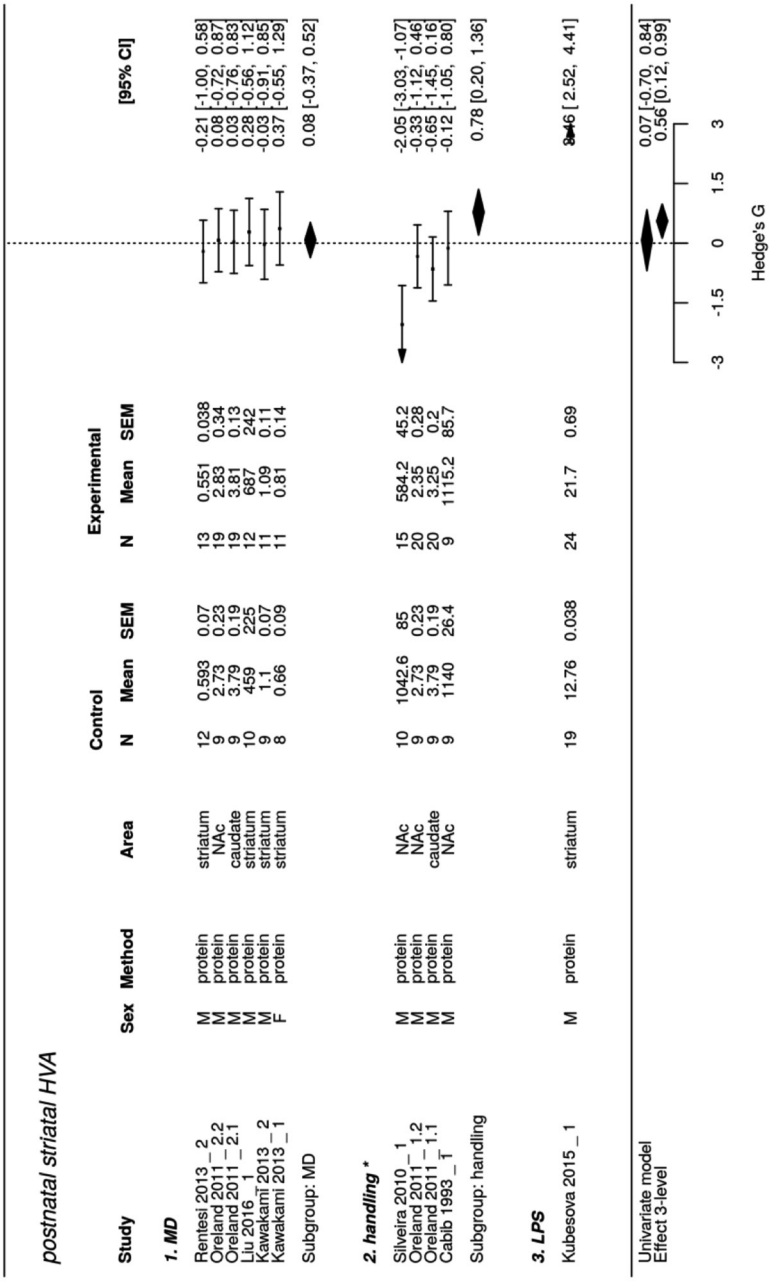
Table S3 - 5 |Summary moderator tests for subgroup analysis in prenatal dataset. No conv = no convergence was achieved in the model.

		Mod. test	
subgroup	**Interaction**	**F**	**p-value**
Model	hypDA	2.671	0.148
	hypDOPAC	3.858	0.107
	limbicDA	3.260	**0.086**
	limbicDOPAC	5.082	**0.054**
	PFCD2R	3.033	0.137
	PFCDA	6.252	**0.008**
	PFCDOPAC	1.076	0.366
	PFCHVA	1.574	0.265
	striatumD1R	4.203	0.033
	striatumD2R	no conv	
	striatumDA	1.216	0.305
	striatumDOPAC	4.721	0.017
	striatumHVA	1.454	0.248
	striatumTh	0.001	0.981
	VTADA	3.988	0.069
	VTADOPAC	1.434	0.265
	VTATh	2.537	**0.097**
subbrain area	hypDA	1.067	0.401
	hypDOPAC	2.090	0.239
	limbicDA	0.055	0.819
	limbicDOPAC	1.185	0.308
	PFCD2R	0.182	0.682
	PFCDA	2.737	0.113
	PFCDOPAC	0.922	0.350
	PFCHVA	0.000	0.985
	striatumD1R	0.519	0.604
	striatumD2R	1.286	0.298
	striatumDA	**4.343**	**0.018**
	striatumDOPAC	0.832	0.445
	striatumHVA	0.050	0.951
	striatumTh	0.578	0.583
	VTADA	1.576	0.256
	VTADOPAC	no conv	
	VTATh	0.067	0.991

Figure S3 - 2 | Forest plot subgroup analysis Th in VTA of the prenatal dataset.

Figure S3 - 3 | Forest plot subgroup analysis DA in striatal area of the prenatal dataset.

Figure S3 - 4 | Forest plot striatal HVA in the postnatal dataset. ** = as explained in the method's section, the effect sizes of the ELS model handling were multiplied by -1 because handling was expected to have an opposite direction from the other ELS models. It follows that in this graph an increase in handling signifies a decrease in protein concentration.*



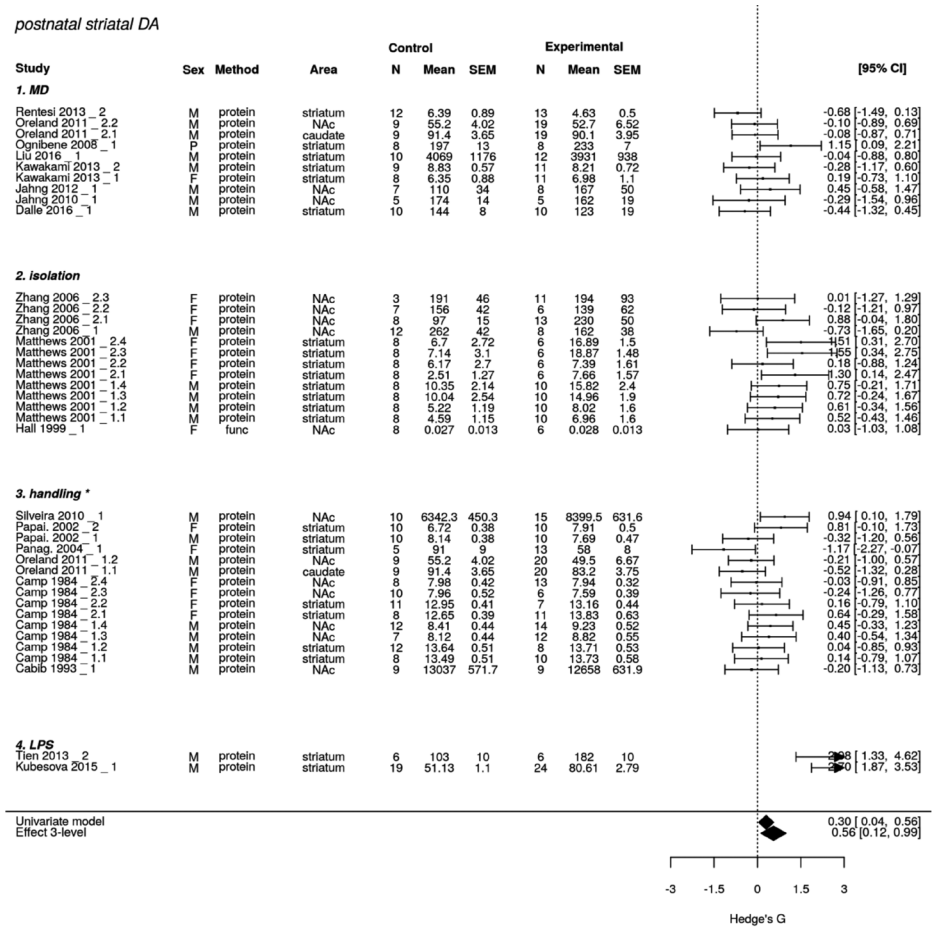


Figure S3 - 5 | Forest plot striatal DA in the postnatal dataset.

S

Table S3 - 6 | Summary subgroup analyses prenatal dataset. The table below provides information about all subgroup analyses of the prenatal dataset. Comp = comparison, ci.lb = confidence interval lower boundary, ci.ub = confidence interval upper boundary, pval = p-value, heter = heterogeneity (Q statistic), het pvalue = p-value Q statistic, hip = hippocampus, MD = separation from the mother

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
hypothalamic	**DA**	Tech Level	protein	9	6	yes	0.484	-0.068	1.036	0.086	16.808	0.032
		Species	Rat	8	5	yes	0.315	-0.191	0.822	0.222	11.314	0.125
		Model	restraint	7	4	yes	0.443	-0.064	0.95	0.087	8.359	0.213
		Sub-brain area	hypothalamus	7	4	yes	0.306	-0.348	0.96	0.359	11.167	0.083
		Sex	Males	6	5	yes	0.487	-0.377	1.35	0.269	15.816	0.007
			Females	3	2	no	0.551	0.08	1.022			
	DOPAC	Tech Level	protein	7	4	yes	0.43	-0.042	0.903	0.074	10.862	0.093
		Model	restraint	6	3	yes	0.253	-0.134	0.639	0.2	6.356	0.273
		Species	Rat	6	3	yes	0.253	-0.134	0.639	0.2	6.356	0.273
		Sub-brain area	hypothalamus	5	2	no	0.168	-0.288	0.624			
		Sex	Males	4	3	yes	0.519	-0.42	1.457	0.279	9.724	0.021
			Females	3	2	no	0.401	-0.068	0.87			
limbic	**DA**	Tech Level	protein	12	5	yes	-0.158	-0.392	0.076	0.185	10.97	0.446
		Sub-brain area	hip	9	5	yes	-0.143	-0.409	0.123	0.291	1.696	0.989
			amygdala	3	2	no	-0.618	-2.059	0.824			
		Model	Poly	7	2	no	-0.086	-0.372	0.2			
			restraint	3	2	no	-0.056	-0.51	0.398			

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
			LPS	2	1	no	-1.442	-3.107	0.224			
		Sex	Males	7	4	yes	-0.379	-0.889	0.131	0.145	9.394	0.153
			Females	5	3	yes	-0.036	-0.373	0.3	0.833	0.6	0.963
		Species	Mouse	7	2	no	-0.086	-0.372	0.2			
			Rat	5	3	yes	-0.393	-1.053	0.267	0.243	9.577	0.048
	DOPAC	Tech Level	protein	10	4	yes	-0.146	-0.421	0.129	0.299	9.284	0.411
		Sub-brain area	hip	8	4	yes	-0.153	-0.525	0.218	0.418	8.545	0.287
			amygdala	2	1	no	-0.015	-0.54	0.509			
		Model	Poly	7	2	no	-0.353	-0.643	-0.063			
		Sex	Males	5	3	yes	-0.198	-0.651	0.255	0.391	6.004	0.199
			Females	5	3	yes	-0.072	-0.518	0.373	0.751	3.18	0.528
		Model	restraint	3	2	no	0.268	-0.188	0.723			
		Species	Mouse	7	2	no	-0.353	-0.643	-0.063			
			Rat	3	2	no	0.268	-0.188	0.723			
prefrontal	**DR2**	Tech Level	protein	8	4	yes	-0.217	-1.213	0.779	0.67	16.917	0.018
		Sub-brain area	PFC	8	4	yes	-0.217	-1.213	0.779	0.67	16.917	0.018
		Model	LPS	5	1	no	-1.025	-1.802	-0.247			
			restraint	2	2	no	0.758	-0.23	1.746			
		Sex	Males	7	3	yes	-0.065	-1.389	1.26	0.924	16.907	0.01
		Species	Rat	7	3	yes	-0.065	-1.389	1.26	0.924	16.907	0.01
	DA	Sub-brain area	PFC	22	12	yes	-0.417	-0.911	0.077	0.098	88.632	0
		Tech Level	protein	20	10	yes	-0.504	-1.051	0.042	0.07	85.04	0
			function	2	2	no	0.22	-0.277	0.717			
		Model	Poly	5	4	yes	0.14	-0.213	0.493	0.437	4.412	0.353
			restraint	14	6	yes	-0.279	-0.84	0.282	0.329	50.423	0
			LPS	3	2	no	-2.489	-4.088	-0.889			
		Sex	Males	14	11	yes	-0.657	-1.357	0.043	0.066	67.172	0
			Females	7	6	yes	-0.13	-1.019	0.758	0.774	19.017	0.004

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
		Species	Mouse	4	3	yes	0.058	-0.329	0.445	0.769	3.462	0.326
			Rat	18	9	yes	-0.57	-1.183	0.043	0.068	83.329	0
	DOPAC	Sub-brain area	PFC	18	8	yes	-0.12	-0.366	0.125	0.337	18.333	0.368
		Tech Level	protein	17	7	yes	-0.146	-0.415	0.124	0.289	17.861	0.332
		Model	Poly	5	4	yes	-0.115	-0.796	0.566	0.74	13.831	0.008
			restraint	11	3	yes	-0.019	-0.284	0.246	0.886	1.695	0.998
			LPS	2	1	no	-0.687	-1.45	0.076			
		Sex	Males	11	7	yes	-0.156	-0.59	0.277	0.479	14.482	0.152
			Females	6	4	yes	-0.124	-0.484	0.236	0.5	3.237	0.663
		Species	Mouse	4	3	yes	-0.181	-1.044	0.682	0.681	13.12	0.004
			Rat	14	5	yes	-0.069	-0.308	0.17	0.572	4.785	0.98
	HVA	Tech Level	protein	11	6	yes	-0.003	-0.322	0.315	0.984	14.074	0.17
		Sub-brain area	PFC	11	6	yes	-0.003	-0.322	0.315	0.984	14.074	0.17
		Model	Poly	3	2	no	-0.287	-0.718	0.145			
			restraint	7	3	yes	0.233	-0.201	0.668	0.293	8.493	0.204
		Sex	Males	6	5	yes	0.036	-0.408	0.48	0.873	6.664	0.247
			Females	4	3	yes	-0.038	-0.821	0.744	0.924	6.751	0.08
		Species	Mouse	3	2	no	-0.287	-0.718	0.145			
			Rat	8	4	yes	0.155	-0.256	0.566	0.459	9.93	0.193
striatal	**DR1**	Tech Level	protein	18	6	yes	0.113	-0.166	0.391	0.428	15.765	0.541
			RNA	2	2	no	0.997	-1.702	3.697			
		Sub-brain area	NAc	10	3	yes	0.025	-0.319	0.368	0.889	3.665	0.932
			striatum	7	6	yes	0.483	-0.315	1.281	0.236	15.863	0.015
			caudate	3	2	no	0.288	-0.337	0.913			
		Model	LPS	6	3	yes	0.521	-0.27	1.311	0.197	6.95	0.224
			Poly	13	3	yes	-0.041	-0.36	0.278	0.8	7.057	0.854
		Sex	Males	10	5	yes	0.64	-0.196	1.475	0.133	14.837	0.096
			Pooled or unclear	10	2	no	-0.019	-0.435	0.396			

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
		Species	Mouse	15	5	yes	0.467	-0.401	1.336	0.291	21.014	0.101
			Rat	5	2	no	0.249	-0.246	0.743			
	DR2	Tech Level	protein	15	6	yes	-0.042	-0.48	0.395	0.85	14.623	0.404
			RNA	8	1	no	0.721	0.31	1.131			
		Sub-brain area	NAc	12	5	yes	0.285	-0.199	0.769	0.248	18.4	0.073
			caudate	7	3	yes	0.578	0.079	1.077	0.023	4.542	0.604
			striatum	4	4	yes	-0.313	-0.877	0.251	0.277	2.766	0.429
		Model	restraint	12	2	no	0.734	0.395	1.073			
			Poly	7	3	yes	-0.276	-0.77	0.218	0.274	3.572	0.734
			LPS	4	2	no	-0.149	-0.706	0.408			
		Sex	Males	19	5	yes	0.267	-0.191	0.725	0.253	25.462	0.113
			Pooled or unclear	4	2	no	-0.366	-1.34	0.609			
		Species	Rat	16	4	yes	0.401	-0.095	0.898	0.113	20.815	0.143
			Mouse	7	3	yes	-0.276	-0.77	0.218	0.274	3.572	0.734
	DA	Tech Level	protein	50	20	yes	0.034	-0.219	0.287	0.793	111.899	0
			function	4	3	yes	0.748	-0.155	1.652	0.104	8.811	0.032
		Sub-brain area	striatum	29	17	yes	-0.122	-0.384	0.14	0.362	57.423	0.001
			NAc	22	13	yes	0.432	0.103	0.76	0.01	54.512	0
			caudate	3	2	no	0.094	-0.356	0.544			
		Model	restraint	24	9	yes	0.244	-0.042	0.529	0.094	48.044	0.002
			LPS	20	9	yes	-0.217	-0.804	0.369	0.467	72.162	0
			Poly	10	5	yes	0.148	-0.108	0.404	0.257	5.644	0.775
		Species	Rat	46	19	yes	0.105	-0.205	0.415	0.507	124.221	0
			Mouse	8	4	yes	0.052	-0.232	0.336	0.719	2.278	0.943
		Sex	Males	36	19	yes	0.197	-0.192	0.585	0.321	99.501	0
			Females	15	8	yes	-0.054	-0.36	0.252	0.73	21.231	0.096
			Pooled or unclear	3	2	no	-0.182	-0.717	0.353			
	DOPAC	Tech Level	protein	31	12	yes	0.513	0.205	0.82	0.001	74.385	0
		Sub-brain area	striatum	15	9	yes	0.764	0.249	1.278	0.004	52.514	0

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
			NAc	14	7	yes	0.293	0.014	0.571	0.039	18.05	0.156
			caudate	3	2	no	0.277	-0.177	0.731			
		Model	restraint	13	5	yes	0.012	-0.386	0.41	0.951	22.913	0.028
			LPS	11	4	yes	0.917	0.478	1.355	0	19.554	0.034
			Poly	8	4	yes	0.635	0.127	1.142	0.014	12.734	0.079
		Sex	Males	22	11	yes	0.55	0.34	0.761	0	26.28	0.196
			Females	9	5	yes	0.17	-0.572	0.911	0.654	23.415	0.003
		Species	Rat	25	10	yes	0.459	0.112	0.806	0.009	60.141	0
			Mouse	7	3	yes	0.697	0.038	1.355	0.038	12.67	0.049
	HVA	Tech Level	protein	37	14	yes	0.239	0.046	0.431	0.015	46.35	0.116
		Sub-brain area	striatum	19	10	yes	0.234	-0.006	0.474	0.056	18.766	0.406
			NAc	15	8	yes	0.21	-0.114	0.533	0.204	26.714	0.021
			caudate	3	2	no	0.347	-0.107	0.801			
		Model	Poly	7	3	yes	0.504	0.186	0.823	0.002	8.005	0.238
			LPS	17	6	yes	0.177	-0.091	0.445	0.195	15.256	0.506
			restraint	13	5	yes	0.126	-0.228	0.48	0.485	18.59	0.099
		Sex	Males	26	12	yes	0.288	0.104	0.473	0.002	23.705	0.536
			Females	10	5	yes	0.065	-0.362	0.493	0.764	16.468	0.058
		Species	Mouse	7	3	yes	0.504	0.186	0.823	0.002	8.005	0.238
			Rat	30	11	yes	0.147	-0.072	0.366	0.188	34.094	0.236
	Th	Tech Level	protein	10	5	yes	-1.569	-2.994	-0.145	0.031	33.16	0
		Sub-brain area	striatum	6	4	yes	-1.446	-3.09	0.197	0.085	22.742	0
			NAc	4	2	no	-1.946	-2.81	-1.082			
		Model	Poly	6	2	no	-1.535	-2.223	-0.848			
			LPS	5	3	yes	-1.711	-3.791	0.369	0.107	22.706	0
		Sex	Males	8	4	yes	-1.614	-3.208	-0.02	0.047	26.53	0
			Pooled or unclear	3	1	no	-1.63	-3.176	-0.084			
		Species	Rat	3	2	no	-3.229	-5.344	-1.115			

Brain Area	Outcome	Moderator	Level Moderator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
			Mouse	8	3	yes	-0.808	-1.854	0.237	0.13	20.401	0.005
VTA	**DA**	Tech Level	protein	10	4	yes	0.116	-0.373	0.604	0.642	19.249	0.023
			function	4	1	no	-1.891	-3.986	0.204			
		Sub-brain area	midbrain	7	2	no	0.1	-0.743	0.942			
			VTA	5	2	no	-1.541	-3.26	0.178			
		Model	restraint	9	3	yes	0.075	-0.505	0.656	0.8	18.65	0.017
			Poly	5	2	no	-1.431	-3.265	0.403			
		Sex	Males	10	4	yes	-0.401	-1.722	0.92	0.552	74.243	0
			Females	4	2	no	-0.059	-0.451	0.333			
		Species	Rat	13	4	yes	-0.367	-1.302	0.568	0.442	75.753	0
	DOPAC	Tech Level	protein	10	4	yes	-0.058	-0.322	0.205	0.664	7.877	0.547
		Model	restraint	9	3	yes	-0.108	-0.384	0.167	0.441	6.443	0.598
		Sub-brain area	midbrain	7	2	no	-0.159	-0.485	0.167			
			VTA	2	2	no	0.074	-0.58	0.727			
		Sex	Males	6	3	yes	0.017	-0.424	0.459	0.938	4.658	0.459
			Females	4	2	no	-0.106	-0.501	0.288			
		Species	Rat	9	3	yes	-0.108	-0.384	0.167	0.441	6.443	0.598
	Th	Tech Level	protein	31	11	yes	-0.16	-0.488	0.168	0.339	77.904	0
		Sub-brain area	SN	21	8	yes	-0.149	-0.591	0.293	0.508	60.438	0
			locuscoerelius	2	1	no	-0.56	-3.386	2.266			
			mesencephalon	2	1	no	-0.025	-0.795	0.744			
			VTA	6	5	yes	-0.06	-0.867	0.747	0.884	10.376	0.065
		Model	LPS	22	6	yes	0.035	-0.308	0.379	0.84	51.853	0
			restraint	7	3	yes	-0.858	-1.743	0.027	0.057	18.036	0.006
			Poly	3	2	no	0.241	-0.374	0.855			
		Sex	Males	26	10	yes	-0.149	-0.521	0.223	0.433	67.844	0
			Females	6	2	no	-0.32	-1.587	0.948			
		Species	Rat	25	8	yes	-0.236	-0.683	0.211	0.3	76.183	0
			Mouse	7	3	yes	0.115	-0.311	0.542	0.596	3.245	0.778

Table S3 - 7 | Heterogeneity of moderators: postnatal dataset. *The table below displays the test of moderators for each dataset as well as the explained heterogeneity by each potential moderator. Test of moderators was considered significant when the p-value was <.10. Percentage of explained heterogeneity was calculated as percentage improvement in heterogeneity score. Het = heterogeneity score, df1 = first degree of freedom for moderator test, df2 = second degree of freedom for moderator test, F = f score for the moderator test*

		Test of moderator				
Moderator	Het	df1	df2	F	p-value	% explained het
Outcome	951.918	7	374	10.495	<.001	10.30
Brain area	1053	4	377	2.035	0.089	0.78
Sex	1032.859	2	379	1.349	0.261	2.68
Species	1061.179	1	380	0.647	0.422	0.01
Age	1057.408	1	380	0.883	0.348	0.36
Method of assessment	1053.578	2	379	.289	.749	0.72

Table S3 - 8 |Summary effects 3-level model of the postnatal ELS dataset. *Hyp = hypothalamic area, limbic = limbic area, striatum = striatal area, VTA = VTA area, Hedges = Hedges G, ci.lb = confidence interval lower boundary, ci.ub = confidence interval upper boundary, com = number of comparisons, exp = number of experiments, papers = number of papers, yes = inclusion in main analysis, no = excluded from main analysis as not enough comparisons are present.*

What	Hedges	ci.ib	ci.ub	Pval	C	Exp	Papers	mainAnalysis
limbicDA	-0.225	-0.589	0.139	0.224	20	12	8	yes
limbicDOPAC	0.186	-0.257	0.630	0.409	10	8	6	yes
limbicHVA	0.424	-0.243	1.090	0.212	4	4	3	yes
PFCD1R	-0.178	-0.673	0.317	0.479	14	5	3	yes
PFCD2R	0.054	-0.510	0.618	0.852	9	6	4	yes
PFCDA	0.020	-0.362	0.401	0.919	16	13	8	yes
PFCDOPAC	0.287	-0.122	0.696	0.169	13	11	7	yes
PFCHVA	0.146	-0.263	0.555	0.483	13	11	7	yes
striatumD1R	**-0.391**	-0.787	-0.005	**0.053**	19	12	8	yes
striatumD2R	0.037	-0.249	0.323	0.801	42	21	14	yes
striatumD3R	0.060	-0.364	0.483	0.781	12	5	4	yes
striatumDA	**0.307**	0.017	0.597	**0.038**	40	24	18	yes
striatumDAT	0.037	-0.309	0.383	0.834	24	13	9	yes
striatumDOPAC	**0.541**	0.135	0.948	**0.009**	13	11	9	yes
striatumHVA	**0.556**	0.123	0.988	**0.012**	11	9	7	yes
VTADA	0.266	-0.286	0.819	0.344	7	5	4	yes
VTATh	-0.223	-0.630	0.184	0.282	27	15	10	yes
hypD1R	-0.050	-0.634	0.534	0.867	10	2	1	no
hypD2R	-0.215	-0.735	0.305	0.416	14	2	1	no
hypDA	0.360	-0.489	1.209	0.405	3	3	2	no
hypDOPAC	0.107	-1.417	1.630	0.891	1	1	1	no
hypTh	-0.589	-1.892	0.715	0.375	3	1	1	no
limbicD1R	-0.094	-0.800	0.612	0.794	6	2	1	no
limbicD2R	-0.380	-1.007	0.247	0.234	7	3	2	no
limbicDAT	-0.028	-1.304	1.249	0.966	2	2	1	no
limbicMT	-2.524	-3.657	-1.391	0.000	1	1	1	no
PFCD3R	0.773	-0.486	2.032	0.228	2	1	1	no
PFCDAT	-0.162	-1.189	0.865	0.756	2	2	2	no
PFCMT	-1.723	-2.602	-0.844	0.000	2	2	2	no
striatumMT	-2.377	-3.317	-1.438	0.000	2	2	2	no
striatumTh	-0.060	-0.991	0.872	0.900	3	3	2	no
VTAD1R	-0.134	-0.941	0.674	0.745	4	2	1	no
VTAD2R	-0.804	-1.310	-0.297	0.002	8	4	2	no
VTAD3R	-0.125	-0.728	0.478	0.683	4	2	1	no
VTADAT	-0.049	-0.626	0.528	0.868	5	3	2	no
VTADOPAC	0.381	-0.217	0.978	0.211	5	3	2	no
VTAHVA	0.287	-0.358	0.932	0.383	4	2	1	no

Table S3 -9 |Summary moderator tests for subgroup analysis in postnatal dataset. *No conv = no convergence was achieved in the model.*

subgroup	interaction	F	p-value
Model	limbicDA	1.654	0.217
	limbicDOPAC	6.619	0.024
	PFCD1R	0.329	0.577
	PFCD2R	0.300	0.601
	PFCDA	5.152	0.016
	PFCDOPAC	9.439	0.005
	PFCHVA	10.051	0.004
	striatumD1R	0.679	0.579
	striatumD2R	1.706	0.182
	striatumD3R	1.190	0.348
	striatumDA	11.179	0.000
	striatumDAT	1.094	0.375
	striatumDOPAC	6.127	0.018
	striatumHVA	15.116	0.002
	VTADA	6.601	0.054
	VTATh	1.484	0.247
spec. brain area	limbicDA	0.000	0.996
	limbicDOPAC	0.090	0.772
	PFCD1R	0.146	0.709
	PFCD2R	0.217	0.655
	PFCDA	0.421	0.527
	PFCDOPAC	0.035	0.855
	PFCHVA	0.002	0.961
	striatumD1R	0.107	0.899
	striatumD2R	0.296	0.745
	striatumD3R	0.004	0.950
	striatumDA	0.542	0.586
	striatumDAT	0.444	0.647
	striatumDOPAC	0.051	0.951
	striatumHVA	0.035	0.966
	VTADA	no conv	
	VTATh	0.298	0.590

Table S3 - 10 | Summary subgroup analyses postnatal dataset. The table below provides information about all subgroup analyses of the postnatal dataset. *Comp = comparison, ci.lb = confidence interval lower boundary, ci.ub = confidence interval upper boundary, pval = p-value, heter = heterogeneity (Q statistic), het pvalue = p-value Q statistic, hip = hippocampus, MD = separation from the mother*

Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
limbic	**DA**	Model	MD	*7*	*5*	*yes*	*-0.09*	*-0.414*	*0.234*	*0.587*	*4.756*	*0.575*
			handling	*4*	*2*	*no*	*-0.347*	*-2.259*	*1.565*			
			isolation	*8*	*1*	*no*	*-0.715*	*-1.39*	*-0.04*			
		Tech Level	*protein*	*20*	*8*	*yes*	*-0.279*	*-0.676*	*0.119*	*0.17*	*66.22*	*0*
		Sex	*Females*	*6*	*3*	*yes*	*0.158*	*-0.93*	*1.246*	*0.776*	*13.92*	*0.016*
			Males	*14*	*8*	*yes*	*-0.304*	*-0.805*	*0.197*	*0.234*	*52.29*	*0*
		Species	*Rat*	*17*	*6*	*yes*	*-0.376*	*-0.84*	*0.088*	*0.112*	*63.48*	*0*
			Mouse	*3*	*2*	*no*	*0.192*	*-0.317*	*0.7*			
		Brain area	*amygdala*	*3*	*2*	*no*	*-0.275*	*-0.735*	*0.184*			
			hip	*17*	*7*	*yes*	*-0.291*	*-0.769*	*0.187*	*0.232*	*64.11*	*0*
	DOPAC	Model	MD	*7*	*5*	*yes*	*0.088*	*-0.236*	*0.412*	*0.593*	*5.048*	*0.538*
			handling	*2*	*1*	*no*	*-0.14*	*-0.698*	*0.417*			
		Tech Level	*protein*	*10*	*6*	*yes*	*0.26*	*-0.157*	*0.677*	*0.222*	*18.53*	*0.029*
		Sex	*Males*	*9*	*6*	*yes*	*0.305*	*-0.155*	*0.766*	*0.194*	*17.98*	*0.021*
		Species	*Rat*	*7*	*4*	*yes*	*0.309*	*-0.265*	*0.882*	*0.291*	*15.3*	*0.018*
			Mouse	*3*	*2*	*no*	*0.167*	*-0.474*	*0.808*			
		Brain area	*hip*	*7*	*5*	*yes*	*0.253*	*-0.273*	*0.779*	*0.346*	*17.85*	*0.007*
			amygdala	*3*	*2*	*no*	*0.082*	*-0.373*	*0.537*			
PFC	**DR1**	Model	MD	*10*	*3*	*yes*	*-0.244*	*-0.605*	*0.116*	*0.184*	*10.82*	*0.288*
			handling	*4*	*1*	*no*	*-0.025*	*-0.709*	*0.659*			
		Tech Level	*protein*	*14*	*3*	*yes*	*-0.192*	*-0.503*	*0.118*	*0.224*	*14.95*	*0.31*
		Sex	*Males*	*14*	*3*	*yes*	*-0.192*	*-0.503*	*0.118*	*0.224*	*14.95*	*0.31*
		Species	*Rat*	*14*	*3*	*yes*	*-0.192*	*-0.503*	*0.118*	*0.224*	*14.95*	*0.31*
		Brain area	*PFC*	*10*	*3*	*yes*	*-0.225*	*-0.755*	*0.305*	*0.406*	*11.78*	*0.226*
			cingulatecortex	*4*	*1*	*no*	*-0.084*	*-1.13*	*0.962*			

Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
	DR2	Model	MD	7	4	yes	0.169	-0.681	1.018	0.697	13.87	0.031
			handling	2	1	no	-0.394	-1.255	0.467			
		Tech Level	protein	7	3	yes	-0.136	-0.818	0.546	0.696	10.16	0.118
			RNA	2	1	no	1.128	0.265	1.992			
		Sex	Males	8	4	yes	0.018	-0.629	0.665	0.956	12.67	0.081
		Species	Rat	9	4	yes	0.078	-0.63	0.786	0.83	16.38	0.037
		Brain area	PFC	7	4	yes	0.125	-0.593	0.844	0.732	13.53	0.035
			cingulatecortex	2	1	no	-0.001	-1.437	1.436			
	DA	Model	MD	8	5	yes	-0.028	-0.374	0.319	0.876	8.952	0.256
			handling	4	2	no	-0.205	-0.659	0.249			
			isolation	3	1	no	-0.245	-1.929	1.439			
		Tech Level	protein	15	8	yes	0.018	-0.408	0.444	0.935	43.29	0
		Sex	Females	4	3	yes	-0.151	-1.175	0.873	0.772	7.736	0.052
			Males	12	8	yes	0.084	-0.332	0.5	0.693	34.08	0
		Species	Rat	13	6	yes	0.03	-0.466	0.526	0.906	41.09	0
			Mouse	3	2	no	0.019	-0.519	0.558			
		Brain area	PFC	14	8	yes	0.079	-0.351	0.51	0.719	41.58	0
			cingulatecortex	2	1	no	-0.227	-0.788	0.335			
	DOPAC	Model	MD	8	5	yes	-0.056	-0.427	0.314	0.765	11.13	0.133
			handling	4	2	no	0.492	0.018	0.967			
		Tech Level	protein	12	7	yes	0.371	-0.044	0.787	0.08	32.88	0.001
		Sex	Males	11	7	yes	0.22	-0.281	0.722	0.389	39.73	0
			Females	2	2	no	0.554	-0.159	1.268			
		Species	Rat	10	5	yes	0.259	-0.299	0.817	0.363	39.06	0
			Mouse	3	2	no	0.258	-0.253	0.768			
		Brain area	PFC	11	7	yes	0.244	-0.273	0.762	0.355	40.13	0
			cingulatecortex	2	1	no	0.361	-0.202	0.925			
	HVA	Model	MD	8	5	yes	-0.154	-0.577	0.269	0.476	13.52	0.06
			handling	4	2	no	0.141	-0.312	0.594			

Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
		Tech Level	*protein*	*12*	*7*	*yes*	*0.178*	*-0.303*	*0.66*	*0.468*	*44.11*	*0*
		Sex	*Males*	*11*	*7*	*yes*	*0.059*	*-0.436*	*0.555*	*0.814*	*40.58*	*0*
			Females	*2*	*2*	*no*	*0.607*	*-0.641*	*1.855*			
		Species	*Rat*	*10*	*5*	*yes*	*0.107*	*-0.438*	*0.653*	*0.699*	*39.41*	*0*
			Mouse	*3*	*2*	*no*	*0.225*	*-0.697*	*1.147*			
		Brain area	*PFC*	*11*	*7*	*yes*	*0.13*	*-0.41*	*0.669*	*0.637*	*44.82*	*0*
			cingulatecortex	*2*	*1*	*no*	*0.16*	*-0.401*	*0.721*			
striatum	**DR1**	Model	MD	*12*	*6*	*yes*	*-0.707*	*-1.219*	*-0.196*	*0.007*	*20.77*	*0.036*
			handling	*3*	*1*	*no*	*0.117*	*-0.578*	*0.811*			
			LPS	*3*	*1*	*no*	*-0.166*	*-0.823*	*0.49*			
		Tech Level	*protein*	*14*	*5*	*yes*	*-0.32*	*-0.798*	*0.158*	*0.189*	*15.01*	*0.307*
			RNA	*5*	*4*	*yes*	*-0.818*	*-1.509*	*-0.126*	*0.02*	*10.05*	*0.04*
		Sex	*Males*	*16*	*7*	*yes*	*-0.219*	*-0.537*	*0.1*	*0.178*	*10.71*	*0.773*
			Females	*2*	*1*	*no*	*-1.657*	*-2.548*	*-0.766*			
		Species	*Rat*	*15*	*7*	*yes*	*-0.386*	*-0.799*	*0.027*	*0.067*	*17.26*	*0.243*
			Mouse	*4*	*1*	*no*	*-0.888*	*-1.947*	*0.172*			
		Brain area	*NAc*	*14*	*7*	*yes*	*-0.544*	*-0.997*	*-0.091*	*0.019*	*24.97*	*0.023*
			caudate	*4*	*2*	*no*	*-0.109*	*-0.693*	*0.475*			
	DR2	Model	MD	*24*	*11*	*yes*	*-0.175*	*-0.507*	*0.157*	*0.303*	*63.73*	*0*
			handling	*8*	*2*	*no*	*0.38*	*0.102*	*0.659*			
			isolation	*7*	*2*	*no*	*-0.301*	*-1.653*	*1.051*			
			LPS	*3*	*1*	*no*	*1.182*	*0.121*	*2.243*			
		Tech Level	*protein*	*27*	*9*	*yes*	*0.21*	*-0.266*	*0.686*	*0.387*	*64.63*	*0*
			RNA	*15*	*8*	*yes*	*-0.412*	*-0.814*	*-0.01*	*0.045*	*31.08*	*0.005*
		Sex	*Females*	*4*	*3*	*yes*	*-0.59*	*-1.181*	*0.001*	*0.05*	*4.458*	*0.216*
			Males	*37*	*12*	*yes*	*0.107*	*-0.232*	*0.446*	*0.536*	*80.91*	*0*
		Species	*Rat*	*37*	*12*	*yes*	*0.005*	*-0.358*	*0.368*	*0.978*	*93.88*	*0*
			Mouse	*5*	*2*	*no*	*-0.38*	*-1.168*	*0.407*			

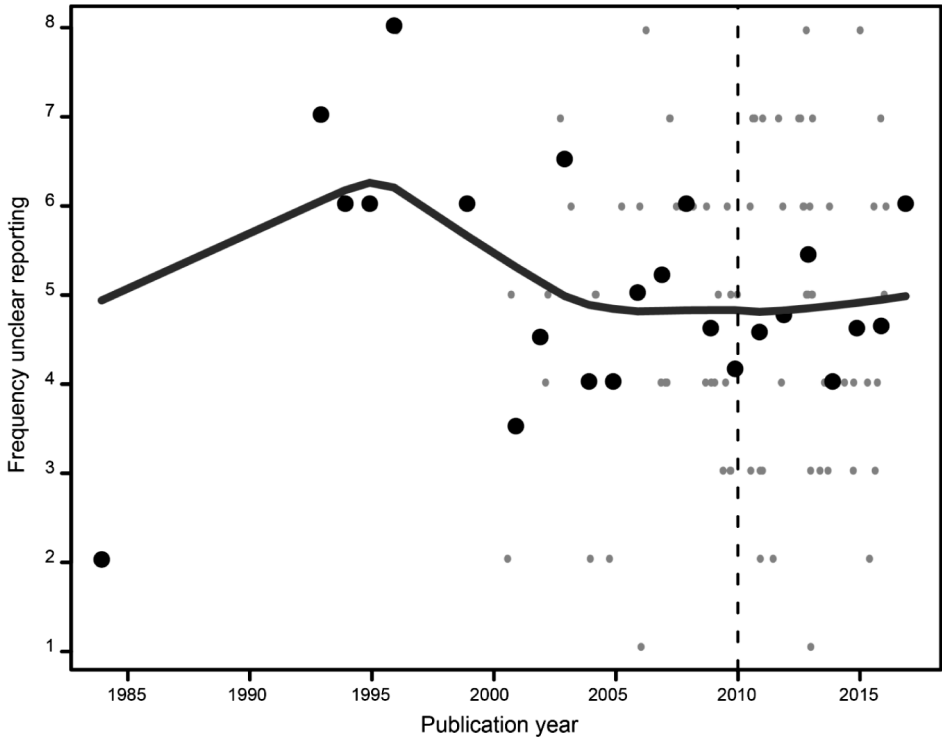
Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
		Brain area	*caudate*	*12*	*5*	*yes*	*0.199*	*-0.148*	*0.546*	*0.261*	*17.1*	*0.105*
			NAc	*26*	*10*	*yes*	*-0.061*	*-0.373*	*0.251*	*0.7*	*52.74*	*0.001*
			striatum	*4*	*4*	*yes*	*-0.037*	*-1.759*	*1.684*	*0.966*	*32.72*	*0*
	DR3	Model	MD	*7*	*3*	*yes*	*-0.179*	*-0.841*	*0.483*	*0.597*	*10.55*	*0.103*
			handling	*4*	*1*	*no*	*0.273*	*-0.039*	*0.584*			
		Tech Level	*protein*	*9*	*2*	*no*	*-0.058*	*-0.754*	*0.638*			
			RNA	*3*	*2*	*no*	*-0.391*	*-1.354*	*0.573*			
		Sex	*Males*	*10*	*3*	*yes*	*0.077*	*-0.464*	*0.618*	*0.78*	*8.445*	*0.49*
		Species	*Rat*	*12*	*4*	*yes*	*-0.195*	*-0.721*	*0.331*	*0.468*	*18.12*	*0.079*
		Brain area	*NAc*	*8*	*3*	*yes*	*-0.255*	*-0.936*	*0.426*	*0.463*	*14.25*	*0.047*
			caudate	*4*	*2*	*no*	*0.149*	*-0.239*	*0.538*			
	DA	Model	handling	*15*	*6*	*yes*	*-0.055*	*-0.437*	*0.328*	*0.779*	*18.52*	*0.184*
			isolation	*13*	*3*	*yes*	*0.363*	*-0.2*	*0.927*	*0.206*	*18.3*	*0.107*
			MD	*10*	*8*	*yes*	*-0.066*	*-0.35*	*0.219*	*0.651*	*9.61*	*0.383*
			LPS	*2*	*2*	*no*	*2.756*	*2.015*	*3.496*			
		Tech Level	*protein*	*39*	*17*	*yes*	*0.25*	*-0.112*	*0.613*	*0.175*	*105.8*	*0*
		Sex	*Females*	*15*	*7*	*yes*	*0.267*	*-0.23*	*0.763*	*0.292*	*26.75*	*0.021*
			Males	*24*	*15*	*yes*	*0.188*	*-0.291*	*0.668*	*0.442*	*75.37*	*0*
		Species	*Mouse*	*5*	*4*	*yes*	*0.18*	*-0.23*	*0.59*	*0.389*	*4.546*	*0.337*
			Rat	*35*	*14*	*yes*	*0.251*	*-0.18*	*0.683*	*0.253*	*101.4*	*0*
		Brain area	*NAc*	*15*	*8*	*yes*	*-0.09*	*-0.362*	*0.182*	*0.517*	*13.9*	*0.457*
			striatum	*23*	*11*	*yes*	*0.439*	*-0.093*	*0.97*	*0.106*	*82.2*	*0*
			caudate	*2*	*1*	*no*	*0.219*	*-0.372*	*0.809*			
	DAT	Model	MD	*16*	*7*	*yes*	*-0.045*	*-0.785*	*0.694*	*0.904*	*41.02*	*0*
			handling	*4*	*1*	*no*	*0.039*	*-0.386*	*0.464*			
			isolation	*2*	*1*	*no*	*0.081*	*-0.54*	*0.701*			
			LPS	*2*	*1*	*no*	*2.256*	*0.111*	*4.4*			
		Tech Level	*protein*	*19*	*6*	*yes*	*0.081*	*-0.539*	*0.701*	*0.798*	*36.51*	*0.006*
			function	*2*	*1*	*no*	*0.177*	*-0.655*	*1.01*			

Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
			RNA	*3*	*2*	*no*	*0.351*	*-1.528*	*2.229*			
		Sex	*Males*	*21*	*8*	*yes*	*0.206*	*-0.354*	*0.766*	*0.471*	*43.37*	*0.002*
			Females	*2*	*2*	*no*	*0.68*	*-0.351*	*1.711*			
		Species	*Rat*	*23*	*8*	*yes*	*0.275*	*-0.192*	*0.742*	*0.248*	*52.63*	*0*
		Brain area	*caudate*	*4*	*3*	*yes*	*0.054*	*-0.317*	*0.426*	*0.774*	*1.553*	*0.67*
			NAc	*14*	*6*	*yes*	*-0.08*	*-0.901*	*0.74*	*0.848*	*43.46*	*0*
			striatum	*6*	*3*	*yes*	*0.535*	*-0.301*	*1.371*	*0.209*	*15.86*	*0.007*
	DOPAC	Model	handling	*5*	*4*	*yes*	*0.846*	*-0.217*	*1.91*	*0.119*	*14.63*	*0.006*
			MD	*7*	*5*	*yes*	*0.198*	*-0.125*	*0.52*	*0.23*	*1.53*	*0.957*
		Tech Level	*protein*	*13*	*9*	*yes*	*0.628*	*0.108*	*1.148*	*0.018*	*41.01*	*0*
		Sex	*Males*	*11*	*8*	*yes*	*0.698*	*0.068*	*1.328*	*0.03*	*40.36*	*0*
			Females	*2*	*2*	*no*	*0.313*	*-0.374*	*1*			
		Species	*Mouse*	*4*	*3*	*yes*	*0.175*	*-0.27*	*0.62*	*0.442*	*0.612*	*0.894*
			Rat	*9*	*6*	*yes*	*0.884*	*0.112*	*1.656*	*0.025*	*36.24*	*0*
		Brain area	*NAc*	*5*	*4*	*yes*	*0.637*	*-0.259*	*1.533*	*0.163*	*16.78*	*0.002*
			striatum	*6*	*5*	*yes*	*0.606*	*-0.124*	*1.336*	*0.104*	*23.01*	*0*
			caudate	*2*	*1*	*no*	*0.363*	*-0.201*	*0.926*			
	HVA	Model	handling	*4*	*3*	*yes*	*0.855*	*-0.245*	*1.955*	*0.128*	*9.572*	*0.023*
			MD	*6*	*4*	*yes*	*0.072*	*-0.267*	*0.412*	*0.676*	*1.189*	*0.946*
		Tech Level	*protein*	*11*	*7*	*yes*	*0.709*	*-0.065*	*1.482*	*0.072*	*55.2*	*0*
		Sex	*Males*	*10*	*7*	*yes*	*0.753*	*-0.12*	*1.627*	*0.091*	*55.04*	*0*
		Species	*Mouse*	*4*	*3*	*yes*	*0.185*	*-0.26*	*0.63*	*0.414*	*0.453*	*0.929*
			Rat	*7*	*4*	*yes*	*1.136*	*-0.205*	*2.477*	*0.097*	*50.93*	*0*
		Brain area	*NAc*	*4*	*3*	*yes*	*0.617*	*-0.267*	*1.502*	*0.171*	*11.42*	*0.01*
			striatum	*5*	*4*	*yes*	*0.763*	*-0.557*	*2.083*	*0.257*	*41.83*	*0*
			caudate	*2*	*1*	*no*	*0.338*	*-0.265*	*0.941*			
VTA	**DA**	Model	*MD*	*4*	*3*	*yes*	*-0.1*	*-0.743*	*0.542*	*0.76*	*3.494*	*0.322*
			handling	*2*	*1*	*no*	*-0.007*	*-0.636*	*0.623*			

Brain Area	Outcome	Moderator	levelModerator	Comp	Papers	Analysis	Hedges	ci.lb	ci.ub	pval	heter	het pvalue
		Tech Level	*protein*	*7*	*4*	*yes*	*0.438*	*-0.893*	*1.769*	*0.519*	*19.51*	*0.003*
		Sex	*Males*	*7*	*4*	*yes*	*0.438*	*-0.893*	*1.769*	*0.519*	*19.51*	*0.003*
		Species	*Rat*	*7*	*4*	*yes*	*0.438*	*-0.893*	*1.769*	*0.519*	*19.51*	*0.003*
		Brain area	*raphenucleus*	*2*	*1*	*no*	*0.095*	*-0.741*	*0.932*			
			VTA	*2*	*1*	*no*	*0.189*	*-0.371*	*0.749*			
			VTASN	*2*	*2*	*no*	*-0.533*	*-1.333*	*0.267*			
	Th	Model	LPS	*8*	*4*	*yes*	*-1.218*	*-2.96*	*0.523*	*0.17*	*56.9*	*0*
			MD	*15*	*6*	*yes*	*0.112*	*-0.245*	*0.469*	*0.538*	*14.69*	*0.4*
			handling	*4*	*1*	*no*	*-0.289*	*-2.702*	*2.124*			
		Tech Level	*protein*	*21*	*7*	*yes*	*-0.351*	*-1.072*	*0.371*	*0.341*	*85.55*	*0*
			RNA	*6*	*3*	*yes*	*-0.064*	*-0.95*	*0.823*	*0.888*	*11.69*	*0.039*
		Sex	*Males*	*23*	*9*	*yes*	*-0.216*	*-0.874*	*0.441*	*0.519*	*92.06*	*0*
			Females	*3*	*2*	*no*	*-0.528*	*-2.07*	*1.013*			
		Species	*Rat*	*25*	*9*	*yes*	*-0.153*	*-0.744*	*0.438*	*0.612*	*92.12*	*0*
			Mouse	*2*	*1*	*no*	*-0.952*	*-1.891*	*-0.014*			
		Brain area	*SN*	*15*	*8*	*yes*	*-0.487*	*-1.613*	*0.638*	*0.396*	*90.46*	*0*
			VTA	*12*	*7*	*yes*	*-0.019*	*-0.361*	*0.323*	*0.915*	*7.575*	*0.751*

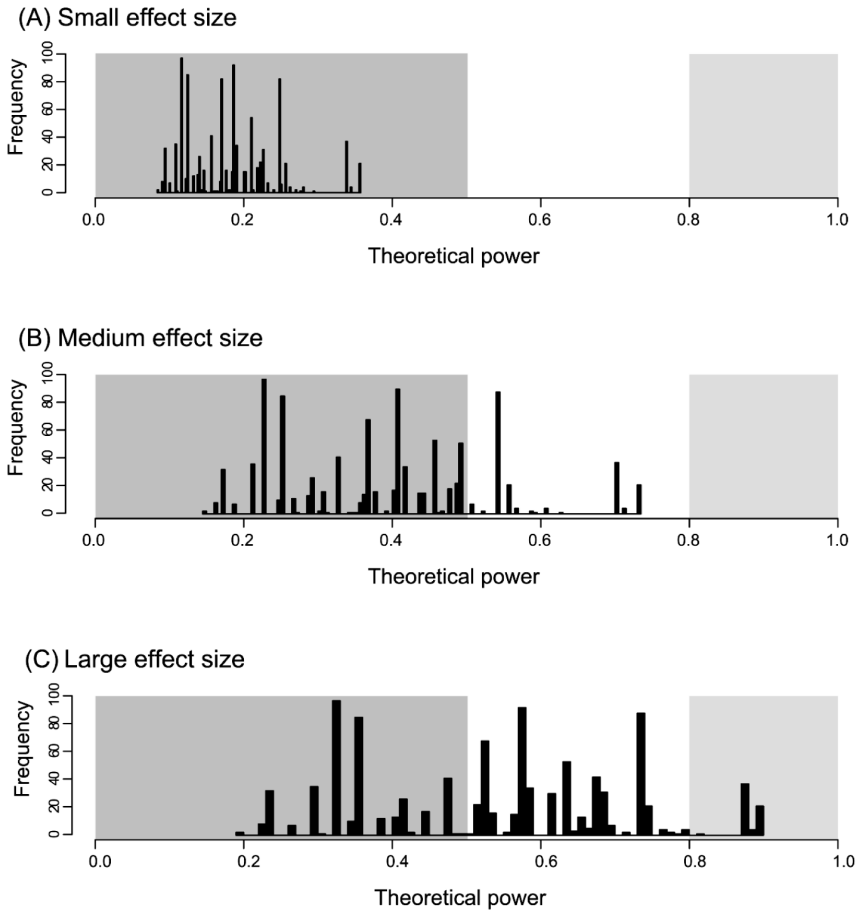
S4 | Discussion appendix

Figure S4-1 | Quality of reporting over time. The figure provides a graphical representation of the quality of reporting over time. Quality of reporting was operationalized by calculating the frequency of "unclear" reporting in each item of the SYRCLE bias assessment guideline. Each blue big dot represents the mean of frequency of unclear reporting for each year. Individual papers are represented by small grey dots. If only one paper was published in a particular year, the blue dot represents that paper. The more elevated the frequency of unclear reporting, the poorer the quality of reporting. The red line represents the best fitting curve given the data points. The dashed black line represents the year of publication of the arrive guidelines (2010). Although this is a rough estimate, it can be appreciated how frequency of unclear reporting plateaued around 2005.



S

Figure S4-2 | Theoretical power. We back-calculated the power that the studies would have had based on the amount of animals used, considering a truly existing effect. The histograms represent how many (frequency) comparisons had a defined power (y axis), when considering a (A) small, (B) medium or (C) large effect size. Shading between 0.8 and 1.0 = power that should be aimed at (>.8); Shading between 0 and 0.5 = theoretical power below chance level.



Supplementary Information for Chapter 6

Changes in monoamine systems after postnatal early life adversity in rodents: a systematic review with meta-analysis.

Contents

Data, scripts and other materials are available at: https://osf.io/4ngu3/

Supplementary notes

Supplementary note 1: search string

Pubmed:

Part 1 - Mice and rats:

("rodentia"[Mesh] OR rodent*[tiab] OR "mus"[Tiab] OR "mice"[Mesh] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[Mesh] OR "rats"[tiab] OR "rat"[tiab])

Part 2 – Postnatal early-life adversity:

("maternal behavior"[MeSh] OR "maternal care"[tiab] OR "early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "ELA"[tiab] OR "early life manipulation"[tiab] OR "early life manipulations"[tiab] OR "early adverse experience"[tiab] OR "early adverse experiences"[tiab] OR "early adversed experience"[tiab] OR "early adversed experiences"[tiab] OR "perinatal stress"[tiab] OR "perinatal adversity"[tiab] OR "perinatal adversities"[tiab] OR "perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "perinatal adverse experience"[tiab] OR "perinatal adverse experiences"[tiab] OR "perinatal adversed experience"[tiab] OR "perinatal adversed experiences"[tiab] OR "postnatal stress"[tiab] OR "postnatal adversity"[tiab] OR "postnatal adversities"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "postnatal adverse experience"[tiab] OR "postnatal adverse experiences"[tiab] OR "postnatal adversed experience"[tiab] OR "postnatal adversed experiences"[tiab] OR "neonatal stress"[tiab] OR "neonatal adversity"[tiab] OR "neonatal adversities"[tiab] OR "neonatal manipulation"[tiab] OR "neonatal manipulations"[tiab] OR "neonatal

adverse experience"[tiab] OR "neonatal adverse experiences"[tiab] OR "neonatal adversed experience"[tiab] OR "neonatal adversed experiences"[tiab] OR "Maternal Deprivation"[Mesh] OR "maternal deprivation"[tiab]OR "maternal separation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR "limited bedding/nesting"[tiab] OR "limited bedding-and-nesting"[tiab] OR "limited nesting/bedding"[tiab] OR "limited nesting-and-bedding"[tiab] OR "early life isolation"[tiab] OR "perinatal isolation"[tiab] OR "postnatal isolation"[tiab] OR "neonatal isolation"[tiab] OR "licking and grooming"[tiab] OR "licking-and-grooming"[tiab] OR "licking/grooming"[tiab] OR "early handling"[tiab] OR "early life handling"[tiab] OR "perinatal handling"[tiab] OR "postnatal handling"[tiab] OR "neonatal handling"[tiab])

Embase search string

Part 1 – Mice and rats:

(rodent*:ab,ti OR mus:ab,ti OR mouse:ab,ti OR mice:ab,ti OR rat:ab,ti OR rats:ab,ti)

Part 2 – Postnatal early-life adversity:

('maternal behavior':ab,ti OR 'maternal care':ab,ti OR 'early life stress':ab,ti OR 'els':ab,ti OR 'early life adversity':ab,ti OR 'early life adversities':ab,ti OR 'ela':ab,ti OR 'early life manipulation':ab,ti OR 'early life manipulations':ab,ti OR 'early adverse experience':ab,ti OR 'early adverse experiences':ab,ti OR 'early adversed experience':ab,ti OR 'early adversed experiences':ab,ti OR 'perinatal stress':ab,ti OR 'perinatal adversity':ab,ti OR 'perinatal adversities':ab,ti OR 'perinatal manipulation':ab,ti OR 'perinatal manipulations':ab,ti OR 'perinatal adverse experience':ab,ti OR 'perinatal adverse experiences':ab,ti OR 'perinatal adversed experience':ab,ti OR 'perinatal adversed experiences':ab,ti OR 'postnatal stress':ab,ti OR 'postnatal adversity':ab,ti OR 'postnatal adversities':ab,ti OR 'postnatal manipulation':ab,ti OR 'postnatal manipulations':ab,ti OR 'postnatal adverse experience':ab,ti OR 'postnatal adverse experiences':ab,ti OR 'postnatal adversed experience':ab,ti OR 'postnatal adversed experiences':ab,ti OR 'neonatal stress':ab,ti OR 'neonatal adversity':ab,ti OR 'neonatal adversities':ab,ti OR 'neonatal manipulation':ab,ti OR 'neonatal manipulations':ab,ti OR 'neonatal adverse experience':ab,ti OR 'neonatal adverse experiences':ab,ti OR 'neonatal adversed experience':ab,ti OR 'neonatal adversed experiences':ab,ti OR 'maternal deprivation':ab,ti OR 'maternal separation':ab,ti OR 'limited bedding':ab,ti OR 'limited nesting':ab,ti OR 'limited material':ab,ti OR 'limited bedding/nesting':ab,ti OR 'limited bedding-and-nesting':ab,ti OR 'limited nesting/bedding':ab,ti OR 'limited nesting-and-bedding':ab,ti OR 'early life isolation':ab,ti OR 'perinatal isolation':ab,ti OR 'postnatal isolation':ab,ti OR 'neonatal isolation':ab,ti OR 'licking and grooming':ab,ti OR 'licking-and-grooming':ab,ti OR 'licking/grooming':ab,ti OR 'early handling':ab,ti OR 'early life handling':ab,ti OR 'perinatal handling':ab,ti OR 'postnatal handling':ab,ti OR 'neonatal handling':ab,ti)

Supplementary note 2: inclusion and exclusion criteria

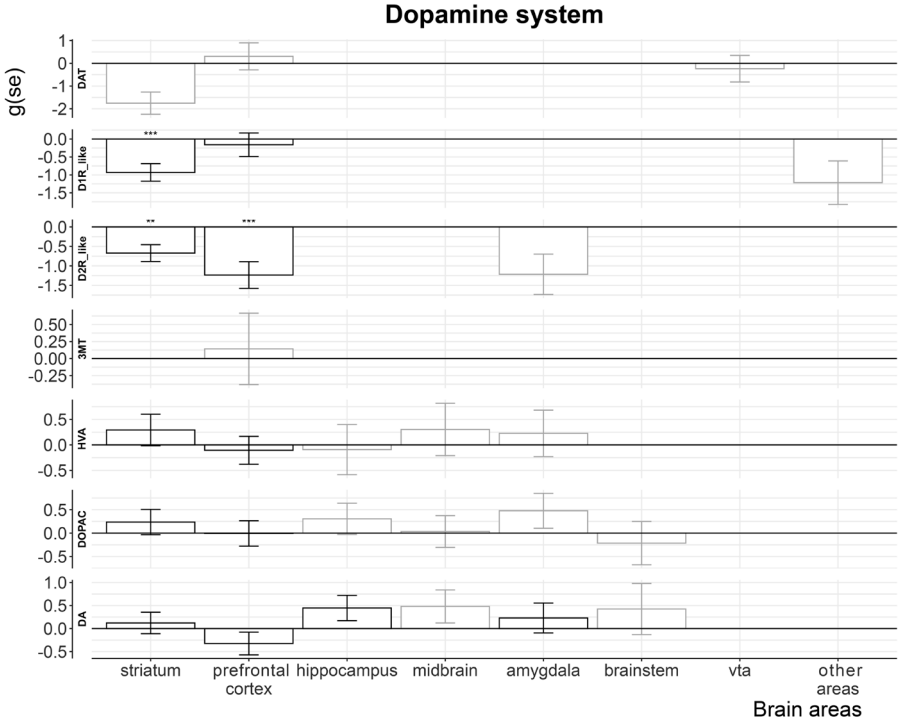
Study selection was performed independently by three (out of 5, see Acknowledgements), who were blinded to the studies' results. The inclusion and exclusion criteria were specified prior to the beginning of the study.

Criteria	Comments
Inclusion	
Peer reviewed original publications in English	
Mice and rats	
ELA starts before P14	ELA model can extend after P14
ELA as alteration of maternal care1 separation of the pup from the mother (maternal separation[2] / deprivation[1]) separation of the pup from mother and siblings (isolation) limited bedding and nesting[3] licking and grooming[4]	We define as 'separation' those models in which the mother was repeatedly separated from the pups (e.g. 3 hours a day for 2 weeks). We define as deprivation those models in which the mother was separated once from the pups for a prolonged time (e.g. 1 time 24 hours, or 2 times 12 hours). In other words, the categorization in maternal separation/deprivation depends on the model used and not the naming used in the papers. The separation/deprivation/isolation models are adaptations of Levine's original model. These adaptations are based on the observation that dams often leave the nest to forage for 15-30 min periods5. For this reason, we consider "adverse" and therefore include only those studies in which the duration of separation/deprivation/isolation time was >1h.
Testing during adult age	Older than 8 weeks but younger than 1 year
Exclusion	
Specific pathogen free animals	
Ovariectomized females	
Sex not specified	Publication is included if sex is retrieved after contacting the authors
Males and females pooled	Publication is included if summary statistics of males and females separately are received after contacting the authors
Handling, gentling and communal nesting as ELA models	
Maternal separation with early weaning 6	Early weaning is defined as separation of the pups from the mother at P17. If early weaning is only in the experimental group, the experiment is excluded. If early weaning occurred in both control and experimental group, the study is included and early weaning is considered a factor that could increase vulnerability
Handling as control group	

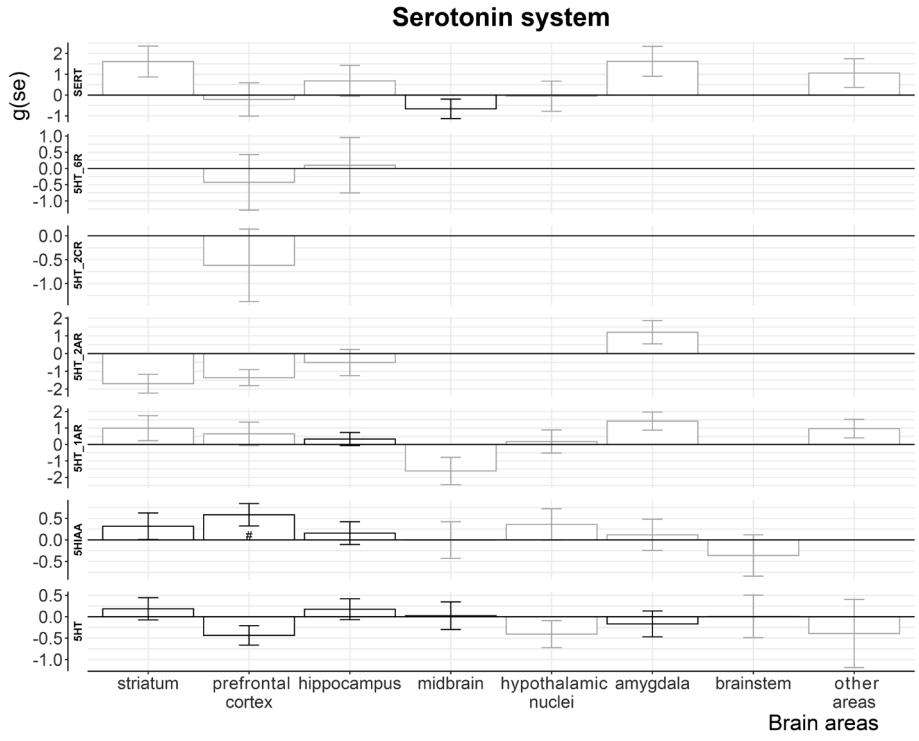
Genetic manipulations	
Animals bred for high/low anxiety-like behavior or novelty response or sensitivity/resilience to depression	
Animals separated in high/low performance	
Administration of any drug or alcohol via any route	e.g. Drug injections before testing, methamphetamine conditioned place preference tests
Any manipulation to previous generations	
Other *	Inclusion/exclusion criteria specific to certain outcomes. See Supplementary Table 1

Supplementary Figures

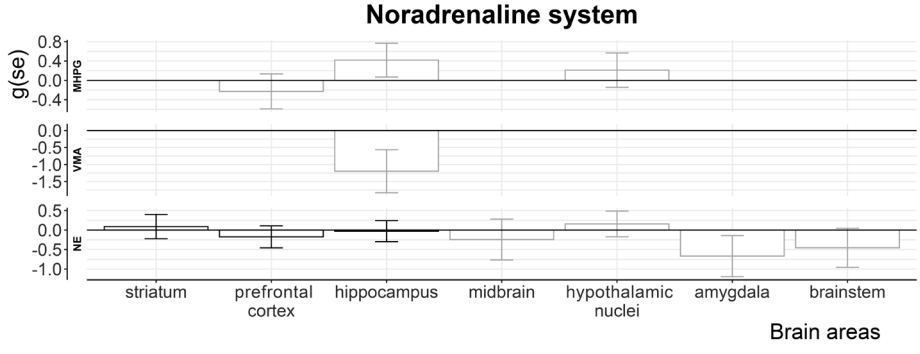
Supplementary Figure 1. The effects of ELA on dopaminergic outcome measures in different brain areas in males at rest. Only comparisons for which a minimal of 3 studies were available are statistically analyzed on significance, comparisons from a lower amount of studies are marked in light grey. *Dopamine receptors of the D1 type (D1like) are decreased in strital tissue of ELS animals, Dopamine receptor D2 like (D2R like) is decreased in both striatal and prefrontal cortex tissue of ELS animals. DA: dopamine. 3-Methoxytyramine (3MT), 3,4-Dihydroxyphenylacetic acid (DOPAC) and homovanillic acid (HVA) are dopamine metabolites. DAT: Dopamine transporter.*



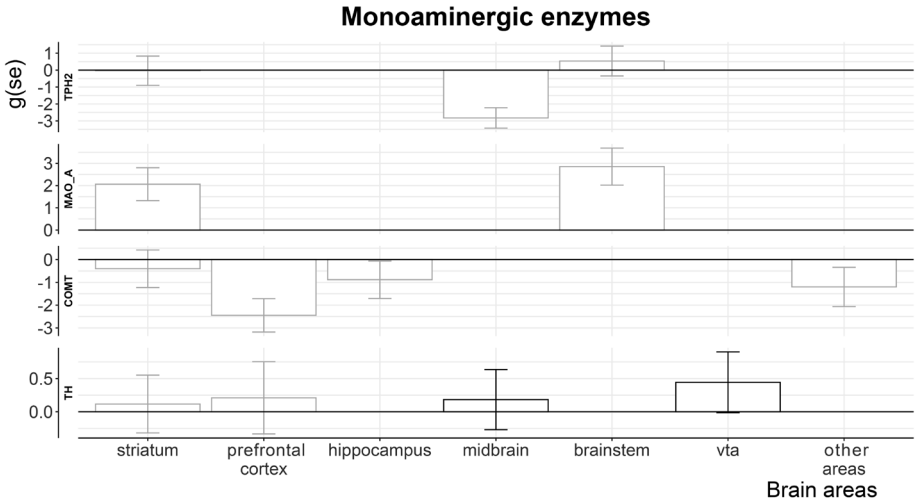
Supplementary Figure 2. The effects of ELA on serotonergic outcome measures in different brain areas in males at rest. Only comparisons for which a minimal of 3 studies were available are statistically analyzed on significance, comparisons from a lower amount of studies are marked in light grey. *In prefrontal cortex, 5-hydroxyindoleacetic acid (5-HIAA, a serotonin metabolite) is increased in ELS animals, while a decrease in serotonin (5-hydroxytryptamine, 5HT) is suggested. 5HT1A, 2A, 2C and 6 are subtypes of serotonin receptor. SERT: serotonine transporter.*



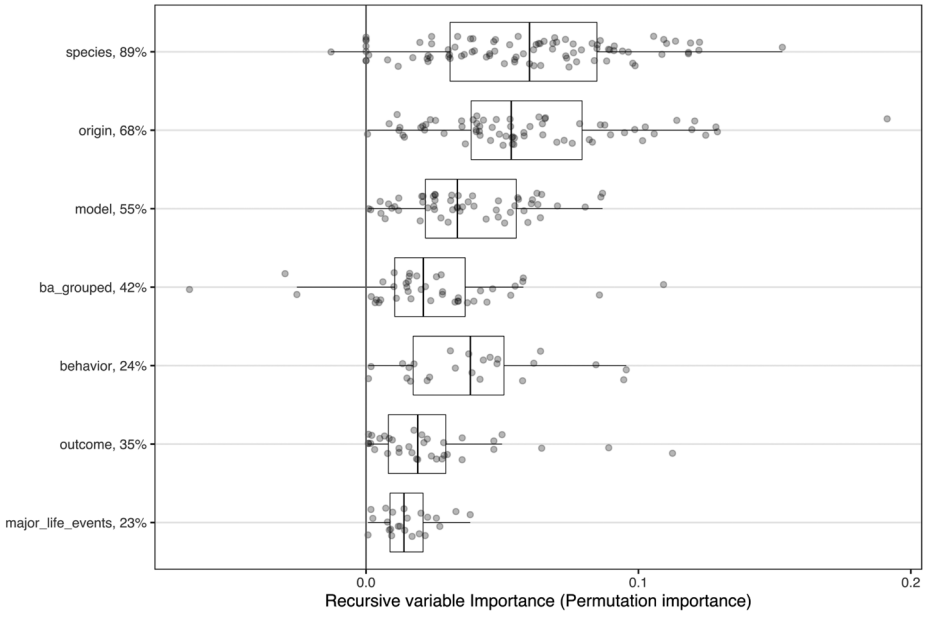
Supplementary Figure 3. The effects of ELA on noradrenergic outcome measures in different brain areas in males at rest. Only comparisons for which a minimal of 3 studies were available are statistically analyzed on significance, comparisons from a lower amount of studies are marked in light grey. *NE: noradrenaline. 3-Methoxy-4-hydroxyphenylglycol (MHPG) and Vanillylmandelic acid (VMA) are both metabolites of NE.*



Supplementary Figure 4. The effects of ELA on enzymes involved in monoamine metabolism in different brain areas in males at rest. Only comparisons for which a minimal of 3 studies were available are statistically analyzed on significance, comparisons from a lower number of studies are marked in light grey. *COMT: Catechol-O-methyltransferase. MAOA: Monoamine oxidase A. TH: tyrosine hydroxylase TPH2: tryptophan hydroxylase 2.*



Supplementary Figure 5. Metaforest variable importance plots (exploratory analysis). *Species* is rat or mice; *origin* categorizes the purchasing condition of dams (bred in house, naïve from provider or pregnant from provider); *model* describes the type of ELA model used (maternal separation, maternal deprivation, isolation or limited nesting); *brain areas* categorizes the brain areas according to Supplementary Table 2; *behavior* categorizes animals as having experienced non-stressfull or stressful tasks at one point in life, no behavioral tasks (but other handling procedurs), or being completely naïve (no proceduces conducted); *outcome* are the outcome measures of monoamines in different brain areas; lastly *major life events* describes the presence or absence of additional major life events (e.g. chronic stress during adolescence or early adulthood, (prolonged) restraint stress, foot-shocks). Each dot represents a different simulation based on bootstrap sampling. The percentages refer to the percentage of simulations where a certain factor was selected.



Supplementary Tables

Supplementary Table 1: included outcomes

Prior the beginning of the study, we identified outcomes that would describe the monoaminergic systems. The table below summarizes which outcomes we aimed to include, vs their actual availability

Monoamine	Group	Outcome	Availability comments
Dopamine-related			
	Dopamine	Dopamine concentration (RNA and protein)	
	Precursors	Tyrosine, L-DOPA	Not available
	Metabolites	DOPAC, HVA, 3-MT	
		Turnovers (i.e. metabolite/DA)	Extracted, but only for systematic review
	Transporters	DAT	
	Receptors	DRD1-DRD5	Only partially available
	Enzymes	*TH, MAO, COMT*	Only partially available
Serotonin related			
	Monoamine	Serotonin (5-HT)	
	Precursors	L-tryptophan, 5-HTP	Not available
	Metabolites	5-HIAA	
		Turnovers (i.e. metabolite/5HT)	Extracted, but only for systematic review
	Transporters	SERT	
		PMAT	Not available
	Receptors	5-HT1 to 7	Only partially available
	Enzymes	*TPH2*, MAO, AAADC*	Only partially available
Noradrenaline related			
	Monoamine	Noradrenaline	
	Metabolites	Noradrenaline aldehyde, MHPG, VMA, NMN	Only partially available
		Turnovers (i.e. metabolite/DA)	Extracted, but only for systematic review
	Transporter	NET	Not available
	Receptors	alpha_1, alpha_2, beta_1, beta_2, beta_3	Only partially available
	Enzymes	*DBH, COMT, MAO, ADH*	Only partially available

* Of note: One publication reported "TPH" rather than specifying the type (1 or 2). Since all other publications were in TPH2 and TPH1 mainly occurs in peripheral and non-neuronal tissues, we considered this publication of the within the same TPH2 group.

Supplementary Table 2: categorization of brain areas

We categorized brain areas in 10 main groups, according to the Allen Brain Atlas 7collecting large amounts of data across modalities, spatial scales, and brain areas. Successful integration of these data requires a standard 3D reference atlas. Here, we present the Allen Mouse Brain Common Coordinate Framework (CCFv3 and the frequency availability of our data. The table below summarizes the final categorizations. Of note, after seeing the frequencies of brain areas in our dataset, we deviated from our pre-planned categorization by making "midbrain" and "VTA" as self-standing categories.

Categorization	Names from publications
amygdala	amygdala amygdala basolateral amygdala basomedial amygdala basoventral amygdala central amygdala cortex
brainstem	brainstem medulla oblongata pons
hippocampus	dentate gyrus dentate gyrus infrapyramidal dentate gyrus suprapyramidal hippocampus hippocampus ca1 hippocampus ca2 hippocampus ca3 hippocampus dorsal hippocampus ventral
hypothalamic nuclei	hypothalamic area dorsal hypothalamic area lateral hypothalamic nucleus dorsomedial hypothalamic nucleus ventromedial hypothalamus hypothalamus anterior hypothalamus arcuate nucleus hypothalamus dorsomedial hypothalamus lateral hypothalamus paraventricular nucleus hypothalamus ventromedial nuclei mammilary nucleus medial preoptic area suprachiasmatic nucleus supraoptic nucleus zona incerta
midbrain	midbrain periaqueductal gray

	raphe nucleus raphe nucleus dorsal raphe nucleus ventral raphe nucleus lateral raphe nucleus medial raphe nucleus median raphe nucleus paramedian raphe nucleus vetralis substantia nigra substantia nigra pars compacta substantia nigra pars lateralis substantia nigra pars reticulata superior colliculus intermediate gray superior colliculus superficial gray
prefrontal cortex	cortex frontal cortex frontal deep cortex frontal layer IV cortex frontal parietal deep cortex frontal parietal superficial cortex frontal superficial cortex frontal layers IV and V cortex prefrontal cortex prefrontal medial cortex prefrontal medial infralimbic cortex prefrontal precentral cortex prefrontal prelimbic frontal cortex prefrontal cortex
striatum	caudate putamen caudate putamen lateral caudate putamen medial globus pallidus lateral nucleus accumbens nucleus accumbens core nucleus accumbens shell pallidum ventral striatum striatum dorsal striatum ventral
thalamic nuclei	habenular nucleus lateral thalamic nucleus laterodorsal thalamus
VTA	ventral tegmental area
other areas	cortex cingulate cortex cingulate anterior cortex cingulate deep cortex cingulate superficial cortex enthorinal cortex limbic deep cortex occipital deep cortex occipital superficial cortex parietal cortex parietal deep cortex parietal layer I to IV cortex parietal intermediate cortex parietal superficial

cortex parietal layer V
cortex perirhinal
cortex retrosplenial deep
cortex retrosplenial superficial
cortex temporal
edinger westphal nucleus
endopiriform nucleus
internal capsule
olfactory bulb external plexiform layer
olfactory tubercle

Supplementary Table 3: categorization of life experiences

Experimental designs were grouped based on the animals' life experiences. Specifically, this was summarized in 4 variables, of which the possible sub-categories are listed below

Variable	Interpretation	Categorization	Considered an additional negative life event ("hit")
Long lasting effects			
Origin	Origin of the breeding animals	Own breeding, dams purchased pregnant, purchased parents, or not specified	Dams purchased pregnant are liable to transportation stress. Therefore, we considered this a prenatal stress for the pups.
Behavior	Whether animals performed behavior tests	Naïve (no life experience besides the early postnatal condition), no behavior (e.g. although no behavior, the animals were handled eg for injections), non-stressful behavior tests (e.g. object in location), or stressful behavior (e.g. fear conditioning)	Stressful behavior tests (eg fear conditioning)
Other major life experiences	Umbrella category for experiences in adolescence and/or adulthood	Yes or no, depending on column "Considered an additional negative life event"	We considered as additional negative life events: (chronic) restraint/immobilization stress, chronic footshock, fox odor, chronic mild/unpredictable stress and combinations, anaesthesia (for mock surgeries), microdialysis, blood sampling
Acute effects			
State	Acute situation of the animal at death	The state of the animal at death was categorized as 1) rest, 2) aroused (after injection, novel environment, fasting, single housing, elevated plus maze or any non-stressful behavior experiments), and 3) stressed (after footshock, restraint/ immobilization, forced swim test, morris water maze, probe implantation, social defeat, resident intruder) *	Not applicable

* Due to frequency of experiments in the various categories, aroused/stressed were merged in only one category.

Supplementary Table 4.

Summary evidence on metabolites' turnovers as reported by respective publications. Publications included:6,8–14*N = sample size, c = control group, ELA = early life adversity group; Sig = increase, decrease or not significant (i.e. ns) as reported by the publication itself.*

					CONTROL			ELA			
Publication	**Species**	**Sex**	**ELA model**	**Brain areas**	**N**	**mean**	**SD**	**N**	**mean**	**SD**	**Sig**
DOPAC/DA											
Desbonnet (2010)[9]	rat	male	maternal separation	amygdala	11	0.78	0.9	7	0.68	1.03	ns
Rentesi (2013)[14]	rat	male	maternal deprivation	amygdala	12	0.54	0.15	13	0.83	0.35	increase
Desbonnet (2010)[9]	rat	male	maternal separation	brainstem	11	1.24	1.49	7	1.11	1.06	ns
Liu (2016)[11]	mice	male	isolation	hippocampus	10	1.06	1.61	12	1.04	1.25	ns
Desbonnet (2010)	rat	male	maternal separation	prefrontal cortex	11	0.41	0.23	7	0.39	0.11	ns
Liu (2016)	mice	male	isolation	prefrontal cortex	10	0.47	0.89	12	0.83	1.25	increase
Matthews (2001)	rat	male	maternal separation	prefrontal cortex	8	1.44	0.23	10	1.08	0.25	decrease
Matthews (2001)	rat	female	maternal separation	prefrontal cortex	8	1.4	0.14	6	0.96	0.17	decrease
Rentesi (2013)	rat	male	maternal deprivation	prefrontal cortex	12	1.04	0.26	13	1.39	0.36	increase
Desbonnet (2010)	rat	male	maternal separation	striatum	11	0.27	0.07	7	0.26	0.05	ns
Liu (2016)	mice	male	isolation	striatum	10	0.09	0.02	12	0.12	0.22	ns
Matthews (2001)	rat	male	maternal separation	striatum	8	0.21	0.14	10	0.34	0.22	ns
Matthews (2001)	rat	female	maternal separation	striatum	8	0.26	0.2	6	0.62	0.15	ns
Rentesi (2013)	rat	male	maternal deprivation	striatum	12	0.42	0.31	13	0.76	0.46	increase
(DOPAC+HVA)/DA											
Récamier-Carballo (2017)	mice	male	maternal separation	amygdala	10	0.82	1.38	10	0.14	0.17	ns
Récamier-Carballo (2017)	mice	male	maternal separation	amygdala	10	0.99	1.27	10	0.12	0.08	ns
Récamier-Carballo (2017)	mice	male	maternal separation	hippocampus	10	0.81	0.66	10	0.38	0.48	ns
Récamier-Carballo (2017)	mice	male	maternal separation	hippocampus	10	0.84	0.87	10	0.09	0.07	ns
Récamier-Carballo (2017)	mice	male	maternal separation	prefrontal cortex	10	1.36	0.66	10	1.2	0.7	ns
Récamier-Carballo (2017)	mice	male	maternal separation	prefrontal cortex	10	0.42	0.52	10	0.94	0.1	increase

Publication	Species	Sex	ELA model	Brain areas	CONTROL N	CONTROL mean	CONTROL SD	ELA N	ELA mean	ELA SD	Sig
HVA/DA											
Rentesi (2013)	rat	male	maternal deprivation	amygdala	12	0.71	0.31	13	0.71	0.03	ns
Liu (2016)	mice	male	isolation	hippocampus	10	1.81	2.25	12	1.68	1.87	ns
Liu (2016)	mice	male	isolation	prefrontal cortex	10	1.6	3.48	12	2.31	4.16	ns
Rentesi (2013)	rat	male	maternal deprivation	prefrontal cortex	12	0.6	0.26	13	0.71	0.32	ns
Liu (2016)	mice	male	isolation	striatum	10	0.16	0.05	12	0.18	0.2	ns
Rentesi (2013)	rat	male	maternal deprivation	striatum	12	0.1	0.03	13	0.12	0.04	increase
5HIAA/5HT											
Desbonnet (2010)	rat	male	maternal separation	amygdala	11	1.34	0.46	7	1.25	0.21	ns
Récamier-Carballo (2017)	mice	male	maternal separation	amygdala	10	0.11	0.08	10	0.05	0.1	ns
Récamier-Carballo (2017)	mice	male	maternal separation	amygdala	10	0.03	0.01	10	0.05	0.05	ns
Rentesi (2013)	rat	male	maternal deprivation	amygdala	12	0.54	0.18	13	0.72	0.17	increase
Desbonnet (2010)	rat	male	maternal separation	brainstem	11	1.98	1.56	7	1.55	0.98	ns
Desbonnet (2010)	rat	male	maternal separation	brainstem	11	1.54	0.5	7	1.55	0.16	ns
Pusceddu (2015)	rat	female	maternal separation	brainstem	10	1.72	0.32	10	1.46	0.16	decrease
Barbosa (2012)	rat	male	maternal deprivation	hippocampus	7	2.17	0.64	7	2.01	0.52	ns
Barbosa (2012)	rat	female	maternal deprivation	hippocampus	7	1.26	0.56	7	1.06	0.38	ns
Desbonnet (2010)	rat	male	maternal separation	hippocampus	11	8.55	5.97	7	4.08	2.62	decrease
Lee (2007)	rat	male	maternal separation	hippocampus	8	1.7	0.3	6	1.97	0.26	ns
Liu (2016)	mice	male	isolation	hippocampus	10	0.93	0.92	12	1.08	1.91	ns
Matthews (2001)	rat	male	maternal separation	hippocampus	8	1.52	0.34	10	1.42	0.28	ns
Matthews (2001)	rat	female	maternal separation	hippocampus	8	2.18	0.48	6	2.26	0.24	ns
Pusceddu (2015)	rat	female	maternal separation	hippocampus	10	12.24	4.78	10	6.76	5.12	decrease
Récamier-Carballo (2017)	mice	male	maternal separation	hippocampus	10	0.1	0.04	10	0.14	0.07	ns
Récamier-Carballo (2017)	mice	male	maternal separation	hippocampus	10	0.08	0.04	10	0.07	0.05	ns

					CONTROL			ELA			
Publication	**Species**	**Sex**	**ELA model**	**Brain areas**	**N**	**mean**	**SD**	**N**	**mean**	**SD**	**Sig**
Daniels (2004)	rat	male	maternal separation	hypothalamic nuclei	7	1.84	0.27	7	1.25	0.27	decrease
Daniels (2004)	rat	male	maternal separation	hypothalamic nuclei	7	2	0.52	7	1.42	0.11	decrease
Lee (2007)	rat	male	maternal separation	midbrain	8	0.73	0.06	6	0.74	0.07	ns
Daniels (2004)	rat	male	maternal separation	prefrontal cortex	7	1.85	0.08	7	1.29	0.27	decrease
Desbonnet (2010)	rat	male	maternal separation	prefrontal cortex	11	1.18	0.36	7	1.17	0.32	ns
Liu (2016)	mice	male	isolation	prefrontal cortex	10	0.3	1.2	12	0.62	1	increase
Matthews (2001)	rat	male	maternal separation	prefrontal cortex	8	1.2	0.4	10	0.85	0.38	ns
Matthews (2001)	rat	female	maternal separation	prefrontal cortex	8	5.3	8.2	6	1.2	6.12	ns
Pusceddu (2015)	rat	female	maternal separation	prefrontal cortex	10	1.33	0.47	10	0.96	0.19	decrease
Récamier-Carballo (2017)	mice	male	maternal separation	prefrontal cortex	10	0.1	0.09	10	0.35	0.09	increase
Récamier-Carballo (2017)	mice	male	maternal separation	prefrontal cortex	10	0.08	0.13	10	0.31	0.06	ns
Rentesi (2013)	rat	male	maternal deprivation	prefrontal cortex	12	0.55	0.17	13	0.78	0.14	increase
Desbonnet (2010)	rat	male	maternal separation	striatum	11	1.66	3.75	7	1.71	0.74	ns
Liu (2016)	mice	male	isolation	striatum	10	0.09	0.06	12	0.18	0.45	ns
Rentesi (2013)	rat	male	maternal deprivation	striatum	12	1.05	0.44	13	1.29	0.54	ns
VMA/NE											
Barbosa (2012)	rat	female	maternal deprivation	hippocampus	7	0	0.13	7	0.41	0.13	ns

Supplementary Table 5.

Summary statistics of all outcomes in male mice at rest investigated by at least 3 publications. *g = Hedge's g; se = standard error; p = p value; sig = significance label, specifically # < 0.05, ** < 0.01, *** < 0.001. Of note, we set our significance threshold to 0.01; N = number; comp = comparison*

outcome	g	se	p	sig	Brain areas	N_{study}	N_{comp}
Dopamine system							
3MT	0.14	0.524	0.786		prefrontal cortex	1	1
D1R like	-1.22	0.606	0.045	#	other areas	1	1
D1R like	-0.16	0.328	0.63		prefrontal cortex	4	4
D1R like	-0.93	0.245	0	***	striatum	5	8
D2R like	-1.22	0.519	0.019	#	amygdala	1	1
D2R like	-1.23	0.343	0	***	prefrontal cortex	4	4
D2R like	-0.67	0.216	0.002	**	striatum	7	13
DA	0.23	0.325	0.48		amygdala	3	3
DA	0.42	0.555	0.444		brainstem	1	1
DA	0.45	0.275	0.104		hippocampus	4	5
DA	0.48	0.359	0.181		midbrain	2	3
DA	-0.32	0.247	0.189		prefrontal cortex	6	7
DA	0.12	0.233	0.601		striatum	7	8
DAT	0.31	0.593	0.606		prefrontal cortex	1	1
DAT	-1.75	0.488	0	***	striatum	2	2
DAT	-0.24	0.584	0.686		vta	1	1
DOPAC	0.48	0.372	0.201		amygdala	2	2
DOPAC	-0.21	0.462	0.644		brainstem	1	1
DOPAC	0.31	0.333	0.359		hippocampus	2	3
DOPAC	0.03	0.339	0.92		midbrain	2	3
DOPAC	-0.01	0.271	0.98		prefrontal cortex	4	5
DOPAC	0.24	0.269	0.38		striatum	4	5
HVA	0.23	0.457	0.62		amygdala	1	1
HVA	-0.09	0.493	0.852		hippocampus	1	1
HVA	0.3	0.515	0.556		midbrain	1	1
HVA	-0.11	0.275	0.701		prefrontal cortex	4	5
HVA	0.29	0.31	0.344		striatum	3	3
Serotonin system							
5HIAA	0.12	0.36	0.746		amygdala	2	2
5HIAA	-0.36	0.479	0.451		brainstem	1	1
5HIAA	0.16	0.264	0.553		hippocampus	4	6
5HIAA	0.36	0.362	0.323		hypothalamic nuclei	1	3
5HIAA	0	0.425	0.992		midbrain	2	2
5HIAA	0.58	0.259	0.025	#	prefrontal cortex	4	6
5HIAA	0.32	0.307	0.303		striatum	3	3
5HT	-0.17	0.303	0.58		amygdala	4	4
5HT	0.01	0.497	0.986		brainstem	1	1
5HT	0.18	0.244	0.472		hippocampus	6	8
5HT	-0.41	0.316	0.198		hypothalamic nuclei	2	4
5HT	0.02	0.323	0.941		midbrain	4	5
5HT	-0.39	0.797	0.622		other areas	1	2

outcome	g	se	p	sig	Brain areas	N_{study}	N_{comp}
5HT	-0.44	0.227	0.055		prefrontal cortex	8	10
5HT	0.19	0.259	0.475		striatum	5	5
5HT 1AR	1.42	0.549	0.01	#	amygdala	2	2
5HT 1AR	0.33	0.398	0.408		hippocampus	4	4
5HT 1AR	0.18	0.703	0.797		hypothalamic nuclei	1	1
5HT 1AR	-1.61	0.832	0.053		midbrain	1	1
5HT 1AR	0.96	0.564	0.088		other areas	2	2
5HT 1AR	0.65	0.714	0.364		prefrontal cortex	1	1
5HT 1AR	0.99	0.761	0.194		striatum	1	1
5HT 2AR	1.2	0.658	0.067		amygdala	1	1
5HT 2AR	-0.51	0.739	0.493		hippocampus	1	1
5HT 2AR	-1.36	0.453	0.003	**	prefrontal cortex	2	2
5HT 2AR	-1.7	0.53	0.001	**	striatum	1	1
5HT 2CR	-0.62	0.759	0.415		prefrontal cortex	1	1
5HT 6R	0.1	0.853	0.908		hippocampus	1	1
5HT 6R	-0.43	0.856	0.617		prefrontal cortex	1	1
SERT	1.62	0.718	0.024	#	amygdala	1	1
SERT	0.68	0.744	0.36		hippocampus	1	1
SERT	-0.05	0.724	0.94		hypothalamic nuclei	1	1
SERT	-0.66	0.469	0.158		midbrain	3	3
SERT	1.05	0.691	0.127		other areas	1	1
SERT	-0.21	0.799	0.793		prefrontal cortex	1	1
SERT	1.61	0.742	0.03	#	striatum	1	1
Noradrenaline system							
MHPG	0.42	0.349	0.23		hippocampus	1	3
MHPG	0.21	0.355	0.551		hypothalamic nuclei	1	3
MHPG	-0.23	0.361	0.527		prefrontal cortex	1	3
NE	-0.67	0.525	0.204		amygdala	1	1
NE	-0.45	0.5	0.364		brainstem	1	1
NE	-0.03	0.269	0.917		hippocampus	4	6
NE	0.16	0.33	0.634		hypothalamic nuclei	2	4
NE	-0.24	0.524	0.643		midbrain	1	1
NE	-0.17	0.282	0.538		prefrontal cortex	3	5
NE	0.09	0.311	0.778		striatum	3	3
VMA	-1.2	0.632	0.058		hippocampus	1	1
Monoaminergic enzymes							
COMT	-0.88	0.823	0.283		hippocampus	1	1
COMT	-1.2	0.861	0.163		other areas	1	1
COMT	-2.45	0.731	0.001	**	prefrontal cortex	2	2
COMT	-0.4	0.823	0.625		striatum	1	1
MAO A	2.86	0.834	0.001	**	brainstem	1	2
MAO A	2.06	0.741	0.005	**	striatum	1	2
TH	0.18	0.452	0.686		midbrain	3	4
TH	0.21	0.544	0.699		prefrontal cortex	1	1
TH	0.12	0.435	0.789		striatum	2	2
TH	0.44	0.458	0.334		vta	3	4
TPH2	0.54	0.88	0.541		brainstem	1	1
TPH2	-2.82	0.601	0	***	midbrain	1	1
TPH2	-0.04	0.865	0.967		striatum	1	1

Supplementary references

1. Levine, S. Regulation of the hypothalamic-pituitary-adrenal axis in the neonatal rat: the role of maternal behavior. *Neurotoxicity research* 4, 557–564 (2002).
2. Sanchez, M. M., Ladd, C. O. & Plotsky, P. M. Early adverse experience as a developmental risk factor for later psychopathology: Eveidence from rodent and promate models. *Development and Psychopathology* 13, 419–449 (2001).
3. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* 149, 4892–4900 (2008).
4. Champagne, F. A., Francis, D. D., Mar, A. & Meaney, M. J. Variations in maternal care in the rat as a mediating influence for the effects of environment on development. *Physiology & behavior* 79, 359–371 (2003).
5. Leon, M., Croskerry, P. G. & Smith, G. K. Thermal control of mother-young contact in rats. *Physiology & Behavior* 21, 793–811 (1978).
6. Carlyle, B. C. et al. Maternal separation with early weaning: A rodent model providing novel insights into neglect associated developmental deficits. *Development and Psychopathology* 24, 1401–1416 (2012).
7. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936-953. e20 (2020).
8. Barbosa Neto, J. B. et al. Stress during development alters anxiety-like behavior and hippocampal neurotransmission in male and female rats. *Neuropharmacology* 62, 518–526 (2012).
9. Desbonnet, L. et al. Effects of the probiotic Bifidobacterium infantis in the maternal separation model of depression. *Neuroscience* 170, 1179–1188 (2010).
10. Lee, J.-H. et al. Depressive behaviors and decreased expression of serotonin reuptake transporter in rats that experienced neonatal maternal separation. *Neurosci Res* 58, 32–39 (2007).
11. Liu, Y.-W. et al. Psychotropic effects of Lactobacillus plantarum PS128 in early life-stressed and naïve adult mice. *Brain Res* 1631, 1–12 (2016).
12. Pusceddu, M. M. et al. n-3 PUFAs have beneficial effects on anxiety and cognition in female rats: Effects of early life stress. *Psychoneuroendocrinology* 58, 79–90 (2015).
13. Récamier-Carballo, S., Estrada-Camarena, E. & López-Rubalcava, C. Maternal separation induces long-term effects on monoamines and brain-derived neurotrophic factor levels on the frontal cortex, amygdala, and hippocampus: differential effects after a stress challenge. *Behav Pharmacol* 28, 545–557 (2017).
14. Rentesi, G. et al. Early maternal deprivation-induced modifications in the neurobiological, neurochemical and behavioral profile of adult rats. *ß* 244, 29–37 (2013).

Supplementary Information for Chapter 7

Effects of early life adversity on immediate early gene expression: systematic review and 3-level meta-analysis of rodent studies.

Contents

Supplementary Methods

Study Protocol

Systematic Review Protocol for Animal Intervention Studies
Format by SYRCLE (www.syrcle.nl)
Version 2.0 (December 2014)

Item #	Section/Subsection/Item	Description	Changes from original protocol
	A. General		
1.	Title of the review	Effect of Early Life Adversity on Immediate Early Gene Expression in Rodents	
2.	Authors (names, affiliations, contributions)	Valeria Bonapersona[1], Heike Schuler[1], Marian Joëls[1,2], R. Angela Sarabdjitsingh[1] [1] Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, The Netherlands [2] University Medical Center Groningen, University of Groningen, The Netherlands	The protocol was written in April 2019 and since then unaltered.
3.	Other contributors (names, affiliations, contributions)	SYRCLE (SYstematic Review Center for Laboratory animal Experimentation), Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands	
4.	Contact person + e-mail address	Valeria Bonapersona; v.bonapersona2@umc-utrecht.nl	
5.	Funding sources/sponsors	The Consortium on Individual Development (CID) is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003).	

6.	Conflicts of interest	None
7.	Date and location of protocol registration	Date: Location: www.crd.york.ac.uk
8.	Registration number (if applicable)	
9.	Stage of review at time of registration	Completed preliminary searches, started with piloting of the study selection process.
	B. Objectives	
	Background	
10.	What is already known about this disease/model/intervention? Why is it important to do this review?	Exposure to adversities during childhood (early-life adversities, ELA) increases the risk to develop psychiatric disorders in adulthood. Building upon the compelling epidemiological evidence, rodent studies have investigated the mechanistic effects of ELA on the brain, ultimately leading to changes in behavior. Modelled as alterations in maternal care, ELA alters brain development on multiple levels, including synaptic organization. Important contributors to synaptic development and cognition are immediate early genes (IEGs). IEGs are expressed directly but transiently upon cell activity; hence, they can be considered a marker for information processing in the brain. Since IEG proteins vary from transcription factors to post-translational proteins, their different functions can highlight different aspects of synaptic development. Systematically reviewing the current literature on the topic can provide insights on long-term changes of ELA on IEGs throughout the brain, thereby providing possible mechanisms for ELA-induced changes in information processing.
	Research question	
11.	Specify the disease/health problem of interest	Childhood maltreatment; Early life adversity; Stress-related psychopathology; Healthy animals
12.	Specify the population/species studied	mice and rats, because they are the most frequently used animal models in stress research; female and male
13.	Specify the intervention/exposure	1) Early life adversity starting before P14; early life adversity defined as alteration in maternal care; models included are maternal separation/deprivation, isolation, limited bedding and nesting, licking and grooming (as measure of variation in maternal care, with final comparisons between offspring receiving low vs high maternal care (for reference: Liu et al., 1997)), handling 2) Acute stressors applied to adult animals; will be restricted to most common ones as based on results from formal screening Both, 1) and 2) need to be applied.

14.	Specify the control population	Control animals differ from experimental animals only by exposure to early life adversity.	
15.	Specify the outcome measures	Immediate early gene mRNA expression, as measured by fold-change or percentage compared to control. Immediate early gene protein expression, as measured by optical density of counts, absolute counts, or optical density of Western Blots.	
16.	State your research question (based on items 11-15)	In (healthy) mice and rats, what is the effect of childhood maltreatment, early life adversity and/or stress-related psychopathology on immediate early gene mRNA expression? Hypothesis-confirming research Questions: • In mice and rats, does early life adversity alter immediate early genes expression after an acute stress challenge? • Is this differential expression amplified by multiple hits? • Secondary exploratory research questions: • Do the brain regions involved in the stress response a different sensitivity with regards to acute stress as seen in immediate early gene expression? • Are the above-mentioned effects sensitive to 1) the type of acute stressor and 2) the choice of early life adversity model? • What is the relationship in mRNA and protein expressions of any given immediate early gene in response to early life adversity?	Change: due to the uneven distribution across subgroups, secondary exploratory questions became sensitivity analyses.
	C. Methods		
	Search and study identification		
17.	Identify literature databases to search (*e.g.* Pubmed, Embase, Web of science)	☑ *MEDLINE via PubMed* ☐ *Web of Science* ☐ *SCOPUS* ☑ *EMBASE* ☐ *Other, namely:* ☐ *Specific journal(s), namely:*	
18.	Define electronic search strategies (*e.g.* use the step by step search guide[15] and animal search filters[20, 21])	When available, please add a supplementary file containing your search strategy: **[insert file name]**	
19.	Identify other sources for study identification	☑ *Reference lists of included studies* ☐ *Books* ☑ *Reference lists of relevant reviews* ☐ *Conference proceedings, namely:* ☐ *Contacting authors/ organisations, namely:* ☐ *Other, namely:*	
20.	Define search strategy for these other sources	Once the second phase of screening is completed, the reference list of the included studies and relevant reviews will be checked by one reviewer (HS). Studies that fit the search criteria identified in Questions 23-30 will be included.	
	Study selection		
21.	Define screening phases (*e.g.* pre-screening based on title/abstract, full text screening, both)	1) Title/abstract screening 2) Full text screening	

22.	Specify (a) the number of reviewers per screening phase and (b) how discrepancies will be resolved	Study inclusion is performed by (at least) two experimenters independently and it consists of two phases. During the first phase, titles and abstracts are screened and studies are excluded if: 1) not primary publication, 2) not in mice or rats, 3) not concerning early life adversity. During the second phase, the full text is screened and studies are selected according to the priority list below. Discrepancies will be resolved by discussion between two experimenters. Should no conclusion be reached between two experimenters (VB & HS), a third researcher (RAS), will be consulted for a solution.	
	Define all inclusion and exclusion criteria based on:		
23.	Type of study (design)	**Inclusion criteria:** primary publications **Exclusion criteria:** reviews; unpublished data; commentaries	
24.	Type of animals/population (*e.g.* age, gender, disease model)	**Inclusion criteria:** adult mice or rats (older than 8 weeks, but younger than 1 year); female and male **Exclusion criteria:** any other species than mice or rats; sexes are pooled; sex is not specified; ovariectomized females; specific pathogen free animals; genetic manipulations; animals bred for high/low anxiety or novelty response or sensitivity/resilience to depression; animals separated in high/low performance; any manipulations to earlier generations; animals with any comorbidities	
25.	Type of intervention (*e.g.* dosage, timing, frequency)	**Inclusion criteria:** 1. Early life adversity starting before P14; early life adversity defined as alteration in maternal care; models included are maternal separation/deprivation, isolation, limited bedding and nesting, licking and grooming (as measure of variation in maternal care, with final comparisons between offspring receiving low vs high maternal care (for reference: Liu et al., 1997)); handling is also considered early life adversity, but will be included only at a systematic review level 2. Acute stressors applied to adult animals; the types will be restricted to most common ones as based on results from formal screening **Exclusion criteria:** pharmacological intervention ("control" injections (of any pharmacological intervention) such as vehicle, saline, sesame oil are instead included); communal nesting as early life adversity model; maternal separation with early weaning, unless early weaning is also applied to control group; the same acute stressor has been applied earlier in life and is therefore not new to the animal	Change: we include also at rest measures. I.e. lack of acute stress is not an exclusion criteria.

26.	Outcome measures	**Inclusion criteria:** IEGs expression as measured by mRNA or protein expression in one of the following brain regions: amygdala, hippocampus, hypothalamus, medial prefrontal cortex, nucleus accumbens, striatum; IEGs expression in other brain regions will be included only at a systematic review level **Exclusion criteria:** brain regions not specified
27.	Language restrictions	**Inclusion criteria:** **Exclusion criteria:** None
28.	Publication date restrictions	None
29.	Other	The second hypothesis-confirming research question asks about the effects of multiple hits on the relationship between ELA and IEGs expression. • The following events are considered second hits: Stressful behavioral test performed previously (e.g., FST, fear conditions) • Footshocks • Chronic (mild) unpredictable stress • Chronic constant light • Chronic restraint • Chronic individual housing • Vaginal balloon distention • Cannula implementation, mock surgeries, blood sampling, isoflurane anaesthesia • Dams transported pregnant • Stress prone strain (BALB/C, wistar Kyoto, DBA) • The following events are not classified second hits: • Intragastric saline • Saline injections • Vaginal smears • Daily handling by experimenter
30.	Sort and prioritize your exclusion criteria per selection phase	**Titles and abstracts selection:** 1. Not primary publications. 2. Did not use mice/rats. 3. Not a model of early life adversity. **Full text selection:** 4. Did not measure IEG products (mRNA or protein expression). 5. Not an acute stressor in adult life. 6. Animals fall into any of the exclusion criteria as specified in question 19. 7. Interventions fall into any of the exclusion criteria as specified in question 20. 8. Outcome measures fall into any of the exclusion criteria as specified in question 24. 9. Intervention specific to control group/experimental group, so that the groups differ by more than just early life adversity exposure.

	Study characteristics to be extracted (for assessment of external validity, reporting quality)	
31.	Study ID (*e.g.* authors, year)	Study ID Author Abstract Year Journal
32.	Study design characteristics (*e.g.* experimental groups, number of animals)	n Control (only differs by exposure to ELS) n Experimental (ELS)
33.	Animal model characteristics (*e.g.* species, gender, disease induction)	Species Strain Origin of the animals (own breeding, purchased pregnant, etc.) Sex Age at Experiment (Acute Stressor)
34.	Intervention characteristics (*e.g.* intervention, timing, duration)	1. **ELA model:** model, duration, start (related to age of animal), end (related to age of animal), litter size 2. **Acute stressor:** type, categorization, duration, intensity (if applicable), time (of day), time before death 3. **Multiple hits:** yes/no, type (if applicable)
35.	Outcome measures	IEG (name; categorical) Brain area (name; categorical) Type (e.g., mRNA or protein; categorical) Measure (e.g., percentage, fold-increase, optical density, counts; categorical)
36.	Other (*e.g.* drop-outs)	/
	Assessment risk of bias (internal validity) or study quality	
37.	Specify (a) the number of reviewers assessing the risk of bias/study quality in each study and (b) how discrepancies will be resolved	Risk of bias will be assessed by two independent researchers. Risk of bias is assessed following SYRCLE guidelines, and it will be distinguished between experimental and study bias. Discrepancies will be resolved by discussion between two experimenters. Should no conclusion be reached between two experimenters, a third researcher (expert in the field of early life adversity), will be consulted for a solution.
38.	Define criteria to assess (a) the internal validity of included studies (*e.g.* selection, performance, detection and attrition bias) and/or (b) other study quality measures (*e.g.* reporting quality, power)	☑ By use of SYRCLE's Risk of Bias tool[4] ☐ By use of SYRCLE's Risk of Bias tool, adapted as follows: ☐ By use of CAMARADES' study quality checklist, e.g [22] ☐ By use of CAMARADES' study quality checklist, adapted as follows: ☐ Other criteria, namely:

	Collection of outcome data	
39.	For each outcome measure, define the type of data to be extracted (*e.g.* continuous/dichotomous, unit of measurement)	Mean Control (continuous) Mean Experimental (continuous) Standard Deviation Control (continuous) Standard Deviation Experimental (continuous) Reported direction of the effect (increase, decrease, non-significant; categorical) Data will be extracted in form of a comparison between a control and an experimental group, which only differ in exposure to early-life adversity. The same animal (group) can be part of multiple comparisons.
40.	Methods for data extraction/retrieval (e.g. first extraction from graphs using a digital screen ruler, then contacting authors)	1. Extraction from numbers provided in the text (means, standard deviations, n). 2. Extraction from graphs. 3. Extraction from statistical analyses. 4. Contacting the authors.
41.	Specify (a) the number of reviewers extracting data and (b) how discrepancies will be resolved	(a) One reviewer will complete data extraction, with a second reviewer checking random samples for agreement. Any numbers presented in the article or supplementary material will be extracted. If data is only presented graphically, then 'WebPlotDigitizer' will be used to extract data from graphs. If results of statistical analyses are given, these will be used to infer summary statistics. Two authors per publication will be contacted in case of missing data, followed by a reminder in case of no reply. Should authors not answer within two months, the comparisons will be reported as missing and will be excluded from analyses. (b) Discrepancies will be resolved by discussion between two experimenters. Should no conclusion be reached between two experimenters, a third researcher (expert in the field of early life adversity), will be consulted for a solution.
	Data analysis/synthesis	
42.	Specify (per outcome measure) how you are planning to combine/compare the data (*e.g.* descriptive summary, meta-analysis)	A quantitative synthesis is planned for results concerning the immediate early genes c-fos, arc, and egr1. Data will be split by sex, and the meta-analysis will be conducted for each dataset separately, since we consider males and females to be two different biological systems that should not be grouped together.
43.	Specify (per outcome measure) how it will be decided whether a meta-analysis will be performed	The decision on which brain regions and acute stressors to include in the quantitative analysis will be made after study selection, with frequency being the determining factor. Remaining immediate early genes, brain regions and acute stressors, as well as the early life adversity model of handling, will be covered in a narrative/descriptive synthesis.
	If a meta-analysis seems feasible/sensible, specify (for each outcome measure):	
44.	The effect measure to be used (*e.g.* mean difference, standardized mean difference, risk ratio, odds ratio)	The standardized mean difference Hedge's g ((mean(Control) - mean(Experimental)) / pooled SD) will be used for all outcome measures.

45.	The statistical model of analysis (*e.g.* random or fixed effects model)	3-level mixed effect meta-analysis (in case of multiple outcomes from the same animals) otherwise random effects meta-analysis, with early life stress predicting IEG mRNA and protein expression. IEG identity, presence of second hits and brain region (if applicable) will be moderators in the model.	
46.	The statistical methods to assess heterogeneity (*e.g.* I^2, Q)	Cochranes Q-test; I^2	
47.	Which study characteristics will be examined as potential source of heterogeneity (subgroup analysis)	Type of IEGs, ELA models, species, types of acute stressors, brain area and outcome measure (mRNA vs protein) will be used for subgroup analyses. These will be considered exploratory.	
48.	Any sensitivity analyses you propose to perform	Specified prior to the analysis, we will assess the influence of the following factors on the outcome measure: 1. Influential cases and outliers. 2. Blinded and randomized studies. 3. Risk of potential bias (bias will be assessed with the SYRCLE Risk of bias tool and an overall score will be used for sensitivity analysis).	
49.	Other details meta-analysis (*e.g.* correction for multiple testing, correction for multiple use of control group)	**Correction for multiple testing:** Bonferroni for family-wise comparisons will be applied for subgroup and exploratory analyses. Primary hypothesis-confirming research questions are considered separate families. **Correction for multiple use of control group:** $n_{(control)} / n_{(comparison)}$	Change: Holm for p-values correction
50.	The method for assessment of publication bias	If sufficient number of studies is achieved and a meta-analysis is conducted, the following methods will be applied: 1. Qualitative assessment of funnel plot 2. Egger's regression, followed by test for funnel plot asymmetry 3. Begg's test 4. Fail and save test 5. Trim and fill	

Search String

Pubmed:

Part 1 - Mice and rats:

("rodentia"[Mesh] OR rodent*[tiab] OR "mus"[Tiab] OR "mice"[Mesh] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[Mesh] OR "rats"[tiab] OR "rat"[tiab])

Part 2 – Postnatal early-life adversity:

("maternal behavior"[MeSh] OR "maternal care"[tiab] OR "early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "ELA"[tiab] OR "early life manipulation"[tiab] OR "early life manipulations"[tiab] OR "early adverse experience"[tiab] OR "early adverse experiences"[tiab] OR "early adversed experience"[tiab] OR "early adversed experiences"[tiab] OR "perinatal stress"[tiab] OR "perinatal adversity"[tiab] OR "perinatal adversities"[tiab] OR

"perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "perinatal adverse experience"[tiab] OR "perinatal adverse experiences"[tiab] OR "perinatal adversed experience"[tiab] OR "perinatal adversed experiences"[tiab] OR "postnatal stress"[tiab] OR "postnatal adversity"[tiab] OR "postnatal adversities"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "postnatal adverse experience"[tiab] OR "postnatal adverse experiences"[tiab] OR "postnatal adversed experience"[tiab] OR "postnatal adversed experiences"[tiab] OR "neonatal stress"[tiab] OR "neonatal adversity"[tiab] OR "neonatal adversities"[tiab] OR "neonatal manipulation"[tiab] OR "neonatal manipulations"[tiab] OR "neonatal adverse experience"[tiab] OR "neonatal adverse experiences"[tiab] OR "neonatal adversed experience"[tiab] OR "neonatal adversed experiences"[tiab] OR "Maternal Deprivation"[Mesh] OR "maternal deprivation"[tiab]OR "maternal separation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR "limited bedding/nesting"[tiab] OR "limited bedding-and-nesting"[tiab] OR "limited nesting/bedding"[tiab] OR "limited nesting-and-bedding"[tiab] OR "early life isolation"[tiab] OR "perinatal isolation"[tiab] OR "postnatal isolation"[tiab] OR "neonatal isolation"[tiab] OR "licking and grooming"[tiab] OR "licking-and-grooming"[tiab] OR "licking/grooming"[tiab] OR "early handling"[tiab] OR "early life handling"[tiab] OR "perinatal handling"[tiab] OR "postnatal handling"[tiab] OR "neonatal handling"[tiab])

Embase search string

Part 1 – Mice and rats:

(rodent*:ab,ti OR mus:ab,ti OR mouse:ab,ti OR mice:ab,ti OR rat:ab,ti OR rats:ab,ti)

Part 2 – Postnatal early-life adversity:

('maternal behavior':ab,ti OR 'maternal care':ab,ti OR 'early life stress':ab,ti OR 'els':ab,ti OR 'early life adversity':ab,ti OR 'early life adversities':ab,ti OR 'ela':ab,ti OR 'early life manipulation':ab,ti OR 'early life manipulations':ab,ti OR 'early adverse experience':ab,ti OR 'early adverse experiences':ab,ti OR 'early adversed experience':ab,ti OR 'early adversed experiences':ab,ti OR 'perinatal stress':ab,ti OR 'perinatal adversity':ab,ti OR 'perinatal adversities':ab,ti OR 'perinatal manipulation':ab,ti OR 'perinatal manipulations':ab,ti OR 'perinatal adverse experience':ab,ti OR 'perinatal adverse experiences':ab,ti OR 'perinatal adversed experience':ab,ti OR 'perinatal adversed experiences':ab,ti OR 'postnatal stress':ab,ti OR 'postnatal adversity':ab,ti OR 'postnatal adversities':ab,ti OR 'postnatal manipulation':ab,ti OR 'postnatal manipulations':ab,ti OR 'postnatal adverse experience':ab,ti OR 'postnatal adverse experiences':ab,ti OR 'postnatal adversed experience':ab,ti OR 'postnatal adversed experiences':ab,ti OR 'neonatal stress':ab,ti OR 'neonatal adversity':ab,ti OR 'neonatal adversities':ab,ti OR 'neonatal manipulation':ab,ti OR 'neonatal manipulations':ab,ti OR 'neonatal adverse experience':ab,ti OR 'neonatal adverse experiences':ab,ti OR

'neonatal adversed experience':ab,ti OR 'neonatal adversed experiences':ab,ti OR 'maternal deprivation':ab,ti OR 'maternal separation':ab,ti OR 'limited bedding':ab,ti OR 'limited nesting':ab,ti OR 'limited material':ab,ti OR 'limited bedding/nesting':ab,ti OR 'limited bedding-and-nesting':ab,ti OR 'limited nesting/bedding':ab,ti OR 'limited nesting-and-bedding':ab,ti OR 'early life isolation':ab,ti OR 'perinatal isolation':ab,ti OR 'postnatal isolation':ab,ti OR 'neonatal isolation':ab,ti OR 'licking and grooming':ab,ti OR 'licking-and-grooming':ab,ti OR 'licking/grooming':ab,ti OR 'early handling':ab,ti OR 'early life handling':ab,ti OR 'perinatal handling':ab,ti OR 'postnatal handling':ab,ti OR 'neonatal handling':ab,ti)

Extracted Variables

To increase subjectivity during data extraction, variables to be extracted were determined a *priori*. The spreadsheet containing all extracted variables and variable coding is available at https://osf.io/qkyvd/.

Entity	Variables
Publication	title; authors; year; journal
Animal	species; strain; origin (e.g., breeding, dams purchased pregnant); sex
Model	model (type, timing, cage (novel cage or home cage); light/dark phase; repetition (e.g., once, twice, predictable, unpredictable)); cross fostering; culling; sex ratio; litter size
Multiple Hits	other life experiences; housing in adulthood
Testing	age at testing; acute stressor (type, duration and novelty); time until perfusion; estrous cycle phase (females only)
Outcome	IEG name and product; measurement (technique, unit of recording (e.g., counts, expression, optical density) and unit of comparison (e.g., raw data, fold change, averages across slices)); brain area and hemisphere
Data	mean, variance and n of control and experimental groups; significant effect

Variables' grouping

f. Brain areas as named in publications and as grouped for the analysis.

Grouped for analysis	Named in publications
Amygdala	Central amygdala; Medial amygdala; Basolateral nucleus; Amygdala central nucleus; basolateral amygdala; Lateral amygdala
Hippocampus	Dorsal CA1; Dorsal CA2; Dorsal CA3; Dorsal dentate gyrus; Ventral CA1; Ventral CA2; Ventral CA3; Ventral dentate gyrus; CA1 subregion of the hippocampus; Dentate gyrus; Central CA3; CA1 region; CA1; CA2; CA3; Hippocampus
Hypothalamus	PVN; mpPVN; mgPVN; lpPVN; dpPVN; Medial mammillary nucleus; medial parvocellular portion of the PVN; paraventricular nucleus of the HAT; Paraventricular nucleus; ventromedial hypothalamic nucleus; anterior hypothalamus; lateral hypothalamus; dorsolmedial hypothalamus
Prefrontal cx	ACC; Cingulate cortex; mPFC; caudal cingulate cortex; rostral cingulate cortex; infralimbic cortex; prelimbic cortex; anterior cingulate cortex; prefrontal cortex; Cingulate cortex; lateral orbital frontal cortex; medial orbital frontal cortex; ventral orbital frontal cortex; medial prefrontal cortex

Thalamus	CM; PV; VPL; Anterodorsal thalamic nuclei; central medial thalamic nucleus; anteroventral thalamus; anteromedial thalamus
Other*	dorsal striatum; Barrel cortex; Piriform cortex; Lateral septum; Caudate putamen; DRN; Pontine region; Cerebellum; vBNST; Nucleus accumbens; ventrolateral periaqueductal gray; dorsolateral periaqueductal gray; DRD; DRV; DRVL; DRI; Retrosplenial cortex; non-preganglionic Edinger-Westphal nucleus; dorsal raphe nucleus; forebrain neocortical tissue; Nacc; VTA; medial orbital frontal cortex; ventral orbital frontal cortex; lateral orbital frontal cortex; insular cortex; dorsolateral striatum; dorsomedial striatum; nucleus accumbens shell; Cortex; Striatum; periaqueductal gray; bed nuclues of the stria terminalis; ventral subiculum; dorsal lateral septum; ventral lateral septum; medial septum; dorsal periaqueductal gray; inferior colliculus; locus coereleus; lateral septum; nucleus accumbens; ventral pallidum

* 'Other' brain areas have not been included in the analysis.

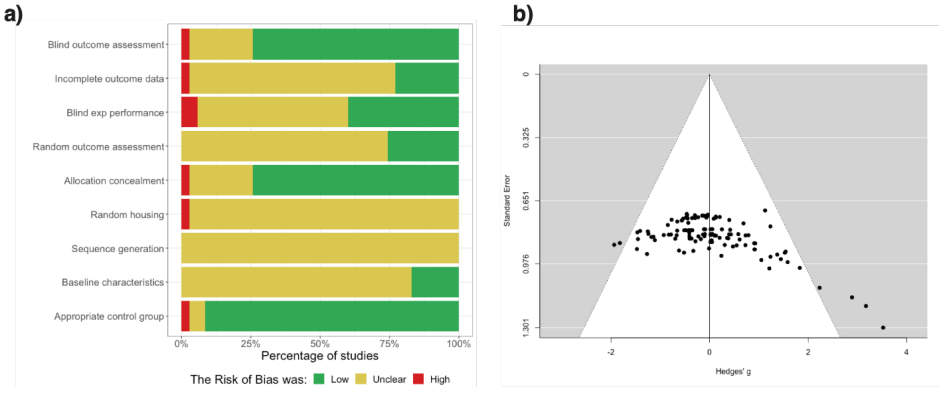
g. Categorization of acute stress in mild and severe.

Intensity	Type
Mild	EPM; OFT; DLB; Competition; Reexposure FC context (no shock); Social Defeat; NE; IGT; Three chamber test; Social interaction after 1d of social isolation
Severe	CRD; RS; FS in inhibitory avoidance task; FST; Shock in shock-probe burial task; MWM

Note on stressor's categorization:
Stressors with a strong memory, social or reward component were assessed on a systematic review level only. Although different types of acute stress can cause differential activation throughout the brain, in this meta-analysis we investigate the difference between controls and ELA animals, rather than the c-fos distribution within/between brain areas. The effects of ELA may interact with other factors for acute stressors with a strong memory, social or reward component. For this reason, we review these experiments only at a systematic review level (S2.4.3). However, for stressors with a physical component, we reasoned that the difference in effects between ELA and control should be comparable.

Supplementary results

Bias assessment

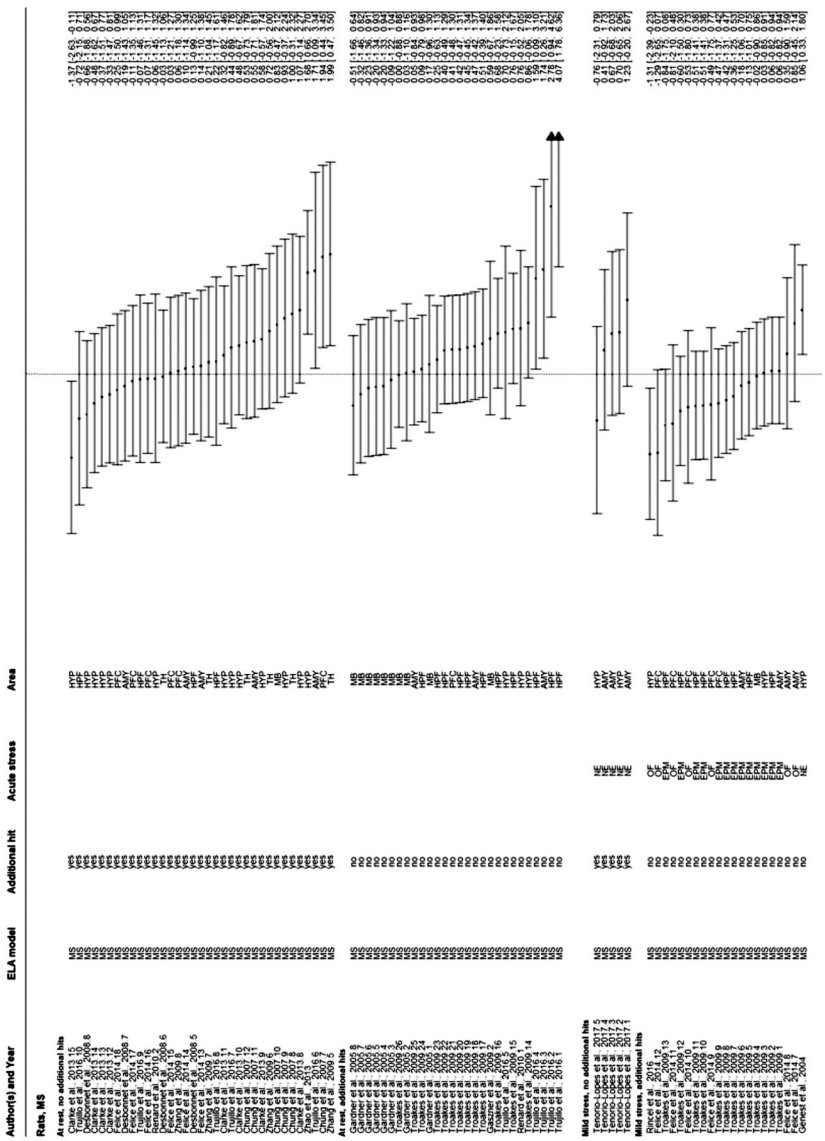


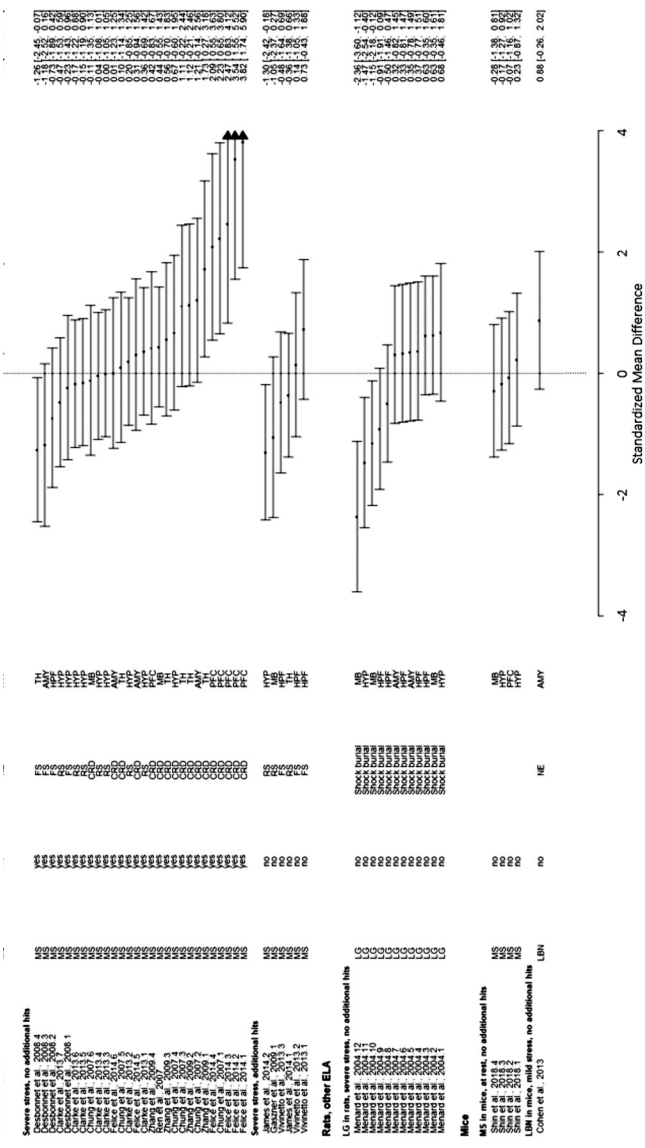
Bias assessment. A) Risk of bias assessment according to SYRCLE's risk of bias tool. B) Funnel plot for publication bias.

Sensitivity analysis species

Given that data from both rats and mice were combined in the meta-analysis, we performed a sensitivity analysis to confirm that the findings are robust to the effect of species. The effect size remains unchanged when analyzing data from rats only (*g[SEM]* = 0.23[±0.075], *z* = 2.983, *p* = 0.003) compared to the full model (*g[SEM]* = 0.223[±0.079], *z* = 2.91, *p* = 0.004). This remains true when looking at the subset of experiments with an acute stressor (rats-only model: *g[SEM]* = 0.084[±0.119], *z* = 0.706, *p* = 0.480; full model: *g[SEM]* = 0.109[±0.116], *z* = 0.938, *p* = 0348).

Forest Plot





Systematic review

cFos in female rodents.

Given fundamental biological differences between males and females [1], we a priori chose to evaluate female cFos data separately from males'. Only ten publications reported on cFos expression in female rodents (n_{comp} = 77). The majority of these studies found no significant differences between cFos levels of ELA versus controls at rest or after an acute stress challenge (n_{comp} = 55; [2–6]).

Only five studies performed the same experiments in both male and female rodents. Among these, Desbonnet et al. [5], Gaszner et al. [6] and Renard et al. [2] reported the same null effects for both male and females. In contrast, James et al. [3] and Genest et al. [4] found no significant ELA effects on cFos levels in females, while

they did report significant differences in males under the same conditions. These sexually dimorphic results could have methodological origins, such as male-focused behavioral paradigms (for ELA or acute stress) or could reflect true biological differences between the sexes [1, 7].

The remaining five studies investigated exclusively females, and all reported at least one significant difference between ELA and control rodents. Auth et al. [8] found significantly increased cFos levels in female mice at rest, but not after acute stress exposure. Interestingly, across two independent baseline cohorts, increased cFos was observed once in the dorso-lateral periaqueductal gray and once in the lateral amygdala, suggesting that the effects do not easily replicate within the same lab. Similarly, Rivarola and colleagues [9, 10] observed an increase in cFos levels in the anterior-dorsal thalamic nucleus of animals with a history of multiple hits in a first [9] but not a second publication [10].

Finally, O'Leary et al. [11] reported decreased cFos levels in the dorsal dentate gyrus and ventral CA3 of female ELA mice after restraint stress, but not in other hippocampal, hypothalamic, prefrontal cortical or amygdalar areas. Banqueri et al. [12] demonstrated differential directionality of effects after the Morris water maze, with ELA females showing increased cFos levels in hippocampal structures, and decreased expression in prefrontal areas. All in all, ELA effects on cFos in females appeared limited. Whether the results are truly sexually dysmorphic remains to be elucidated.

cFos and other brain areas Five studies investigated the effect of ELA on cFos expression in brain areas of male rodents other than those reviewed in the meta-analysis, including the striatum, sensory cortices, hindbrain nuclei and the cerebellum of male rodents. Out of 24 comparisons, 16% displayed a significant difference between ELA and control animals (ncomp = 4) at systematic review level. Troakes et al. [13] showed that cFos levels of ELA males are significantly decreased in the piriform cortex in comparison to controls after acute exposure to a mild stressor, but not at rest. Early research indicated that cFos levels in the piriform cortex are highly responsive to acute stressors, and its role in the sensory integration of olfactory stimuli suggests that the reduced cFos expression could correspond to decreased information processing abilities under stressful [14, 15].

In addition, Menard et al. [16] found decreased cFos expression in ELA males in the lateral septal complex and the ventral subiculum after performing a shock-probe burial task, but not in other striatal areas or hindbrain nuclei. Given that the lateral septal complex relays reward and fear information for contextualization of the experience, the decreased cFos expression here potentially presents a task-specific effect related to spatial mapping of the buried probe [17]. However, Shin et al. [18] report upregulation of cFos after ELA in the lateral septal complex as well as the ventral tegmental area in a social interaction task, suggesting a broader task-specific involvement of striatal areas.

Finally, neither Clarke et al. [19] nor Desbonnet et al. [5] could find significant differences between ELA males and controls in the bed nuclei of the stria terminalis, neither at rest nor after acute stress, suggesting that cFos expression in this area is not or only minimally changed after ELA. All in all, these results suggest that areas with task-specific effects are worth exploring, and that cortical areas involved in sensory processing and information integration potentially display altered transcriptional activity as well. Yet, considering that the most frequently areas under investigation are also those areas considered to be sensitive to the effects of stress, it is likely that the main results of interest are covered by the meta-analytic outcomes.

cFos and alternative behavioral paradigms. Acute stressors that included a strong memory, reward or social component were excluded from the meta-analysis. They involve cognitive processes other than the response to stress, which recruit brain-areas depending on the task requirements.

Daskalakis et al. [20] investigated cFos expression in rats placed back into a fearful context after a fear-conditioning paradigm, thereby probing memory processes in addition to stress-related functions. cFos expression in the medial amygdala and basolateral amygdala was increased in rats placed into a novel cage during the maternal separation (MS) procedure, while an increase was only observed in the medial amygdala in MS animals that remained in the home cage [20]. This study highlights how that choices of study characteristics (i.e., home cage vs novel cage) can influence the outcome investigated.

Two studies further investigated the effects of ELA on cFos expression after exposure to a rodent version of the Iowa Gambling Task [21, 22]alters cognitive functioning and in humans is thought to increase the vulnerability to psychopathology-e.g. depression, anxiety and schizophrenia- later in life. Here we investigated whether subtle natural variations among individual rat pups in the amount of maternal care received, i.e. differences in the amount of licking and grooming (LG. This task depends not only on spatial memory, but also contains a strong reward component [22]including the formation of contextual memory; it is also (transiently. In 2012, van Hasselt et al. [21] correlated percentage of licking and grooming with cFos expression in a wide range of brain areas in male and female rats and found a negative correlation in the shell of the nucleus accumbens and the agranular insular cortex when sex was pooled. However, using the same task, MS did not alter cFos expression in these areas in the 2017 study, but rather decreased cFos expression in the right CA1, right CA3, left infralimbic area and left agranular insula [22]including the formation of contextual memory; it is also (transiently. While inconsistent, these studies highlight that reward-based processes also likely result in differential activation of IEGs after ELA exposure, thus, warranting further investigations in the future.

Under several social paradigms, no differences between ELA and control animals

were observed in medial PFC areas [18, 23, 24]the social dimension has rarely been incorporated into the analysis due to methodological limitations. This study characterized the effects of neonatal social isolation (early deprivation, ED, the central amygdala [23]the social dimension has rarely been incorporated into the analysis due to methodological limitations. This study characterized the effects of neonatal social isolation (early deprivation, ED, the dorsal raphe nucleus [25], or striatal and hypothalamic areas [18]. On the other hand, Benner et al. [23]the social dimension has rarely been incorporated into the analysis due to methodological limitations. This study characterized the effects of neonatal social isolation (early deprivation, ED observed an increase in the basolateral amygdala and a decrease in CA1 of cFos expression in ELA mice compared to controls after 40-days social competition task. A possible explanation is that the differences observed in the study by Benner and colleagues are due to the memory component, rather than the stress/social component of the task. In addition, Shin et al. [18] observed an increase in cFos expression in the lateral septal complex and the ventral tegmental area after social interaction in mice previously exposed to social isolation, suggesting that multiple adverse experiences may be required to observe altered IEG expression after ELA in social tasks. Overall, social behaviors in isolation seem less inducive of activity-regulated transcription than the above-discussed reward-based and memory-based paradigms.

ELA and IEG other than cFos *Arc* is a post-synaptic protein, which plays an essential role in regulating the homeostatic scaling of AMPA receptors, thereby directly modifying plasticity at the synapse [26]. Arc expression has been investigated in five publications under varying conditions in male and female mice and rats. While two publications did not find any alterations in the mPFC, hippocampal, or amygdaloid areas at rest or after acute stress [23, 27], another study reported a significant decrease in CA1, CA3 and dentate gyrus Arc levels in male ELA animals at rest [28]. Interestingly, animals in this study were exposed to maternal separation for one week longer (PND 1-21) than animals in the studies reporting no significant alterations, suggesting that the duration of the *ELA* experience could be essential in causing long-term effects on *Arc* expression. It is noteworthy that a decrease in *Arc* expression results in increased synaptic plasticity [26], thus, following in line with the findings of increased cFos expression at rest in the male meta-analysis.

In contrast, McGregor et al. [29] found increased *Arc* expression at rest in the dorsal striatum of male rats with and without a history of second hits. As this publication is the only one reporting on IEG levels in the dorsal striatum, it is unclear whether the finding is a result of the study design or presents a genuine area specific IEG response. Rincel et al. [30] suggest that ELA effects on *Arc* expression are sex-specific, showing evidence that ELA leads to decreased *Arc* expression in the mPFC of male mice, but to increased Arc levels in the mPFC of female mice. These contradictory findings could

be a strain-specific, as C3H/HeNRj mice were used [30]. All in all, reported *Arc* levels appear to be in coherence with cFos effects on synaptic plasticity at rest, and thereby further support the notion of at rest sensitization of activity-regulated transcription.

Early-growth response (*Egr*) proteins are a family of transcription factors with a zinc-finger motif, which allows all *Egr* factors to connect to identical DNA binding sites [31]. We identified three studies investigating Egr expression after ELA exposure at rest; one investigated Egr-1 [32], another investigated Egr-4 only [30], and one other investigated Egr-2 and Egr-4 [29].

Egr-1 mRNA expression was decreased in the cortex, but only in Balb/c and not C57Bl/6 male mice [32]. This is in line with the general notion of Balb/c mice as a stress-sensitive strain [33]. In contrast, McGregor et al. [29] report increases in Egr-2 and Egr-4 expression in the dorsal striatum, with Egr-2 levels only significantly increased in animals experiencing a second hit during adolescence. Finally, Rincel et al. [30] highlight that ELA alters Egr-4 expression in a sex-specific manner in the mPFC of mice, observing an downregulation in males but an upregulation in females.

Since it is expected that proteins of the Egr-family behave similarly [31], the discrepancy between findings are likely the result of differences in study design, such as the brain areas investigated. Considering that IEGs of the Egr-family, and in particular Egr-1, have been shown to be associated with the development and treatment of those psychiatric disorders, which individuals with a history of early life stress are more likely to develop, Egr-family proteins are an understudied, yet important candidate for investigating activity-regulated transcriptional alterations after ELA in the future [34].

FosB is an IEG of the *Fos* family, and - similarly to *cFos* - if binds to members of the Jun family to form the AP1 transcription factor [35]. Of particular interest in stress research is its isoform Δ*FosB*, whose extended half-life makes Δ*FosB* an exceptional marker for chronic stress [35].

Three publications reporting on the expression of Δ*FosB* at rest in ELA and control animals were identified. Kim et al. [36] reported a reduction of Δ*FosB* expression in the nucleus accumbens of ELA females in comparison to controls, whereas Wang et al. [37] report elevated Δ*FosB* levels in the mPFC of ELA rats of unspecified sex, pointing towards opposite effects of ELA in these two areas. Interestingly, and in line with these findings, previous results suggest that overexpression of FosB in the nucleus accumbens accompanied by reduced expression of Δ*FosB* in mPFC promote a phenotype resilient to the effects of chronic stress [38, 39]. It should still be highlighted, that Lippmann et al. [40] found no significant alterations in either of these areas in male rodents, neither induced by maternal separation nor by handling. Due to the low number of studies investigating Δ*FosB*, we cannot conclude whether these null findings are attributable to sex or a result of study design heterogeneity. Yet the outlined potential of a more stable IEG in researching chronic alterations in transcriptional activity emphasizes the relevance of investigating ELA modifications on Δ*FosB* expression.

Supplementary references

[1] Shansky RM, Murphy AZ. Considering sex as a biological variable will require a global shift in science culture. *Nat Neurosci* 2021; 24: 457–464.

[2] Renard GM, Rivarola MA, Suárez MM. Gender-dependent effects of early maternal separation and variable chronic stress on vasopressinergic activity and glucocorticoid receptor expression in adult rats. *Dev Neurosci* 2010; 32: 71–80.

[3] James MH, Campbell EJ, Walker FR, et al. Exercise reverses the effects of early life stress on orexin cell reactivity in male but not female rats. *Front Behav Neurosci* 2014; 8: 244.

[4] Genest SE, Gulemetova R, Laforest S, et al. Neonatal maternal separation and sex-specific plasticity of the hypoxic ventilatory response in awake rat. *J Physiol* 2004; 554: 543–557.

[5] Desbonnet L, Garrett L, Daly E, et al. Sexually dimorphic effects of maternal separation stress on corticotrophin-releasing factor and vasopressin systems in the adult rat brain. *Int J Dev Neurosci* 2008; 26: 259–268.

[6] Gaszner B, Jensen KO, Farkas J, et al. Effects of maternal separation on dynamics of urocortin 1 and brain-derived neurotrophic factor in the rat non-preganglionic Edinger-Westphal nucleus. *Int J Dev Neurosci* 2009; 27: 439–451.

[7] Shansky RM. Are hormones a "female problem" for animal research? Outdated gender stereotypes are influencing experimental design in laboratory animals. *Science* 2019; 364: 825–826.

[8] Auth CS, Weidner MT, Popp S, et al. Differential anxiety-related behaviours and brain activation in Tph2-deficient female mice exposed to adverse early environment. *Eur Neuropsychopharmacol* 2018; 28: 1270–1283.

[9] Rivarola MA, Dalmasso C, Valdez DJ, et al. Early maternal separation and chronic variable stress as adults differentially affect fos expression in the anterodorsal thalami nuclei. *Int J Neurosci* 2008; 118: 735–748.

[10] Rivarola MA, Suárez MM. Early maternal separation and chronic variable stress in adulthood changes the neural activity and the expression of glucocorticoid receptor in limbic structures. *Int J Dev Neurosci* 2009; 27: 567–574.

[11] O'Leary OF, Felice D, Galimberti S, et al. GABAB(1) receptor subunit isoforms differentially regulate stress resilience. *Proc Natl Acad Sci U S A* 2014; 111: 15232–15237.

[12] Banqueri M, Méndez M, Arias JL. Why are maternally separated females inflexible? Brain activity pattern of COx and c-Fos. *Neurobiol Learn Mem* 2018; 155: 30–41.

[13] Troakes C, Ingram CD. Anxiety behaviour of the male rat on the elevated plus maze: Associated regional increase in c-fos mRNA expression and modulation by early maternal separation. *Stress* 2009; 12: 362–369.

[14] Cullinan WE, Herman JP, Battaglia DF, et al. *Pattern and time course of immediate early gene expression in rat brain following acute stress.* Epub ahead of print 1995. DOI: 10.1016/0306-4522(94)00355-9.

[15] Blazing RM, Franks KM. Odor coding in piriform cortex: mechanistic insights into distributed coding. *Current Opinion in Neurobiology* 2020; 64: 96–102.

[16] Menard JL, Champagne DL, Meaney MJP. Variations of maternal care differentially influence 'fear' reactivity and regional patterns of cFOS immunoreactivity in response to the shock-probe burying test. *Neuroscience* 2004; 129: 297–308.

[17] Wirtshafter HS, Wilson MA. Lateral septum as a nexus for mood, motivation, and movement. *Neuroscience and Biobehavioral Reviews* 2021; 126: 544–559.

[18] Shin S, Pribiag H, Lilascharoen V, et al. Drd3 Signaling in the Lateral Septum Mediates Early Life Stress-Induced Social Dysfunction. *Neuron* 2018; 97: 195–208.

[19] Clarke M, Cai G, Saleh S, et al. Being suckled in a large litter mitigates the effects of early-life stress on hypothalamic-pituitary-adrenal axis function in the male rat. *J Neuroendocrinol* 2013; 25: 792–802.

[20] Daskalakis NP, Diamantopoulou A, Claessens SEF, et al. Early experience of a novel-environment in isolation primes a fearful phenotype characterized by persistent amygdala activation. *Psychoneuroendocrinology* 2014; 39: 39–57.

[21] van Hasselt FN, de Visser L, Tieskens JM, et al. Individual variations in maternal care early in life correlate with later life decision-making and c-Fos expression in prefrontal subregions of rats. *PLoS One* 2012; 7: e37820.

[22] Loi M, Sarabdjitsingh RA, Tsouli A, et al. Transient prepubertal mifepristone treatment normalizes deficits in contextual memory and neuronal activity of adult male rats exposed to maternal deprivation. *eNeuro*; 4. Epub ahead of print 2017. DOI: 10.1523/ENEURO.0253-17.2017.

[23] Benner S, Endo T, Endo N, et al. Early deprivation induces competitive subordinance in C57BL/6 male mice. *Physiol Behav* 2014; 137: 42–52.

[24] Hidaka C, Kashio T, Uchigaki D, et al. Vulnerability or resilience of motopsin knockout mice to maternal separation stress depending on adulthood behaviors. *Neuropsychiatr Dis Treat* 2018; 14: 2255–2268.

[25] Gardner KL, Thrivikraman K V, Lightman SL, et al. Early life experience alters behavior during social defeat: Focus on serotonergic systems. *Neuroscience* 2005; 136: 181–191.

[26] Bramham CR, Worley PF, Moore MJ, et al. The Immediate Early Gene Arc/Arg3.1: Regulation, Mechanisms, and Function. *J Neurosci* 2008; 28: 11760–11767.

[27] Benekareddy M, Goodfellow NM, Lambe EK, et al. Enhanced function of prefrontal serotonin 5-HT2 receptors in a rat model of psychiatric vulnerability. *J Neurosci* 2010; 30: 12138–12150.

[28] Solas M, Aisa B, Mugueta MC, et al. Interactions between age, stress and insulin on cognition: Implications for alzheimer's disease. *Neuropsychopharmacology* 2010; 35: 1664–1673.

[29] McGregor NW, Dimatelis JJ, Van Zyl PJ, et al. A translational approach to the genetics of anxiety disorders. *Behav*

Brain Res 2018; 341: 91–97.
[30] Rincel M, Aubert P, Chevalier J, et al. Multi-hit early life adversity affects gut microbiota, brain and behavior in a sex-dependent manner. *Brain Behav Immun* 2019; 80: 179–192.
[31] Beckmann AM, Wilce PA. Egr transcription factors in the nervous system. Neurochem Int 1997; 31: 477–510.
[32] Navailles S, Zimnisky R, Schmauss C. Expression of glucocorticoid receptor and early growth response gene 1 during postnatal development of two inbred strains of mice exposed to early life stress. *Dev Neurosci* 2010; 32: 139–148.
[33] Brinks V, van der Mark M, de Kloet R, et al. Emotion and cognition in high and low stress sensitive mouse strains: A combined neuroendocrine and behavioral study in BALB/c and C57BL/6J mice. *Front Behav Neurosci*; 1. Epub ahead of print 2007. DOI: 10.3389/neuro.08.008.2007.
[34] Duclot F, Kabbaj M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders. *Front Behav Neurosci* 2017; 11: 35.
[35] Nestler EJ. ΔfosB: A transcriptional regulator of stress and antidepressant responses. *Eur J Pharmacol* 2015; 753: 66–72.
[36] Kim JY, Lee JH, Kim D, et al. Beneficial effects of highly palatable food on the behavioral and neural adversities induced by early life stress experience in female rats. *Int J Biol Sci* 2015; 11: 1150–1159.
[37] Wang H, Tao X, Huang ST, et al. Chronic stress is associated with pain precipitation and elevation in DeltaFosB expression. *Front Pharmacol* 2016; 7: 138.
[38] Vialou V, Robison AJ, LaPlant QC, et al. FosB in brain reward circuits mediates resilience to stress and antidepressant responses. *Nat Neurosci* 2010; 13: 745–752.
[39] Vialou V, Bagot RC, Cahill ME, et al. Prefrontal cortical circuit for depression- and anxiety-related behaviors mediated by cholecystokinin: Role of ΔFosB. *J Neurosci* 2014; 34: 3878–3887.
[40] Lippmann M, Bress A, Nemeroff CB, et al. Long-term behavioural and molecular alterations associated with maternal separation in rats. *Eur J Neurosci* 2007; 25: 3091–3098.

Supplementary Information for Chapter 8

Structural changes after early life adversity in rodents: a systematic review with meta-analysis.

Contents

Supplementary Notes

Supplementary Note 1: search string

Pubmed:

Part 1 - Mice and rats:

("rodentia"[Mesh] OR rodent*[tiab] OR "mus"[Tiab] OR "mice"[Mesh] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[Mesh] OR "rats"[tiab] OR "rat"[tiab])

Part 2 – Postnatal early-life adversity:

("maternal behavior"[MeSh] OR "maternal care"[tiab] OR "early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "ELA"[tiab] OR "early life manipulation"[tiab] OR "early life manipulations"[tiab] OR "early adverse experience"[tiab] OR "early adverse experiences"[tiab] OR "early adversed experience"[tiab] OR "early adversed experiences"[tiab] OR "perinatal stress"[tiab] OR "perinatal adversity"[tiab] OR "perinatal adversities"[tiab] OR "perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "perinatal adverse experience"[tiab] OR "perinatal adverse experiences"[tiab] OR "perinatal adversed experience"[tiab] OR "perinatal adversed experiences"[tiab] OR "postnatal stress"[tiab] OR "postnatal adversity"[tiab] OR "postnatal adversities"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "postnatal adverse experience"[tiab] OR "postnatal adverse experiences"[tiab] OR "postnatal adversed experience"[tiab] OR "postnatal adversed experiences"[tiab] OR "neonatal stress"[tiab] OR "neonatal adversity"[tiab] OR "neonatal adversities"[tiab] OR "neonatal manipulation"[tiab] OR "neonatal manipulations"[tiab] OR "neonatal adverse experience"[tiab] OR "neonatal adverse experiences"[tiab] OR "neonatal adversed experience"[tiab] OR "neonatal adversed experiences"[tiab] OR "Maternal

Deprivation"[Mesh] OR "maternal deprivation"[tiab]OR "maternal separation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR "limited bedding/nesting"[tiab] OR "limited bedding-and-nesting"[tiab] OR "limited nesting/bedding"[tiab] OR "limited nesting-and-bedding"[tiab] OR "early life isolation"[tiab] OR "perinatal isolation"[tiab] OR "postnatal isolation"[tiab] OR "neonatal isolation"[tiab] OR "licking and grooming"[tiab] OR "licking-and-grooming"[tiab] OR "licking/grooming"[tiab] OR "early handling"[tiab] OR "early life handling"[tiab] OR "perinatal handling"[tiab] OR "postnatal handling"[tiab] OR "neonatal handling"[tiab])

Embase search string

Part 1 – Mice and rats:

(rodent*:ab,ti OR mus:ab,ti OR mouse:ab,ti OR mice:ab,ti OR rat:ab,ti OR rats:ab,ti)

Part 2 – Postnatal early-life adversity:

('maternal behavior':ab,ti OR 'maternal care':ab,ti OR 'early life stress':ab,ti OR 'els':ab,ti OR 'early life adversity':ab,ti OR 'early life adversities':ab,ti OR 'ela':ab,ti OR 'early life manipulation':ab,ti OR 'early life manipulations':ab,ti OR 'early adverse experience':ab,ti OR 'early adverse experiences':ab,ti OR 'early adversed experience':ab,ti OR 'early adversed experiences':ab,ti OR 'perinatal stress':ab,ti OR 'perinatal adversity':ab,ti OR 'perinatal adversities':ab,ti OR 'perinatal manipulation':ab,ti OR 'perinatal manipulations':ab,ti OR 'perinatal adverse experience':ab,ti OR 'perinatal adverse experiences':ab,ti OR 'perinatal adversed experience':ab,ti OR 'perinatal adversed experiences':ab,ti OR 'postnatal stress':ab,ti OR 'postnatal adversity':ab,ti OR 'postnatal adversities':ab,ti OR 'postnatal manipulation':ab,ti OR 'postnatal manipulations':ab,ti OR 'postnatal adverse experience':ab,ti OR 'postnatal adverse experiences':ab,ti OR 'postnatal adversed experience':ab,ti OR 'postnatal adversed experiences':ab,ti OR 'neonatal stress':ab,ti OR 'neonatal adversity':ab,ti OR 'neonatal adversities':ab,ti OR 'neonatal manipulation':ab,ti OR 'neonatal manipulations':ab,ti OR 'neonatal adverse experience':ab,ti OR 'neonatal adverse experiences':ab,ti OR 'neonatal adversed experience':ab,ti OR 'neonatal adversed experiences':ab,ti OR 'maternal deprivation':ab,ti OR 'maternal separation':ab,ti OR 'limited bedding':ab,ti OR 'limited nesting':ab,ti OR 'limited material':ab,ti OR 'limited bedding/nesting':ab,ti OR 'limited bedding-and-nesting':ab,ti OR 'limited nesting/bedding':ab,ti OR 'limited nesting-and-bedding':ab,ti OR 'early life isolation':ab,ti OR 'perinatal isolation':ab,ti OR 'postnatal isolation':ab,ti OR 'neonatal isolation':ab,ti OR 'licking and grooming':ab,ti OR 'licking-and-grooming':ab,ti OR 'licking/grooming':ab,ti OR 'early handling':ab,ti OR 'early life handling':ab,ti OR 'perinatal handling':ab,ti OR 'postnatal handling':ab,ti OR 'neonatal handling':ab,ti)

Supplementary Tables

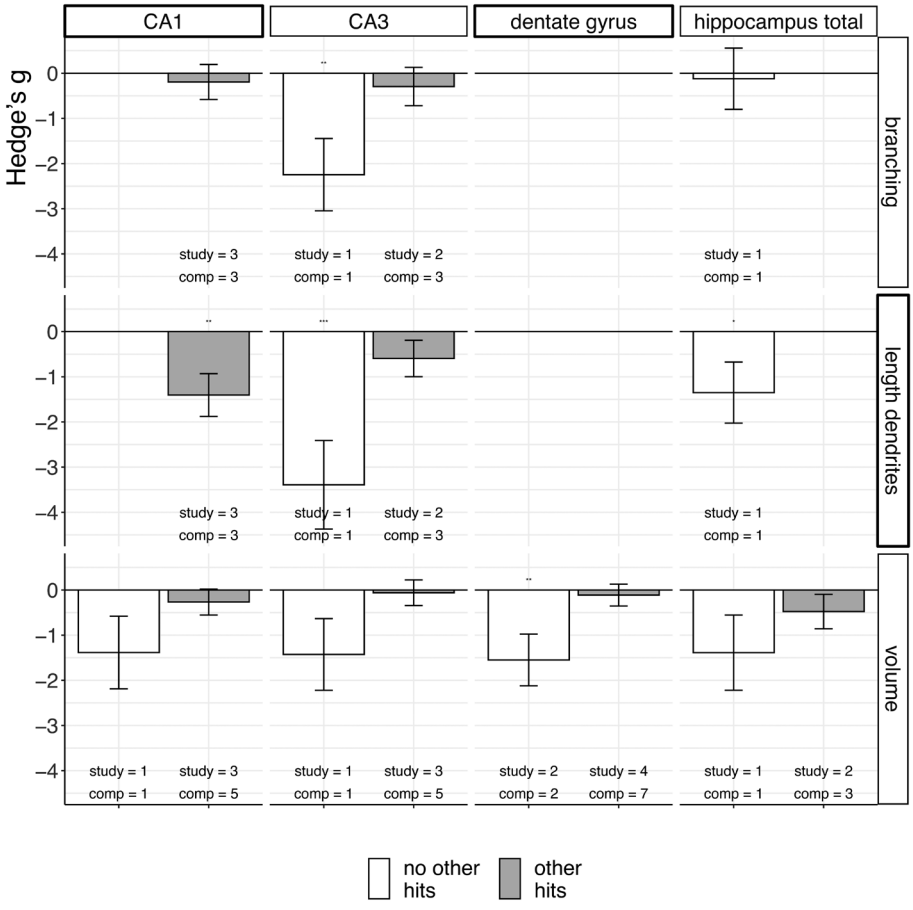
Supplementary table 1: model building considerations

Prior the beginning of the study, we identified several factors that may be important moderators of the effects of ELA on structural plasticity, namely: 1) specific outcome parameters, 2) brain area(s), 3) experience of other traumatic events, 4) product measured (mRNA or protein, only for the outcome BDNF), 5) state of the animal at death (only for BDNF and neurogenesis), 6) delay between the start of the experimental manipulation and measuring the outcome (only for the neurogenesis-related parameter brdu). Prior the beginning of the study, we chose that these moderators needed to be addressed in our study, as moderators, filtering variables, subgroup analysis or sensitivity analysis. The choice was based on the distribution of the factors of these variables in the dataset, so that to maximize interpretability (max interaction of 3 moderators) and to minimize the number of tests used. Specific outcome parameters and brain areas were important variables used for filtering. In particular, we performed out analysis only on the hippocampus, because it was the brain area most investigated. We excluded from the quantitative synthesis those outcomes reported by a limited number of publications (for specifics, see table below). With this filtering, we were able to maximize the homogeneity of our dataset for quantitative analysis. The table below reports the analytical considerations for each of the final models.

Structural plasticity:	Final model	Excluded data (filtering)	Considerations
Morphology	Interaction between sub-part of the hippocampus and experience of other traumatic events	Complexity as a composite measure (npaper = 1), spine density (npaper = 3, ncomp = 6), number of dendrites (npaper = 3, ncomp = 6)	
Neurogenesis	Experience of other traumatic events. Sub-parts of hippocampus not applicable (only dentate gyrus included). Subgroup analysis for brdu with short induction and ki67 for interaction of state of the animals at death.	Data not specific to the dentate gyrus (npaper = 1)	Brdu with short (<1 day) vs long (> 1 day) induction time are considered two separate outcomes to decrease the number of moderators used.
BDNF	Interaction between other traumatic events, state of the animal at death and product measured. These moderators were selected because they explained a significant proportion of the variance in univariate models.		State of the animal at death is relevant only for RNA outcomes because for most studies there was a short interval between the induction of the arousal/stress state and decapitation. Sub-parts of the hippocampus were not included in the final model because the majority of observations were measured in the whole hippocampus (58.3% of BDNF comparisons).

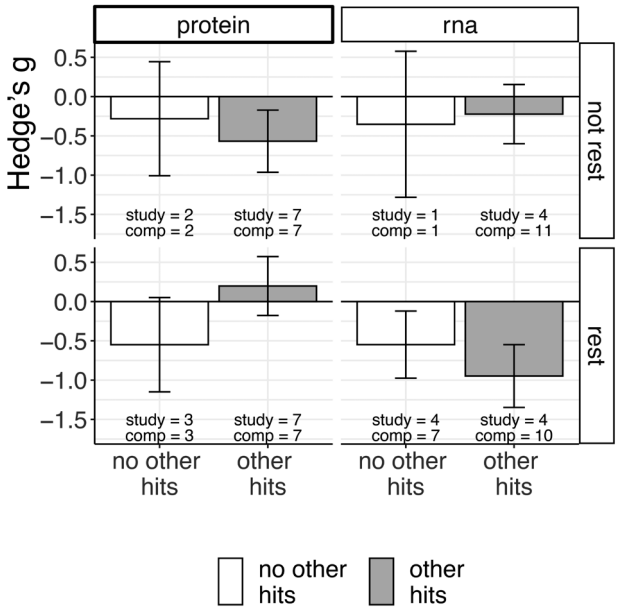
Supplementary Figures

Supplementary Figure 1 - Morphology



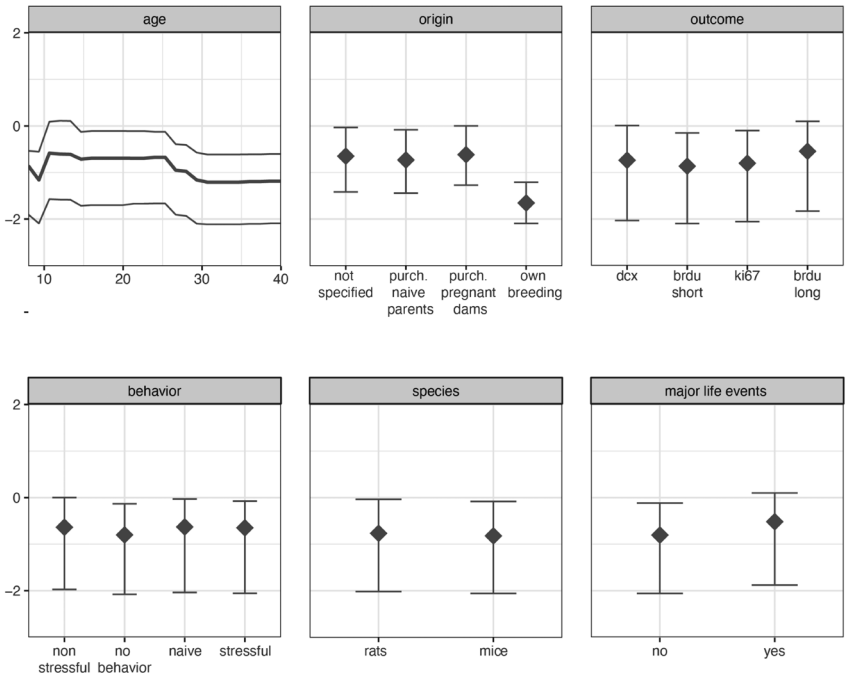
Effect estimates of morphology divided by sub-part of the hippocampus (vertical facets), outcome (horizontal facets) and the experience of additional life traumas (white and grey bars, see legend). *Study = # of publications for the specific outcome; comp = # of comparisons of the specified outcome. * = p < 0.05; ** = p < 0.01; *** = p < 0.001. The p-values reported are not adjusted and should be interpreted as exploratory only.*

Supplementary Figure 2 – BDNF analysis



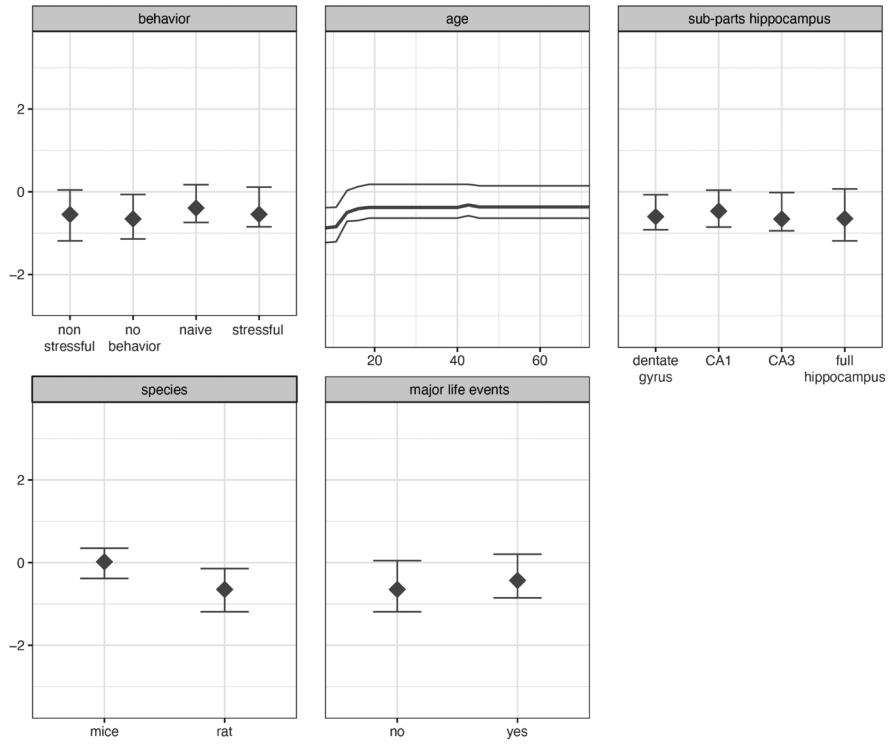
Effect estimates of BDNF divided by RNA/protein (vertical facets), state of the animal at death (horizontal facets) and the experience of additional life traumas (white and grey bars, see legend). *Study = # of publications for the specific outcome; comp = # of comparisons of the specified outcome. * = p < 0.05. The p-values reported are not adjusted and should be interpreted as exploratory only.*

Supplementary Figure 3



Neurogenesis partial dependency plots of those variables that in at least 50% of the replications had a positive variable importance (MetaForest analysis). Interval corresponds to prediction interval. The changes between sub-groups of each factor appear minor, with exception of origin "own breeding".

Supplementary Figure 4



BDNF partial dependency plots of those variables that in at least 50% of the replications had a positive variable importance (MetaForest analysis). Interval corresponds to prediction interval. The changes between sub-groups of each factor appear minor.

Supplementary Information for Chapter 9

Seven principles of early life adversity in rodents.

Contents

Supplementary note 1

Strings used for the systematic search. The first two elements (i.e. about ELA and about rodents) are identical for the behavior and neurobiology systematic searches. Additionally, the behavior search string had an added element to filter on the behaviors of interest. These strings have previously been published in (Bonapersona et al., 2019; Schuler et al., 2022).

PubMed

("early life stress"[tiab] OR "ELS"[tiab] OR "early life adversity"[tiab] OR "early life adversities"[tiab] OR "early life adversity*" OR "early stress"[tiab] OR "neonatal stress"[tiab] OR "postnatal stress"[tiab] OR "perinatal stress"[tiab] OR "neonatally stressed"[tiab] OR "early adverse experience"[tiab] OR "perinatally stressed"[tiab] OR "early adverse experiences"[tiab] OR "postnatal manipulation"[tiab] OR "postnatal manipulations"[tiab] OR "perinatal manipulation"[tiab] OR "perinatal manipulations"[tiab] OR "maternal separation"[tiab] OR "maternal deprivation"[tiab] OR "maternal care"[tiab] OR "isolation"[tiab] OR "limited bedding"[tiab] OR "limited nesting"[tiab] OR "limited material"[tiab] OR licking and grooming"[tiab] OR "licking-grooming"[tiab] OR "licking/grooming"[tiab])

AND

("murine"[tiab] OR "rodentia"[tiab] OR "rodent"[Tiab] OR "rodents"[Tiab] OR "rodentia"[tiab] OR mus[Tiab] OR murinae[Tiab] OR muridae[Tiab] OR "mice"[MeSH Terms] OR "mice"[tiab] OR "mouse"[tiab] OR "rats"[MeSH Terms] OR "rat"[tiab] OR "rats"[tiab])

AND

("Behavior, Animal"[Mesh] OR "behaviour"[tiab] OR "behavior"[tiab] OR "behaviours"[tiab] OR "behaviors"[tiab] OR "behav*"[tiab] OR "behavioural test"[tiab] OR "behavioural tests"[tiab] OR "behavioral test"[tiab] OR "behavioral tests"[tiab] OR "test, behavioral"[tiab] OR "test, behavioural"[tiab] OR "tests, behavioral"[tiab] OR "tests,behavioural"[tiab] OR "anxiety"[tiab] OR "fear"[tiab] OR "anxiety/fear"[tiab] OR "anxiety-fear"[tiab] OR "emotional learning"[tiab] OR "non-stressful learning"[tiab] OR "non stressful learning"[tiab] OR "social behaviour"[tiab] OR "social behavior"[tiab] OR "sexual behaviour"[tiab] OR "sexual behavior"[tiab] OR "radial arm"[tiab] OR "T maze"[tiab] OR "Ymaze"[tiab] OR "what where which task"[tiab] OR "what-where-which task"[tiab] OR "object in location"[tiab] OR "object in context"[tiab] OR "object recognition"[tiab] OR "object discrimination"[tiab] OR "barnes maze"[tiab] OR "holeboard"[tiab] OR "circular maze"[tiab] OR "Morris water maze"[tiab] OR "spontaneous alteration task"[tiab] OR "maze learning"[tiab] OR "active avoidance"[tiab] OR "spring test"[tiab] OR "inhibitory avoidance"[tiab] OR "passive avoidance"[tiab] OR "defensive withdrawal"[tiab] OR "fear conditioning"[tiab] OR "cat box"[tiab] OR "elevated plus maze"[tiab] OR "EPM"[tiab] OR "cross maze"[tiab] OR "open field"[tiab] OR "concentric square field test"[tiab] OR "place preference"[tiab] OR "place avoidance"[tiab] OR "light/dark test"[tiab] OR "light dark test"[tiab] OR "light-dark test"[tiab] OR "light/dark box"[tiab] OR "light dark box"[tiab] OR "light-dark box"[tiab] OR "object exploration"[tiab] OR "square field test"[tiab] OR "shuttle box"[tiab] OR "social interaction"[tiab] OR "three chambers"[tiab] OR "3 chambers"[tiab] OR "three chamber"[tiab] OR "3 chamber"[tiab] OR "1 chamber"[tiab] OR "one chamber"[tiab] OR "emotional witness stress"[tiab] OR "social play"[tiab] OR "social approach test"[tiab] OR "social encounter test"[tiab] OR "social interaction test"[tiab] OR "social preference test"[tiab] OR "social learning"[tiab] OR "social preference"[tiab] OR "social hierarchy"[tiab] OR "dominance"[tiab] OR "tube test"[tiab] OR "resident test" [tiab] OR "intruder test"[tiab] OR "resident intruder test"[tiab] OR "resident/intruder test"[tiab] OR "resident-intruder test"[tiab] OR "competitive behaviour"[tiab] OR "competitive behaviour"[tiab] OR "play fighting behaviour"[tiab] OR "play fighting behaviour"[tiab] OR "play-fighting behaviour"[tiab] OR "play-fighting behavior"[tiab] OR "play/fighting behaviour"[tiab] OR "play/fighting behavior"[tiab])

WebOfScience

"early life stress" OR "ELS" OR "early life adversity" OR "early life adversities" OR

"early life adversity*" OR "early stress" OR "neonatal stress" OR "postnatal stress" OR "perinatal stress" OR "neonatally stressed" OR "early adverse experience" OR "perinatally stressed" OR "early adverse experiences" OR "postnatal manipulation" OR "postnatal manipulations" OR "perinatal manipulation" OR "perinatal manipulations" OR "maternal separation" OR "maternal deprivation" OR "maternal care" OR "isolation" OR "limited bedding" OR "limited nesting" OR "limited material" OR "licking and grooming" OR "licking-grooming" OR "licking/grooming"

AND

"murine" OR "rodentia" OR "rodent" OR "rodents" OR "rodentia" OR mus OR murinae OR muridae OR "mice" OR "mouse" OR "rat" OR "rats"

AND

"behaviour" OR "behavior" OR "behaviours" OR "behaviors" OR "behav*" OR "behavioural test" OR "behavioural tests" OR "behavioral test" OR "behavioral tests" OR "test, behavioral" OR "test, behavioural" OR "tests, behavioral" OR "tests,behavioural" OR "anxiety" OR "fear" OR "anxiety/fear" OR "anxiety-fear" OR "emotional learning" OR "non-stressful learning" OR "non stressful learning" OR "social behaviour" OR "social behavior" OR "sexual behaviour" OR "sexual behavior" OR "radial arm" OR "T maze" OR "Ymaze" OR "what where which task" OR "what-where-which task" OR "object in location" OR "object in context" OR "object recognition" OR "object discrimination" OR "barnes maze" OR "holeboard" OR "circular maze" OR "Morris water maze" OR "spontaneous alteration task" OR "maze learning" OR "active avoidance" OR "spring test" OR "inhibitory avoidance" OR "passive avoidance" OR "defensive withdrawal" OR "fear conditioning" OR "elevated plus maze" OR "EPM" OR "cross maze" OR "open field" OR "concentric square field test" OR "place preference" OR "place avoidance" OR "light/dark test" OR "light dark test" OR "light-dark test" OR "light/dark box" OR "light dark box" OR "light-dark box" OR "object exploration" OR "square field test" OR "shuttle box" OR "social interaction" OR "three chambers" OR "3 chambers" OR "three chamber" OR "3 chamber" OR "1 chamber" OR "one chamber" OR "emotional witness stress" OR "social play" OR "social approach test" OR "social encounter test" OR "social interaction test" OR "social preference test" OR "social learning" OR "social preference" OR "social hierarchy" OR "dominance" OR "tube test" OR "resident test" OR "intruder test" OR "resident intruder test" OR "resident/intruder test" OR "resident-intruder test" OR "competitive behaviour" OR "competitive behaviour" OR "play fighting behaviour" OR "play fighting behaviour" OR "play-fighting behaviour" OR "play-fighting behavior" OR "play/fighting behaviour" OR "play/fighting behavior"

Supplementary note 2

Systematic review comment on morphology and excitability in amygdala The volume of the amygdala was reported to be significantly decreased after maternal separation in one publication (Aleksić et al., 2016). For other aspects of morphology, there were evident study-specific effects. Krugers and colleagues reported a decreased dendritic length, branching and consequently complexity in amygdalar neurons following ELA (Krugers et al., 2012). In the same laboratory, Pillai could not reproduce the changes in neuronal structure, and did not identify changes in amygdalar excitability. This was measured as the ratio between NDMA to AMPA receptor-mediated excitatory postsynaptic currents, as well as glutamate release probability. Lastly, Koe and colleagues identified an increased dendritic length and spine density in the amygdalar of animals exposed to ELA (Koe et al., 2016). Of note, the effect sizes reported are much larger than what one would expect : between 5 and 9 hedge's g rather than the common hedge's g smaller than 1 (Bonapersona et al., 2021).

Supplementary note 3

Details on meta-analytical re-analysis of the dataset. All meta-analysis were conducted using the same analytical approach. We used a three-level mixed effect model with restricted maximum likelihood estimation. This model accounts for the anticipated heterogeneity between studies, as well as the dependency of effects within experiments (Cheung, 2014). In other words, the mixed effect model is built with 1) a random effect between experiments, and 2) a fixed effect within experiments. In our experimental design, the 3 levels correspond to variance of effect size between 1) animals, 2) outcomes and 3) experiments. To test for sub-group differences, we used a Wald-type test, following the recommendations from Viechtbauer (Viechtbauer, 2010). We set the significance level alpha = 0.01 rather than 0.05. This conservative measure was chosen to decrease the probability of a Type I error, i.e. false positive findings. Therefore, we aimed to identify robust and reproducible effects of ELA on behavior and neurobiology. Absence of statistically significant results should not be interpreted as evidence that there are no effects. Throughout the text, we specify how many experiments a certain conclusion is based on to facilitate the reader in interpreting the results.

Supplementary note 4

Methodological details on exploratory analyses

We performed an exploratory analysis to investigate whether the acute situation at testing could mediate the effects of ELA on biochemical and functional outcomes in the hippocampal region and prefrontal cortex. We excluded from this analysis

outcomes related to morphology and neurogenesis, since it is unlikely that the effects of an acute situation could be observed with a short time delay on these outcomes. Furthermore, we did not include basal fEPSP since this outcome is unlikely to change in acute stress circumstances.

For details on the model used, see Supplementary Note 3. As a rule of thumb, we meta-analyzed only those outcomes reported by at least 3 independent publications. This value was chosen to provide the most detail description of the data; however, the conclusions should be interpreted with caution when only a small number of publications is available. We set our *alpha* level to account false positive findings to 0.01.

Supplementary note 5

Methodological details on meta-analyses on CVR The Coefficient of Variation Ratio (CVR) is a measure of difference in variability between the ELA and the control group (Nakagawa et al., 2015). While *Hedge's G* is a measure of difference in means, CVR is a measure of difference in variation. To meta-analyze CVR, we use the same analytical approach used for g, described in Supplementary Note 3. Briefly, for each individual comparison (i.e. a difference between the ELA and control groups) we calculated CVR. We meta-analyzed all data with a 3-level mixed effect model, where the life experiences was used as a moderator. Of note, here we categorized "life experience" rather than "additional negative life experiences (hits)" according to the cumulative stress theory. See Supplementary Table 3 for an in-depth explanation. We calculated the effect size for each subgroup (i.e. no other life events, non-stressful behavior, +1 / + 2 / +3 hits) separately. These estimates were then analyzed against 0 (alpha = 0.05) to test whether in each subgroup of life experiences, the ELA group had increased variability.

Supplementary table 1

The inclusion and exclusion criteria were specified prior to the beginning of the studies. Here below, we report inclusion exclusion criteria for the population, intervention and control group. For outcome-specific inclusion/exclusion criteria, please see each individual publication. *Table adapted from: (Bonapersona et al., 2019).*

Criteria	Comments
Inclusion	
Peer reviewed original publications in English	
Mice and rats	
ELA starts before P14	ELA model can extend after P14
ELA as alteration of maternal care(Levine, 2002) • separation of the pup from the mother (maternal separation(Sanchez et al., 2001) / deprivation(Levine, 2002)) • separation of the pup from mother and siblings (isolation) • limited bedding and nesting(Rice et al., 2008) • licking and grooming(Champagne et al., 2003)	We define as 'separation' those models in which the mother was repeatedly separated from the pups (e.g. 3 hours a day for 2 weeks). We define as deprivation those models in which the mother was separated once from the pups for a prolonged time (e.g. 1 time 24 hours, or 2 times 12 hours). In other words, the categorization in maternal separation/deprivation depends on the model used and not the naming used in the papers. The separation/deprivation/isolation models are adaptations of Levine's original model. These adaptations are based on the observation that dams often leave the nest to forage for 15-30 min periods(Leon et al., 1978). For this reason, we consider "adverse" and therefore include only those studies in which the duration of separation/deprivation/isolation time was >1h.
Exclusion	
Specific pathogen free animals	
Ovariectomized females	
Sex not specified	Publication is included if sex is retrieved after contacting the authors
Males and females pooled	Publication is included if summary statistics of males and females separately are received after contacting the authors
Fasting before behavioral test (unless part of the test itself)	
Handling, gentling and communal nesting as ELA models	
Maternal separation with early weaning(Carlyle et al., 2012)	Early weaning is defined as separation of the pups from the mother at P17. If early weaning is only in the experimental group, the experiment is excluded. If early weaning occurred in both control and experimental group, the study is included and early weaning is considered a factor that could increase vulnerability
Handling as control group	
Genetic manipulations	
Animals bred for high/low anxiety-like behavior or novelty response or sensitivity/resilience to depression	
Animals separated in high/low performance	
Administration of any drug or alcohol via any route	e.g. Drug injections before testing, methamphetamine conditioned place preference tests
Any manipulation to previous generations	

Supplementary table 2

Results on Principle #2 on acute situation.

Acute = refers to the acute situation, g = Hedge's g, sampling error, z = z value, p = p value, CI = confidence interval, # = number of, exp = experiments, comp = comparisons, 5HT = serotonin, 5HTR = serotonin receptors

Outcome	Acute	g	se	z	p	CI	# studies	# exp	# comp
Hippocampal region									
5HIAA	rest	-0.098	0.604	-0.162	0.871	-1.28, 1.085	3	3	3
5HT	rest	-0.041	0.458	-0.09	0.929	-0.938, 0.856	4	4	7
5HTR inhibitory	rest	0.159	0.59	0.27	0.787	-0.997, 1.316	3	3	8
ampa	rest	-0.519	0.402	-1.292	0.196	-1.306, 0.268	5	5	15
bdnf	stressed	0.196	0.397	0.494	0.621	-0.582, 0.973	7	8	48
bdnf	rest	-0.245	0.232	-1.054	0.292	-0.7, 0.211	19	24	74
DA	rest	0.055	0.434	0.127	0.899	-0.795, 0.905	4	5	8
GABA a Receptor	rest	0.407	0.645	0.631	0.528	-0.856, 1.67	3	3	4
LTP fepsp slope	rest	-1.46	0.649	-2.25	0.024	-2.732, -188	5	5	5
nmda	rest	-1.466	0.357	-4.107	0	-2.166, -0.767	8	8	23
vgat	rest	-1.095	0.599	-1.827	0.068	-2.269, 0.08	3	3	5
vglut	rest	-0.924	0.532	-1.735	0.083	-1.967, 0.119	3	3	8
Prefrontal cortex									
5HIAA	rest	0.136	0.591	0.229	0.819	-1.023, 1.294	3	3	3
5HT	rest	-0.682	0.399	-1.707	0.088	-1.465, 0.101	8	8	9
5HTR excitatory	rest	-0.253	0.649	-0.39	0.696	-1.524, 1.018	3	3	4
bdnf	stressed	-0.332	0.588	-0.564	0.573	-1.485, 0.821	4	4	4
bdnf	rest	0.04	0.435	0.092	0.927	-0.813, 0.893	4	4	15
D1-like receptors	rest	0.022	0.478	0.046	0.963	-0.915, 0.959	5	5	10
D2-like receptors	rest	0.249	0.481	0.518	0.604	-0.693, 1.191	5	5	7
DA	rest	0.03	0.428	0.071	0.944	-0.809, 0.87	6	7	8
DOPAC	rest	-0.216	0.495	-0.435	0.663	-1.187, 0.755	4	5	5
GAD	rest	-0.844	0.585	-1.444	0.149	-1.991, 0.302	3	4	6
HVA	rest	-0.23	0.496	-0.463	0.643	-1.202, 0.742	4	5	5
nmda	rest	-1.666	0.489	-3.406	0.001	-2.624, -707	3	3	10

Supplementary table 3

Categorization of life experiences as "additional negative life experiences", i.e. "hits".

This table has been adapted from (Bonapersona et al., 2019)e.g. depression. Whereas most human studies are limited to correlational conclusions, rodent studies can prospectively investigate how ELA alters cognitive performance in several domains. Despite the volume of reports, there is no consensus on i. Briefly, we defined in the study protocol of each meta-analysis elements that could be interpreted as "additional negative life experiences". In each meta-analysis previously conducted, we solely categorize the presence / absence of other negative life events, due to the complexity and subjectivity of the categorization. In the current manuscript, we dare to take it one step further, and we categorize stress during life as a rather continuous variable (e.g. severity), according to the cumulative stress theory. In this respect, we distinguish between animals that did not experience any other life event (i.e. are otherwise naïve), animals that performed non-stressful behavior tasks (i.e. have been handled by experimenters), and negative life experiences (i.e. cumulatively, +1, +2 or +3 hits). Animals (of both control and ELA group) were considered having experienced "additional negative life events" if they experienced at least one of elements described in the table below. Of note, within each comparison, control and ELA animals differ only in the presence/absence of ELA. Therefore, animals of both groups must have experienced an "additional negative life experience" for the comparison to be included in the study.

Considered multiple hits	Not considered multiple hits
Stressful behavioral test performed previously (e.g. FST, fear conditioning)	Intragastric saline
Footshocks	Saline injections
Chronic (mild) unpredictable stress	Vaginal smears
Chronic constant light	Daily handling by experimenter
Chronic restraint	
Chronic individual housing	
Vaginal balloon distention	
Cannula implantation, mock surgeries, blood sampling, isofluorane anaesthesia	
Dams transported pregnant	

****Note:** manipulated genetic background were excluded from the meta-analysis and therefore could not be included in the definition of vulnerability, despite it being an important factor. Furthermore, solitary housing could be considered as stressful life event, depending on sex, species and strain. In our dataset, all animals that were solitarily housed also experienced other negative life events; therefore, the categorization of housing was not necessary to define the experience of additional negative life events.

Supplementary Information for Chapter 10

Increasing the statistical power of animal experiments with historical control data.

Contents

Supplementary notes

Supplementary Note 1: Systematic review

Systematic review To identify meta-analyses on rodent primary studies, a systematic literature search was conducted on April 12th 2019 in Embase. By screening titles, the search string *(rodent* OR mice OR mouse OR rat*)* AND (*meta-analys** OR *metaanalys**) identified 170 publications, while one additional record was identified via other sources. The articles' full-texts were screened by two authors (VB and RAS) and included if it matched the pre-defined inclusion criteria (Supplementary Table 2). For a flow chart of the methodology, see Supplementary Fig. 1.

The identified meta-analytic articles (n_{ma} = 69) were used to select primary publications in mice and rats. Of all included primary publications in each meta-analysis ($n_{primary\_study}$ = 1935, "Data A" from Supplementary Fig. 1), we extracted the sample sizes of the two largest groups (equally split if only pooled quantities were reported), independent of the complexity of the experimental design, number of experiments and outcomes reported. We assumed that at *least* the comparison between these two groups would have been sufficiently powered.

Of the 69 identified meta-analyses, 8 meta-analyses matched our additional criteria for effect size estimation. These belonged to the fields of Neuroscience and Metabolism. From the resulting 482 primary studies, we extracted the summary statistics (mean, standard deviation or standard error of the mean, sample size) of all available comparisons between two independent groups, from which we calculated 2738 *Hedge's G* (n_{summ_stat}, "Data B" from Supplementary Fig. 1).

Supplementary Note 2: References of meta-analytic studies included

Identified with Embase: [1–69]
Added via other sources: [70]

1. Antoniuk, S., Bijata, M., Ponimaskin, E. & Wlodarczyk, J. Chronic unpredictable mild stress for modeling depression in rodents: Meta-analysis of model reliability. *Neuroscience and Biobehavioral Reviews* 99, 101–116 (2019).
2. Reis, D. J., Casteen, E. J. & Ilardi, S. S. The antidepressant impact of minocycline in rodents: A systematic review and meta-analysis. *Scientific Reports* 9, 1–11 (2019).
3. Zhou, L. yun et al. Riluzole promotes neurological function recovery and inhibits damage extension in rats following spinal cord injury: a meta-analysis and systematic review. *Journal of Neurochemistry* 150, 6–27 (2019).
4. Eschke, R. C. K. R. et al. Impact of physical exercise on growth and progression of cancer in rodents-a systematic review and meta-analysis. *Frontiers in Oncology* 9, (2019).
5. Skinner, J. A., Campbell, E. J., Dayas, C. V., Garg, M. L. & Burrows, T. L. The relationship between oxytocin, dietary intake and feeding: A systematic review and meta-analysis of studies in mice and rats. *Frontiers in Neuroendocrinology* 52, 65–78 (2019).
6. Qin, C. et al. Induced pluripotent stem cell transplantation improves locomotor recovery in rat models of spinal cord injury: A systematic review and meta-analysis of randomized controlled trials. *Cellular Physiology and Biochemistry* 47, 1835–1852 (2018).
7. Zhang, Y. et al. Isoquinoline alkaloids and indole alkaloids attenuate aortic atherosclerosis in apolipoprotein E deficient mice: A systematic review and meta-analysis. *Frontiers in Pharmacology* 9, (2018).
8. Qin, C. et al. Myelotomy promotes locomotor recovery in rats subjected to spinal cord injury: A meta-analysis of six randomized controlled trials. *Neural Regeneration Research* 13, 1096–1106 (2018).
9. He, F., Zhou, A. & Feng, S. Use of human amniotic epithelial cells in mouse models of bleomycin-induced lung fibrosis: A systematic review and meta-analysis. *PLoS ONE* 13, 1–17 (2018).
10. Jiao, N. et al. Gut microbiome may contribute to insulin resistance and systemic inflammation in obese rodents: A meta-analysis. *Physiological Genomics* 50, 244–254 (2018).
11. Zhang, Y., Ge, M., Hao, Q. & Dong, B. Induced pluripotent stem cells in rat models of Parkinson's disease: A systematic review and meta-analysis. *Biomedical Reports* 8, 289–296 (2018).
12. Chang, J. li et al. Chinese Herbal Medicine for Osteosarcoma in the Mouse: A Systematic Review and Meta-Analysis. *Chinese Journal of Integrative Medicine* 25, 370–377 (2019).
13. Kara, N. Z., Stukalin, Y. & Einat, H. Revisiting the validity of the mouse forced swim test: Systematic review and meta-analysis of the effects of prototypic antidepressants. *Neuroscience and Biobehavioral Reviews* 84, 1–11 (2018).
14. Wang, C. et al. The Variable Regulatory Effect of Arsenic on Nrf2 Signaling Pathway in Mouse: a Systematic Review and Meta-analysis. *Biological Trace Element Research* 190, 362–383 (2019).
15. Wang, W. et al. Down-regulated microRNA-34a expression as a prognostic marker for poor osteosarcoma in mice: A systematic review and meta-analysis. *Journal of Cancer* 9, 4179–4186 (2018).
16. Abunab, H., Dator, W. L. & Hawamdeh, S. Effect of olive leaf extract on glucose levels in diabetes-induced rats: A systematic review and meta-analysis. *Journal of Diabetes* 9, 947–957 (2017).
17. Viswanatha, G. L., Shylaja, H. & Moolemath, Y. The beneficial role of Naringin- a citrus bioflavonoid, against oxidative stress-induced neurobehavioral disorders and cognitive dysfunction in rodents: A systematic review and meta-analysis. *Biomedicine and Pharmacotherapy* 94, 909–929 (2017).
18. Cai, W. et al. Effects of glyphosate exposure on sperm concentration in rodents: A systematic review and meta-analysis. *Environmental Toxicology and Pharmacology* 55, 148–155 (2017).
19. Wassenaar, P. N. H., Trasande, L. & Legler, J. Systematic review and meta-analysis of early-life exposure to bisphenol a and obesity-related outcomes in rodents. *Environmental Health Perspectives* 125, (2017).
20. Hou, Q. liang et al. Adipose tissue-derived stem cell therapy for erectile dysfunction in rats: a systematic review and meta-analysis. *International Urology and Nephrology* 49, 1127–1137 (2017).
21. Phie, J. et al. Flavonols reduce aortic atherosclerosis lesion area in apolipoprotein E deficient mice: A systematic review and meta-analysis. *PLoS ONE* 12, 1–13 (2017).
22. Barha, C. K., Falck, R. S., Davis, J. C., Nagamatsu, L. S. & Liu-Ambrose, T. Sex differences in aerobic exercise efficacy to improve cognition: A systematic review and meta-analysis of studies in older rodents. F*rontiers in Neuroendocrinology* 46, 86–105 (2017).
23. Lang, C. I. et al. Cardiac cell therapies for the treatment of acute myocardial infarction: A meta-analysis from mouse studies. *Cellular Physiology and Biochemistry* 42, 254–268 (2017).
24. Xu, M. et al. Oxidative Damage Induced by Arsenic in Mice or Rats: A Systematic Review and Meta-Analysis. *Biological Trace Element Research* 176, 154–175 (2017).
25. Zhang, J. X. et al. Hemorheology Index Changes in a Rat Acute Blood Stasis Model: a Systematic Review and Meta-Analysis. African journal of traditional, complementary, and alternative medicines : *AJTCAM* 14, 96–107 (2017).

26. Wassenaar, P. N. H. & Legler, J. Systematic review and meta-analysis of early life exposure to di(2-ethylhexyl) phthalate and obesity related outcomes in rodents. *Chemosphere* 188, 174–181 (2017).
27. Hu, S. et al. Long non-coding RNAs for osteosarcoma in the mouse: A metaanalysis. *Oncotarget* 8, 100533–100544 (2017).
28. Lu, Y. P., Tsuprykov, O., Vignon-Zellweger, N., Heiden, S. & Hocher, B. Global Overexpression of ET-1 Decreases Blood Pressure - A Systematic Review and Meta-Analysis of ET-1 Transgenic Mice. *Kidney and Blood Pressure Research* 41, 770–780 (2016).
29. Besson, A. A., Lagisz, M., Senior, A. M., Hector, K. L. & Nakagawa, S. Effect of maternal diet on offspring coping styles in rodents: a systematic review and meta-analysis. *Biological Reviews* 91, 1065–1080 (2016).
30. Ben-Eliezer, D. & Yechiam, E. Hypericum perforatum as a cognitive enhancer in rodents: A meta-analysis. *Scientific Reports* 6, 1–8 (2016).
31. Xu, J. et al. Bone mass improved effect of icariin for postmenopausal osteoporosis in ovariectomy-induced rats. *Menopause* 23, 1152–1157 (2016).
32. Toop, C. R. & Gentili, S. Fructose beverage consumption induces a metabolic syndrome phenotype in the rat: A systematic review and meta-analysis. *Nutrients* 8, (2016).
33. Klement, R. J., Champ, C. E., Otto, C. & Kämmerer, U. Anti-tumor effects of ketogenic diets in mice: A meta-analysis. *PLoS ONE* 11, 1–16 (2016).
34. Li, M., Li, H., Ruan, Y., Wang, T. & Liu, J. Stem cell therapy for diabetic erectile dysfunction in rats: A meta-analysis. *PLoS ONE* 11, 1–14 (2016).
35. ZHANG, G. et al. Extracellular vesicles derived from mesenchymal stromal cells may possess increased therapeutic potential for acute kidney injury compared with conditioned medium in rodent models: A meta-analysis. *Experimental and Therapeutic Medicine* 11, 1519–1525 (2016).
36. Deminice, R. & Rosa, F. T. Creatine supplementation decreased homocysteine plasma levels in rats but not humans: A critical review with meta-analysis. Journal of Nutrition and Intermediary Metabolism 3, 50–57 (2016).
37. Chang, J. et al. MicroRNAs for osteosarcoma in the mouse: a meta-analysis. *Oncotarget* 7, 85650–85674 (2016).
38. Yang, L. et al. Schwann Cells Transplantation Improves Locomotor Recovery in Rat Models with Spinal Cord Injury: a Systematic Review and Meta-Analysis. *Cellular Physiology and Biochemistry* 37, 2171–2182 (2015).
39. Xu, X. et al. Beneficial Effects of Caloric Restriction on Chronic Kidney Disease in Rodent Models: A Meta-Analysis and Systematic Review. *PLOS ONE* 10, e0144442 (2015).
40. Lagisz, M. et al. Little appetite for obesity: meta-analysis of the effects of maternal obesogenic diets on offspring food intake and body mass in rodents. *International Journal of Obesity* 39, 1669–1678 (2015).
41. YANG, H. et al. Meta-analysis on protective effect of electroacupuncture at "Nèiguān" (PC 6) in rats with reperfusion injury induced myocardial ischemia. *World Journal of Acupuncture - Moxibustion* 25, 43–53 (2015).
42. Durg, S., Dhadde, S. B., Vandal, R., Shivakumar, B. S. & Charan, C. S. Withania somnifera (Ashwagandha) in neurobehavioural disorders induced by brain oxidative stress in rodents: a systematic review and meta-analysis. *Journal of Pharmacy and Pharmacology* 67, 879–899 (2015).
43. Rui, Z. The Protective Effect of Puerarin on Myocardial Infarction Reperfusion Injury (MIRI): A Meta-Analysis of Randomized Studies in Rat Models. *Medical Science Monitor* 21, 1700–1706 (2015).
44. Khatib, M. N. et al. Effect of Ghrelin on Mortality and Cardiovascular Outcomes in Experimental Rat and Mice Models of Heart Failure: A Systematic Review and Meta-Analysis. *PLOS ONE* 10, e0126697 (2015).
45. Shan, H. et al. Stem Cell Therapy for Erectile Dysfunction of Cavernous Nerve Injury Rats: A Systematic Review and Meta-Analysis. *PLOS ONE* 10, e0121428 (2015).
46. Yao, M. et al. Neurological Recovery and Antioxidant Effects of Curcumin for Spinal Cord Injury in the Rat: A Network Meta-Analysis and Systematic Review. *Journal of Neurotrauma* 32, 381–391 (2015).
47. Cao, X.-F. et al. Effect of glutamine enriched nutrition support on gut barrier of rats with obstructive jaundice: A Meta-analysis. *World Chinese Journal of Digestology* 23, 1500 (2015).
48. Dzhambov, A. & Dimitrova, D. Chronic noise exposure and testosterone deficiency — meta-analysis and meta-regression of experimental studies in rodents. *Endokrynologia Polska* 66, 39–46 (2015).
49. Lou, J., Xie, T.-Y. & Chen, X.-Y. Meta analysis of VEGF periodic expression in intraocular tissue of diabetic rats. *International eye science* (2015) doi:10.3980/j.issn.1672-5123.2015.10.19.
50. Montgomery, S. E., Sepehry, A. A., Wangsgaard, J. D. & Koenig, J. E. The Effect of S-Adenosylmethionine on Cognitive Performance in Mice: An Animal Model Meta-Analysis. *PLoS ONE* 9, e107756 (2014).
51. Takase, K. et al. Meta-Analysis of Melanin-Concentrating Hormone Signaling-Deficient Mice on Behavioral and Metabolic Phenotypes. *PLoS ONE* 9, e99961 (2014).
52. Zhang, D. & He, X. A Meta-Analysis of the Motion Function through the Therapy of Spinal Cord Injury with Intravenous Transplantation of Bone Marrow Mesenchymal Stem Cells in Rats. *PLoS ONE* 9, e93487 (2014).
53. Wu, W., Geng, H., Liu, Z., Li, H. & Zhu, Z. Effect of curcumin on rats/mice with diabetic nephropathy: A systematic review and Meta-analysis of randomized controlled trials. *Journal of Traditional Chinese Medicine* 34, 419–429 (2014).
54. Xiang, L. et al. Meta analysis of olfactory ensheathing cell transplantation promoting functional recovery of motor nerves in rats with complete spinal cord transection. *Neural Regeneration Research* 9, 1850 (2014).
55. Yu, M., Yang, Z., Zhu, Y. & Lu, N. Efficacy of glucocorticoids in rodents of severe acute pancreatitis: a meta-analysis. *International journal of clinical and experimental pathology* 7, 3647–61 (2014).

56. Oliveri, R. S., Bello, S. & Biering-Sørensen, F. Mesenchymal stem cells improve locomotor recovery in traumatic spinal cord injury: Systematic review with meta-analyses of rat models. *Neurobiology of Disease* 62, 338–353 (2014).
57. Fu, S. et al. Systematic review and meta-analysis of the bone protective effect of phytoestrogens on osteoporosis in ovariectomized rats. *Nutrition Research* 34, 467–477 (2014).
58. Todd, T. E. et al. Meta-analysis of variables affecting mouse protection efficacy of whole organism Brucella vaccines and vaccine candidates. *BMC Bioinformatics* 14, S3 (2013).
59. Sun, G. et al. Adoptive infusion of tolerance dendritic cells prolongs survival of small intestine allografts in rats: systematic review and meta-analysis. *Journal of Evidence-Based Medicine* 6, 185–196 (2013).
60. Simons, M. J. P., Koch, W. & Verhulst, S. Dietary restriction of rodents decreases aging rate without affecting initial mortality rate - a meta-analysis. *Aging Cell* 12, 410–414 (2013).
61. Swindell, W. R. Dietary restriction in rats and mice: A meta-analysis and review of the evidence for genotype-dependent effects on lifespan. *Ageing Research Reviews* 11, 254–270 (2012).
62. Wheaton, P., Mathias, J. & Vink, R. Impact of pharmacological treatments on outcome in adult rodents after traumatic brain injury: a meta-analysis. *Journal of Psychopharmacology* 25, 1581–1599 (2011).
63. Knowles, S. C. L. The effect of helminth co-infection on malaria in mice: A meta-analysis. *International Journal for Parasitology* 41, 1041–1051 (2011).
64. Schlüter, K.-D., Schreckenberg, R. & da Costa Rebelo, R. M. Interaction between exercise and hypertension in spontaneously hypertensive rats: a meta-analysis of experimental studies. *Hypertension Research* 33, 1155–1161 (2010).
65. Kook, S., Kim, G.-H. & Choi, K. The Antidiabetic Effect of Onion and Garlic in Experimental Diabetic Rats: Meta-Analysis. *Journal of Medicinal Food* 12, 552–560 (2009).
66. Jonasson, Z. Meta-analysis of sex differences in rodent models of learning and memory: a review of behavioral and biological data. *Neuroscience & Biobehavioral Reviews* 28, 811–825 (2005).
67. Dirx, M. J. M., Zeegers, M. P. A., Dagnelie, P. C., van den Bogaard, T. & van den Brandt, P. A. Energy restriction and the risk of spontaneous mammary tumors in mice: A meta-analysis. *International Journal of Cancer* 106, 766–770 (2003).
68. Brown, K. G. & Strickland, J. A. Utilizing data from multiple studies (meta-analysis) to determine effective dose–duration levels. Example: rats and mice exposed to hydrogen sulfide. *Regulatory Toxicology and Pharmacology* 37, 305–317 (2003).
69. Bonapersona, V., Joels, M. & Sarabdjitsingh, R. A. Effects of early life stress on biochemical indicators of the dopaminergic system: A 3 level meta-analysis of rodent studies. *Neuroscience and biobehavioral reviews* 95, 1–16 (2018).
70. Bonapersona, V. et al. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neuroscience and Biobehavioral Reviews* 102, 299–307 (2019).

Supplementary Note 3: The math of priors

Studying whether two numbers (A and B) are distinct is equivalent to investigating whether their difference (A – B) is different from 0. To calculate the difference distribution, we estimated the populations from the respective samples by using (un) informative priors.

In the control group

$$y_i \sim N(\mu_{con}, \sigma^2_{con}) \quad (1)$$

where y_i denotes the score on the outcome variable in the control group for $i=1,\ldots,n_{con}$ animals. Similarly, in the experimental group

$$y_i \sim N(\mu_{exp}, \sigma^2_{exp})_! \quad (2)$$

S

for i=3,...,n_{exp}. Dropping the subscripts *con* and *exp*, the posterior distribution of μ and σ^2 in the control and experimental groups is given by (*section 3.2 and 3.3* of Gelman, A. *et al.* Bayesian data analysis. Chapman & Hall, 1995.)

$$g(\mu, \sigma^2) =$$

$$g(\mu|\sigma^2)\, g(\sigma^2) = N\left(\mu \middle| m_{post}, \frac{\sigma^2}{n_{post}}\right) Inv - \chi^2\left(\sigma^2 | n_{post}, \sigma^2_{post}\right) \qquad (3)$$

where the posterior mean is

$$m_{post} = \frac{n_{prior}}{n_{prior}+n} m_{prior} + \frac{n}{n_{prior}+n} \bar{y} \qquad (4)$$

the posterior variance is

$$\sigma^2_{post} = (n_{prior}\sigma^2_{prior} + (n-1)s^2 + \frac{n_{prior}n}{n_{prior}+n}\left(\bar{y} - m_{prior}\right)^2)/n_{post} \qquad (5)$$

and the posterior degrees of freedom is

$$n_{post} = n_{prior} + n \qquad (6)$$

For the experimental group, the posterior distribution is based on an uninformative prior distribution, that is, the prior sample size n_{prior}= 0. For the control group, an informative prior distribution based on a previous study is used, where n_{prior} denotes the sample size of the previous study, m_{prior} denotes the mean of the scores on the outcome variable in the previous study and s^2_{prior} the variance. In fact, the posterior distribution for the control group is based on p=1,...,P prior studies. Bayesian updating is used to obtain the posterior distribution for the control group:

- **Step 1.** Use equation (3) to update an uninformative prior distribution (n_{prior}= 0) with the data from the first prior study p = 1 weighted with n_{prior}= n_1 *index*$_1$, where n_1 denotes the sample size of the first prior study, resulting in a posterior distribution.
- **Step 2.** This posterior distribution becomes the current prior distribution.
- **Step 3.** For p=2,...,P, that is, the remaining prior studies, use equation (3) to update the current prior distribution with the data from prior study p weighted with n_{prior}= n_p *index*$_p$, where n_p denotes the sample size of the pth prior study,

resulting in a posterior distribution that will have the role of prior distribution in the next step.

- **Step 4.** In this last step, using equation (3) the prior distribution resulting from Steps 1 through 3 is updated with all the data from the current study, rendering the posterior distribution in the control group.

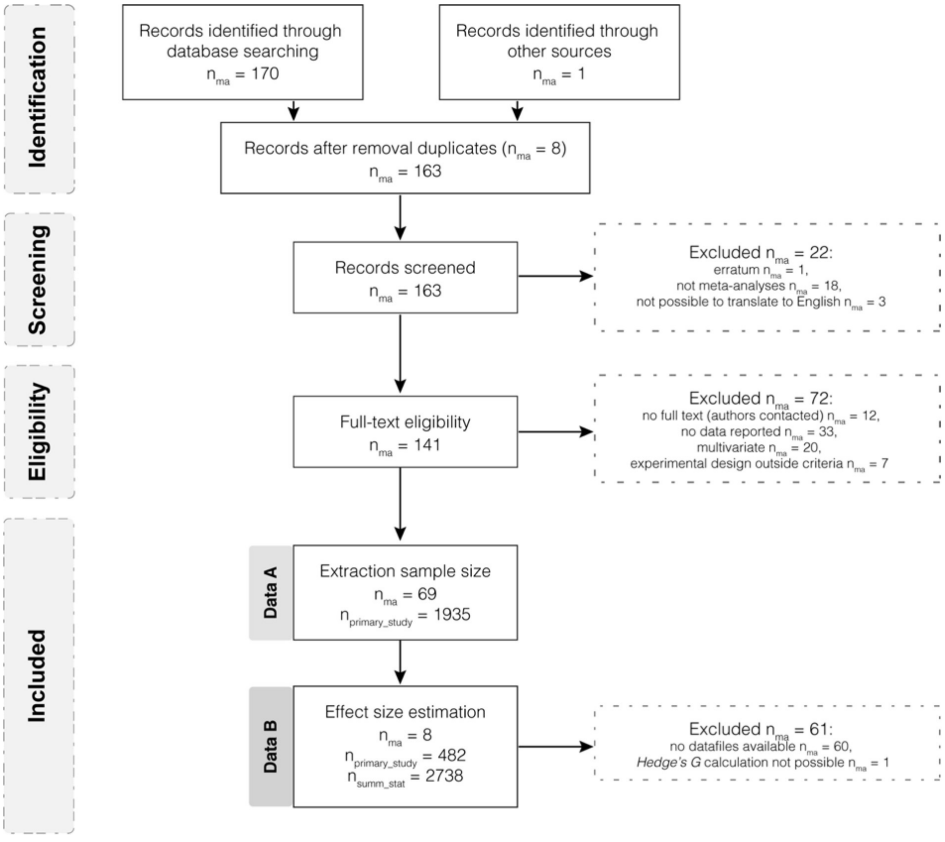
The confidence interval for μ_{exp}-μ_{con} is obtained by sampling t=1,..,10000 values μ^t_{exp} and μ^t_{con} from the respective posterior distributions and computing their difference $\delta^t = \mu^t_{exp} - \mu^t_{con}$. The 2.5th and 97.5th percentile of the distribution of δ^t for t=1,...,T constitute respectively the lower and upper bound of the 95% confidence interval for $\mu^t_{exp} - \mu^t_{con}$. When the value 0 is/is not contained in the confidence interval, the null-hypothesis that the experimental and control means are equal is not/is rejected.

Supplementary Note 4: Establishing the RELACS consortium

The consortium consists of laboratories working on the effects of ELA on neutral memory, identified with a systematic literature search[70]. To this we added unpublished data that we gathered through our own network; we also asked the laboratories identified through the search if they had any unpublished data available. Next, each lab received an excel document, in which the relevant information could be imported. In cases where the information provided was unclear to us, authors were contacted for clarification. Of note, the use of data collected through this consortium is not limited to this particular publication and in fact the database encompasses more information than presently used; therefore, we specified in the current paper the additional inclusion/exclusion criteria (Supplementary Table 4) that we applied to come to the present data set. The criteria were specified a priori and blindly to the outcomes and the laboratories. Issues such as availability of the data to external parties (particularly when data are still unpublished) and authorships were addressed prior to data gathering in a formal consortium agreement.

Supplementary Figures

Supplementary Figure 1



Flowchart methodology for data collection. From the EMBASE systematic literature search, 69 meta-analyses met our pre-specified inclusion criteria (Supplementary Table 2). From the 1935 articles used in these meta-analyses, we extracted the sample size of the two largest groups ("Data A"), which was used for theoretical power calculations. 8 meta-analyses met our additional inclusion criteria (Supplementary Table 2, "additional inclusion criteria"). From the 482 primary publications used in these meta-analyses, we extracted the available effect sizes, for a total of 2738 ("Data B"), which were used for calculations of achieved power (Fig.1a) and range of effect sizes (Fig. 1b). n_{ma} = number of meta-analyses; $n_{primary\_study}$ = number of unique publications in mice and rats; n_{summ_stat} = number of summary statistics (mean, SD or SEM, sample size) extracted; Data A = data extracted at this level from dataset A; Data B = data extracted at this level from dataset B. All data is available at https://osf.io/wvs7m/.

Supplementary Tables

Supplementary Table 1

Summary of analyses to verify performance of *RePAIR* on RELACS dataset. This summary should be interpreted in relation to the results shown in Fig. 2b.

Fig.2b	Aim	Experiment	Prior
t-test	RePAIR can achieve the same conclusion as a Welch t-test	RELACS	Non informative
underpower	When n_{con} is decreased, the test is no longer significant	ELA group from RELACS; control group 30% of RELACS (randomly selected)	
Literature Prior RePAIR	Prior of literature can substitute prior from same dataset	Same as underpower	From literature as selected by VB

Supplementary Table 2

Inclusion and exclusion criteria for systematic literature search to identify relevant meta-analyses perform on data from mice and rats.

Inclusion	
Meta-analyses of literature	
Mice and rats as population investigated	
Comparison between (at least) two independent groups	
Additional inclusion criteria	Only for range effect size estimation Fig. 1b
Data available as data file	
Hedge's G can be calculated with the available information	
Exclusion	
Language not English or not translatable with google translate	
Multivariate analyses (e.g. gene expression)	

Supplementary Table 3

Factors varied in the simulation study.

Factors	Values and interpretation	Comments
	0 = noninformative 10 = pilot from own lab 20 = experiment from own lab 50 = routinely performed in own lab 100 = from literature or with data from other labs 200 = common outcome across labs	Values arbitrarily selected to simulate real-life situations
Effect sizes	*Hedge's G* 0.2 = small 0.5 = medium 0.9 = large	See Fig. 1b
Control population parameters	Standardized values: $\mu_{con} = 0$ $\sigma^2_{con} = 1$	
Experimental population parameters	For mean: Calculated from *Hedge's G* definition $$Hedge's\ G = \frac{\mu_{exp} - \mu_{con}}{\sigma^*_{pooled}}$$ Where $$\sigma^*_{pooled} = \frac{\sigma^2_{exp} + \sigma^2_{con}}{2}$$ Therefore, $$\mu_{exp} = \mu_{con} + (Hedge's\ G^* \sigma^*_{pooled})$$	Of note, μ and to the σ^2 population parameters.
	For variance: Same variance as control, $\sigma^2_{exp} = \sigma^2_{con}$ Larger variance than control, $\sigma^2_{exp} = 2^*\sigma^2_{con}$	For results of larger variance, see R script. [1]
Index	$index = 1$	Not varied.[2]

[1] Experimental groups are often more variable than controls; therefore, we performed the simulation also with unequal variances (equal variances are not assumed). Since the results of same and larger variances are extremely comparable, only same variances are reported in text. For larger variances, please see the R script.

[2] In the simulation, the index was not varied as it would simply decrease nprior. For example, $n_{prior} = 50$ is equivalent to $n_{prior} = 100$ with an index = 0.5.

Supplementary Table 4

Inclusion/exclusion criteria for selection for RELACS dataset.

Inclusion	Comments
Male mice	Insufficient data on females, males more frequently used
Adult	Older than 8 weeks of age but younger than 1 year
LBN as ELA model	Amount of bedding material of ELA animals could be both ¼ and ½ of controls.
Object in location task performed, with available exploratory time of both objects	Necessary to operationalize memory of each animal as: $discrimination = \frac{time_{novel}}{time_{old} + time_{novel}}$ Where *time* refers to the time spent exploring an object in either the novel or old location.
Animals habituated to test cage prior to the learning phase	To avoid novelty-induced stress effects on memory
Exploration time of each object > 0s	Exploration of both objects is present
Preference of objects/locations avoided	Objects and locations were experimentally balanced or no preference was observed in previous experiments
Retention time between learning and test phase at least 1 hour	Working memory excluded
Experiments were performed and analyzed blindly and randomly	To ensure good experimental quality

Exclusion	Comments
Metal grid not used in the LBN model	
Animals not habituated to test cage prior to the learning phase	To avoid novelty-induced stress effects on memory
Control group is unable to discriminate the novel location	Discrimination index unequal to 50% at a group level to exclude possible problems in the set-up of the experiment

UMC Utrecht

Universiteit Utrecht

ISBN 978-90-393-7494-8