Multidisciplinary assessment of cumulative experience in laboratory rhesus macaques

Janire Castellano Bueno

A thesis submitted for the degree of

Doctor of Philosophy

Faculty of Medical Sciences

Institute for Biosciences

Newcastle University

May 2022

Abstract

Rhesus macaques (*Macaca mulatta*) are widely used in biomedical research due to their phylogenetic proximity to humans and their ability to perform complex cognitive tasks. Optimising their welfare is essential from both an ethical and a scientific perspective. However, welfare concern due to their extensive stays at research facilities is increasing and led to the introduction of the concept of cumulative experience (defined as the net impact of all events that affect adversely, positively and by way of amelioration, the welfare of the animal over its lifetime). The importance of assessing cumulative experience in laboratory macaques has been highlighted in research regulation.

To address this need, my thesis aimed to assess the cumulative experience of laboratory rhesus macaques involved in neuroscience experiments and to identify factors that could be impacting it. To accomplish this, I used new, validated behavioural and neuroimaging indicators of cumulative experience, as well as currently used indicators of welfare based on body weight and alopecia. I refined the use of the behaviour *Inactive not alert* to maximise its sensitivity and specificity as a marker of cumulative experience by testing different durations as well as estimating the number of sessions required to assess reliably the frequency of the behaviour.

Neither the neuroimaging indicator nor the weight measures showed any signs of cumulative experience in the study subjects. However, I found an increase in alopecia and the frequency of *Inactive not alert* behaviour over time. The behavioural data suggest that this latter effect is likely to be caused by experimental procedures. However, the effect is small, compared to other factors known to have a detrimental impact on the welfare of macaques.

This is the first study revealing a cumulative negative effect of experimental procedures on laboratory macaques. Implications of these results and next steps are discussed.

Acknowledgements

These past four years have been both incredibly challenging and rewarding. I have learned, evolved and found many things about myself that I would have never found out otherwise. None of it would have been possible without a number of incredible people that stayed by my side across the journey. First, I would like to take the opportunity to thank the immense support of my supervisors Colline Poirier and Melissa Bateson. Thank you so much for the guidance, the teachings, and the patience over these past years. A special huge thanks to Colline, for all the help with the many drafts, the long and insightful discussions over the years and particularly for granting me this opportunity.

This research would have not been possible without Alexander Thiele, Christopher Petkov and Michael Schmid who kindly granted us access to their animals and data, as well as the Barbour Foundation that funded this research. I also would like to thank all the CBC staff at Newcastle, that made this research possible, especially Emma Woolgar and Ahsley Waddle for kind replies to abundant emails. From the Centre for Macaques, I would like to thank Claire Witham for contributing to this project even with my very last-minute requests.

This project would not be half what it now is if it wasn't for all the students and colleagues that helped me with the hundreds of hours of video encoding over the years. A special mention to Nathan and Chris for not only the help with the data but also the laughs and good moments shared.

I huge shoutout to all the people that made the *good old* IoN and the fresh new HWB such a welcoming and inspiring environment: Maria, Holly, Jen, Ben, Courtney, Matthew, Rachael, Alex… and many more. Thank you, I will miss you terribly. Beshoy, for the technical, emotional, and nutritional support, shokran.

No puedo concluir este trabajo sin dar reconocimiento al inmenso apoyo de mi ama, mi aita, mis hermanos, mis tíos, y todos mis amigos desde la lejanía. Completar una tesis doctoral nunca es sencillo, pero hacerlo lejos de casa y durante la situación que nos ha tocado vivir añade una capa de dificultad que no me imaginaba. No podría haberlo hecho sin vosotros. Eskerrik asko.

Finally, to the monkeys.

Table of contents

Abbreviations

%	Percentage
Δ	Delta (difference)
AIC	Akaike's information criterion
ANOVA	Analysis of variance
ß	Estimate
CfM	Centre for macaques
DSTL	The Defence Science and Technology Laboratory
EU	European union
Kg	Kilogram
n	Sample size
m	Metre
mg	Milligram
min	Minutes
ml	Millilitres
MRC	Medical Research Council
MRI	Magnetic resonance Imaging
NcL	Newcastle
NHP	Non-human primate
P	P-value
SD	Standard deviation
SE	Standard error
T	Tesla (unit of measurement to define the magnetic flux density)
UK	United Kingdom
USA	United States of America
VBM	Voxel Base Morphometry
yo	Years old

List of figures

List of tables

Chapter 1: General introduction

1.1. Importance of rhesus macaques in biomedical research

Humans have been relying on the use of animals since the very beginning of society, from their use in agriculture, clothing and leisure, to transport and medicine. The development of industry and technology has completely changed how humans live, but the dependency on animals has remained. Our society uses animals for a multitude of services almost as much as ancient societies did, however, the public opinion on their use has evolved since then, and animals' extensive use has led to concerns about their wellbeing. One of the fields that raises the most concerns is the use of animals, as models of human beings, in research.

Concerns about the welfare of laboratory animals have come a long way since some of the first reported animal experiments started in the 17th century (Weatherall, 2006). The evaluation of the ethical issues concerning animal experimentation used to be delegated to theologians and philosophers, outside science and far from public interest (Mellor & Reid, 1994). However, the balance between the harm and the benefit of animal experimentation has always been questioned, and the concern has been slowly growing. In the last few decades, there has been a huge increase in the general public concern about the welfare of laboratory animals which is reflected in the recent changes in the legislation that regulates their use (Animal Welfare Act, 2006; Ares, 2022; Directive 2010/63/EU). This increase is a result of the growing understanding, from both the general public and the scientific community, of animal emotions and their ability to suffer (Webster, 1995).

Among the different species used in scientific experiments, the biggest public and scientific concern is related to the welfare of non-human primates (NHPs) (Bateson, 2011). The number of NHPs used in research is difficult to calculate since some subjects are re-used, used for breeding purposes, or involved in projects that are not published (Carlsson et al., 2004). Carlsson and colleagues (2004) estimated that the annual worldwide use of non-human primates is between 100,000 and 200,000, which would represent less than 0.1% of all the animals used in research. NHPs involved in research include Prosimians, New World and Old-World monkeys, with the latter being the most used worldwide (65%) (Pickard, 2013) (Figure 1). In the United Kingdom (UK), the most common species are macaques (*Macaca mulatta* and *Macaca cynomolgus*), and common marmosets (*Callithrix jacchus*); Prosimians and baboons (*Papio sp*), have not been used in the UK since 1991 and 1998, respectively (Weatherall, 2006). Moreover, the use of great apes is banned in the UK. This ban was

announced in November 1997 and followed the argument that the greater susceptibility to suffering of great apes compared to other NHPs is not compensated for by the added benefit of having models closer to humans. For other primates, their susceptibility to suffering is considered to be outweighed by the added benefit of having a model closer to humans compared to non-primate models. According to the Annual Statistics of Scientific Procedures on Living Animals Great Britain (Home Office, 2020) in 2020, the total number of procedures carried out in NHPs in the UK was the lowest since 2008 with a total of 2393 procedures. Most of the NHPs are used in microbiology studies (26%), followed by neuroscience (19%), biochemistry (12%) and pharmacology (11%) (Carlsson et al., 2004).

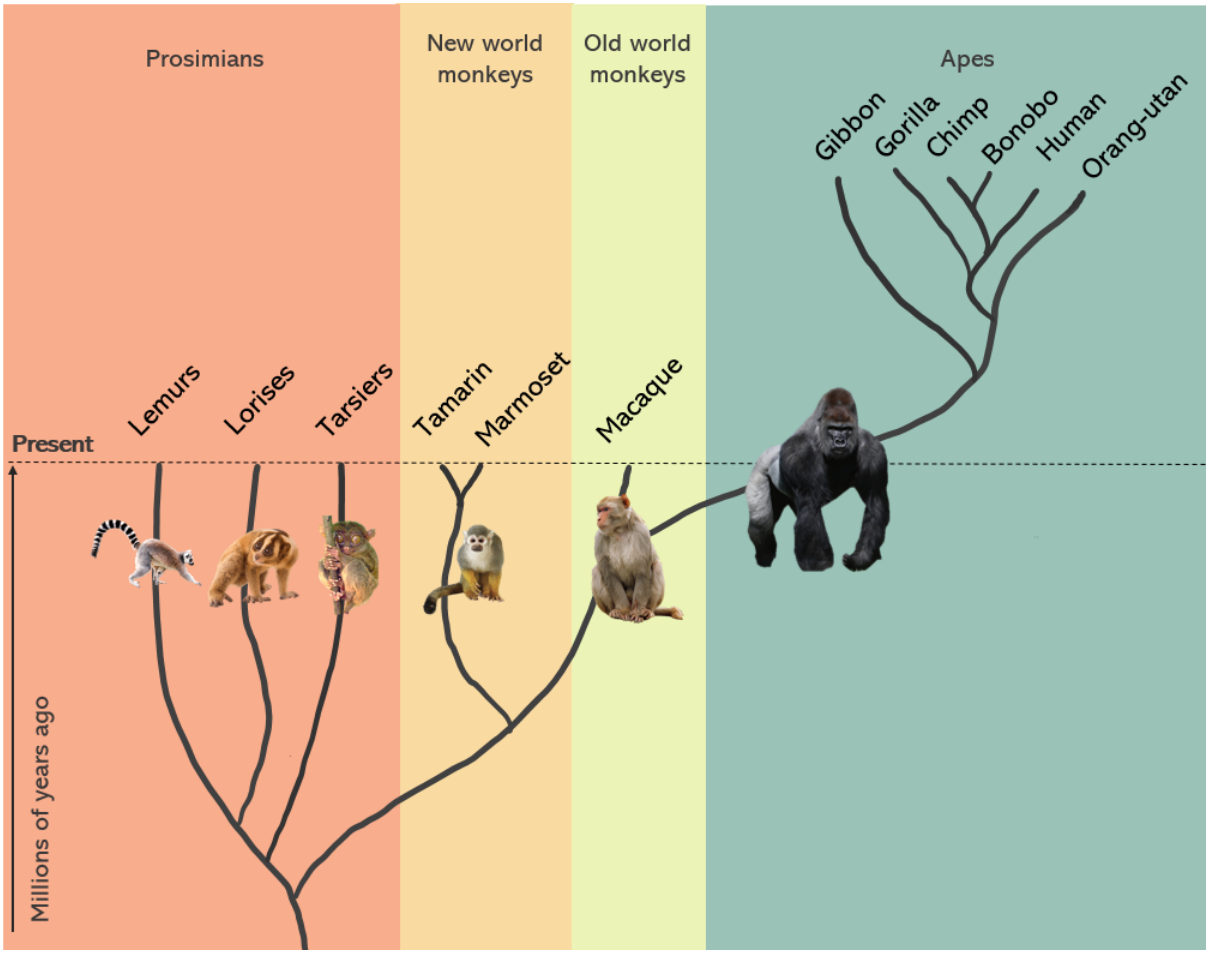

Figure 1. Simplified family tree diagram of primate evolution.

Although the number of NHPs used in research is low compared with other commonly used animal models, the concern about their welfare is significantly higher due to their evolutionary proximity to humans. This proximity is key for some biomedical research; the similarities in physiology, neuroanatomy, reproduction, development, cognition and social abilities with humans allow advances in crucial research fields. But this proximity also suggests high

sentience, self-awareness, and human-like suffering (Hau & Schapiro, 2006; Summerhoff, 1990), which raises more ethical concerns than with most other animal species (Pickard, 2013).

Consequently, installing stricter regulations on the use of NHPs that will decrease their use and potentially increase their welfare in the EU and UK has been recently studied (SCHEER, 2017). However, implementing stricter regulations is recognised to endanger the welfare of the research animals, as it conceivably prompts a transfer of the research with NHPs to countries with lower animal welfare standards. Therefore, stricter restrictions in the regulation of animal model use have been discarded for the near future.

Accordingly, the scientific case for the use of NHPs in the UK has been thoroughly studied and a need for a regular update on potential alternatives was reflected in the Directive 2010/63/EU. The alternatives to NHP models considered are human studies, *in vitro* methods (i.e., organ in a chip), *in silico* techniques (i.e., computer modelling) and the use of other animal models considered to entail fewer welfare concerns (Weatherall, 2006). In the specific case of neuroscience research, NHPs are used when other animals lack the brain structures or neurotransmitter systems being investigated, or their visual or motor abilities, functional connectivity or cognitive and behavioural abilities are too dissimilar from those of humans (SCHEER, 2017). One of the most promising alternatives to NHPs in neuroscience is the use of computational models; these are still in early stages of development and are currently being improved with the use of animal studies. As for *in vitro* alternatives, researchers have been studying the use of brain organoids which has proved to be useful for the study of disease modelling, transcriptomics, drug screening, brain evolution and electrophysiological studies (Weatherall, 2006). Although the field is progressing rapidly, these organoids are still far from simulating the complexities of the human brain (Chiaradia & Lancaster, 2020). Moreover, on top of the current methodological limitations, ethical concerns will need to be considered if the organoids were to be developed to reach a close resemblance to the human brain.

In conclusion, there have been major and promising advances in the development of non-invasive approaches to study many areas of medicine without the need for NHPs, but all these methods are still in the early stages of development. Therefore, at least in the near future, the use of NHPs in research is the only approach possible for addressing certain crucial medical and scientific challenges. As such, the animal welfare community's focus must be on ensuring the best possible welfare for all the non-human primates used in research (Weatherall, 2006).

Late reports also highlight the acceptance of the general public of using animals for biomedical research when the so-called 3Rs principle is followed, and with a full replacement of the NHPs as a future aim (SCHEER, 2017). The 3Rs stand for Replace, Reduce, Refine and represent more humane animal research. The framework was developed over 50 years ago and, in the UK, the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) is the national organization responsible for advancing the 3Rs.

As there are currently no alternatives to replace NHPs in many areas of neuroscientific research, the remaining 2Rs (i.e., Reduction and Refinement) are critical (Bonini, 2019). Reduction is already fully implemented around the UK and EU countries, with many studies using two or three animals (SCHEER, 2017). Therefore, the focus should be on refinement. Husbandry, experimental procedures, and data collection technologies are the main areas highlighted in need of refinement (Bonini, 2019).

On top of the 3Rs principle and the strict legislation assuring that NHPs are only used when no other method is available, the use of these animals in biomedical research is also regulated with a "harm-benefit assessment". This is a case-by-case approach based on a utilitarian argument of balancing harm versus benefit. The main justification follows the idea that the number of NHPs used in the research is very small compared to the large number of humans who will benefit from it. This assessment is a critical step in the authorisation process under the (Directive 2010/63/EU) to grant licenses to projects involving animal research (Quigley, 2007). This approach has been challenged (Quigley, 2007; Rossi, 2009) but it is still part of the current regulation.

To sum up, researchers who want to use NHPs in their research must consider all the points summarised above and demonstrate that the use of these animals is fully justified. Their case must be accepted by two independent bodies: 1) the Animal Welfare and Ethical Review Body (AWERB) of the institution where they conduct their research and 2) the Home Office's Animals (Scientific Procedure) Inspectorate which will get advice from the Animals in Science Committee (ASC). On top of those, funders of the science might also ask for an additional body such as the NC3Rs to ethically assess the project proposals.

Ensuring the optimal welfare state of the NHPs used in research is not just important from an ethical perspective, but it is also essential from a scientific point of view (Poole, 1997). Poor welfare of an animal model decreases the quality of the research due to the variability in

welfare states induced between and within the subjects. Additionally, it might be generating results that may not be representative of a healthy human being (Bayne & Würbel, 2014).

As in neuroscience experiments using NHPs the sample size is commonly limited to a few individuals, and differences between subjects are assumed to not pose an important impact on the results. Nevertheless, ensuring the welfare of each subject in order to reduce the within-subject variability caused by potential welfare issues can be critical to the validity of the results. Many neuroscience tasks rely on behavioural measures as proxies for a variety of neural processes. As welfare is known to impact the behaviour of the animals, variability of the behaviours within subjects could have an impact on the results obtained from the neuroscience tasks. Moreover, the welfare of the animals can have a direct impact on their ability to learn, which could slow down and bias the results of neuroscience tasks.

Poor welfare of the animal model can even increase the financial expenses of the study, as a result of unexpected veterinary procedures, increased duration of the experiment or increased minimum sample size required (Pickard, 2013).

1.2. Definition of cumulative experience

Traditionally, NHP welfare has been monitored by assessing the potentially detrimental effect of individual stressors (Pickard, 2013). However, from research fields such as medicine, we know that repeated exposure to stressful events can have a cumulative impact on humans' health. For example, repetition of minor concussions can have major long-term effects on rugby players' health (Thornton et al., 2008). Depressive disorders can result from repeated exposure to stressful events in humans, even if each one, in isolation, seems harmless (Young & Korszun, 2010). As such, it should be expected that the cumulative effect of minor aversive stressors might potentially have harmful consequences on the welfare of laboratory animals (Pickard, 2013). Research facilities can have many aversive stressors that can affect the animal models housed there. These can come from both the experiments themselves (e.g., fluid restrictions or surgeries) as well from the general husbandry of the animals (e.g., veterinary procedures or housing). To avoid underestimation of potentially harmful stressors, their long-term impact should be taken into consideration.

In neuroscience research, the underestimation of the harmful effects of procedures could be even greater than in other fields. The tasks which NHPs have to perform for neuroscience research commonly require extensive training. Therefore, the scientific value of the subjects increases with the time they have spent in the facility and the training received. As

a consequence, the NHPs in this field can stay for several years in research facilities (Jennings et al., 2009), which increases the probability of suffering from the cumulative effect of repeated stressors (Pickard, 2013).

To acknowledge this cumulative effect, a new concept was introduced by the EU (European Union) Directive (Directive 2010/63/EU), termed cumulative experience. Cumulative experience is defined as "the net impact of all events (procedurally and husbandry-based) and the effects that affect adversely, positively and by way of amelioration, the welfare of the animal over its lifetime" (ASRU, 2015). The cumulative experience, therefore, includes not only all the negative events throughout the animal's lifetime but also the positive events. Positive events such as social interaction or a healthy lifestyle could improve their welfare and buffer the effect of the negative events. Including them will give a more holistic view of the life experience of the NHPs.

The cumulative experience concept has already been established and its use is growing. However, the potential cumulative effect of repeated procedures that may be affecting NHPs involved in research has not yet been investigated. The first study investigating this effect used qualitative data provided by self-reporting practitioners, with the only objectively measurable indicators of the NHPs' welfare being their body weight (ASC, 2015; Pickard, 2013).

In a recent advice note from the Animals in Science Regulation Unit (ASRU, 2015), the need for objective, long-term studies that investigate the cumulative effect of repeated husbandry and experimental procedures was highlighted. The main obstacle to the completion of such studies is the lack of a standardised method to assess the cumulative experience of NHPs.

1.3. Animal welfare and affective states

Up to this point, we have been referring to welfare as it is currently understood by modern society. Nevertheless, the definition of welfare (or wellbeing) has evolved profoundly during the last decades, and it is thus essential to explain what we will regard as welfare here. Traditionally, welfare in both human and non-human animals has been used as a synonym for physical health. For many years, the lack of observable diseases and injuries was enough to acknowledge an individual as healthy and in good welfare. However, insight into mental health has changed the understanding of welfare in humans, and consequently in non-human animals (Dawkins, 1980; Departmental Committee on Experiments in Animals, 1965).

The early Animal Act of 1911 already recorded the mental side of welfare by including words such as “infuriate” and “terrify” when addressing the animals’ suffering. However, it wouldn’t be until decades later that the UK legislation would focus on the psychological side of animal welfare, by separately defining mental suffering and stress caused by non-physical origin (Committee on Cruelty to Wild Animals, 1951; Departmental Committee on Experiments in Animals, 1965). A cornerstone for this change in regulation was the publication of Animal Machines (1964) by Ruth Harrison, a book that aimed to expose the reality of animals’ lives in intensive poultry and livestock farming.

In 1965, the “Report of the Technical Committee to Inquire into the Welfare of Animals Kept Under Intensive Livestock Husbandry Systems”, known as the Brambell report, stated that “An animal should at least have sufficient freedom of movement to be able without difficulty, to turn around, groom itself, get up, lie down and stretch its limbs”. These recommendations later expanded into a more detailed list of needs, known as “Brambell’s Five Freedoms”, and they were the initial step for a wider range of frameworks and definitions of animal welfare (Mellor & Beausoleil, 2015).

In 1979 (FAWC), the Five Freedoms framework, as currently known, was created to ensure the mental and physical needs of all animals are met: freedoms from (1) Hunger and Thirst, (2) Discomfort, (3) Pain, Injury or Disease; (4) freedom to Express Normal Behaviour and (5) freedom from Fear and Distress.

The Five Freedoms framework was then extended to further develop what is included in the last term: fear and distress (Mellor & Beausoleil, 2015). To develop the extended list of terms within the “distress” domain, behavioural, physiological and neuroscience studies were considered, and it is divided into two groups: states linked with internally generated inputs and states linked with external circumstances. These frameworks contributed to the origin of research into the internal states of animals.

Dawkins (1980, 2006, 2017) went a step further by, not just including mental well-being in the definition of welfare, but also stating that good welfare requires that animals experience positive emotions and have what they want.

Over the last two decades, multiple other frameworks and definitions of animal welfare have been developed depending on the aim of interest (McCulloch, 2012; Mellor & Reid, 1994). Fraser (1997, 2008) combined some of these with a three-dimensional approach by defining welfare according to three components: (1) the animal’s normal biological function,

which would refer to its physical state, (2) the animal's emotional state, and (3) the animal's ability to express natural behaviours.

For the present study, the reasoning was made as to the following: since poor physical health is usually accompanied by negative mental well-being (Wolfensohn & Honess, 2005) while negative mental well-being is not always accompanied by poor physical health (McFarland, 1989), the sole dependence on physical health as an indicator for welfare is not enough to discard poor welfare. Therefore, our welfare focus will be on the mental well-being or, so-called, affective states of the animals.

The affective states are internal states described by humans as feelings. Human studies have shown that these states have a valence (they can be positive or negative) and can also differ in arousal (reviewed in Mendl et al., 2010). Therefore, describing affective states with two underlying dimensions (i.e., arousal and valence) (Figure 2) can help researchers conceptualize the structure of internal experiences (Mendl et al., 2010).

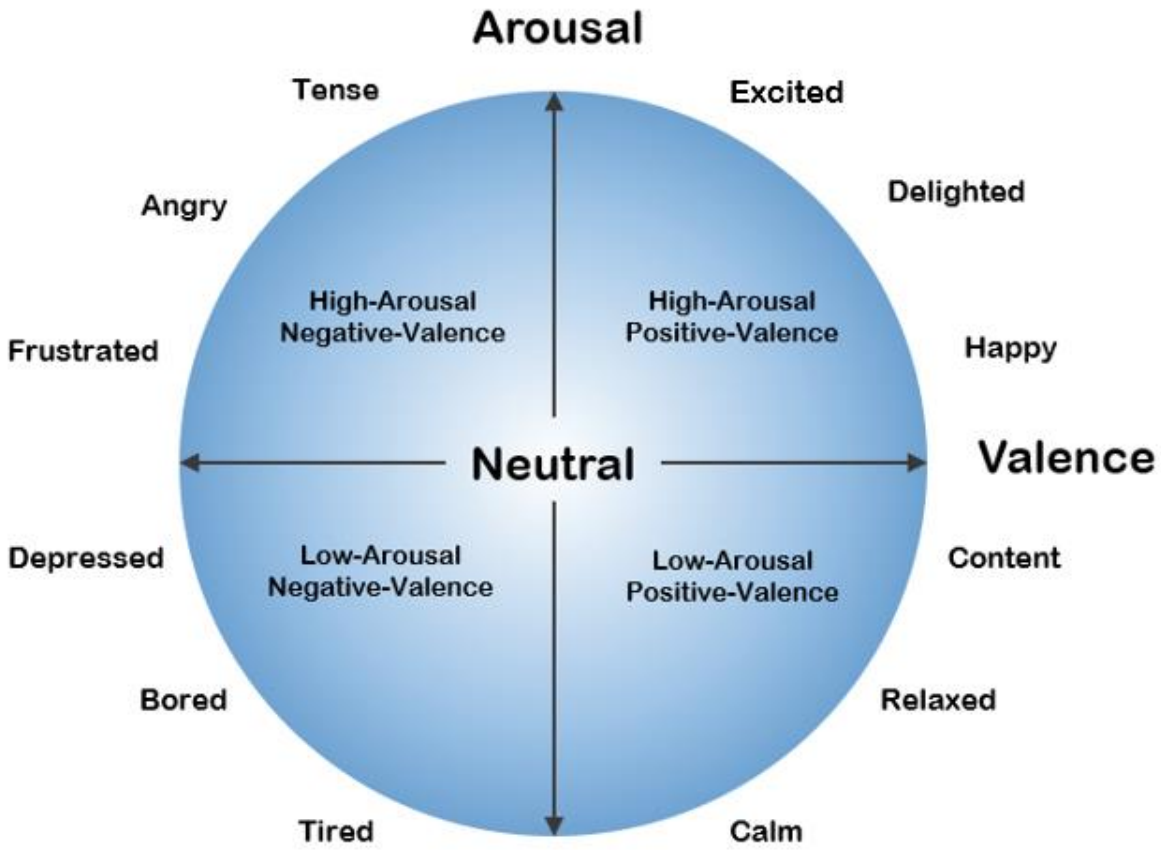

Figure 2. Two-dimensional diagram of affective state. The left side of the figure represents a negative affective state and the right side represents a positive affective state. The upper side

of the figure represents a high arousal level of the affective state, while the lower side represents a low arousal level. The words outside the blue circle indicate possible locations of specific affective states. Adapted from Mendl et al., 2010.

The affective state of the animals can also be divided depending on the time window of interest: acute vs long-lasting affective states. Acute affective states can last between seconds and hours, and long-lasting affective states can be anything from days to years after the exposure to the stressor.

The cumulative experience of an individual can be assessed through the variation of the long-lasting affective states during the lifetime of the animals. Alternatively, when the only available information derives from the acute affective states, investigation of their variation over time can also provide valuable information about the cumulative experience (i.e., sensitisation/habituation effects).

1.4. Welfare indicators for non-human primates

Affective states are private subjective experiences fully accessible only to the individual experiencing them (Dawkins, 1980; McFarland, 1989); these affective states are unique to those individuals. Even when all individuals are exposed to the same stressors in the same environment during the same period of time, different ultimate affective states are expected from each individual. This difference is explained by differences in genetics, personality/temperament, and coping styles (e.g., active vs passive; escape, remove, search or wait) between the individuals impeding reliable assessment of the potential affective states based on the environmental factors.

Thus, the most reliable method to assess the affective state in humans is to use linguistic self-report (Mendl et al., 2010). The inability to use this method is one of the main struggles when assessing the affective states of non-human animals. As an alternative, animal welfare scientists use welfare indicators. The main indicators proposed for assessing NHP welfare are (1) body weight (2) alopecia (3) physiological indicators, (4) behavioural indicators, (5) hippocampal indicators and (6) telomere biomarkers.

1.4.1. Body weight

Body weight is a commonly used indicator of long-lasting affective states. Most NHPs are weighed on a weekly or even daily basis. It is an easy, fast, and non-invasive method used in most animal research facilities. A sharp change in the weight of an animal can indicate a variety of mental and physical conditions and being overweight has been related to several

disorders in NHPs (Wolfensohn & Honess, 2005). However, the affective state of the animal can be compromised with no perceivable changes to its body weight, reflecting a lack of sensitivity (defined as the measure for how strong a stimulus has to be, before the indicator reacts to it). Similarly, a change in the body weight might be caused by factors unrelated to the affective state of the animal (e.g., age, diet), manifesting the lack of required specificity (defined as the narrowness of range of factors by which it will be affected). Consequently, although body weight can provide useful information, it is not a reliable indicator of cumulative experience in NHPs.

1.4.2. Alopecia

Alopecia, which is the lack or loss of hair in otherwise hair-covered parts of the body, is another indicator used to assess long-lasting affective states. Alopecia is a common condition in captive NHPs colonies. A study by Kramer and colleagues (2010) found that 48% of the rhesus macaques in the colony investigated had alopecia at some point in their medical history, with some studies showing an even higher prevalence (Steinmetz et al., 2005). Some of the suggested causes of alopecia include hormonal dysregulations, bacterial infections, nutritional deficiencies, stress and age (Honess et al., 2005; Lutz et al., 2013). Alopecia is therefore often considered an indication of poor psychological well-being (Luchins et al., 2011). Standardised alopecia indices (Bellanca et al., 2014; Honess et al., 2005) can provide the tools for an easy, relatively fast, visual, non-invasive assessment of the alopecia status of an animal. Despite the potential advantages of its use, the sensitivity and specificity to the changes in the affective state of the animal might not be sufficient to be used as a unique indicator of cumulative experience. The sensitivity to minor changes in the affective state is unknown and it is not specific to factors related to cumulative experience (e.g., seasonal effect). Therefore, results obtained with this indicator will require careful interpretation.

1.4.3. Physiological approaches

For many years, physiologists have been the leading scientists dealing with the concept of “stress” (Dawkins, 1980). Consequently, previous research placed a great focus on the physiological measures of affective state (Novak et al., 2016). The most used physiological biomarker to evaluate animals’ affective state is the level of glucocorticoids. Glucocorticoids are steroid hormones commonly known as stress hormones. In many mammals including NHPs, the main glucocorticoid is cortisol and therefore it is the one commonly measured by researchers. Cortisol synthesis in NHPs is increased in response to acute stressful events such as exposure to unfamiliar environments or restraint (Reinhardt et al., 1991). Studies have also

found increases in cortisol levels after exposure to chronic stressors such as decreased photoperiod, relocations to unknown habitats and maternal separations (Davenport et al., 2006; Fairbanks et al., 2011; Qin et al., 2015). Researchers can measure cortisol levels with a variety of techniques depending on the time window of interest, retrospectively, in a non-invasive manner (Novak et al., 2013). Saliva samples can be taken to measure acute cortisol levels (hours), faeces samples can be used for daily cortisol measurements and hair samples can be collected to measure the long-term (annual) accumulation of cortisol levels. This allows for assessing both the acute affective states and cumulative experience of the animals.

Nevertheless, multiple studies have found conflicting results. Cortisol levels have been observed to increase in positively valenced arousing situations such as reward, excitement, or sexual intercourse, as well as to remain at baseline levels after exposure to chronic stressors (Camus et al., 2013; Hennessy et al., 2014; Miller et al., 2007). Therefore, since the cause for the increase of the cortisol level may not be solely related to a negative affective state, it should not be used as a singular indicator of poor welfare in animals (Ralph & Tilbrook, 2016).

Another physiological indicator of affective states proposed for NHPs is the adrenal secretion of dehydroepiandrosterone sulphate (DHEAS). This hormone is unique to primates, and it has not been as deeply studied as cortisol. Nonetheless, the variation of DHEAS is stated to parallel cortisol during acute stress responses and to be more sensitive to chronic stress (Maninger et al., 2010). However, trying to assess the affective states of the NHPs with the concentrations of DHEAS entails the same difficulty in terms of the interpretation of the valence of the affective state.

In conclusion, even though physiological indicators are objectively measurable and can give a reliable measure of the arousing level of the affective state, they are not specific to negative valence.

1.4.4. Behavioural approaches

Changes in the behaviour of NHPs have also been proposed as indicators of affective states. Until the last few decades, a part of the scientific community was reluctant to use behaviour as an indicator of affective states due to the potential anthropocentric misinterpretation (Dawkins, 1980). However, many advances have since been made in the understanding of individual behaviours.

A common approach to assessing affective states by behavioural methods consists of measuring all possible behaviours in search of significant changes (Camus, et al., 2013; Gray

et al., 2016; Smith et al., 2006). This approach has been used for both assessments of acute and long-lasting affective states. However, the use of this approach can lead to important biases and result in both false positives and false negatives.

The first limitation of this approach is inherent to the study design of behavioural studies. Behavioural studies have a specific time window when the behaviours are observed. Because time is limited, the number of behaviours that the animal can display, and the time dedicated to each of them is also limited. Consequently, when multiple behaviours are investigated, any type of perturbation, associated with negative or positive valence, is going to induce changes in behaviour frequencies. Therefore, understanding how changes in different directions of all behaviours are indicative of changes in the affective state of the animal can prove very challenging and lead to false positives.

A second limitation originates from the loss of sensitivity during the statistical analysis of many behaviours due to the correction for multiple comparisons needed. This can lead to false negatives by discarding changes in behaviour that indicate changes in the affective state but are not statistically significant due to the low alpha value imposed by the correction for multiple comparisons.

Accordingly, the field is moving in the direction of identifying specific behaviours as indicators of negative and positive affective states.

1.4.4.1. Stereotypic behaviours

One common example of these indicators of negative affective states are stereotypic behaviours. Stereotypies are repetitive and invariant behaviours mostly observed in captive animals (Mason & Rushen, 2008). They are a group of heterogeneous behaviours, some of which have been hypothesised to indicate negative affective states (Wolfensohn & Honess, 2005). In NHPs, pacing is the most common stereotypy and has been used to assess both acute and long-lasting affective states.

Pacing is a locomotive stereotypy mainly defined as repetitive walking. Pacing has been observed at a higher frequency in animals in captivity than in wild populations (Kroshko et al., 2016). Consequently, this stereotypy is often interpreted as an indicator of a negative affective state in many animal species. However, recent studies in macaques have found that pacing frequency decreased after the animals were exposed to stressors (Peterson et al., 2017; Poirier, Oliver, et al., 2019), while another study with capuchin monkeys found no relation between

pacing and negative affective states (Pomerantz et al., 2012). The inconsistency between the results emphasizes the need for further study of this stereotypic behaviour in NHPs. The use of pacing as an indicator of affective states before the full understanding of its cause could lead to false positive and false negative results (Poirier et al., 2019).

1.4.4.2. Displacement behaviours

Another commonly used indicator of affective states are displacement behaviours. Displacement behaviours in non-human primates are a group of self-directed behaviours that are displayed with no apparent ecological context or that seem irrelevant to the ongoing activity of the animal (Maestripieri et al., 1992; Schino et al., 1996). The main behaviours considered as displacement behaviours in NHPs are body-shaking, self-grooming and self-scratching (Schino et al., 1996). These behaviours fulfil a function when displayed in an appropriate context, and thus it is the context that will determine if the behaviours constitute displacement behaviours. Increases in these behaviours have been observed in primates after high levels of neighbour vocalizations, social rank uncertainty or spatial proximity of a dominant male (Baker & Aureli, 1997; Castles et al., 1999; Schino et al., 1990), suggesting that they might be indicators of negative short-lasting affective states associated with high arousal similar to anxiety in humans (Figure 2). This interpretation is reinforced by pharmacological studies using anxiogenic and anxiolytic treatment as a response to acute stress. Researchers observed an increase in the display of the behaviours when the subjects were treated with anxiogenic drugs and exposed to an acute stressor; and the opposite response was observed when treated with anxiolytic drugs (Maestripieri et al., 1992; Schino et al., 1996). The effect found was dose-dependent, meaning that there was a significant difference between the control and the treatment group and that, additionally, the dose differences had a significant effect on the frequency of the behaviours in the expected direction. This constitutes a full pharmacological validation.

However, although the response to acute stress has been studied, we have no information regarding the assessment of the integration of long-term experiences in these behaviours. Therefore, displacement behaviours can only be used as indicators of acute affective states and not as indicators of cumulative experience in NHPs.

Scratching behaviour has also been proposed to be used independently as an indicator of acute stress. Scratching has been found to increase when animals were exposed to stressful situations such as separation from an infant, unreconciled conflict, and aggression (Aureli et

al., 1997; Castles et al., 1999; Maestripieri, 1993; Majolo et al., 2009; Schino et al., 1990). A more recent study's preliminary results found increases in scratching behaviour in Geoffroy's spider monkeys (*Ateles geoffroyi*) when the animals were isolated, in an uncertain situation and when the mothers were not in close proximity to their infants (Dell'Anna et al., 2022). In pharmacological studies, scratching has been observed to decrease in female macaques when treated with lorazepam anxiolytic drug (Schino et al., 1991). An increase in the frequency of this behaviour was also found when macaques were treated with the anxiogenic drug FG7142 (Major et al., 2009). However, the increase was not dose-dependent and the difference between doses was just significant at the highest dose. In conclusion, scratching, though not fully pharmacologically validated gathers favourable evidence for its use as a potential indicator of acute stress in NHPs. Furthermore, this behaviour could be especially useful when the use of displacement behaviours is not feasible, such cases include comparisons between socially housed and single-housed animals, in which case the frequency of self-grooming behaviour (one of the displacement behaviours) is expected to increase due to the absence of allo-grooming in the single-housed animals, which could cause biases in the measures.

Yawning tends to be considered a displacement behaviour by many researchers (Troisi, 2002). Thus, it has also been suggested as an indicator of acute stress. Behavioural studies revealed animals display yawning after acute social stressors such as the proximity of a dominant male and risk of aggression (Carpenter & Schultz, 1940; Darwin, 1872; Hadidian, 1980; Schino et al., 1990; Zuckerman, 1932). A pharmacological study using Beta CCE (an anxiogenic drug) failed to observe an increase in yawning systematically (Lagarde et al., 1990). On the other hand, in another study, the anxiogenic drug FG7142 was found to increase the frequency of yawning when animals were treated with a medium dose of the drug (Major et al., 2009). However, no main effect of the drug was found, and the low and high doses of the drug failed to significantly change the frequency of the behaviour, which may suggest a false-positive result. Therefore, more evidence is needed to understand the role of yawning in connection to acute stress. As for the integration of long-lasting experiences, there is no evidence suggesting that yawning could be used as an indicator of cumulative experience.

1.4.4.3. Affiliative behaviours

Non-human primates, as social species, may be involved in affiliative behaviours for up to 20% of their day (Henazi & Barrett, 1999). Thus, understanding the relationship between the affective states and affiliation can play a pivotal role in a better understanding of the

affective states of the animals. Hence, researchers have been investigating affiliative behaviours in a number of different scenarios. The display of affiliative behaviours in NHPs has been observed to increase after being exposed to social stressors. This phenomenon is often called "social buffering" (Kikusui et al., 2006) and has been proposed to act as a coping mechanism. Grooming was found to increase after the animals were exposed to the stressor of food competition in a large group of rhesus macaques (de Waal, 1984) and chimpanzees (de Waal & van Roosmalen, 1979). In captivity, Affiliative behaviours were also found to increase in rhesus macaques when reunited with their cage mate after being separated during the night, a situation suggested (but not confirmed) to cause acute stress as it increases the frequency of displacement behaviours (Cassidy et al., 2020). However, pharmacological studies found no change in affiliative behaviours after treatment with either anxiolytic or anxiogenic drugs (Schino et al., 1996). On the other hand, in a study with vervet monkeys, separation anxiety increased affiliative behaviours and antidepressants prevented such increase (Marais et al., 2006). To our knowledge, affiliative behaviours have not been found to increase after exposure to a stressor in a non-social context. Therefore, current evidence suggests affiliative behaviours might be an indication of exposure to social stressors, instead of representing a general acute negative affective state. Another element to consider is linked to the coping hypothesis. If the coping behaviours are efficient and deal fully with the effect of the stressor presented, the net impact on the affective state of the animal is neutral. In this case, affiliative behaviours would not be useful markers of a negative affective state, but potentially they would indicate exposure to a stressor. Nevertheless, more research will need to be undertaken to understand the factors affecting the behaviours.

1.4.4.4. Inactive behaviour

Long-lasting negative affective states can also be assessed by investigating depressive-like behaviours. This type of behaviour was first observed in NHPs when researchers were seeking to create an animal model of human depression (Harlow & Zimmermann, 1959). The animal models displayed inactivity and withdrawn behaviour, with the latter being defined as low interest in the surrounding environment. The behaviour of inactivity was often linked with a hunched posture (Harlow & Zimmermann, 1959). Depressive-like behaviour has subsequently been experimentally induced in NHPs using stressors that also cause depression in humans, such as social isolation and the shortening of the photoperiod (trigger of seasonal depression) (Hennessy et al., 2014, 2017; Li et al., 2013). The behaviour has also been pharmacologically validated with anti-depressant drugs (Perera et al., 2011; Qin et al., 2015).

For instance, Perera and colleagues (2011) have shown that treating macaques with antidepressants prevented the display of depressive-like behaviour in subjects that were repeatedly exposed to social isolation (two days a week for 15 weeks).

In another study by Qin and colleagues (2015), researchers attempted to create an animal model of Seasonal Affective Disorder. This disorder, often known as seasonal depression, is a common mood disorder that affects humans, and it is related to the shortening of the photoperiod. After three months of exposure to a short photoperiod (5 hours per day instead of 12), the frequency of depressive-like behaviour was found to have increased in all subjects. Subsequent treatment with an antidepressant drug returned the frequency of the behaviour to baseline levels.

The display of depressive-like behaviours is not limited to a result of experimental manipulation. NHPs in many types of laboratory facilities have been observed to display this behaviour spontaneously. The appearance of this behaviour has been observed in indoor single-housed animals (Camus et al., 2013; Shively et al., 2005), indoor socially housed animals (Camus et al., 2014; Shively et al., 2005; Xu et al., 2015) and outdoor socially housed animals (Camus et al., 2014; Hennessy et al., 2017). The wide display of the behaviour in the absence of any experimental stressor suggests that it may be sensitive to subtle stressors.

1.4.4.5. Cognitive bias tests

One suggested approach for the study of cumulative experience using behaviour is by using cognitive bias measurements. This approach is based on the fact that decision-making is known to be influenced by background long-lasting affective states. Ambiguous situations will tend to be judged positively by an animal with a positive affective state, and negatively by an animal in a negative affective state. The derivations in judgement dependant on the affective states are known as cognitive biases. These derivations can be measured to obtain information about the long-lasting affective state of the animal. In a study by Bethell and colleagues (2012), The subjects that had been exposed to events that could potentially affect their affective state were found to experience shifts in their cognitive bias. These results suggest that the use of cognitive bias is a promising method for assessing cumulative experience. However, the main obstacles to the wide use of this method are the requirement for extensive training of the subjects and the difficulty to repeat the measure over time to track changes in the long-lasting affective states (Paul et al., 2005).

As a solution to the extensive training required by judgement bias tests, another type of cognitive test has been proposed: attention bias tests (Bethell et al., 2012). These new tasks rely on emotionally charged stimuli, thus significantly reducing the need for training. Human studies suggest that individuals stare for longer at an aggressive face when they are in an anxious-like state (Bar-Haim, 2007). Therefore, monkeys can be presented with an image of an aggressive and a non-aggressive monkey face, while their attention is being measured. Although these tasks could give us some insight into the animals' affective states, they access distinct emotion-cognition interactions compared to judgement bias tests (Bethell et al., 2016). A recent study investigated the changes in attention after exposing the NHPs to stressors and found very low reproducibility in attention bias scores (Howarth et al., 2021). Moreover, it is currently unknown the time window in which a stressor can influence the attention of the animals, as well as the type of stressors they are sensitive to (i.e., social stressors). Consequently, more research will need to be conducted before attentional bias tests can be used as indicators of cumulative experience in NHPs.

Hence, from all the behaviours summarised above, displacement behaviours seem to be the most reliable indicators of acute negative stress, and depressive-like behaviours appear to be the most reliable indicators of cumulative experience.

1.4.5. Hippocampal markers

Novel approaches to assess the cumulative experience are also being considered. One of them is using a hippocampal biomarker. A recent paper by Poirier and colleagues (2019) has shown that hippocampal biomarkers are valid measures of cumulative experience in non-human mammals and humans unable to self-report their well-being.

Stress affects the hippocampus at different spatial scales. At a microscopic level, the rate of new neuron incorporation and the structure of mature neurons have been observed to be sensitive to stress (Fanselow & Dong, 2010; Moser & Moser, 1998; Strange et al., 2014). At a macroscopic level, the local amount of grey matter (i.e., where the cell bodies of neurons are more highly concentrated) in the anterior region of the hippocampus and its relative size, seem to be sensitive too (Abbott et al., 2014; Arnone et al., 2012, 2013; Frodl et al., 2008; Koolschijn et al., 2009).

Comparative studies between behavioural tests using long-lasting affective states and rates of neurogenesis, local amount of grey matter and hippocampal volume in non-human animals, demonstrated a correlation between the hippocampal markers and the behavioural

studies. Moreover, the hippocampal biomarkers followed the hypothesized direction when animal models were treated with anti-depressant drugs or exposed to anxiogenic or depressogenic situations (Morais et al., 2017; Santarelli et al., 2003; Snyder et al., 2009).

Experimental animal studies using chronic stressors resulted in a decrease in the different hippocampal biomarkers of the subjects studied (Jackowski et al., 2011; Mitra et al., 2006; Perera et al., 2011). Furthermore, when animals were chronically exposed to situations that induce positive affective states, the hippocampal biomarkers were observed to increase (Kozorovitskiy et al., 2005). In addition, the hippocampal biomarkers have been observed to depend on the interaction between the exposure to the event and the genetic background of the animal exposed (Ieraci et al., 2016; Mitra et al., 2006). Therefore, the markers seem to reflect the individual response to the event, considering the differences in vulnerability to stress. Finally, the markers seem to be successfully integrating the experiences of the animal subjects over time, at least during a considerably long-time window (months/years) (Kim et al., 2013; Li et al., 2017; Morais et al., 2017).

Accordingly, the hippocampal biomarkers seem to match theoretical assumptions and encompass all facets of the construct, validating them as proper cumulative experience indicators (Poirier et al., 2019).

Several factors unrelated to the affective states are known to influence the hippocampal biomarkers reviewed. Some of these factors include intrinsic individual differences (i.e., age, sex, total brain volume and genotype) that can be controlled for with statistical analysis and an appropriate study design. Other confounding factors such as the potential effect of acute stressors (Schoenfeld & Gould, 2012) or learning processes (Fanselow & Dong, 2010) could be dealt with during the processing of the images and statistical analysis (Poirier et al., 2019).

This approach also involves certain limitations. The use of this method requires MRI equipment and experts in neuroimaging to measure hippocampal plasticity at a macroscopic level, which most facilities will not have access to. On top of this, the animals involved in the research need to be trained and endure invasive surgical procedures or be anaesthetised and fully supervised by a veterinary team. Finally, the processing and analysis of the obtained data require time and specialist skills. In conclusion, hippocampal biomarkers, despite their limited accessibility, are a very promising approach to assessing cumulative experience and can be used to assess the impact of procedures routinely used in neuroscience experiments.

1.4.6. Telomere markers

Another novel approach to assessing the cumulative experience of NHPs consists of examining telomeres. Telomeres are DNA-protein complexes that are located at the end region of chromosomes. The main function of these structures is to protect the coding regions of DNA from damage. The length of these complexes shortens over time with every cell division. The rate of telomere shortening is increased by several mechanisms linked to stressful events: oxidative stress, inflammation and cortisol levels (Bateson & Poirier, 2019). In addition, some evidence suggests that exposure to some events can induce increases in the telomere length (Bateson & Nettle, 2016). It has been suggested that the telomere shortening rate could be used to estimate the cumulative experience of NHPs (Bateson et al., 2015).

However, most of the studies investigating this potential marker produce cross-sectional, correlational data, and might be underpowered (Pepper et al., 2018). In addition, recent studies show that there is a high measurement error when investigating the change in the telomere length which will be an obstacle to achieving the required sensitivity (Nettle et al., 2021). Therefore, although promising, more research should be done on the validation of telomere length as a marker for cumulative experience.

1.5. Aims and conclusions

The main three aims of my project were the following:

1. Determine whether the welfare of laboratory macaques used as an animal model of healthy humans decreases over years (negative cumulative experience).
2. Identify the procedures having a long-lasting effect on the welfare of the subjects.
3. Assess the effect of these procedures (negative versus positive).

To achieve the stated aims, I chose two main indicators to assess the cumulative experience of the animals. From all the indicators discussed above, several achieved the criteria for use as indicators of cumulative experience: depressive-like behaviour and the hippocampal markers. The depressive-like behaviour marker has the features to be an adequate method to assess the negative affective states of laboratory NHPs. Accordingly, depressive-like behaviour will be the behavioural indicator chosen for the current project.

Hippocampal biomarkers also seem promising as indicators of cumulative experience and have been chosen as such in this study. From the different markers proposed, we decided to select a macroscopic hippocampal marker that would allow us to assess the cumulative

experience *in vivo*, in a relatively non-invasive manner repeatedly through the subjects' lifetime. Specifically, the local amount of grey matter in the anterior hippocampus will be the marker used.

In addition to these two indicators, weight and alopecia were also used as traditional indicators of cumulative experience. The hypothesis was made that these two latter indicators would not be sensitive and specific enough to assess subtle changes in the cumulative experience of the animals, as will be compared to the results obtained with the depressive-like behaviour and hippocampal biomarker.

These four indicators were used during the project to study the potential cumulative effect of repeated husbandry and experimental procedures in NHPs.

Chapter 2: General methods and materials

2.1. Ethical statement

All studies were conducted at the Comparative Biology Centre at Newcastle University and approved by the Home Office for regulated work, as appropriate (PPL: 70/7976; PC6981D63, 60/4431; PABAD450E; PA2C18B73). All work was performed in accordance with the EU Directive (2010/63/EU), ASPA (1986) and the NIH Guidelines for Care and Use of Animals for Experimental Procedures (National Institutes of Health, 2011). The animals used in this study were kept at the research facility as a result of their involvement in other existing experimental projects. None of the animals was kept solely for the purpose of the present study.

2.2. Subjects of study and husbandry

A total of 31 laboratory rhesus macaques (*Macaca mulatta*) participated in the project (11 females and 20 males). The age range was between 3 and 17 years old, and the weights were between 4.23 and 19.3kg (Table 1).

All the animals involved in the project belonged to the Newcastle University research facility which complies with the NC3Rs Guidelines for "Primate Accommodation, Care and Use" (NC3Rs, 2017). Typically, animals were housed as pairs or trios in cages. Single-housed macaques are exceptional cases in the Newcastle University research facility. Social isolation caused by single housing is known to have detrimental consequences on the welfare of NHPs linked with pronounced changes in their behaviour (Charbonneau et al., 2022; Perera et al., 2011). However, due to veterinary and husbandry reasons, some animals were singly housed for a period of time. Some of the veterinary reasons for single housing included injuries requiring temporary protection from contact with cage mates, or instances of resistant bacteria infections resulting in isolation to limit the spread. As for the husbandry reasons, some animals were hostile to all attempted pairings with a cage mate with instances of severe fights that would endanger both animals' wellbeing. As a result, on those occasions, isolation was considered the best viable housing option. The dimensions of the cages (2.1 x 3.0 x 2.4 m) exceeded the minimal space requirement under the UK legislation of 1.8 m^3 per animal. The vertical dimension of the cages enabled the animals to jump, run and climb, and exhibit inherent behaviours like escaping during playful or agonistic interactions (Reinhardt, 2008). The cages were located in a facility with over 20 individuals with whom the subjects had visual and auditory contact. Cages had a plain floor covered with wood shavings and they were enriched

with swings, ropes, shelves, and natural light (see Figure 3 for reference). Subjects were provided with daily foraging opportunities as the food offered was spread over the floor, as recommended by NC3Rs primate welfare guidelines (2017) and LAREF (Reinhardt et al., 2007). Several studies found that macaques prefer food that requires effort to obtain (Line et al., 1989; Reinhardt, 1994a). Spreading their food over the floor encourages their cognitive and motor skills as they move when looking for food and allows them to display foraging and manipulative activities that are part of their natural environment (Bonini, 2019). On top of this, the use of this method can help reduce the tension and aggressive behaviours between cage mates as it disperses them around the enclosure (Bryant et al., 1988; Wolfensohn & Honess, 2005). This feeding method has also been observed to decrease detrimental self-directed behaviours like hair-pulling (Blois-Heulin & Jubin, 2004; Bryant et al., 1988).

Novel floor objects were introduced regularly in all the cages, with 2-week rotations. As well as this, aerial enrichment was changed monthly. In addition to natural light, the facility had artificial light on a 12h light/dark cycle. The humidity was set at approximately 24% and the temperature at 20 °C to better resemble conditions of native habitat (i.e., India, Bangladesh, Pakistan, Nepal, Myanmar, Thailand, Afghanistan, Vietnam and southern China).

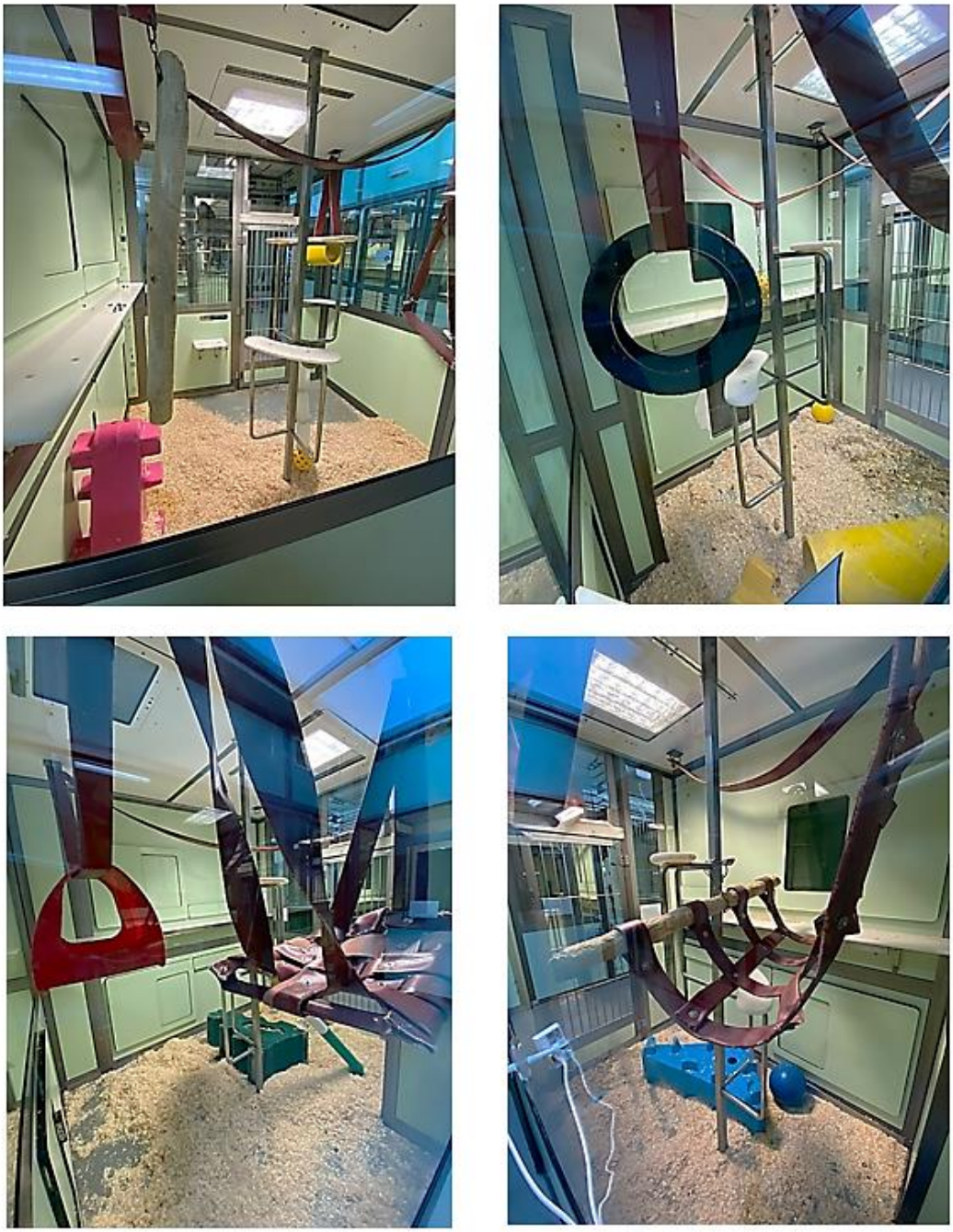

Figure 3. Enrichment at Newcastle University's non-human primate facility. Swings, ropes, shelves, and plain floor covered in wood shavings can be observed in the cages prepared for rhesus macaques.

All the subjects were raised in UK breeding centres: Medical Research Council's Centre for Macaques (CFM) and The Defence Science and Technology Laboratory (DSTL). The macaques were housed with their mothers for at least six months (mean weaning age= 1.7yo) and with other juvenile individuals until they were adolescents (Table 1).

Table 1. Subject information compiled by: Subject number, Sex (F: Females and M: Males), breeding centre, date of birth (DOB), arrival at Newcastle, date of first procedure, weaning date, weight range during the length of the present project (in kilograms), date of first collected record for the study, the participation dates in which each subject was involved in the present project and the group the subjects are part of ("Early" vs. "Late").

Subject	Sex	Breeding centre	DOB	Arrival at NcL	Date of first procedure	Weaning age	Weight (kg)	Date of first record	Participation dates	Group
1	M	DSTL	01/04/2002	17/06/2005	01/10/2005	02/10/2002	15.9 to 19.26	14/11/2014	27/05/2014 to 05/06/2017	Late
2	M	DSTL	08/04/2002	17/06/2005	23/10/2005	09/09/2003	16.42 to 17.87	11/11/2022	23/04/2014 to 18/07/2014	Late
3	M	DSTL	17/02/2004	17/01/2007	11/11/2022	07/12/2004	11.11 to 15.74	02/06/2014	02/06/2014 to 11/12/2014	Late
4	M	DSTL	15/12/2005	28/08/2008	01/04/2009	05/09/2006	11.11 to 16.61	11/11/2014	12/05/2014 to 23/10/2016	Late
5	M	DSTL	18/12/2005	28/08/2008	01/04/2009	05/09/2006	8.26 to 16.46	13/11/2014	02/06/2014 to 16/10/2018	Late
6	M	DSTL	03/03/2005	28/08/2008	01/04/2009	05/09/2006	7.08 to 10.675	31/05/2014	10/06/2014 to 11/01/2021	Late
7	M	DSTL	19/12/2005	28/08/2008	01/04/2009	05/09/2006	11.82 to 14.5	25/11/2014	03/06/2014 to 07/12/2015	Late
8	M	DSTL	12/12/2005	28/08/2008	01/04/2009	05/09/2006	8 to 11.12	25/05/2014	12/05/2014 to 06/11/2017	Late
9	M	DSTL	14/05/2006	25/03/2009	01/06/2009	18/01/2007	12.87 to 16.48	22/04/2016	22/04/2016 to 22/10/2019	Late

Subject	Sex	Breeding centre	DOB	Arrival at NcL	Date of first procedure	Weaning age	Weight	Date of first procedure	Participation dates	Group
10	M	DSTL	06/02/2006	25/03/2009	01/06/2009	07/12/2006	8 to 17.35	24/05/2014	24/05/2014 to 07/12/2020	Late
11	M	DSTL	13/12/2006	22/07/2009	01/12/2009	15/11/2007	11 to 18	21/05/2014	21/05/2014 to 24/09/2019	Late
12	M	DSTL	01/06/2006	22/07/2009	20/10/2009	13/06/2007	4.23 to 12.46	20/10/2009	13/11/2017 to 12/10/2018	Late
13	M	CfM	22/07/2008	12/05/2011	01/06/2011	11/05/2011	10.69 to 12.99	29/05/2014	10/11/2014 to 29/04/2016	Late
14	M	DSTL	15/07/2008	18/08/2011	10/10/2011	29/06/2009	10.99 to 15.64	14/04/2015	14/04/2015 to 25/01/2021	Late
15	M	DSTL	02/11/2009	08/01/2013	23/01/2013	15/09/2010	10.93 to 13.58	30/05/2014	30/05/2014 to 16/11/2019	Early
16	M	CfM	04/07/2010	22/10/2013	27/01/2014	09/05/2013	4.23 to 12.46	31/05/2014	31/05/2014 to 03/03/2020	Early
17	M	CfM	27/01/2011	24/03/2014	01/05/2014	23/03/2014	5.64 to 13.8	17/11/2014	17/11/2014 to 25/11/2020	Early
18	F	CfM	26/07/2010	19/01/2015	01/09/2015	09/05/2013	5.88 to 8.89	14/10/2015	14/10/2015 to 12/12/2019	Early
19	F	CfM	16/06/2009	19/01/2015	01/09/2015	09/05/2013	6.6 to 11.6	29/10/2015	29/10/2015 to 27/11/2020	Early
20	F	CfM	11/04/2012	14/07/2015	23/10/2017	03/07/2014	5.38 to 9.5	27/11/2017	27/11/2017 to 14/05/2021	Early
21	F	CfM	19/02/2012	14/07/2015	03/01/2017	03/07/2014	6.26 to 9.86	12/12/2017	30/03/2018 to 21/11/2020	Early

Subject	Sex	Breeding centre	DOB	Arrival at NcL	Date of first procedure	Weaning age	Weight	Date of first procedure	Participation dates	Group
22	M	CfM	14/05/2012	14/07/2015	01/02/2017	10/04/2014	5.32 to 12.2	23/10/2015	23/10/2015 to 24/10/2020	Early
23	F	DSTL	23/04/2011	08/03/2016	06/03/2017	25/06/2012	6.88 to 10.78	24/04/2017	24/04/2017 to 15/02/2020	Early
24	F	DSTL	27/05/2011	08/03/2016	06/03/2017	25/06/2012	6.4 to 9.9	27/04/2017	27/04/2017 to 12/09/2019	Early
25	F	CfM	30/05/2013	18/04/2017	05/02/2018	01/05/2015	6.3 to 9	12/12/2017	03/06/2018 to 23/02/2020	Early
26	F	CfM	04/02/2012	18/04/2017	25/06/2018	23/09/2014	5.6 to 8.19	12/12/2017	12/06/2018 to 23/02/2020	Early
27	M	CfM	10/10/2015	06/12/2018	03/03/2021	01/02/2017	6.6 to 10.72	18/12/2018	07/07/2019 to 31/01/2021	Early
28	F	CfM	16/10/2015	06/12/2018	26/04/2021	25/01/2017	5.01 to 8.33	18/12/2018	29/07/2019 to 01/02/2021	Early
29	M	CfM	18/11/2014	06/12/2018	26/04/2021	04/07/2016	7.61 to 14.26	18/12/2018	01/08/2019 to 04/02/2021	Early
30	F	CfM	19/06/2015	30/05/2019	10/09/2020	20/11/2017	6.13 to 7.18	19/01/2020	19/01/2020 to12/11/201	Early
31	F	CfM	12/08/2014	30/05/2019	10/09/2020	24/08/2015	4.67 to 6.69	20/01/2020	20/01/2020 to 14/11/2021	Early

Animals at the facility underwent daily visual checks by the technicians. Any concerns about their activity level, fur condition, faeces, food and fluid intake were referred to the veterinary team. Additionally, the veterinarian staff collected blood samples annually from all the animals at the facility for health screening.

The subjects of this study were involved in neuroscience and behavioural experiments as animal models of humans throughout the duration of the study. Some of these studies included acute single and multi-unit electrophysiological recordings and brain imaging (MRI and fMRI), electroencephalography (EEG), transcranial ultrasound stimulation (TUS), behavioural tasks requiring eye-fixation, etc. The inclusion criteria were the following:

1. Subjects were animal models of healthy human beings. This criterion excludes animals used as a model of human diseases or injuries, since the former were the animals with long stays in research facilities, and whose cumulative experience is in question.
2. Subjects were expected to stay at the research facility for several years. A short stay in the research facility might not allow for the investigation of the cumulative experience of the subjects.

The project started in 2014 with the subjects that fitted the explained criteria. However, new subjects were incorporated during the project which led to the creation of two distinct groups of subjects within our sample. The two groups were differentiated according to the time between their first procedure at the Newcastle facility and the first data we collected from them. The groups were labelled “Late” and “Early” (Figure 4).

1. Group “Late”: The subjects included in this group were investigated during the middle-to-late stage of their experimental life. The animals incorporated at the beginning of the project (2014) comprise most of the subjects in this group, unintentionally, resulting in older animals which had been at the facility for some months or years. This group included 14 subjects, all of them males (Table 1).
2. Group “Early”: This group of subjects was studied during their early up to late experimental life and included animals that arrived at the facility once the project had already started. As a result, these subjects tended to be, on average, younger than the subjects in the “Late” group. Changes in neuroscientific research trends and legislation also had an impact on the specific characteristics of the sample. This resulted in a higher number of females in this group, as well as an increased weaning age for subjects in this group. Finally, the animals included in the “Early” group were more likely to have been

bred at the CfM breeding centre than at the DSTL. This group included 17 subjects (6 males and 11 females) (Figure 4).

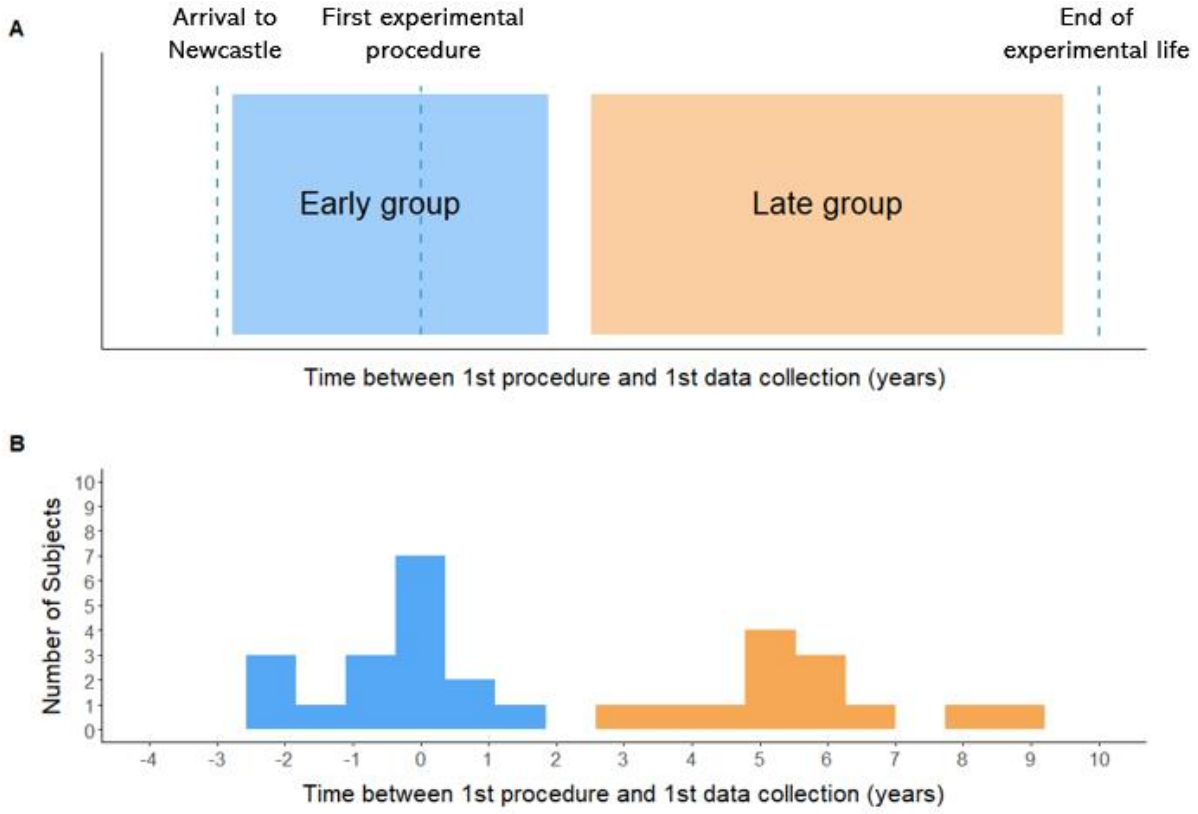

Figure 4. Diagram of the two groups of subjects included in the dataset according to the time between the 1st procedure and the 1st data collected for the present study. Subjects from the "Early" group are represented in blue and subjects in the "Late" group in orange. Panel A displays the diagram for the division between the "Early" and the "Late" groups of subjects. Panel B shows the number of subjects according to the time in years between the 1st procedure and the 1st data collected for the present study.

2.3. Behavioural recording

The behaviours were recorded using remotely controlled video cameras (Cube HD 1080, Y-cam Solutions Limited, Twickenham, UK; Axis HD Cube and M1065-L, Axis Communications AB, Lund, Sweden). Blue Iris Video Management Software (https://blueirissoftware.com/) was used to automate the recording times. The cameras were located outside the enclosures of the animals at a distance of 1.5m-2.5m from the cages to provide a wide view of the cage as well as avoid being in reach of the subjects.

Recordings started 15 minutes after the light was switched on in the primate facility and lasted for 45 minutes in each session. The lights were switched on in the facility automatically from 6:15 to 7 am and were switched off 12 hours and 30 minutes later.

The early morning was chosen as the ideal time for the behavioural data collection for the following reasons:

1. It was the only period during daylight when the focal subjects and their cage-mates were guaranteed to be together in their home cage.
2. The human presence was significantly lower than at any other time during daylight hours.
3. The 15-minute lapse between the light switching on and the start of the data recording aimed to ensure that the monkeys were all fully awake. The macaques were not expected to be asleep during the early morning (Reite et al., 1965).

For analysis purposes, the 45 minutes per session were divided into 1-minute time bins, in which the presence (1) or absence (0) of a pre-defined list of behaviours was encoded. The frequency of the behaviour was calculated as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. Due to the enrichment devices and the location of the cameras in the facility, the focal subject was not always visible to the observer. On the occasions in which the focal subject was not visible for more than two seconds, the time bin was encoded as *out-of-sight*.

The data were encoded by a total of 11 observers, each of them encoding differing numbers of sessions. All observers underwent extensive video training, followed by a formal assessment. During this assessment, the accuracy of the new observer encoding was compared with the encoding of an expert observer and an inter-rater reliability score was calculated. For observers who encoded video recording for many months, the formal assessment was repeated after 6 months, and an intra-rater reliability score was calculated to ensure there was no drift in the identification of the behaviours over time. All the observers encoded the recordings while having no prior knowledge regarding the life history of the focal subjects.

Cohen's Kappa score, commonly used to measure inter- and intra-rater reliability, is known to be highly sensitive to the frequency of the behaviour, resulting in unrealistically high reliability for the commonly observed behaviours, and low reliability for the rare ones (Cicchetti & Feinstien, 1990). Due to the nature of our data, extreme frequencies are common. Therefore, we used the Maxwell R.E. coefficient (Maxwell, 1977), which has been shown to deal efficiently with extreme values and to be more sensitive (Feng, 2013). R.E. coefficient of more than 80% was achieved both after the training phase and at the end of the encoding, indicating a strong inter- and intra-rater reliability (mean R.E. coefficient range per session 81.9-100; absolute R.E. mean=96.18).

2.3.1. Refinement of Inactive not alert behaviour as an indicator of cumulative experience

Following the literature review (see section 1.4.4 on behaviours as welfare indicators in non-human primates), the best behavioural indicator of cumulative experience was determined to be the so-called depressive-like behaviour. Previous studies have used a wide range of definitions for this behaviour, with some problematic aspects such as the need for intervention or specific environmental stimuli (e.g. 'rejecting social advances', 'socially withdrawn', 'lack of responsivity to environmental events'), or requiring some baseline measure to identify the behaviour (e,g, 'Increased levels of immobility', 'Reduced alertness') (Table 2). We aimed to create an objective, operational definition of the depressive-like behaviour to be used as an indicator of cumulative experience based on definitions that were previously used in the literature (Table 2).

Table 2. Compilation of previous definitions used in the literature that represents a depressive-like Inactivity behaviour.

Reference	Label	Definition
Baker et al., 2012	Inactive	Passive, awake or asleep
Camus et al., 2013	Slumped	Seated head lower than shoulder's line body posture in which monkeys appear to be withdrawn from the environment
Camus et al., 2014	Depressive- or anxiety-like behaviours	Increased levels of immobility, a slumped body posture in which they appear to be withdrawn from the environment
Clarke et al., 1988	Depressed posture	Huddling over or lying down
Hennessy et al., 2014	Withdraw	Sitting in hunched position with head below shoulders and eyes open for at least 30 s while not engaging in any behaviours
	Hunched posture	Sitting with head the same level or lower than the shoulders; arms and limbs huddled to the centre of the body; no movement of the body or the four limbs; eyes open or unable to determine whether the eyes are open or not. When huddling, the animal can yawn or scratch.

Reference	Label	Definition
Hennessy et al., 2014	Lie	Relaxed posture with body resting on a horizontal surface. Weight is not supported by limbs; eyes are open.
	Day time sleep	Sitting or lying with eyes closed (observations made during daytime); if sitting, head must be above shoulders to differentiate from hunched posture.
Hinde & Rowell, 1962	Hunched sitting	The back is strongly curved, with the point of maximum curvature slightly more anterior than in relaxed sitting. The legs and arms are tucked close to the body, the forearms resting on the knees, hands hanging in front. The head is low, slightly withdrawn, and often resting between the knees.' The tail is curled round the side of the body. This is the common resting position of animals of middle rank. Small groups of monkeys sit together to sleep in this posture, the chin of one animal resting on the back of its neighbour. If the animal is alerted, its face is raised: the more subordinate animals tend to keep the head low and look out from under their brows, while the more dominant ones raise the head above the shoulders. The arms may be put to the ground as an intention movement of locomotion.
Li et al., 2013	Huddling	Sitting alone on the floor with head down, displaying no interest in the external environment; a self-directed behaviour
Perera et al., 2011	Anhedonia	Alone and immobile, slumped or collapsed body posture, lack of purposeful eye movements or responsiveness to environmental stimuli, rejecting social advances.
Polanco, 2021	Withdrawn	Facing a corner for an extended period of time; usually acting socially withdrawn, for at least 30 seconds

Reference	Label	Definition
Reinhardt, 1994	Depression	Reduced alertness and interest in otherwise favoured food treats
Shively & Willard, 2012	Behaviour termed depressive	Sat in a slumped or collapsed body posture, with open eyes, accompanied by a lack of responsivity to environmental events
Shively et al., 1997	Depressed	Slumped or collapsed body posture, relatively unresponsive to environmental stimuli to which other monkeys are attending
Shively et al., 2005	Depressed	Slumped or collapsed body posture accompanied by a lack of responsivity to environmental events, eyes open
Smith et al., 2006	Potential life-threatening signs	Not alert/not responding
	Potential signs of clinical issues	Hunched posture Lying on cage floor (responsive)/lethargic
Truelove et al., 2017	Withdrawn or self-directed	Inactivity: Animal is asleep or is stationary and not engaged in any active behaviours other than directing attention through looking, and so forth. Self-directed stereotypy: Repetitive and ritualistic behaviours directed at animal's own body that do not involve locomotion. Examples include eye behaviours (for example, eye poking), self-clasping, rocking, swaying, digit sucking, mouthing, skin pulling, and tail rubbing or grasping. Huddle: Inactive, hunched seated position with head at lower than shoulder level; may include the chin close to or touching the chest.

Therefore, the final definition of the *Inactive not alert* behaviour for the present study was the following: sitting or lying stationary and alone, while not looking at objects or individuals (eyes may be open or closed), and not doing anything else. This behaviour will subsequently be referred to as *Inactive not alert* behaviour. The decision regarding whether eyes open or closed is a pragmatic one to allow quantifying the behaviour even when the eyes of the monkeys are not visible.

The chosen method to measure the behaviour (i.e., presence or absence) required a minimum duration to consider the behaviour as displayed. Since the minimal duration is rarely mentioned in the previously used definitions, and when it is, no justification is provided, a preliminary analysis was conducted. The duration of the behaviour required a long enough duration to be specific to the affective state of interest and avoid false-positive results, such as including instances when the subject is observing an object on the floor. On the other hand, a relatively long duration is likely to detect only the most extreme cases of negative affective states, leading to flooring effects, and reducing the sensitivity of the behaviour as a welfare indicator. Therefore, considering that the duration of the unit of analysis is 1 minute, durations of 2 seconds, 5 seconds and 10 seconds were investigated. The frequency, defined as the proportion of 1-minute time bins in which the behaviour was observed (on a scale from 0 to 1) of the three behaviours was highly correlated (Pearson's r>0.8). However, the 2-second duration did not allow for distinguishing between the behaviour of interest and a direct gaze at an object on the ground (which is normally a rapid gaze within 5 seconds of duration followed by a movement toward the object) and was thus rejected. As a result, the 5-second and 10-second durations were initially selected. The minimum durations of 5 and 10 seconds of the behaviour showed similar variability across the subjects that displayed them (Figure 5).

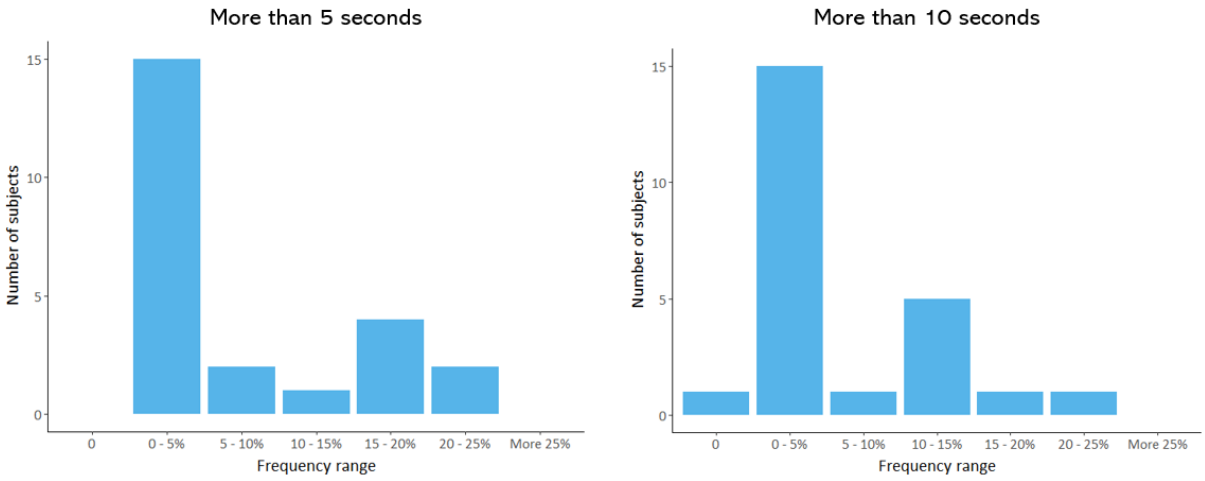

Figure 5. Variability between the display of the two durations of *Inactive not alert* behaviour. The figure shows how many subjects displayed on average the behaviour *Inactive not alert* at the following frequency and frequency ranges: 0%, 0-5%, 5-10%, 10-15%, 15-20%, 20-25% and more than 25% of the investigated time when the duration of the behaviour was encoded as left) >5 seconds; right) >10 seconds at every 1-minute time bin. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

Informed by this analysis (Figure 5), a conservative approach was followed. The longest duration (10 seconds) was chosen for having a higher likelihood of being specific to the affective state of interest (i.e., negative valence).

In addition to *Inactive not alert*, *Inactive alert* behaviour was also encoded. This behaviour was defined as sitting or lying stationary while looking at objects or individuals inside or outside the cage, and not doing anything else, for more than 10 seconds. The use of this behaviour aimed to offer information regarding the general inactivity of the subjects. This type of inactive behaviour is not associated with the affective state of the animals and is displayed when the animals are engaged in collecting information about their physical and social environment. However, an increase in this behaviour over time could also be indicative of a general ageing process (Mechling & Netz, 2009). Previous studies have encoded *Inactive not alert* and *Inactive alert* as one combined behaviour. Since *Inactive alert* behaviour is not associated with a negative affective state, combining the two behaviours might have decreased the sensitivity of the behaviour as a welfare indicator.

In the same manner, as with *Inactive not alert*, no reference duration was found in the literature for *Inactive alert* behaviour. In this case, the 10-second duration was chosen to mirror the duration of the *Inactive not alert* behaviour as a control.

See Figure **6** for a visual distinction between "*Inactive not alert"* (Figure 6.A) and "*Inactive alert*" (Figure 6.B) behaviours.

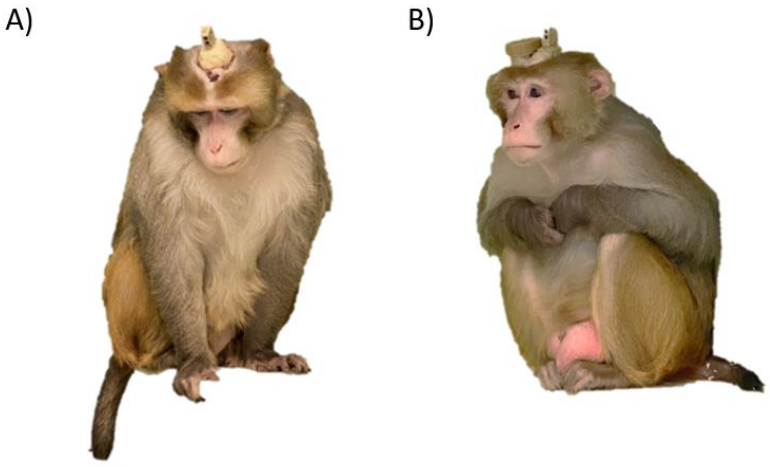

Figure 6. Subject displaying a) *Inactive not alert* behaviour; b) *Inactive alert* behaviour.

During preliminary control analysis, the behaviours *Inactive alert* and *Inactive not alert* were found not to be correlated suggesting, as predicted, that they reflect two distinct behaviours (Pearson's r=0.294) (Figure 7).

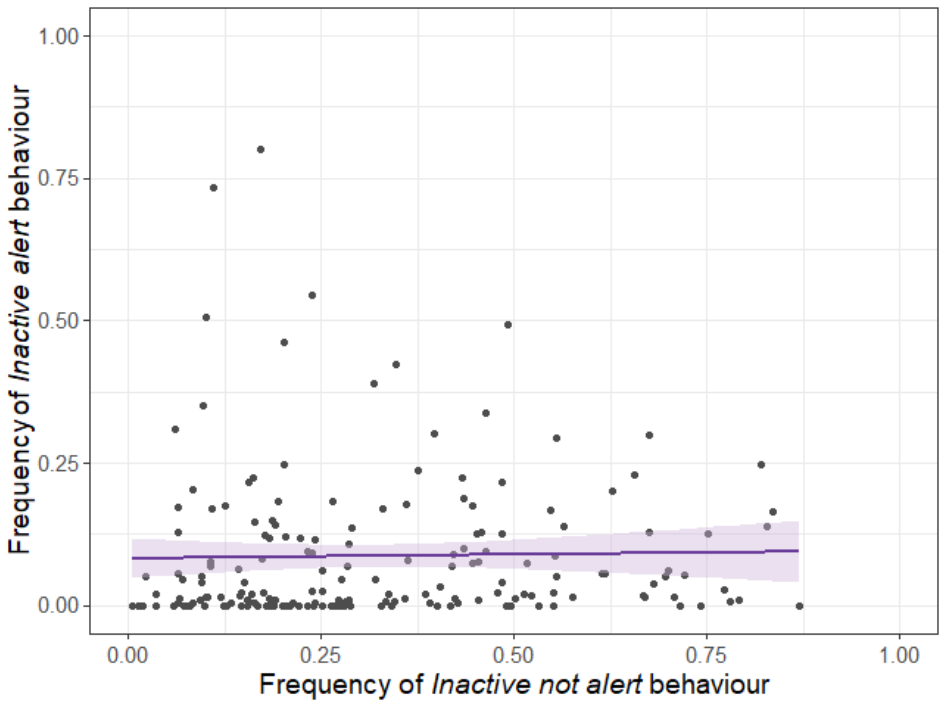

Figure 7. Correlation between *Inactive not alert* and *Inactive alert* behaviours. The observed frequency of the *Inactive not alert* behaviour is represented in the x-axis and the frequency for the *Inactive alert* behaviour in the y-axis. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

The behaviour *Inactive not alert* can be displayed both alone and in physical contact with a cage mate. Previous studies have interpreted the physical contact with a cage mate as affiliation behaviour (Aureli, 1992; Baker et al., 2014; Das et al., 1998; Judge et al., 2011; NC3Rs, 2022.) and, thus, associated it with a positive affective state. Since the affective state associated with the *Inactive not alert* behaviour when displayed with a cage mate is unknown, the decision was to encode the two sub-types of behaviours separately. In order to preserve the specificity of *Inactive not alert* behaviour as an indicator of a negative affective state, the subsequent analyses of the behavioural indicator were restricted to the instances when the focal subject displayed the behaviour alone, not in physical contact with the cage mate(s). Moreover, by excluding the instances when the behaviour is displayed in contact with a cage mate, the indicator can be used with both single-housed and social-housed macaques (as well as to compare them).

2.3.2. Other behaviours investigated

The present project has been active since 2014, gathering a large amount of data that can potentially be used for other projects. Therefore, when analysing the video recordings, additional behaviours to *Inactive not alert* behaviour were encoded. Encoding multiple

behaviours would allow us to assess the welfare of the animals with other behavioural indicators in the case that studies arise suggesting that other behaviours could be indicative of a negative affective state. Furthermore, these additional behaviours may assist with the investigation of other scientific questions. The behaviours used throughout my thesis are listed subsequently (Table 3).

Table 3. Ethogram for the relevant behavioural categories encoded in the behavioural approach of the project. The behavioural indicator of a long-lasting negative affective state, *Inactive not alert*, is underlined.

Behavioural item	Description
Inactive not alert	Sitting (sometimes with head lower than the shoulders) or lying stationary and alone, while not looking at objects or individuals (eyes may be open or closed), and not doing anything else, for more than 10 seconds.
Inactive alert	Sitting or lying stationary and alone, while looking at objects or individuals inside or outside the cage, and not doing anything else, for more than 10 seconds.
Inactive not alert with a cage mate	Sitting (sometimes with head lower than the shoulders) or lying stationary with a cage mate, while not looking at objects or individuals (eyes may be open or closed), and not doing anything else, for more than 10 seconds.
Inactive alert with a cage mate	Sitting or lying stationary with a cage mate, while looking at objects or individuals inside or outside the cage, and not doing anything else, for more than 10 seconds.
Self-groom	Stroking, picking, or otherwise manipulating own body surface (excluding head post and margins).
Self-scratching	Scratching the skin with nails.
Body shake	A dog-like shake of the whole body.
Yawning	Open the mouth widely, teeth exposed, lips retracted without vocalisation.
Pacing	Repetitive (at least twice) walking the same path in the cage with no interruption of more than 2 sec.
Out-of-sight	The subject is not visible for more than 2 sec.

2.4. Factors suspected to impact the cumulative experience of the subjects

I aimed to investigate the impact of environmental factors on the long-lasting affective state of the animals. As the subjects were housed in a controlled environment, the number of factors that could have had an impact was limited. After discussion with the researchers, technicians, and veterinarians in charge of the laboratory rhesus macaques, a consensus was reached, and five husbandry and experimental factors were identified as potentially impacting the long-lasting affective state of the animals (Table 4). In addition to these five factors, the potential effect of ageing was also investigated. These factors were suspected to be associated with a neutral, negative or positive effect on the long-lasting affective state of the subjects.

Table 4. Summary of the hypothesised impact of environmental, husbandry and experimental factors on the welfare of the NHPs. The symbol "+" refers to a positive impact on the welfare of the NHPs. The symbol "-" refers to a negative impact on the welfare of the NHPs. The term "neutral" refers to neither a positive nor a negative impact on the welfare of the NHPs.

Factor	Definition	Predicted impact on the welfare
Experimental involvement	Number of days the subject was involved in an experiment	-
		+
NSAID	Number of days under non-steroidal anti-inflammatory drug treatment	-
		+
Ketamine	Number of days the subject was anaesthetised with Ketamine drug	-
		+
Dura scrapes	Number of dura scrapes performed on the subject	-
Antibiotics	Number of days under antibiotic treatment	-
Ageing	Age of the subject in days	neutral
		-
		+

The first factor we suspected to have an impact on the cumulative experience of the animals is their involvement in experiments. This factor integrates a multitude of events that could potentially have a negative impact on the welfare of the animal such as social separation from the group (with the deprivation of visual and olfactory contact with conspecifics), head restraint or transport. However, other events might have a beneficial impact such as social interaction with the researchers, control over rewards or cognitively enriching tasks. The long-term effect of all these events has not been studied and thus, the net potential effects of this factor on the cumulative experience of the animals are unknown. Finally, as a result of involvement in experiments, the animals will be exposed to other more invasive and potentially detrimental factors that may impact their welfare (i.e., the need for antibiotics, surgeries, or anti-inflammatory drugs). These aspects will be covered by the other factors investigated, which will be explained subsequently.

When it comes to the predictors of *NSAID* and *Ketamine*, no *a priori* hypothesis was made of their impact on the long-lasting affective state of the animals since previous studies supported both directions. The NSAID factor integrates the number of days the animals received a non-steroidal anti-inflammatory drug. Since depressive disorders are suspected to be related to inflammatory processes (Hennessy et al., 2017), the treatment with NSAIDs could have an anti-depressant impact on the affective state of the animals. However, the NSAID factor does not only integrate the potential effect of the drug but also the association with events hypothesised to be detrimental such as fights and pathologies. Consequently, no prediction was made on the direction of the effect of this factor.

Similarly, ketamine, a sedative commonly used to induce anaesthesia, would normally be hypothesised to have a negative impact on the long-lasting affective state of the animals, since extended exposure to anaesthetics might cause neurodegeneration and behavioural deficits in NHPs (Paule et al., 2011). Nonetheless, ketamine has also been observed to hold neuroprotective (Hudetz & Pagel, 2010) and anti-depressant effects in humans (Browne & Lucki, 2013). As the literature supports both a positive and a negative impact of ketamine on the long-lasting affective state of the animal, no *a priori* predictions were made.

Scrapes and *Antibiotics* factors were hypothesised to have a detrimental impact on the welfare of the subjects. Antibiotics were given to the subjects because of events such as surgeries, infections or fight-related injuries which are considered negative for the welfare of the animals. Therefore, this factor was predicted to have a negative impact on the long-lasting affective state of the animals. The number of dura scrapes (Scrapes), an invasive surgical procedure required for intra-cerebral electrophysiological recordings, was also predicted to

pose a detrimental impact on the welfare of the subjects. This procedure might not only be detrimental by itself (Pfefferle et al., 2018), but is also highly correlated with the number of intra-cranial electro-physiological recordings performed on the animal, a factor which causes micro-injuries of the brain tissue and may also have a detrimental impact on the cumulative experience of the subjects.

Additionally, since I aimed to investigate the change in affective states of the animals over time, the potential effect of ageing was also considered in the analysis. Ageing is a natural event and, therefore, the potential impact on the long-lasting affective state of the animals could be considered neutral. However, this factor could also integrate the effect of other husbandry-related factors (e.g., captivity, cage-cleaning) that increase over time and are not included in the other predictors. Captivity is predicted to induce a negative effect on the cumulative experience of the subject, as it limits the range of behaviours they can express as well as the social relationships they can form (Animals in Science Committee, 2017). For the other events integrated into this category, such as cage-cleaning, no *a priori* hypothesis of their impact was made, as there are no known studies investigating their effect. On the other hand, the time at the research facility could also have induced habituation to potential negative stressors, resulting in an ultimately positive impact on the welfare of the animal. This factor is also predicted to cause non-welfare related variations in some of the indicators used. Ageing is known to increase inactivity in humans (Mechling & Netz, 2009) and is expected to have the same impact on other animals like rhesus macaques. Therefore, an increase in the *Inactive not alert* behavioural indicator could be explained by a general increase in inactivity due to natural ageing. Alopecia is known to increase with age in many species with macaques being one of them (UNO, 1986); therefore, an increase in alopecia with age, unrelated to a decrease in welfare, is predicted. However, for these two indicators, we have no detailed data regarding the exact age at which the ageing process might start impacting them. Therefore, we cannot discard that for the specific age range present in this study, ageing has no non-welfare impacts. As for weight, rhesus macaques are expected to grow and thus increase in weight until the ages of 6 to 8. Consequently, an increase in weight unrelated to the affective state of the animals is expected during those ages. Moreover, an absence of an increase in weight during those ages would be indicative of a suboptimal welfare status. As for the other indicators, no non-welfare related effect is expected to be caused by ageing.

These factors were extracted retrospectively from veterinary and technicians' records.

2.5. Statistical analysis

All our analyses focused on investigating within-subject effects. We aimed to compare the difference in the frequency within each subject when exposed to the two conditions (Chapter 3) and over time (Chapters 4, 5 and 6). To focus on the within-subject effect, "Subject" was declared as a random effect in all the models of this project. However, this method is not enough to eliminate the between-subject effects. Therefore, on top of this, all the independent variables used in the models were within-subject mean centred. This method consisted of subtracting the mean subject value from each data point (Figure 8). This method was the best option to analyse the current dataset as found during simulation analyses.

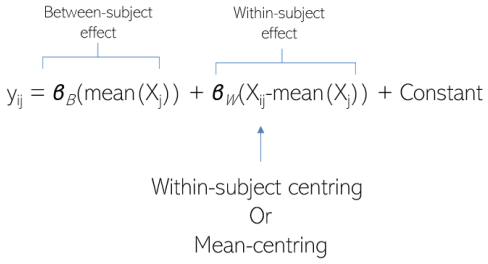

Figure 8. The formula describes within-subject mean-centring. The *X* indicates the value of each independent variable, and the *ß* is the slope (coefficient) value of each independent variable. The *Y* represents the dependent variable, in this case, the welfare indicator.

The statistical analysis of the data of all the studies was performed using the statistical programme R, version 4.0.5, with the exception of the analysis for the MRI data (Chapter 6). The p-value was calculated with the *lmerTest* (Kuznetsova et al., 2017) package, implementing Satterthwaite's method for denominator degrees of freedom which are found to produce fewer Type 1 errors (Luke, 2017). The statistical significance was set at an α-value of 0.05 unless stated otherwise. The normality of the data was assessed by using the *descdist* function in R (Delignette-Muller & Dutang, 2015), which assessed the distribution of the residuals of the model and determined which type of probability distribution should be used. All the statistical coefficient results will be shown with no standardization unless stated otherwise.

When assessing the fit of the models that investigated the factors responsible for the variability in the indicators, we used a model comparison approach. All the predictors were included in an original model, and subsequently, the model comparison was done using Akaike's information criterion (AIC) approach allowing for the comparison of all the possible combinations of predictors simultaneously. This approach estimates the rank of the models based on their AIC value with the lowest value indicating a better fit for the data. Therefore, it is based on the goodness of fit of each model, and not on the significance level of the predictors.

A small difference between the AIC values of different models implies that the models are competing in fit, with an AIC difference of fewer than 2 units indicating that the models are considered to be equally as good in explaining the variability of the data. This method also favours the model with fewer predictors to control for oversaturation (Symonds & Moussalli, 2011). Once the best (or multiple best) model is identified, this model is tested against a null model.

Chapter 3: Acute effect of fluid control on the affective state of rhesus macaques

3.1. Abstract

Rhesus macaques are widely used in biomedical research due to their phylogenetic proximity to humans and their ability to perform complex cognitive tasks. Some studies require a large number of trials and therefore a high motivation from the subjects. To achieve this, fluid availability can be controlled in the home cage, and subsequently used as a reward during experimental tasks. Whilst previous studies have shown no significant physiological impact on the animals, the potential psychological effect of this routine procedure is still debated.

Using an unprecedented sample size of 28 socially housed adult macaques (11 females and 17 males), this study aimed to investigate the acute effect of fluid control on macaques' welfare using a behavioural approach. We recorded the home-cage behaviour of animals enrolled in fluid control protocols due to their participation in unrelated neuroscientific and behavioural experiments. The protocols involved a maximum of 5 consecutive days of fluid control followed by a minimum of 2 days of free access to fluid. During the fluid control condition, animals had controlled access to fluid in their home cages and unlimited access to fluid as a reward during the experiment. Behaviour was recorded in the early morning for 45 minutes/day. Two conditions were investigated: after at least 12 hours of fluid control and after access to fluids *ad libitum*. Animals' affective state was assessed by quantifying the frequency of (1) pharmacologically-validated behavioural indicators of anxiety (i.e. self-scratching, body shaking and self-grooming); and (2) behaviours suspected to indicate a negative acute affective state but still lacking proper validation (e.g. pacing, yawning, *Inactive not alert* behaviour). 700 hours of video footage were analysed, covering up to 6 years of intermittent fluid control. Using generalised linear mixed models and a within-subject mean-centring approach, we disentangled the between- and within-subject effects of fluid control and control for the effect of time. Bonferroni correction was applied to control false-positive results (Bonferroni, 1936).

We found no effect of fluid control on the frequency of any behaviour, nor any indication of habituation or sensitisation over years. This study suggests that the fluid control protocol, as implemented in this study, does not have a negative impact on the acute affective state of macaques.

3.2. Introduction

3.2.1. Background

Non-human primates have the ability to perform complex cognitive and behavioural tasks that require high manual dexterity and behavioural flexibility. Together with their phylogenetic proximity to humans, these abilities make NHPs highly valuable animal models in neuroscience research (See Chapter 1, section 1.2). Many behavioural tasks implemented in neuroscience experiments require a large number of consecutive trials to obtain scientifically reliable results. However, such a high number of trials is only reached when the NHPs are highly motivated to perform them. Neuroscientists generally use one of two methods to increase the motivation of the NHPs: fluid control or food control. Depending on the nature of the task that needs to be performed by the NHP, one method will be favoured over the other. These methods consist of using small amounts of fluid or food during the experimental tasks as positive reinforcement (i.e., *rewards*). The use of fluid control is favoured when (1) the nature of the experiment does not allow chewing movements that can induce electrical noise and motion artefacts (e.g., electrophysiological recordings, neuroimaging experiments), (2) the reward needs to be automatically and accurately quantifiable, (3) the reward needs to be forwarded without the animal seeing it, (4) the reward is automatically forwarded with a controlled and standardized equipment often with a specific time resolution (5) the experimental sessions are long and require an extensive number of small rewards to be forwarded (Prescott et al., 2010; Willems, 2009).

In the UK, most food control protocols will schedule the food intake, while the fluid control protocols might require limiting the amount of fluid the animals can consume outside the experiments per day (Prescott et al., 2010; Willems, 2009). Although fluid control paradigms have been found to pose no negative welfare impact at a physiological level (Yamada et al., 2010; Gray et al., 2016), the psychological impact of the procedure is still debated (Desimone et al., 1992; Orlans, 1991; Prescott et al., 2010; Westlund, 2012).

3.2.2. Previous studies

Gray and colleagues (2016) investigated the effect of fluid control on the welfare of four male laboratory rhesus macaques using 13 different behaviours as welfare indicators. They compared free access to fluids to a fluid control protocol of two differing durations: 5 versus 7 consecutive days. They hypothesised that if there was a negative impact of fluid control on the animal's welfare, the frequency of the behavioural indicators would increase significantly. They found a significant decrease in the frequency of the behavioural categories referred to as *inactivity, social, locomotion, self-grooming, self-directed behaviours, body shaking* and

yawning; and no significant change in the rest of the measured behavioural categories when the animals were fluid-controlled with either duration. Therefore, they concluded that the investigated fluid control protocols posed no welfare concern to the laboratory macaques. However, there are several limitations to consider. The sample size was small and included only male macaques. Additionally, the free access condition period occurred during a time when none of the animals at the facility were taking part in experiments and the human presence was lower than usual, which might have compromised the comparability between the control condition and fluid-controlled conditions. Finally, some of the behavioural categories chosen as welfare indicators were not validated as indicators of short-term negative affective states and might have not been sensitive enough to detect changes in the affective state of the animals.

In 2014, Hage and colleagues also investigated the effect of fluid control, in this case, on the welfare of seven laboratory male macaques. They compared a 12 consecutive day fluid control protocol with a free access condition of two different durations: 2 days and 9 days of free access to fluid. They used 11 behavioural categories as welfare indicators that were measured at the beginning (day 1 to 5) and the end (day 8 to 12) of the fluid control periods. The behaviours were measured while the animals were in their home cage, just after they had been involved in experimental tasks. They found no significant change in the frequency of the investigated behaviours between the two durations of free access condition neither for the beginning of the fluid control period nor for the end, concluding consequently, that the fluid control used had no negative impact on the welfare of the animals. In this study, there were also some limitations to consider. Firstly, the behaviours were always recorded during periods of fluid control, and if the effect of fluid control on the welfare was transient, this comparison might not be sensitive to detect any effect of the differences in the duration of the free access periods. As in the case of Gray and colleagues (2016), the sample size was small and did not include any female macaques. The behavioural categories chosen as welfare indicators might have also not been sensitive enough to detect changes in the affective state of the animals. Lastly, the behaviours were encoded using a human observer instead of a video recording device, which is known to bias the behaviour displayed by the animals (Iredale et al., 2010).

Using cortisol as the welfare indicator instead of behavioural measures, Pfefferle and colleagues (2018) investigated the effect of fluid control on 16 male laboratory macaques before and after the animals were exposed to a stressor. In this case, their fluid control protocol restricted the access to fluids for the experimental task, with no extra fluid available in the home cage, regardless of the amount consumed during the experiment. In this study, they found a significant increase in the cortisol measures before the exposure to the stressors when the

animals were fluid-controlled, concluding that the fluid control protocol might be having a detrimental impact on the welfare of the animals. Nonetheless, as previously explained (See section 1.4.3), cortisol measures might be reflecting a change in the arousal of the affective state of the animals and not in the valence, which would hinder the interpretation of the obtained results.

On top of discussed concerns, all of the studies above focused on the effect of fluid control over a short time period. Gray's (2016) study recorded the behaviours over a total of 4 months and Hage's (2014) study over 5 months, while Pfefferle and colleagues (2018), conducted the collection of the samples for each subject within the same day. None of the three studies considered potential habituation or sensitisation to the fluid control by implementing a time dimension into their analysis. Considering that NHPs involved in neuroscience research can be in fluid control protocols for several consecutive years, investigating the effect of fluid control protocols over a long time seems key to ensuring the welfare of the animal models.

Consequently, due to the contradictory results and limitations found in previous research, this study aimed to investigate the acute effect of fluid control on the welfare of a large number of laboratory macaques using behavioural indicators. As a secondary aim, we planned on investigating potential habituation and sensitisation to fluid control over the years.

3.3. Methods and materials

3.3.1. Animals and husbandry

A subset of the dataset described in Chapter 2.2 was used for this study. The animals that were singly housed during the recording were excluded from this analysis. These data were excluded as in the current dataset most chronic single housing occurred during the last months of the animal's experimental life. Therefore, the non-homogeneous distribution of a factor known to have a detrimental impact on the welfare of the animals (See Chapter 2, section 2.2) could bias the analysis of the data.

The dataset was also restricted to animals for which behavioural data were recorded in both conditions (*ad libitum* water vs. fluid control). All included animals had at least two sessions where the individuals had free access to fluid and two sessions where they were subjected to fluid control for at least the previous 12 hours.

3.3.2. Fluid control protocol

Fluid control protocols differ between institutions as well as between projects. The animals in this study were all subjected to the same fluid control protocol.

For this protocol, a daily minimum amount of fluid is identified per individual. Considering the *ad libitum* water intake as the starting point, the researchers slowly decrease the daily fluid available in the home cage and quantify the number of trials the animals perform until the animal reaches the motivation required to perform the necessary number of trials in each experimental task. An absolute minimum computed according to the weight of each subject also limited the lower limit of daily fluid that could not be reduced. Across the sample, the minimum varied between 150 and 400 ml/day.

This minimum must always be reached and can be surpassed during the experiments. If the animals do not reach the minimum during the experiments, the amount left to reach it is provided at the end of the experimental day when the animals are back at their home enclosures. Animals can only be fluid-controlled for five consecutive days, which must be followed by at least two days of *ad libitum* water access at their home enclosures.

3.3.3. Indicators of acute negative affective state

As described in section 1.4.4.2 the only validated indicators of acute negative affective state for rhesus macaques are displacement behaviours (self-scratching, self-grooming and body shaking). Therefore, this indicator was the primary outcome measure used in this chapter to investigate the acute effect of fluid control on the welfare of laboratory macaques.

However, in case of a negative result in the primary analysis, secondary analyses, using potential indicators of acute negative affective state, were planned. These potential indicators were Yawning, Pacing and *Inactive not alert* behaviour. Research done with these three behaviours is described in sections 1.4.4.1, 1.4.4.2 and 1.4.4.4.

To avoid the limitations described in section 1.4.4, regarding the risk of false positives and false negatives as well as the loss of sensitivity, we only used potential indicators with strong evidence supporting their use and which are independent of each other. For this reason, we did not include self-scratching and affiliative behaviours.

Self-scratching would not be independent nor more sensitive than the category of displacement behaviours (see section 1.4.4.2). Therefore, it would not pose any benefit to using it in the present analysis.

Affiliative behaviours were not included for several reasons. This behavioural category has been interpreted as an indication of both good and bad welfare status (see section 1.4.4.3), considered in most research facilities as an indication of a healthy relationship between the NHPs displaying them. Moreover, all the studies revealing increases in affiliative behaviours

after the subjects have been exposed to a stressor show (1) initial increases in displacement behaviours, implying that affiliative behaviours would not offer additional information to displacement behaviours, and (2) they have always been related to stressors of a social nature, which is not the case for our study. Finally, affiliative behaviours have also been suggested as coping mechanisms, which could hinder a reliable interpretation of the results (see section 1.4.4.3).

3.3.4. Statistical analyses

The data were modelled using linear mixed models with the *lme* function from the *nlme* package (Pinheiro et al., 2011). The distribution of residuals was found to be non-normal for all the following models of yawning behaviour (kurtosis > 7), pacing behaviour (kurtosis > 7) and *Inactive not alert* behaviour (skewness > 2, kurtosis > 7). Therefore, the dependent variable was transformed via a square root for all the models that include those behaviours.

The analysis of the data was done in three steps that will be explained subsequently.

3.3.4.1. Effect of condition

The first analysis consisted of investigating the effect of fluid control on the frequency of displacement behaviours. This effect of the condition was investigated by using general linear mixed models in which the independent variable was the condition (i.e., *ad libitum* vs. fluid-controlled) and the dependent variable was the frequency of the behaviour.

If fluid control was having an acute detrimental impact on the affective state of the subjects, we predicted that the fluid control condition would increase the frequency of the behavioural indicators used.

3.3.4.2. Effect of consecutive days of fluid control

During the second analysis, we investigated the effect of consecutive days of fluid control on the frequency of displacement behaviours. For this analysis, only sessions during the fluid control condition were included. The maximum number of consecutive days of fluid control in the investigated protocol was five.

This effect was investigated by using general linear mixed models in which the independent variable was the number of consecutive days under fluid control (i.e., from one day to five days) and the dependent variable was the frequency of the behaviour.

In this case, if there was a detrimental impact of consecutive days of fluid control on the affective state of the animal, we predict to observe a significant increase in the behaviour over the consecutive days of fluid control.

3.3.4.3. *Effect of years of intermittent fluid control*

The third step of the analysis investigated the acute effect of fluid control when animals are subjected to it over years. With this analysis, we aimed to investigate any potential habituation and/or sensitisation processes.

The effect of years of intermittent fluid control was investigated by using general linear mixed models in which the independent variable was the total number of days since the first fluid control of the subject in interaction with the condition, and the dependent variable was the frequency of the behaviour.

A second analysis was done to investigate potential sensitisation and/or habituation effects on fluid control. For this, we used the two groups of subjects present in our data set: subjects who had been investigated years after their first fluid control (i.e., involvement in any experiments) and subjects who were investigated weeks or months after their first exposure to fluid control (see section 2.2). This gave us the opportunity to investigate the effects of fluid control in two time windows: an early effect and a late effect. This in turn could unmask potential sensitisation and/or habituation effects to the fluid control.

For this second analysis, the model will consist of a triple interaction between three independent variables: the condition, the days since the first fluid control and the group of the subject (“Early” vs “Late”).

The diagram below shows the predicted changes in the frequency of the behavioural indicator in the case of habituation and sensitisation (Figure 9). Panel A illustrates the scenario in which subjects are habituated to fluid control, which is displayed by a decrease in the frequency of the behaviour over time during the fluid-controlled condition and a decrease in the frequency difference between the two fluid control conditions (Figure 9.A). However, this scenario can only be confirmed in panel B (Figure 9.B), where the effect of the fluid control condition is investigated per group of subjects (“Early” vs “Late”). This scenario would be illustrated in the following manner: the “Early” group of subjects would display a decrease in the frequency of the behavioural indicator over time when they are fluid-controlled, together with a significant interaction between the frequencies of the two fluid control conditions. This latter interaction could be visualized with a significant decrease in the frequency difference between the two fluid control conditions over time. In addition, we would observe no change

in the frequency of the behavioural indicator over time for the "Late" group in any of the conditions, nor any significant difference between the frequencies in each condition.

In the case of a sensitisation scenario, the first analysis of the effect of years of intermittent fluid control would reveal an increase in the frequency of the behavioural indicator overtime during the fluid control condition, together with an increase in the frequency difference between the two fluid control conditions (Figure 9.C). For confirmation, the analysis illustrated in panel D (Figure 9.D) would be performed. In this analysis, the "Early" group would display an increase in the frequency of the behavioural indicator during the fluid control condition and a significant interaction between the two conditions. Once more, this latter interaction could be observed visually in the illustration by an increase in the frequency difference between the two conditions over time. Finally, the "Late" group would display no change in the frequency of the indicator over time and a significant difference in the frequency of the behaviour between the two conditions.

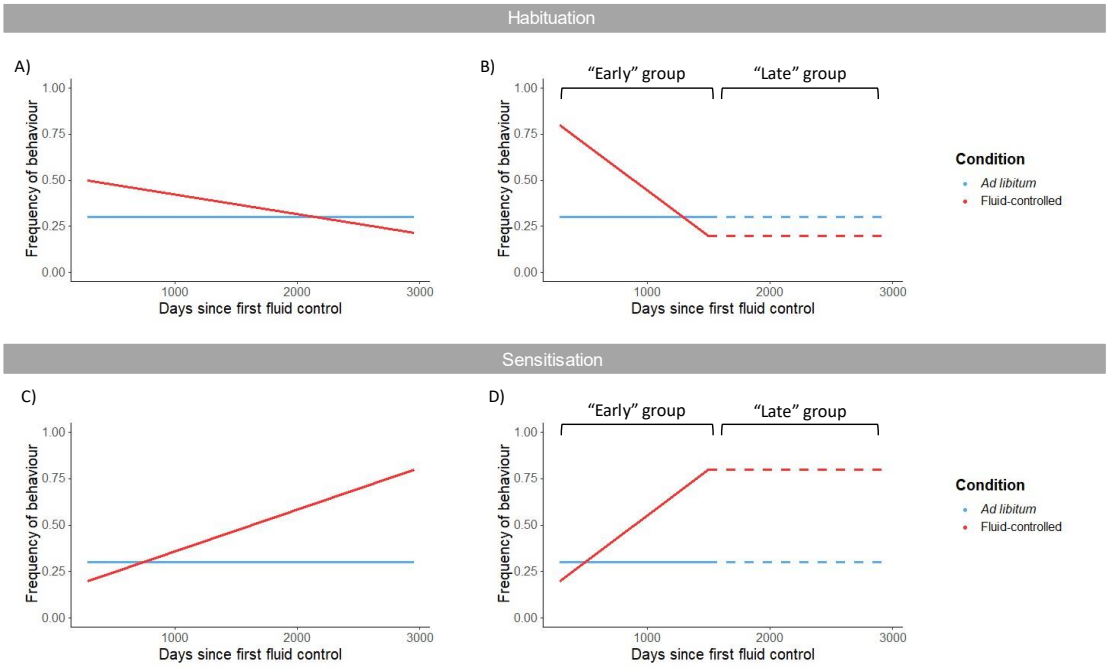

Figure 9. Diagram of predicted changes in the frequency of the behavioural indicator in the case of habituation and sensitisation to fluid control. Habituation scenario when investigating all subjects together (panel A). Habituation per group of subjects (panel B). Sensitisation scenario when investigating all subjects together (panel C). Sensitisation per group of subjects (panel D).

For the two models of this analysis, the results will be shown after scaling due to the small size of the unit in these analyses (e.g., days).

3.4. Results

The final data set consisted of 932 daily sessions from a total of 28 laboratory rhesus macaques: 17 males and 11 females. 53.7% of the sessions were part of the *ad libitum* condition.

3.4.1. Displacement behaviours

The difference in the frequency of displacement behaviour between the "Fluid-controlled" and *ad libitum* condition was not significant (ß=-0.006, SE=0.01, DF=903, F=0.358, P=0.549) (Figure 10) suggesting no acute detrimental impact of fluid control on the welfare of the subjects.

Figure 10. Effect of condition on the frequency of displacement behaviours. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. The *ad libitum* condition is shown in blue and the "Fluid-controlled" condition in red. Each coloured dot represents a session, and the data is jittered for illustration purposes. The black dots represent the average per condition.

Fluid control could have had an impact only after several consecutive days. Therefore, the change in the frequency of the behaviour over the five maximum days of fluid control was investigated. This analysis included only fluid controlled sessions (n=430). The frequency of the behaviour did not increase over the consecutive days of fluid control (ß=-0.013, SE=0.005, DF= 403, F= 7.294, P=0.007), but it decreased significantly (Figure 11). The average frequency after five days of consecutive fluid control was lower than the average frequency during the *ad libitum* condition.

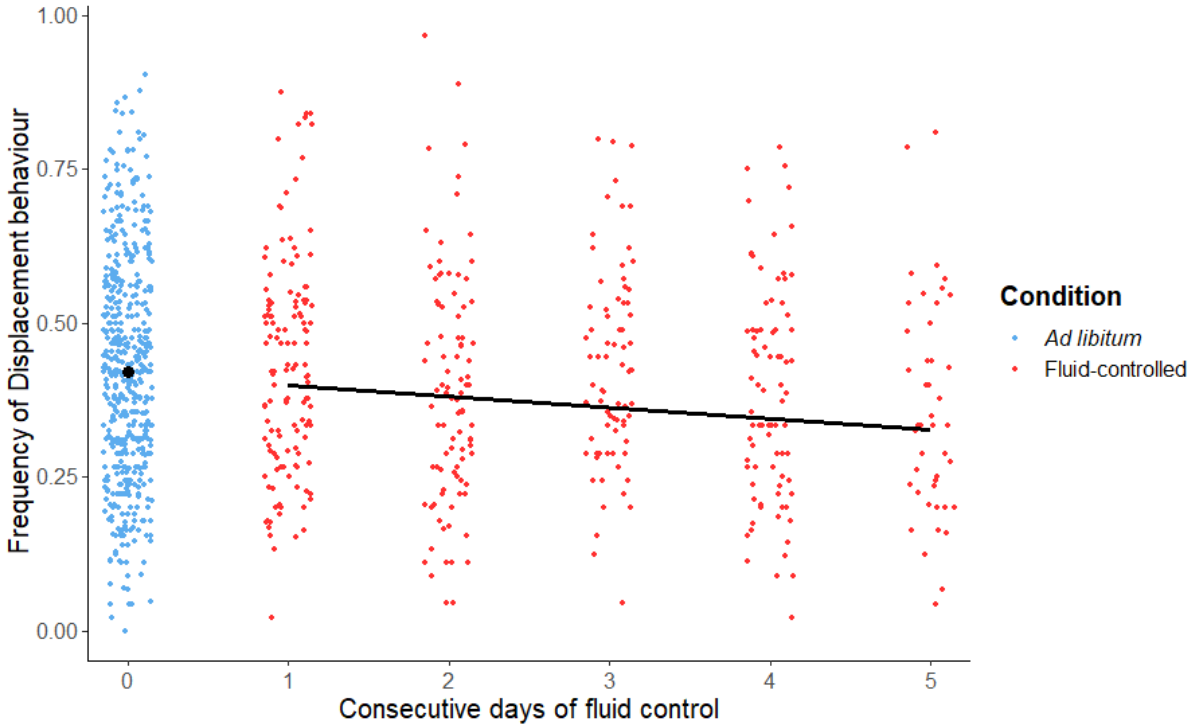

Figure 11. Effect of consecutive days of fluid control on the frequency of displacement behaviours. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. The *ad libitum* condition is shown in blue and the "Fluid-controlled" condition in red. Each coloured dot represents a session, and the data is jittered for illustration purposes. The black dot represents the average frequency of the *ad libitum* condition, and the black line represents the change in average frequency for the behaviour.

Next, the change in the frequency of the behavioural indicator was investigated over the years of intermittent fluid control. No interaction was found between the years since the first fluid and the fluid control condition of the subjects ("Fluid-controlled" vs *ad libitum*) (ß=-0.006, SE=0.005, DF=901, F=1.959, P=0. 162) (Figure 12.A).

Lastly, we investigated a potential habituation and sensitisation effect of fluid control by investigating the effect of fluid control over the years separately for the two groups of subjects ("Early" vs "Late") (Figure 12.B). No significant three-way interaction was found between the days since the first fluid control, the effect of the condition and the group of the subjects (ß=0.009, SE=0.006, DF=898, F=2.404, P=0.121).

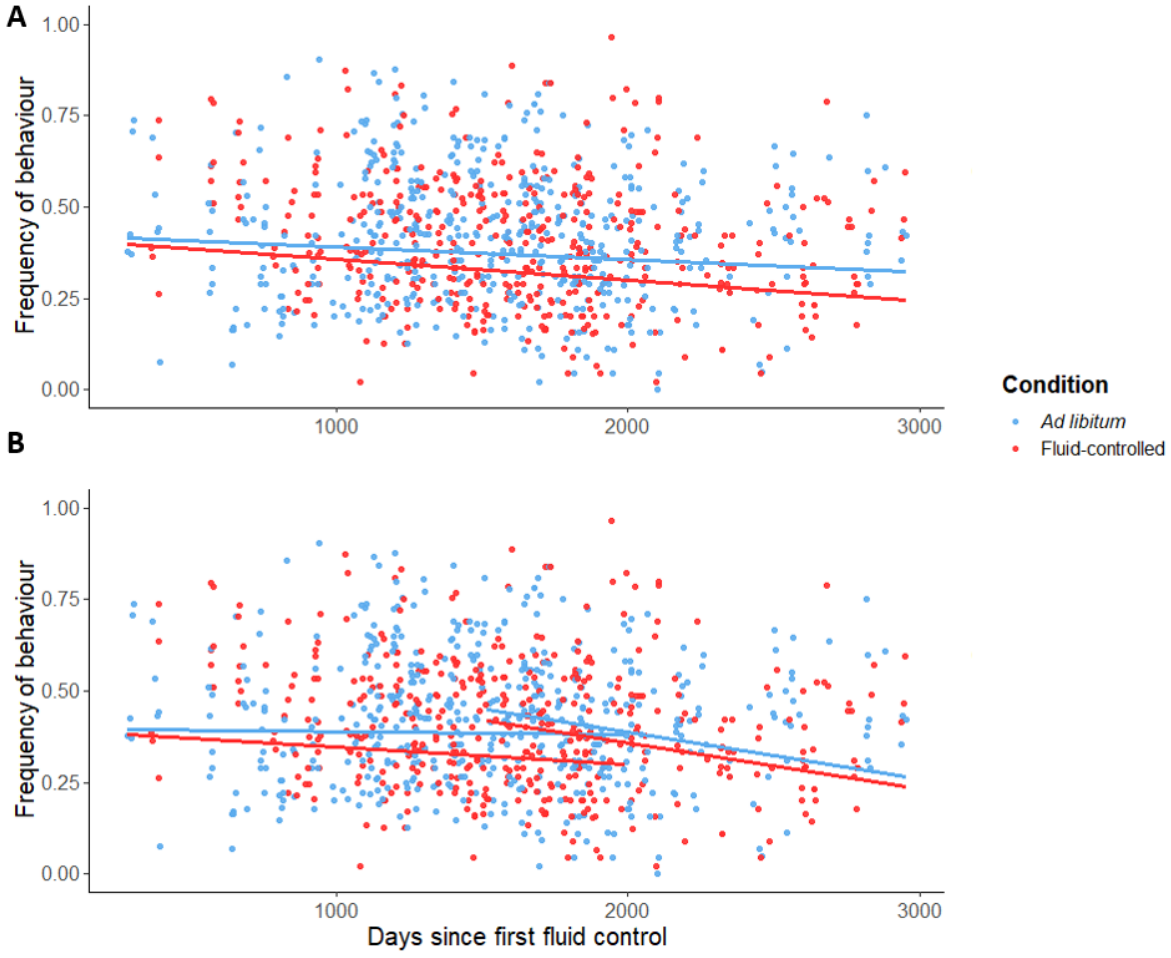

Figure 12. Effect of fluid control on the frequency of displacement behaviours over days since the first fluid control (A) and when divided into the "Early" (left side of the graph) and "Late" (right side of the graph) groups (B). The *ad libitum* condition is shown in blue and the "Fluid-controlled" condition in red. Each coloured dot represents a session. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

As no significant negative effect of fluid control was found on displacement behaviours, the three steps would be repeated for each potential indicator of acute negative welfare proposed (i.e., Yawning, Pacing and *Inactive not alert* behaviour). For the analysis of the potential indicators, I corrected for multiple comparisons using the Bonferroni correction (Bonferroni, 1936). After the correction, the α value was reduced to 0.016.

3.4.2. Potential indicators of acute negative affective state

3.4.2.1. Yawning

Yawning behaviour frequency did not change according to the fluid control condition (ß=0.002, SE=0.008, P=0.806) (Figure 13.A). The consecutive days of fluid control did not increase the frequency of the behaviour (ß=-0.002, SE=0.004, DF=403, F=0.412, P=0.520) (Figure 13.B). The analysis of the effect of years of intermittent fluid control showed no significant interaction between fluid control and the days since the first fluid control (ß=-0.004, SE=0.004, DF=901, F=1.15, P=0.282), (Figure 13.C). Finally, the triple interaction between the days since the first fluid control, the effect of condition and the group of the subjects revealed no significant results (ß=0.003, SE=0.005, DF=898, F=0.592, P=0.441) (Figure 13.D).

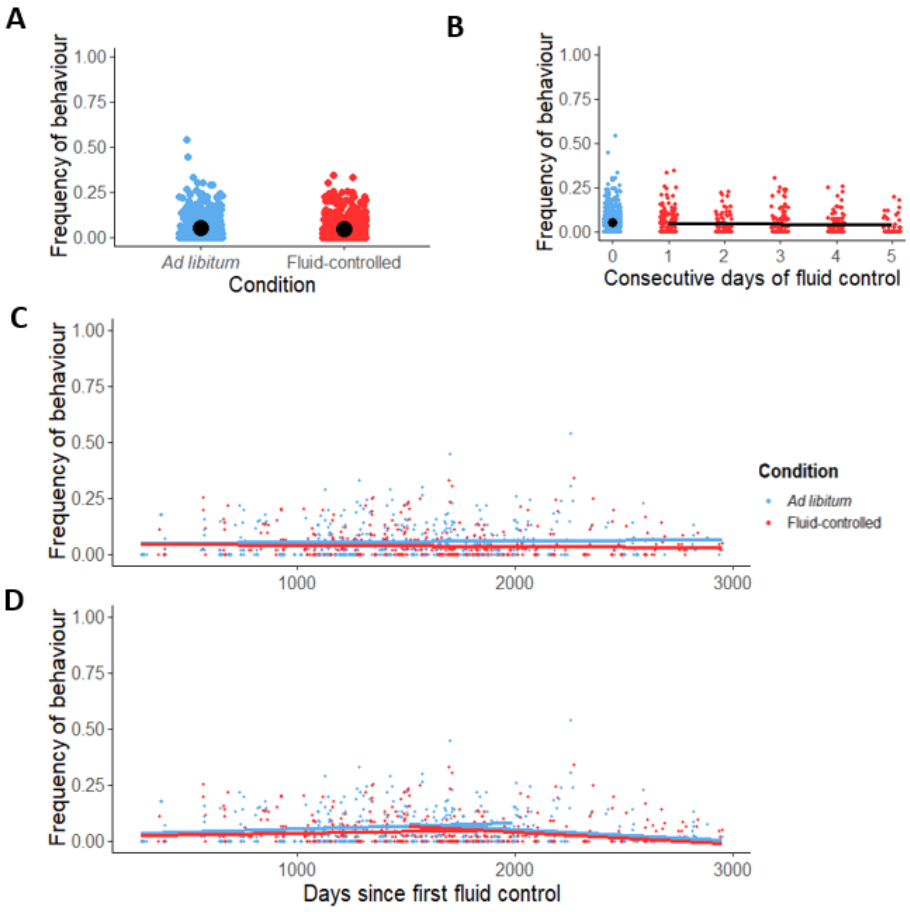

Figure 13. Changes in the frequency of Yawning behaviour. (A) Effect of condition (fluid-controlled vs *ad libitum*). (B) Over the consecutive days of fluid control. (C) Over time, starting with the first date of fluid control. (D) Over time, with the two groups of subjects. Each coloured dot represents a session, and in panels A and B the data is jittered for illustration purposes. The

frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

3.4.2.2. Pacing

For the analysis of the Pacing behaviour, the data set was restricted to the animals that displayed Pacing behaviour at any of the investigated sessions. As a result, the data was reduced to 665 sessions with 19 subjects (7 females and 12 males).

The frequency of Pacing behaviour was not found to be influenced by the fluid control condition of the subjects (ß=-0.009, SE=0.019, DF=645, F=0.256, P=0.612) (Figure 14.A). The number of consecutive days under fluid control showed no signs of having an impact on the frequency of Pacing behaviour (ß=-0.017, SE=0.009, DF=273, F=3.117, P=0.078) (Figure 14.B). The frequency of the behaviour in both conditions is decreasing over the years, with no significant interaction between the two conditions (ß= 0.008, SE= 0.008, DF=643, F=1.03, P=0.310) (Figure 14.C). When the two groups of subjects were divided, the interaction between conditions, days since the first fluid control and the group of the subject was not significant (ß= 0.008, SE= 0.010, DF=640, F=0.644, P=0.422) (Figure 14.D).

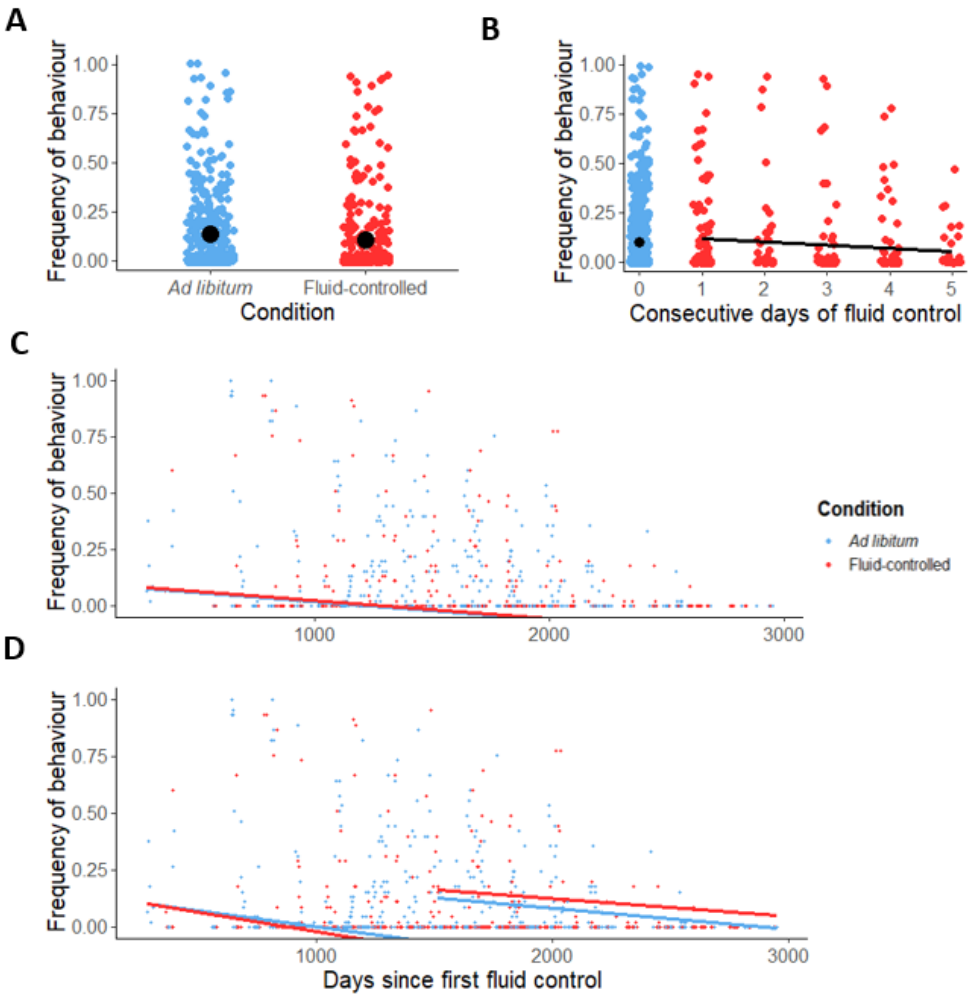

Figure 14. Changes in the frequency of Pacing behaviour. (A) Effect of condition (fluid-controlled vs *ad libitum*). (B) Over the consecutive days of fluid control. (C) Over time, starting with the first date of fluid control. (D) Over time, with the two groups of subjects. Each coloured dot represents a session, and in panels A and B the data is jittered for illustration purposes. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

3.4.2.3. Inactive not alert

Inactive not alert behaviour frequency was not significantly higher in the "Fluid-controlled" condition than in the *ad libitum* condition (ß=0.009, SE=0.013, DF=903, F=0.438, P =0.508) (Figure 15.A). The frequency of *Inactive not alert* behaviour did not increase significantly over the consecutive days of fluid control (ß=0.014, SE=0.007, DF=103, F=3.956, P =0.047) (Figure 15.B). The analysis of the effect of years of intermittent fluid control revealed no significant interaction between the fluid control condition and the days since the first fluid control (ß= -0.003, SE= 0.006, DF=901, F=0.304, P =0.581) (Figure 15.C). The last analysis on the effect of intermittent fluid control with the two groups of subjects showed no significant interaction between the condition, the groups of subjects and the days since their first fluid control (ß= -0.01, SE= 0.007, DF=898, F=1.671, P=0.196) (Figure 15.D).

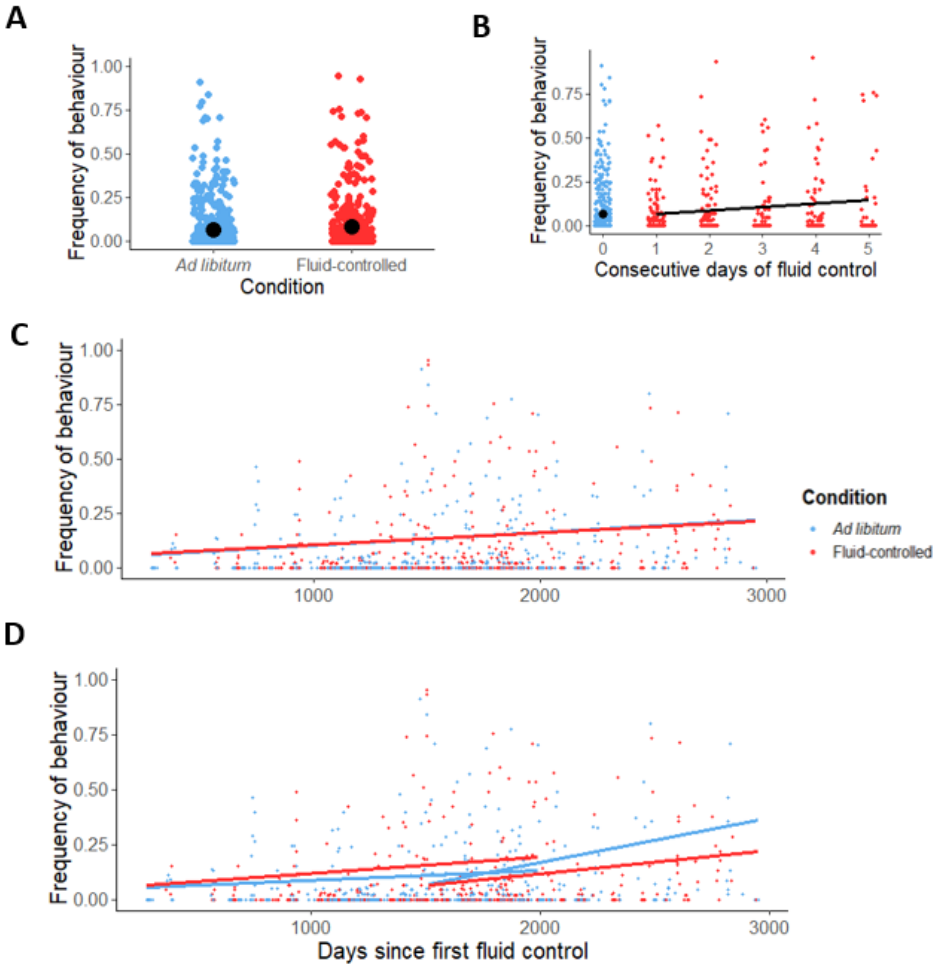

Figure 15. Changes in the frequency of *Inactive not alert* behaviour. (A) Effect of condition (fluid-controlled vs *ad libitum*). (B) Over the consecutive days of fluid control. (C) Over time, starting with the first date of fluid control. (D) Over time, with the two groups of subjects. Each coloured dot represents a session, and in panels A and B the data is jittered for illustration

purposes. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1.

3.5. Discussion

Fluid control protocols are broadly used in a multitude of research facilities with laboratory rhesus macaques. If this protocol was to have a negative impact on the welfare of the animals the consequences for the research community could be crucial. It is therefore imperative to reach a solid conclusion about the potential impacts of this protocol.

The fluid control condition showed to have no impact on displacement behaviours; nor did it on the potential indicators of acute negative affective state. This result suggests that fluid control has no impact on the acute affective state of the investigated animals.

Our analysis of the effect of consecutive days of fluid control, revealed similar results, with no significant increases in the welfare indicators over the consecutive days of fluid control. These results suggest that the use of fluid control consecutively, up to a maximum of five consecutive days, poses no detrimental impact on the affective state of the animals.

Finally, the intermittent use of fluid control over an extensive period (up to 11 years) had no significant impact on the frequencies of the indicators of acute negative affective state.

In conclusion, our results suggest that the fluid control protocols, used at Newcastle University, have no acute negative impact on the affective state of the laboratory rhesus macaques. These results concur with most of the results found in previous studies that investigated this effect by using behavioural welfare indicators (Gray et al., 2016; Hage et al., 2014).

In the study by Gray and colleagues (2016) the frequency of most of the behavioural indicators they investigated did not change when the animals were subjected to fluid control protocols. Moreover, fluid control protocols were associated with decreases in the frequency of some of the behaviours (i.e., yawning, self-grooming and body shaking). This is consistent with the results we obtained with yawning and displacement behaviours with decreases in the frequency of the behaviours during the fluid control condition. This decrease in frequency over time may be explained by a decrease in uncertainty which could affect positively the affective state of the animals. This interpretation is explained subsequently. When the animals are fluid-controlled, it is almost certain that they will be participating in experiments the following day. However, when they have access to fluid *ad libitum*, there are multiple possible scenarios of what will happen the following day (e.g., no involvement in experimental tasks, veterinary

check, change of home cage, an experimental task at the home cage). Rhesus macaques are known to prefer the routine, thus, having the certainty of what they can expect the following day might be beneficial for the valence of their affective state.

Our results contrast, however, with the conclusions reached by Pfefferle and colleagues (2018), in which they interpreted increases in cortisol in fluid-controlled animals as signs of exposure to mild stressors. Nonetheless, it is important to highlight, that the fluid control protocol used in Pfefferle's study did not allow the NHPs access to fluid at the home cage even when the daily limit was not reached during the experimental task. This differs from the protocol used in the present study and the protocols of both Gray and colleagues (2016) and Hage and colleagues (2014). This singularity in Pfefferle's (2018) protocol might negatively impact the affective state of the animals and explain the conflicting results between our results and the ones obtained by their study. On the other hand, those contrasting results might arise from the difficulty in interpreting cortisol measures due to the limitations described in section 1.4.3. Cortisol measures have shown conflicting results with both increases and decreases when the animals were exposed to stressors known to have a detrimental impact (Camus et al., 2013; Hennessy et al., 2017; Miller et al., 2007). Therefore, the interpretation of cortisol measures as a sole indicator of a stress response might not be reliable.

The validated indicator of acute negative affective states that we used, displacement behaviours, is predicted to be more sensitive to negative affective states with high arousal (see section 1.4.4.2). As a consequence, solely with displacement behaviours, it is not possible to reject the hypothesis that fluid control might be having a negative impact on the affective state of the subjects since it might be impacting the subjects' negative affective state with a low level of arousal. Nevertheless, the results obtained with the potential indicator *Inactive not alert* behaviour might support the lack of a detrimental impact of fluid control. *Inactive not alert* behaviour is commonly considered an indicator of long-lasting negative affective states at a low arousal level (see section 1.4.4.4). However, the considerable day-to-day variability in the frequency of this behaviour we found in our results suggests that it might also be sensitive to acute changes in the affective state of the subjects. Therefore, the combined results of both Displacement behaviours and *Inactive not alert* behaviour point in the direction of a lack of impact of fluid control protocols on the subjects' acute affective state of both high and low arousal levels.

Recent advances in animal welfare science have helped with the development of cognitive bias tests for the assessment of the valence of the affective state of NHPs (see section 1.4.4.5). These tests can provide valuable information regarding the affective state of the

animals; thus, their use could be proposed to elucidate further potential concerns about the use of fluid control protocols. However, their use is not compatible with the present study for three reasons: (1) studies that have investigated judgement bias tests indicate that the affective state being measured is long-lasting, often described as mood; (2) judgement bias tests require training, which on top of being time-consuming and not suitable for some research facilities, might also require the use of additional fluid control; (3) lastly, as previously discussed (see section 1.4.4.5), judgement bias test cannot be used to test the impact of repeated procedures, which is the case of fluid control protocols. Finally, the results obtained by the present study failed to find any indication of a potential negative impact of fluid control on the subject's welfare. Therefore, even if the explained obstacles could be overcome, we do not anticipate discording results.

3.6. Conclusion

In this study, we used an unprecedented number of subjects to address the need for conclusive data on the effect of fluid control protocols on the welfare of laboratory rhesus macaques used in neuroscience research (Prescott et al., 2010). The availability of longitudinal data offered us additional information about how this protocol could impact the animal's welfare in the long term. We used the most reliable and pharmacologically validated indicators of acute negative affective state, as well as additional potential indicators of this affective state.

These data failed to show any detrimental impact on the Newcastle University fluid control on the welfare of laboratory rhesus macaques.

Chapter 4: Traditional methods as indicators of long-lasting negative affective state (weight and alopecia)

4.1. Abstract

Research facilities require indicators of welfare that are reliable but also practicable. Body weight and the level of alopecia are two indicators that are often considered as such.

In the present chapter, we used body weight and alopecia levels as indicators of cumulative experience of 28 laboratory rhesus macaques. The indicators were collected between 2014 and 2021. Weight measurements were taken weekly, and alopecia scores were annual. Alopecia was scored as the presence or absence of any hair loss in the whole body of the animals.

Body weight did not indicate any detrimental weight loss and results of weight gain were not quantifiable due to limitations of the study design. We found a significant increase in developing alopecia over time, which does not seem to be caused by age, but by the involvement of the subjects in the experiments. Our results, therefore, suggest a potential negative cumulative experience caused by involvement in experiments.

Nevertheless, the limitations to be considered are discussed further in the present chapter.

4.2. Introduction

The welfare of NHPs in scientific facilities is pivotal to the research conducted. However, the assessment of this welfare is not straightforward (see section 1.3). In large facilities with hundreds of animals, it is important to have measures of welfare that don't require specialist skills or equipment and are non-invasive and fast. Two indicators that fulfil these criteria are weight and alopecia, which are recorded in most biomedical NHP facilities.

4.2.1. Weight

Body weight has been traditionally used as an indicator of welfare for rhesus macaques, and its use has been recently recommended (Prescott et al., 2022; Truelove et al., 2020). In a facility with hundreds of individuals, it is one of the only animal-based indicators considered feasible to measure by staff (Truelove et al., 2020).

Body weight has been suggested to be affected by a multitude of welfare-related factors including anxiety (Meishvili & Chalyan, 2021), depression (Jarczok et al., 2018; Qin et al., 2015; Shively et al., 2005), early life experiences (Contia et al., 2012), poor immunological function (Wolfensohn & Honess, 2005), social rank (Zehr et al., 2005) and fluid control (Prescott et al., 2010).

Traditionally, weight loss has been considered the main measure of weight concern in animal models. However, with a growing literature body on the risks of being overweight in humans (Rand & Macgregor, 1990; Wadden & Bray, 2018; World Health Organization, 2021), and the benefits of weight loss (Swencionis et al., 2013), weight increases are now also being considered as welfare indicators. Therefore, currently both underweight and overweight are considered indicators for the welfare of macaques (Kemnitz & Francken, 1986; Shively & Clarkson, 1987), with underweight requiring a 10% loss of body weight to be normally associated with detrimental welfare (Lanz et al., 2013; Smith et al., 2006). As for overweight, there is no current agreement in the literature on the limits considered to be indicative of negative welfare in NHPs.

Regarding psychological wellbeing, both anxiety and depression have been observed to affect the body weight of macaques in opposing directions. In a study by Meishvili & Chalvan (2021) overweight was associated with state anxiety in cynomolgus macaques (*Macaca fascicularis*). On the other hand, rhesus macaques with induced seasonal affective disorder, which causes depressive symptoms, were observed to decrease in weight in the study by Qin and colleagues (2015).

Other factors were found to affect the weight of NHPs such as the impact of stress-related to hierarchy and social rank. However, such factors have revealed conflicting results with observed lower weights for subordinate individuals (Wolfensohn & Honess, 2005), higher food intake by subordinate females when there is food availability and specific diets (Michopoulos et al., 2012; Wilson et al., 2008), as well as lack of difference in body weight between ranks (Shively & Day, 2015). Therefore, while those factors might explain some between-subject variability, there is no clear trend in within-subject variability. Another factor contributing to between-subject variability is early life experiences. Early life experiences have been found to create disturbances and reductions in the sexual dimorphism that causes the weight difference between males and females in macaques (Contia et al., 2012; Wallen, 1996). Finally, fluid control has been suggested to cause weight decreases due to dehydration as well as weight gains as a consequence of binge eating (Prescott et al., 2010). However, no weight gain nor detrimental weight loss was found to be associated with fluid control when investigated in laboratory male rhesus macaques by Gray and colleagues (2016).

Nevertheless, the following must be taken into consideration: The animals at the Newcastle research facility, as at many other facilities, have limited access to food, as it is controlled by the facility staff based on the recommendations of the veterinarians. As a consequence, the following analysis has an important limitation for interpreting a potential overweight scenario. With limited availability of food, the subjects are not able to eat *ad libitum*, which will restrict their potential to become overweight, restricting, as a consequence, the reliability of overweight as an indicator of a negative affective state. Moreover, it is common practice at the facility for the veterinarians to recommend dietary restrictions to individual animals when it is considered they have reached unhealthy weights.

We predicted that if there was cumulative experience caused by husbandry and/or experimental factors, we would observe underweight tendencies in the subjects over time.

4.2.2. Alopecia

Alopecia, aka hair loss, is a common condition found in NHPs. This condition can be caused by multiple factors such as ageing (Beisner & Isbell, 2009; Steinmetz et al., 2006), genetics (Ratterree & Baskin, 1992), seasonality (Isbell, 1995; Steinmetz et al., 2006), friction due to housing characteristics (Beisner & Isbell, 2008, 2009; Novak et al., 2014), hormonal variation (Lair et al., 1999) or pregnancy (Lutz, 2021; Lutz et al., 2019). However, alopecia can also be caused by pathological and stress-related factors. This, together with the evidence showing that alopecia has been more frequently observed in captive than wild populations of macaques, reaching up to half of the captive population, (Lutz et al., 2013), leads to it being

commonly associated with negative welfare (Novak et al., 2017) and used as an indicator of long-lasting negative affective state.

Some of the welfare-related mechanisms suggested as influencing alopecia are nutritional imbalances (Novak & Meyer, 2009), infections (Baker et al., 1971) and certain behaviours that induce hair loss (Beisner & Isbell, 2008). Potential nutritional imbalances and infections are routinely assessed by the veterinary team at the Newcastle facility to ensure the correct course of treatment. As for the latter mechanism, the main behaviour investigated concerning alopecia in laboratory NHPs is the so-called hair-pulling or hair-plucking behaviour (Luchins et al., 2011). Hair-pulling is the removal of hair with either hands or teeth and can be self-directed or directed towards another individual (Heagerty et al., 2017; Reinhardt et al., 1986). This behaviour differs from grooming as it pulls out the hair of the animals instead of solely brushing it and cleaning it of parasites or external particles (Heagerty et al., 2017; Luchins et al., 2011). This behaviour has been suggested to be caused by stress and therefore related to welfare concerns (Pomerantz et al., 2013).

As part of the stressors behind the direct mechanisms causing alopecia the following have been suggested: social rank (Beisner & Isbell, 2009), single-housing (Lutz et al., 2003) and early life stress (Contia et al., 2012). However, it has been hypothesised that many other stressors can also induce this condition.

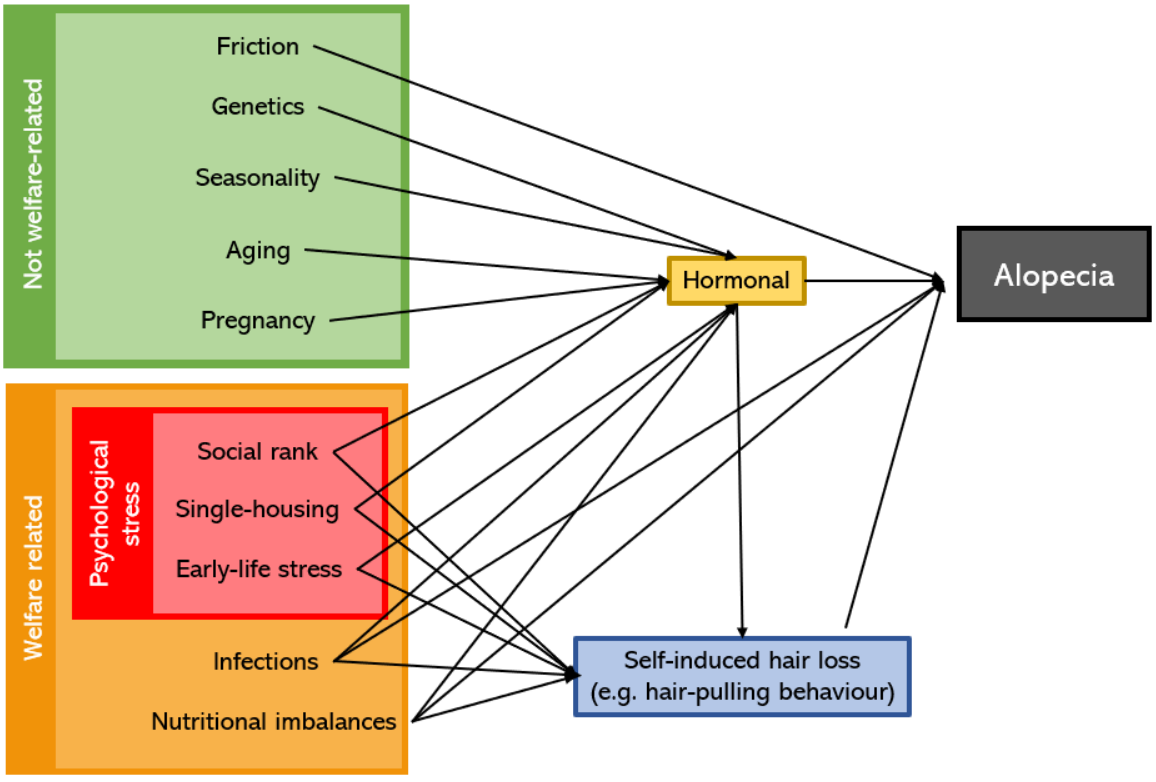

Figure 16. Diagram of mechanisms and causes for alopecia in laboratory rhesus macaques.

In conclusion, the current literature body on NHP alopecia is quite ambiguous, with diverse causes and mechanisms suggested as being responsible for the variability in alopecia (Figure 16). However, many of these factors can be controlled to ease the interpretation of the results (Novak et al., 2017). In the current study, alopecia will be used as an indicator of a negative long-lasting affective state. Non-welfare related factors will be controlled for either in the experimental plan (e.g., friction, seasonality, pregnancy) or in the statistical analysis (e.g., ageing, genetics).

4.3. Methods and materials

4.3.1. Weight

The weight of all rhesus macaques at the Newcastle University animal facility is registered by technicians weekly. In addition, animals' weights are also recorded by the researchers before being transported to the laboratory rooms for performing experimental tasks. Consequently, all the subjects have one to five weight data entries a week recorded in the diaries at the NHPs unit.

For this project, the individual subject weight data was extracted from the unit diaries for the individual's time window of interest, considered as the years when the animals were part of the current project (see Table 1).

In addition to this dataset, I also had access to the weight data from The Medical Research Council Centre for Macaques (CfM) primate breeding unit. This breeding unit is one of the two units providing the rhesus macaques to the Newcastle research facility. The animals at CfM live in groups and are not involved in any experimental procedures. They are commonly weighed at least once a year during routine health checks. The access to the CfM data brings this project an optimal opportunity to compare baseline weights of animals within the same genetic pool since most of the animals at the Newcastle facility (51%) were bred at this breeding centre. As the changes in weight with age were expected to be non-linear, they could not be controlled statistically. Instead, I restricted my comparative analysis to age-matched groups within each sex. The pregnant animals were also excluded from the analysis. In summary, the CfM data was used as a control to investigate the baseline weight curve that laboratory rhesus macaques of a similar genetic pool should follow.

4.3.2. Alopecia

Alopecia was scored annually, during November, in a visual manner while the animals were in their home cages. November was chosen for two reasons: (1) the coat of the animals is full during the winter months which would allow us to avoid hair loss due to the seasonal effect

and (2) NHPs undergo annual health checks in late November – early December, during which some parts of the animal's hair is shaved which could bias the alopecia scores.

The scoring was done with a combined version of the protocols developed by Honess and colleagues (2005) and Bellanca and colleagues (2014). The former study scored the coat condition of the animals' back qualitatively on an ordinal scale from 1 to 5, where a score of 1 is defined as "Very good coat condition" and 5 as "Back completely bald (more skin visible than hair-coat)" (Honess et al., 2005) . Bellanca and colleagues (2014) favoured a more complex approach, considering the percentage of the body affected by hair loss and its pattern of it (i.e., Thinning, patching, bare). A preliminary data collection with each scoring system was conducted which revealed that the intra-rater reliability with the Bellanca method did not reach desirable standards. However, although the intra-rater reliability of the Honess method was favoured, the absence of information for the hair condition of the limbs, head and tail that this protocol excluded, was considered suboptimal. Therefore, the final decision was to use the Honess protocol with an additional score per subject for the whole-body coat condition. The main reasons for including the whole body were the following:

1. The hair in the animal's back is longer than in other areas of the body. The length of the hair can prevent the observer from visualising bald patches of the skin.
2. The animals do not have full access to their backs. If the loss of hair was caused due to self-hair-pulling, the back would not be the most reliable indicator of this behaviour as it is furthest from their reach.

The home cages of the animals are highly enriched which could potentially cause difficulties in reliably scoring the alopecia visually from outside the cages. As a control, in the last data collection year (2021) the animals were scored in alopecia twice on the same day; once in their home cages before being taken for the annual health checks, and a second time while there were anaesthetised minutes later. Scoring the alopecia while the animals were anaesthetised enabled us to observe closer the whole-body coat of the animal.

The data set was restricted to subjects with at least 3 alopecia values over the investigated period of time.

4.3.3. Statistical analysis

The data analysis of all the studies that investigated the cumulative experience of the laboratory rhesus macaques (Chapters 4, 5 and 6) followed approximately the same main analysis structure. First, the change of the indicators over time was investigated, in addition to some control analysis for the effect of ageing that differed between the indicators. If a

significant change in the indicator was found over time, which suggested a negative impact on the long-lasting affective state of the animals, we proceeded to investigate which factors could be responsible for that effect.

4.3.3.1. Weight

The data were modelled using Linear Mixed Effects models with the *lmer* function from the *lme4* package (Bates et al., 2022). "Subject" was declared as a random effect in all the models. Despite the fact that the aim was to investigate changes in weight over time within each subject, in this case, as the change in weight does not follow a linear trend but a sigmoid curve (Wolfensohn & Honess, 2005), the within-subject mean centring of the predictor variable was not possible. Results, therefore, reflect a mixture of within- and between-subject effects.

4.3.3.2. Alopecia

The data were analysed using generalized linear mixed models with the *glmer* function in the *lme4* package (Bates et al., 2022), applying random intercept (Subject). A binomial family with a maximum-likelihood estimation was used to assess significance values when a given fixed effect was excluded from the model. The dependent variable was the binarized alopecia score (i.e., 0 vs. 1) and the independent variable was the within-subject mean-centred time in years that the subject has spent in the facility.

4.4. Results

4.4.1. Weight

10860 weight values from 28 subjects at the Newcastle facility were extracted covering a total of seven years of data (January 2014 – December 2021). The data set included 11 females and 17 males. Subjects had an average of 386 entries (min=16, max=519) depending on the length of time each animal was part of the project and the number of visits each animal had to the experimental room. The high number of data points might entail overpowering during the statistical analysis which might increase the probability of achieving significance levels for very small effect sizes.

The average weight of the animals was 10.1kg (SD =2.82kg). However, there was variability between subjects, as the minimum recorded weight was 4.23kg and the maximum was 19.26kg. As expected, sexual dimorphism played a major role in this variability. Within the females, the average weight was 7.8kg, compared to 11.1kg among the males.

The CfM data included fewer data points per subject (mean=9, min=2, max=45) with a total of 2999 weight values, but a much larger number of subjects, (n=319) of which 237 were

females and 82 were males. The average weight was 8.28kg (SD=2.55) and it ranged from 2.9kg to 19.6kg. The average weight for females was 7.48kg and for males 10.3kg.

The first step of the analysis included investigating the effect of the facility of origin (CfM vs Newcastle) on the weight of the animals while controlling for the predicted effect of Sex. Subjects at Newcastle facility were revealed to be significantly heavier (ß=1.45, SE=0.316, DF=1037, F=21.025, P=5.08e-06) even when controlling for Sex differences (ß=1.14, SE=0.306, DF=932, F=13.93, P=0.0002) (Figure 17).

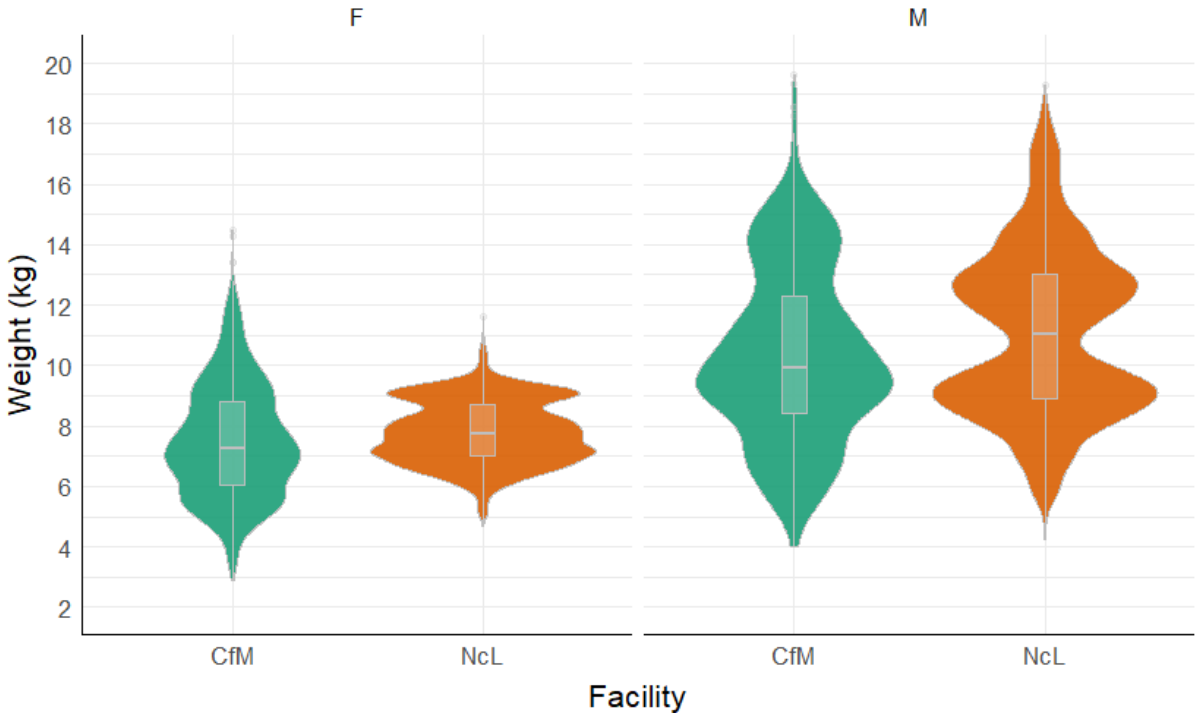

Figure 17. Weights of the animals at each facility. The subjects from the Centre for Macaques (CfM) facility are shown in green and subjects from the Newcastle (NcL) as shown in orange. Females are on the left and males on the right side of the figure.

The Newcastle data for the male macaques showed a binomial distribution of weights (Figure 17), which we suspected to be caused by the difference in the breeding centre origin of the animals (e.g., genetics, early life experience differences, etc.). Therefore, the same analysis was repeated by removing the data from the animals that came from the DSTL breeding centre (Figure 18). After removing these subjects, the weight difference between the CfM and the Newcastle subjects is not significantly different (ß=0.459, SE=0.354, DF=1086, F=1.678, P=0.195). As it cannot be discarded that the differences in change of weight over time might be caused due to the breeding centre of origin (CfM vs DSTL), the subsequent models were investigated solely on subjects bred at the CfM breeding centre.

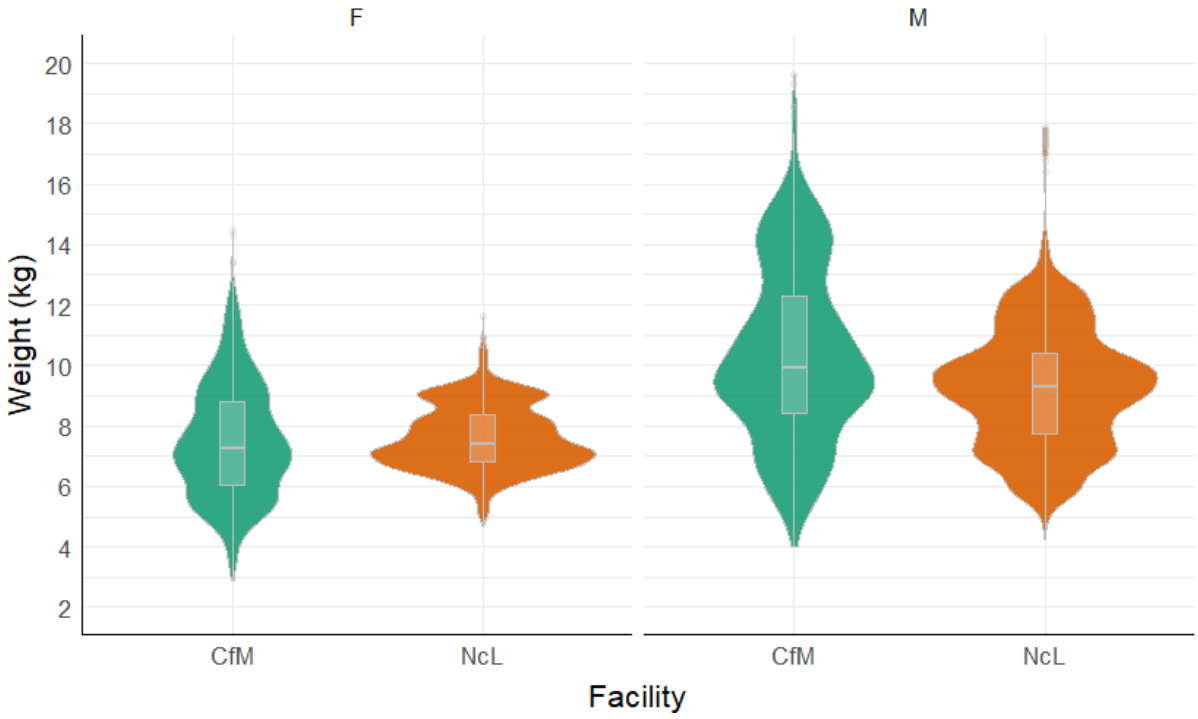

Figure 18. Weights of the animals at each facility when the subjects from the Newcastle facility bred at the DSTL breeding centre are excluded. The subjects from the Centre for Macaques (CfM) facility are shown in green and subjects from the Newcastle (NcL) as shown in orange. Females are on the left and males on the right side of the figure.

The following step was to investigate the change of weight over time at the facility. Most of the subjects included in the data set (Table 1) were within an age range when they are expected to increase in weight as part of their natural growing pattern (Wolfensohn & Honess, 2005). Therefore, an increase in weight was expected regardless of any potential changes in the long-lasting affective state. As a consequence, the measure used to assess the welfare was not the absolute weight of the animals, but the difference in change of weight between Newcastle, where animals are exposed to experimental procedures, and a baseline growth curve of captive laboratory rhesus macaques not exposed to experimental procedures. To do this, the CfM data set was used to investigate the fluctuations that the weight data followed for their rhesus macaque population. General linear models were built using the time they have spent at the facility (in years) as a single predictor of the change in Weight. To determine the optimal degree of the polynomial function, a model comparison was performed between models including various degrees, from 1 to 5. The use of larger polynomial degrees was avoided to prevent overfitting the model.

The analysis with the CfM data set which aimed to investigate the curve that best fitted the data showed that the polynomial of 4^{th} degree was best. This model produced an AIC of

7971.3 which was significantly lower (P<0.001) than the model with the orthogonal polynomial degree of 3 (AIC=8002.3).

Sexual dimorphism is known to be present in rhesus macaques (Wolfensohn & Honess, 2005) and to have an impact on the weight difference between sexes. Therefore, the interaction of time with the sex of the subject was investigated. As predicted, sex was found to have a significant impact on the change of weight (ß= 2.87, SE= 0.176, DF=363, F=267.126, P< 2.2e-16). Therefore, the subsequent models were investigated independently for males and females.

The next step was to investigate the difference in weight between the Newcastle and the CfM animals over time. To do this a model was created with weight as the dependent variable and with an interaction between two predictors: (1) Time evaluated at the polynomial degree found to be the best fit with the CfM data, and (2) the breeding centre from which the data point comes from (i.e., CfM vs Newcastle). This model revealed significant interactions between age and the origin of the facility for the male (ß= 77.524, SE= 7.37, DF=1568, F=137.82, P< 2e-16) (Figure 18 right), and female subjects (ß= - 4.781, SE= 1.32, DF=4152, F=31.12, P< 2e-16) (Figure 18 left). However, as illustrated in Figure 19, the differences were extremely small and limited in time, with all NcL weights within the normal range of CfM data. The statistical significance of the results is therefore likely to be due to the analysis being overpowered.

Figure 19. Weight of female (left) and male (right) macaques at Newcastle facility (red) and CfM breeding centre (green).

Due to the lack of a clear and biologically significant indication of being underweight, no further analyses were performed.

4.4.2. Alopecia

The alopecia dataset was composed of a total of 118 scores from 25 subjects (15 males and 10 females). 8 out of 25 subjects did not reveal any type of alopecia, neither with the whole body nor with the back scoring.

First, the reliability of the alopecia scores was investigated by calculating the correlation between the scores obtained during the observation of the animal's coat when the animals were awake in their home cages with the scores taken while the animals were anaesthetised (Whole body r=0.88, Back r=1).

As an exploratory analysis, the correlation between the back and whole-body scores was calculated (r= 0.799). 25% of the scores revealed some degree of alopecia in the back of subjects while 42% of the scores revealed alopecia over the whole body of the subjects. Therefore, as predicted, the scores collected at the whole-body level seem to be more sensitive to alopecia and thus, the following analyses were restricted to this measure.

The distribution of the scores across the subjects and data points was visualized (Figure 20). Most of the scores were 1 (perfect coat) and the maximum score collected was 3. Due to the lack of variability, the alopecia scores were binarized with 0 being defined as "No signs of alopecia" and 1 being defined as "Alopecia present".

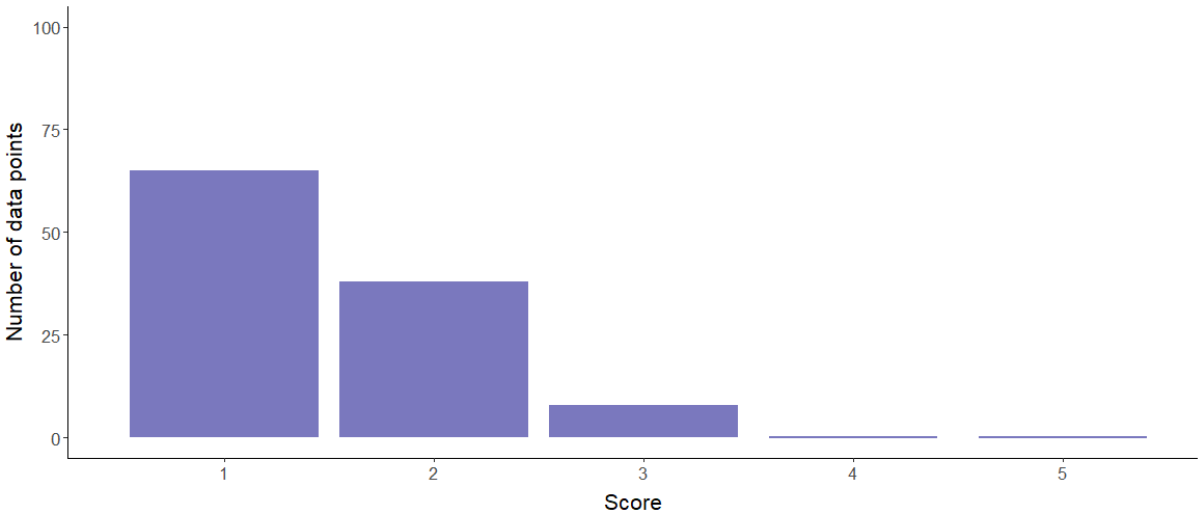

Figure 20. Variability of alopecia scores across different years and subjects. The x-axis represents the score of the alopecia incidence per collected sample, and the y-axis represents the number of subjects for which that score was collected.

The probability of developing alopecia increased significantly over time (ß=0.316, SE=0.1456, DF=1, LogLik= -76.615, P=0.023) (Figure 21).

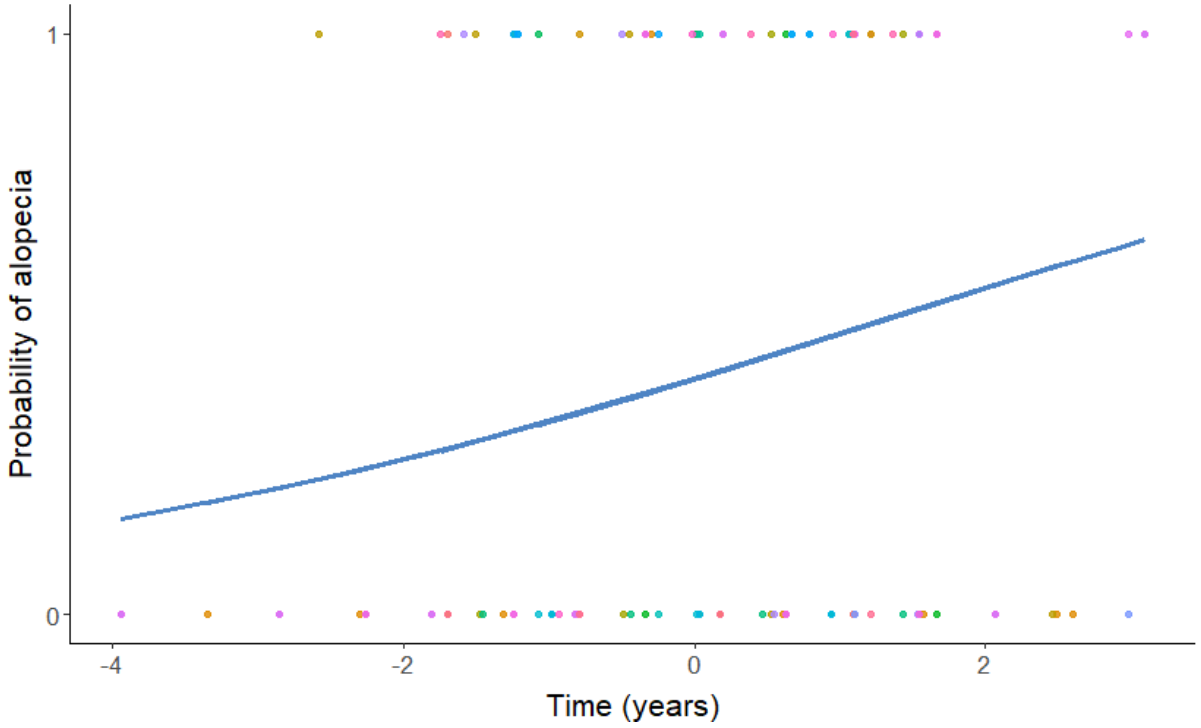

Figure 21. Probability of being scored as presenting alopecia at the whole-body level (from 0 to 1) over within-subject mean centred time (in years).

To investigate the factors responsible for the change of alopecia over time, a subsequent model was created. The dependent variable was again the binarized alopecia score, and the independent variables were the within-subject mean centred factors suspected to impact the cumulative experience of the NHPs (Table 5). A multi-model inference approach was used with the AIC method to assess the influence of each of the factors using the function *model.sel* from the *MuMIn* R package (Barton, 2022). Models had low multicollinearity (Variance Inflation Factor <10) when tested using function *vif*, (package car) (Fox, John & Weisberg, 2011). The process of model selection revealed six models that should be considered equally good to explain the change in alopecia scores (Δ AIC<2) (Table 5). The factor Scrapes was not kept in any of these models. The only factor included in all the six models was the involvement of the subjects in experiments. This predictor was revealed to induce a statistically significant increase in alopecia in all the best models. The other factors included in the best models were Age, Ketamine, Antibiotics and NSAIDs. None of these predictors reached the significance level of α=0.05.

Table 5. Models found to best fit the increase in the incidence of alopecia over time. The parameters estimate (ß), Standard error (SE) and p-value (P) are provided for each variable. The models are ordered according to their Δ AIC value calculated from the difference between the top-most model's AIC (lowest AIC from all possible combinations) and each models' AIC value.

Involvement in experiments	Age	Ketamine	Antibiotic	NSAID	Δ AIC
ß=1.375 SE=0.503 P=0.006					
ß=2.531 SE=0.982 P=0.010	ß=-0.440 SE=0.072 P=0.162				0.13
ß=2.326 SE=0.905 P=0.010		ß= -0.098 SE=0.073 P=0.181			0.32
ß=2.269 SE=0.899 P=0.011		ß=-0.167 SE=0.099 P =0.090	ß=0.017 SE=0.016 P =0.282		1.11
ß=1.628 SE=0.70485 P=0.020				ß=-0.007 SE=0.014 P=0.591	1.87
ß=2.674 SE=1.024 P=0.009	ß= -0.292 SE=0.400 P=0.465	ß=-0.055 SE=0.093 P=0.554			1.98

4.5. Discussion

4.5.1. Weight

The fluctuations of the subject's weight were explored by using the data from the Centre for Macaques (CfM) breeding centre as a baseline growth curve. The hypothesis was that the

subjects' weight would divert from the baseline growth curve with exposure to husbandry and/or experimental procedures over time.

The obtained results indicate no signs of being detrimentally underweight in subjects investigated when compared to the baseline CfM growth curve.

Regarding potential overweight concerns, the results cannot be reliably interpreted. As subjects at Newcastle facility, and at most (if not all) research facilities, do not have *ad libitum* access to food sources, the maximum weights they can reach is dependent on the food provided. In addition, the animals that veterinarians considered to have reached the level of being unhealthily overweight are routinely put under dietary restrictions to control their maximum food intake. Consequently, the weights the animals can reach in the higher end are controlled. Furthermore, the starting date of the dietary restrictions recommended by the veterinarian staff coincides with the decrease in weights of the Newcastle subjects.

On the other hand, the animals at CfM are also exposed to similar controls to their diet, with the additional factor of inhabiting large groups, which, depending on their social rank, can further limit their access to food resources.

Finally, it should be considered that the effect at the group level could be masking individual negative impacts. Weight could reflect negative cumulative experiences both with weight loss and weight gain. There is no indication suggesting that all the subjects investigated would display a negative cumulative experience in weight in the same direction, thus, hindering the reliability of interpreting visually the net results at a group level.

These results match with the data obtained by Pickard (2013) who investigated for the first time the cumulative experience of laboratory macaques at multiple research facilities using body weight. They found no signs of negative cumulative experience and no association of body weight with a specific surgical procedure in 9 out of 10 subjects.

Regarding the subjects bred at the DSTL breeding centre, a reliable baseline dataset is needed to statistically analyse the potential cumulative experience reflected in body weight over time. Therefore, at the moment, the data of such subjects cannot be interpreted.

4.5.2. Alopecia

The quantification of alopecia was used as an indication of negative cumulative experience in the investigated sample size, with the hypothesis that if there was any detrimental impact of husbandry and/or experimental procedures on the welfare of the animals, an increase in the probability of developing alopecia would be observed. The results found in the present

study indicated that the probability of being assessed as displaying alopecia was found to increase over time. This is the first reported evidence suggesting negative cumulative experience in laboratory rhesus macaques.

There are several non-welfare related factors found to influence the probabilities of developing alopecia. These factors include genetics, ageing, seasonality, friction with the cage substrate and/or other objects, and pregnancy (Beisner & Isbell, 2009; Isbell, 1995; Lair et al., 1999; Lutz, 2021; Lutz et al., 2019; Novak et al., 2014; Ratterree & Baskin, 1992; Steinmetz et al., 2006). All of these known factors were controlled for in the present analysis. The first of these factors, genetics, was controlled for by restricting the analysis to a within-subject analysis. This was performed during the statistical analysis where the factor subject was considered as a random effect and all the predictors were within-subject mean-centred. The effect of ageing was investigated by adding age as one of the predictors in the model which investigated the factors responsible for the variability in the probability of developing alopecia. Confounds with seasonality were avoided by collecting samples only during late autumn when the coat of the animals is expected to be at its fullest (Novak & Kessler, 2009). As all the home cages shared the same features (e.g., substrate, enrichment and materials), none of these factors should have an impact on the results of the analysis, even when the subjects changed their cage. Finally, none of the animals at the facility were pregnant nor had been pregnant during their stay at Newcastle and thus, the known effect of the reproductive status in females was controlled for. Therefore, as the factors known to influence the probability of alopecia in rhesus macaques were controlled for, the potential changes in the alopecia scores could be explained by changes in the welfare of the animals. Furthermore, a recent study by Polanco and colleagues (2021) found hair-pulling (referred to as self-plucking), one of the main mechanisms known to cause alopecia in NHPs, to be the best indicator of cumulative experience for laboratory rhesus macaques out of the behaviours investigated.

Our result contrasted with the study by Novak and colleagues performed in 2017. In this study, they photographed the coat condition of the animals and categorized them into <5% hair loss and >30% hair loss and found no changes in the alopecia status of rhesus macaques over time from the former to the latter status. The ages of their subjects ranged from 4 to 24 years old (mean=9.8yo). They suggested that the severity and not only the presence of alopecia should be considered when using this indicator to assess welfare in NHPs. However, this study had only two sampling periods 6 to 8 months apart, which might have not been enough to observe differences in alopecia. Furthermore, due to the sampling periods, a potential effect might have been masked by a seasonal effect. Finally, the division of the alopecia status chosen (<5% hair

loss vs >30% hair loss) might not be specific enough to assess small changes in the cumulative experience of the animals.

The observed increase in the probability of displaying alopecia over time was not unforeseen since ageing is known to increase hair loss (Beisner & Isbell, 2009; Steinmetz et al., 2006). However, the ages at which this effect begins to be revealed are not delineated. Therefore, to control for the potential effect of ageing, the effect of this factor was also investigated during the statistical analysis in addition to potentially detrimental husbandry and experimental procedures. The results showed no significant effects of ageing on the probability of displaying alopecia in the group of subjects investigated. A matching lack of effect of ageing in alopecia was found in baboons (age range 4-23yo, mean = 9.6yo) (Lutz, 2021), where a similar alopecia scoring protocol with binarized scores for the statistical analysis was utilised.

A study by Beisner & Isbell (2009) found increases in alopecia in captive rhesus macaques caused by ageing (age range 3 to >15yo, mean =8.2 yo). However, this study only included females, which had had previous and recent pregnancies, a factor known to increase alopecia and found to interact significantly with age in the same study. This study controlled for the current reproductive status of the subjects, as well as the seasonal effect, the rank, and the floor substrate of the home cage of the subjects. The differences in past reproductive status (i.e., if the subjects have been recently pregnant) might explain the differences between ours and the results found by Beisner and Isbell (2009).

In the study by Steinmetz et al. (2006), they also reported increases in alopecia with age. The association between alopecia and age was investigated with a Pearson's correlation analysis and included infants who showed almost no signs of alopecia, juveniles, and adults. The inclusion of infants might cause differences with the analysis performed in the present study, where all of the subjects were above 3 years old. But more importantly, the Steinmetz and colleagues (2006) correlation study did not control for the multitude of factors that could be impacting the welfare of the animals over time, which would therefore be integrated into their age factor. As a consequence, the results that were reported might be representing an increase in alopecia over time, which might not be caused by ageing.

In the current study, the best factor found to explain the increase in hair loss over time was the involvement of the subjects in experiments. Involvement in experiments includes various events that could be perceived as stressful and therefore induce a detrimental impact on the welfare of the animals. Some of these events include transport between the cage and the

experimental room; separation from conspecifics; and head restraint. Further studies will be required to disentangle the effects of the numerous events integrated into this factor.

As for the other factors included in the models found to best explain the increase in alopecia (i.e., age, ketamine, antibiotics and NSAID), none of them was found to have a significant impact on the probability of developing alopecia over time. A bigger sample size might be beneficial to better understand the potential effect of these other factors.

Finally, the size of the effect found for the involvement in experiments in alopecia appears to be small (Chen et al., 2010).

In conclusion, a small negative cumulative experience was revealed by the investigation of alopecia incidence, and such a negative effect seems to be driven by the subject's involvement in experiments. Further considerations discussing the interpretation of the alopecia results and its implications are discussed in Chapter 7.

Chapter 5: *Inactive not alert* as an indicator of long-lasting negative affective states in laboratory rhesus macaques

5.1. Abstract

The lack of an objective method to assess enduring negative affective states is a current obstacle to improving the welfare of non-human primates in research. We propose the behaviour *Inactive not alert* as a welfare indicator of enduring negative state in rhesus macaques.

The behaviour is defined as sitting or lying stationary with no visual contact with objects or individuals and not engaging in any other action for at least 10 seconds. In the past, this behaviour has been experimentally induced by exposing macaques to stressors known to cause depression in humans and has been pharmacologically validated with antidepressants. Therefore, the literature suggests that the display of this behaviour is related to a long-lasting negative affective state associated with low arousal (depressive-like state). We refined the use of the behaviour *Inactive not alert* to maximise its sensitivity and specificity as a marker of cumulative experience by testing different durations, as well as estimating the number of sessions required to assess reliably the frequency of the behaviour.

Using video recordings of home-cage behaviour of 25 laboratory rhesus macaques housed in a relatively enriched environment, we report an increase in this behaviour frequency, over the years. Further analyses suggest that this effect is caused by experimental procedures. However, the effect is small, compared to other factors known to have a detrimental impact on the welfare of macaques. This is the first study revealing a cumulative negative effect of experimental procedures on laboratory macaques.

5.2. Introduction

As discussed in depth in Chapter 1, rhesus macaques are widely used animal models in biomedical research. Due to the nature of the experiments they are involved in, they tend to stay for years in research facilities which consequently exposes them repeatedly to potential stressors. Both from an ethical and a scientific perspective, it is essential to investigate the impact that repeated exposure to potential stressors can have on the welfare of the animals in the long term.

One of the main difficulties in tackling this question has been the lack of a reliable indicator of a long-lasting affective state. As described in Chapter 1, inactive behaviours (i.e., depressive-like behaviours) have been used in previous studies as indicators of long-lasting negative affective states in NHPs. However, this behavioural indicator has not been standardized, which hinders its reliability, sensitivity and specificity. In this study, I refined the use of the inactivity behaviours as indicators of a long-lasting affective state by (1) standardizing the definition of the behaviour, (2) testing different durations and (3) estimating the minimum number of sessions required to assess reliably the frequency of the behaviour as seen in Chapter 2 (section 2.3.1).

Using the refined inactivity behaviour, I aimed to study the long-lasting affective state of rhesus macaques staying at research facilities and to identify factors that could be impacting it.

5.3. Methods and materials

5.3.1. Behavioural indicator

In this study *Inactive not alert* was determined as the behavioural indicator of a long-lasting negative affective state. As previously described (see 2.3.1), *Inactive not alert* behaviour is defined as sitting (sometimes with head lower than the shoulders) or lying stationary and alone, while not looking at objects or individuals (eyes may be open or closed), and not doing anything else, for more than 10 seconds. In addition, *Inactive alert* behaviour was used as a control for general inactivity. This latter behaviour was defined as sitting or lying stationary and alone, while looking at objects or individuals inside or outside the cage, and not doing anything else, for more than 10 seconds (see 2.3.1).

5.3.2. Dataset: Inclusion and exclusion criteria

For the present study, the dataset described in section 2.2 was used. The unit of analysis was Time point, defined as a 5-week period approximately every six months. To obtain the

frequency of *Inactive not alert* behaviour per Time point, the frequency of all the sessions that were analysed within each Time point, was averaged within a 5-week period.

A pilot study was performed to determine the number of sessions needed to accurately estimate the frequency of the behaviours at each Time point. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. In this pilot study, eight sessions of 45 minutes corresponding to 8 different days were analysed. The frequency of the behaviour was plotted according to the total number of sessions analysed, showing that the frequency obtained remained stable after the inclusion of five sessions in the analysis (Figure 22). Based on the data from the pilot study, the target number of sessions to be analysed per Time point was set to five, while Time points where less than 3 sessions could be recorded were discarded.

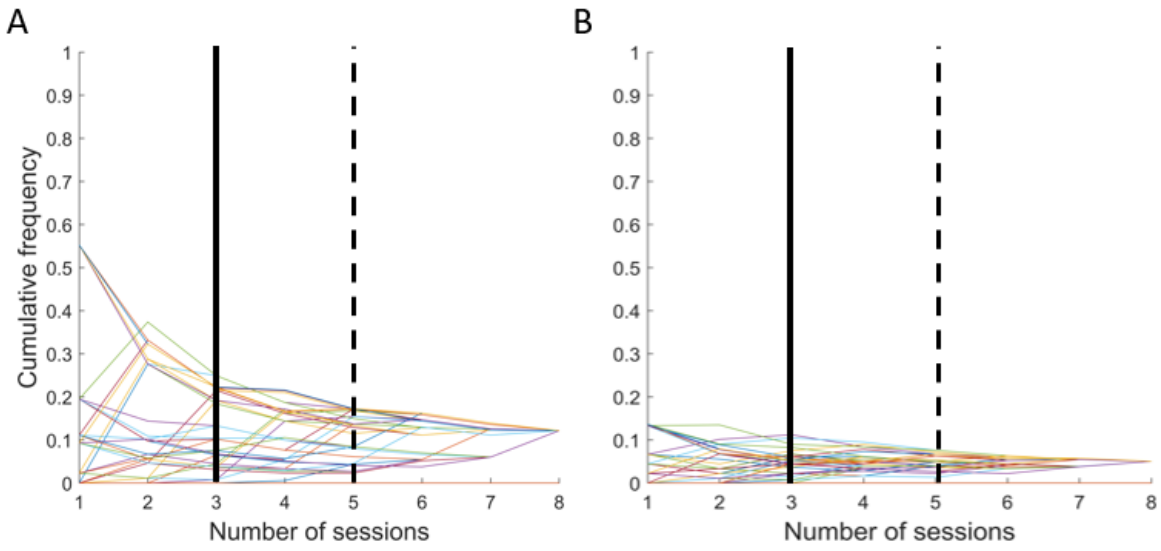

Figure 22. Frequency of the *Inactive not alert* behaviour, defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1, according to the total number of sessions analysed for, A) Subject 21 during October/November 2014 and B) Subject 8 during October/November 2016. The black vertical line represents the minimum number of sessions used as a criterion to obtain a reliable estimation of the frequency of the behaviour, and the dashed black vertical line represents the ideal number of sessions for the latter purpose. The rest of the coloured lines represent the frequencies of the behaviour for each possible combination of sessions.

In order to increase the quality of the analysed data, the dataset was restricted according to six criteria:

1. For each session, the focal animal had to be visible during at least 50% of the one-minute time bins. This choice was made to limit underestimating the frequency of behaviours that might potentially be displayed disproportionally in parts of the cage not easily visible in the video recordings.
2. All the sessions were within a 5-week range of each Time point. The long-lasting affective state of the animal is not expected to change over this time scale, therefore by including sessions within a 5-week range we aimed to have an accurate representation of the affective state during this period of time.
3. Sessions occurring less than 3 days after the subject was anaesthetised were removed. Anaesthesia could affect the behaviour of the animals, by decreasing their mobility for example. Therefore, by removing the sessions recorded after anaesthesia we aimed to reduce the potential changes in the behavioural indicator which are not related to the affective state of the animals, but the physiological changes caused by the anaesthetics.
4. All Time points included at least 3 sessions. This choice was made based on the pilot data, which suggested that a smaller number of sessions would make the estimation of the frequency unreliable (Figure 22).
5. All subjects included a minimum of three Time points. As our interest lies in a within-subject analysis of the change in the frequency of the behaviour over time, fewer than three Time points could not offer us enough information about the variability of the behaviour over time.
6. None of the subjects had been singly housed for longer than a month consecutively previous to the recorded sessions.

In addition to this set of data, I also used a secondary data set from three singly housed subjects that had been singly housed for longer than a month due to veterinary or husbandry reasons. This additional data set was used to determine the effect size of a factor known to have a negative impact on the welfare of the animal (i.e., chronic single housing) when assessed by the *Inactive not alert* indicator known. This dataset followed the same criteria presented above except for criterion number 6.

Finally, an additional analysis to increase the quality of the present study was to control for factors that could impact the acute affective state of the subjects. This was achieved by investigating the potential effect of fluid control on the welfare of the included subjects, which is the main acute repeated factor expected to have a detrimental welfare impact. No effect of

fluid control was found on the acute affective state of the animals, discarding this factor as a driver not only for any changes in the acute affective state of the animals but also for changes in the frequency of *Inactive not alert* behaviour.

5.3.3. Environmental factors

We investigated the effects of factors identified as potentially detrimental to the cumulative experience of the laboratory rhesus macaques in the variability of the behavioural indicator (Table 4). The total number of each factor (e.g., number of days under antibiotic treatment or number of days being involved in experiments) was extracted per Time point. To compute this number, first, the dates of all the sessions included within each Time point were averaged. Then, the cumulative number of each factor was calculated until that newly computed average date.

5.3.4. Statistical analyses

We used Linear Mixed Effects models with the *lme* function from *nlme* package (Pinheiro et al., 2011). The dependent variable of the models was the frequency of *Inactive not alert* behaviour, defined from 0 to 1. The dependent variable was transformed via square root to follow normality.

As the aim of the analyses was to investigate the within-subject effect of potentially detrimental factors on the frequency of the behavioural indicator, I declared "Subject" as random effects and I subject-mean-centred all the explanatory variables (see section 3.3.4).

5.4. Results

5.4.1. Descriptive statistics

The combined data set included a total of 989 sessions across 179 Time points with subjects from both groups ("Early" and "Late"). A total of 25 subjects were included, of which 15 were males and 10 were females.

The average number of sessions per Time point was 5.3 (range 3 to 9) (Figure 23). Most of the Time points included in the analysis (% 82) were based on at least 5 sessions, which was previously determined as a reliable number of sessions to obtain a consistent frequency of the *Inactive not alert* behaviour per Time point (Figure 23).

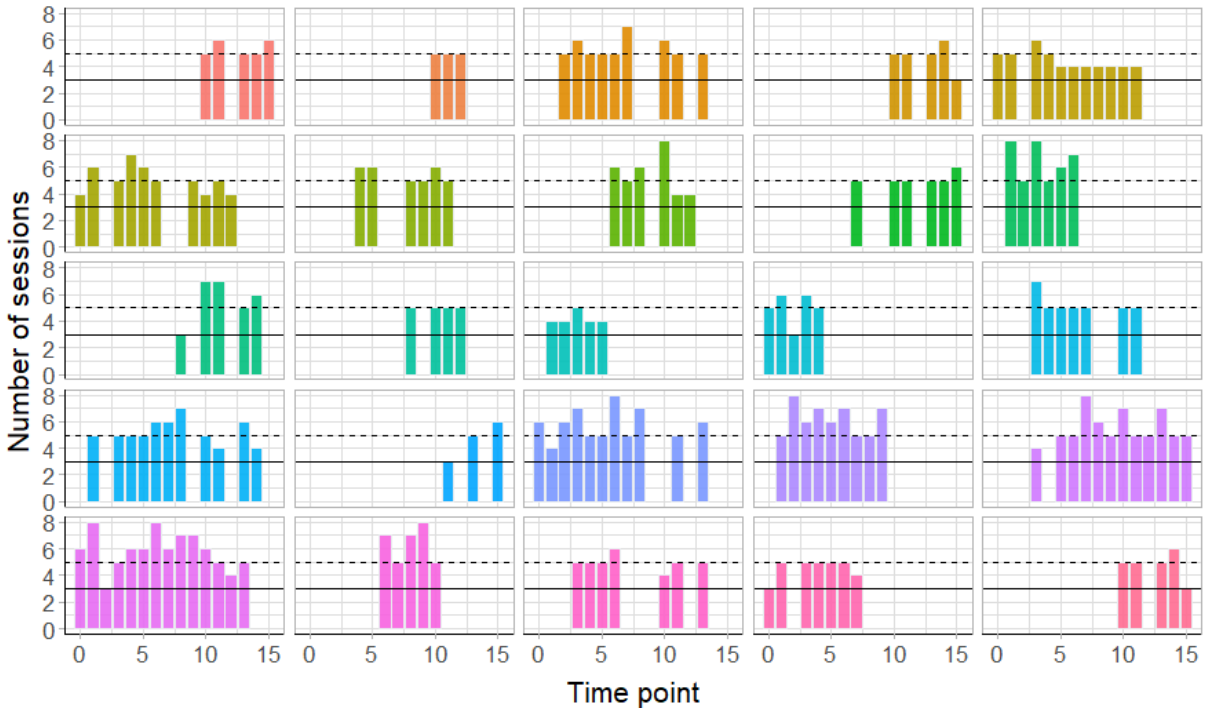

Figure 23. Number of sessions per Time point (approximately every six months) included in the statistical analysis. Each panel represents a subject.

The frequency of *Inactive not alert* ranged from 0 to 0.802 (mean=0.086, SD=0.131), with 24 out of 25 subjects displaying the behaviour at least during one of the sessions (Figure 24).

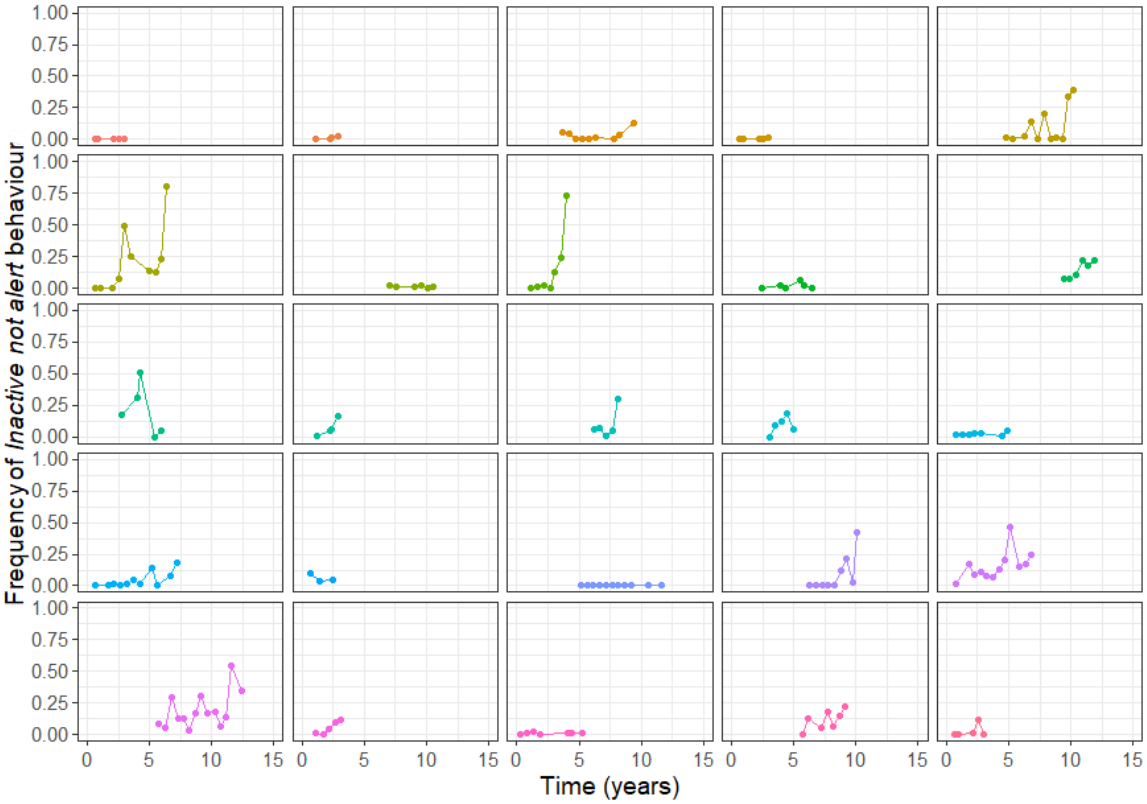

Figure 24. Frequency of *Inactive not alert* behaviour over time (in years). The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. Zero represents the subject's arrival at the Newcastle facility. Each colour panel represents a subject.

The behaviour *Inactive alert* was displayed by all the subjects with a frequency ranging from 0.006 to 0.869 (mean= 0.3197, SD= 0.212).

5.4.2. Effect of time

We first investigated the change in the behaviour frequency over time, with time defined as the number of years since arriving at the Newcastle facility (Figure 25). The linear mixed model revealed a significant increase in the behaviour over time (ß=0.046, SE=0.007, P<.0001). In order to test whether this effect could be explained by ageing, we investigated the change over time in the combined frequency of Inactivity behaviour, which showed no significant increase (Figure 26, *Inactive alert* combined with *Inactive not alert* behaviour when the subjects were alone and with the cage mate) (ß =0.01, SE=0.008, P=0.226).

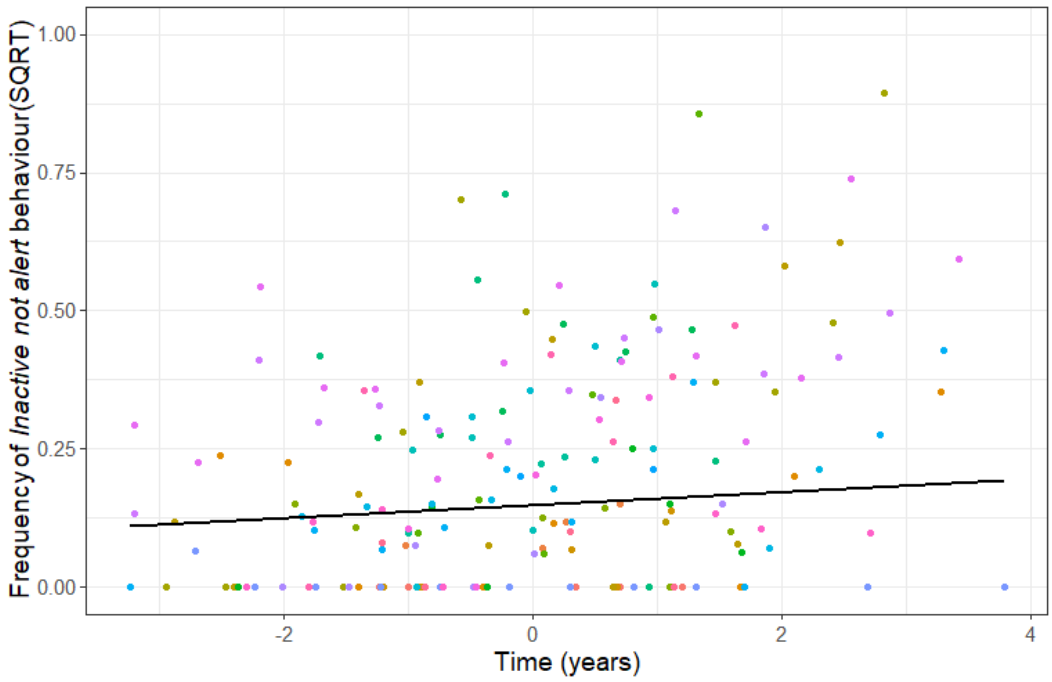

Figure 25. Frequency of *Inactive not alert* behaviour over the years at the Newcastle facility. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. The x-axis is within-subject mean-centred. Each colour represents a subject.

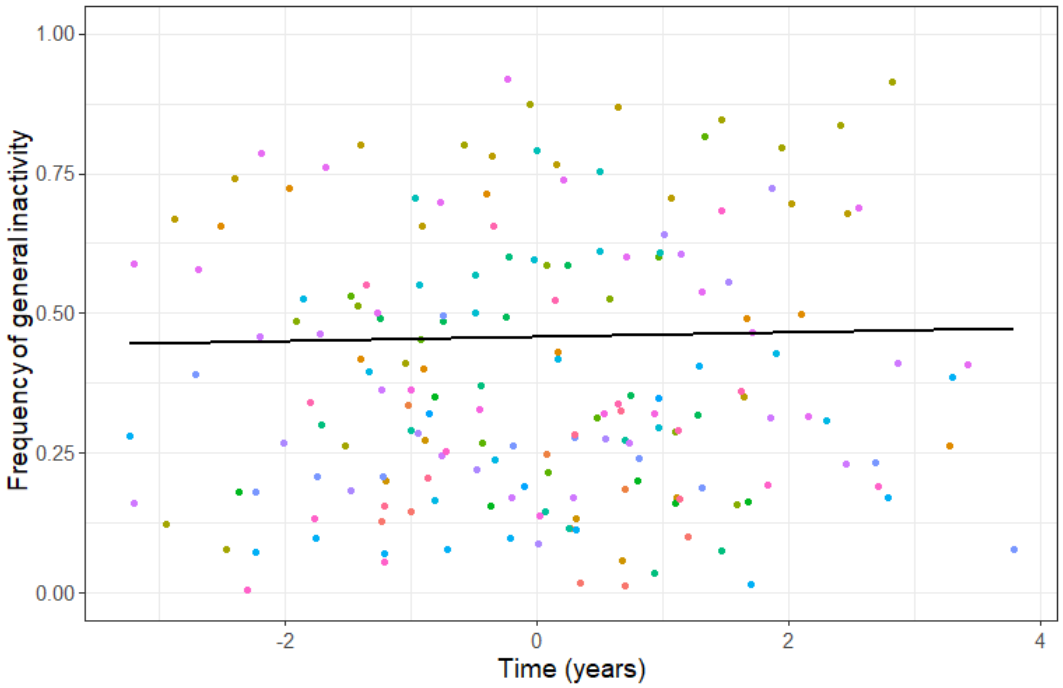

Figure 26. Frequency of the combined inactivity behaviour over the years at the Newcastle facility. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. The x-axis is within-subject mean-centred. Each colour represents a subject.

To enable me to interpret biologically the size of the effect of time, I investigated the effect of chronic single-housing (known to have a detrimental impact on the long-lasting affective state of the animals) in a small dataset with subjects that had been consecutively single-housed for longer than a month. This data set included a total of 14 Time points of 3 male subjects. The frequency of the *Inactive not alert* behaviour was found to increase at a rate of 0.11% per year (Figure 27).

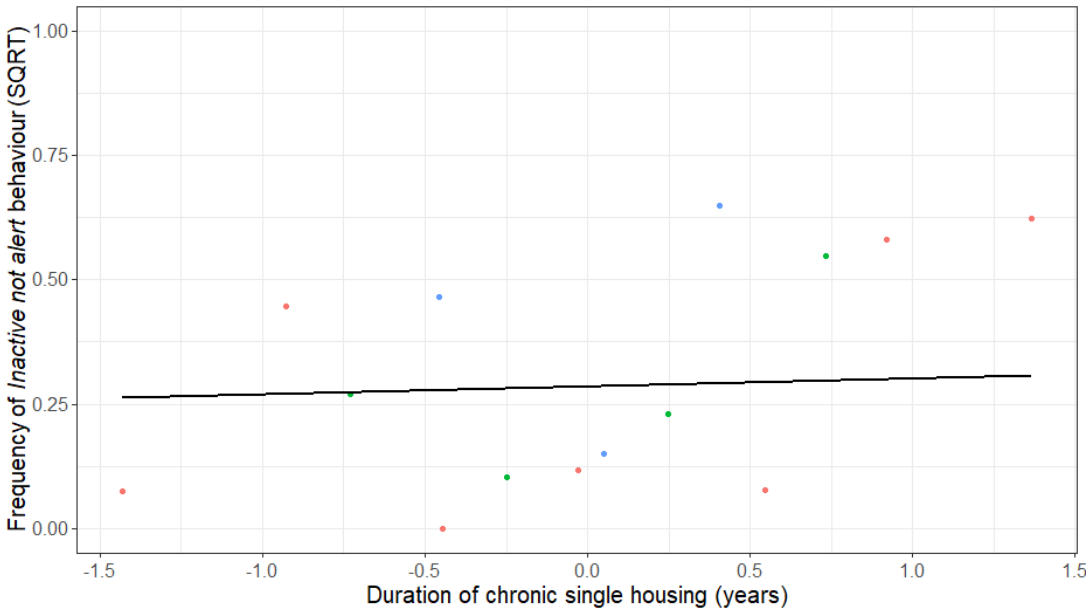

Figure 27. Frequency of *Inactive not alert* over the duration of the chronic single housing in years. The frequency was defined as the proportion of 1-minute time bins in which the behaviour was observed, on a scale from 0 to 1. The x-axis is within-subject mean-centred. Each colour represents a subject.

5.4.3. Effect of environmental factors

The subsequent analysis aimed to investigate the factors responsible for the observed increase in the behavioural welfare indicator (*Inactive not alert*) over time. The effect of these factors was tested in the combined data set after the removal of the sessions where the subject was singly housed (see section 4.3.3). This data set included 158 Time points of 24 subjects from both groups.

A linear mixed model was performed with the six factors suspected to impact the cumulative experience as predictor variables (Table 4). In this model the VIF factor of the predictor Age was above 10, suggesting high multicollinearity. As the main potential multicollinearity concern derives from the factors of Aging and Involvement in experiments, restricting the analysis to subjects who were investigated before and after being involved in

experiments was predicted to reduce the multicollinearity. Thus, we repeated this analysis with the “Early” group of subjects only.

This group included a total of 91 Time points and 15 subjects. The average frequency of *Inactive not alert* across the subjects was 0.067 (SD=0.105). The frequency of the behaviour *Inactive not alert* also increased significantly over time for this group (ß =0.043, SE=0.009, P<.0001). We conducted a model which integrated all the identified factors suspected to impact the cumulative experience of the NHPs (Table 4). The new VIF calculation on the model revealed that all the predictors were below 10. The full linear mixed model was followed by model selection using the function *dredge* from the *MuMIn* R package (Barton, 2022).

The model selection approach revealed four models considered to be an equally good fit to explain the increases in the frequency of the behavioural indicator (Δ AIC<2) (Table 6). The factor involvement in experiments was present in all these models and was associated with a significant increase in the behavioural indicator. The other factors included in the models were Antibiotics, Ketamine and NSAIDs. These three factors were associated with a decrease in the behaviour over time, and none of them reached statistical significance. The factors Scrapes and Age were not kept in any of the models considered to be best at explaining the variability in the behavioural indicator.

Table 6. Models found to best fit the increase in the frequency of *Inactive not alert* over time. The parameters estimate (ß), Standard error (SE) and p-value (P) are provided for each variable. The models are ordered according to their ΔAIC value calculated from the difference between the top-most model's AIC (lowest AIC from all possible combinations) and each models' AIC value.

Involvement in experiments	Antibiotics	Ketamine	NSAID	Δ AIC
ß = 0.140 SE = 0.028 P < .0001				
ß = 0 .161 SE = 0.035 P < .0001	ß = -0.001 SE = 0.001 P = 0.286			1.10
ß = 0.188 SE =0.061 P < .0001		ß = -0.004 SE = 0.005 P = 0.3607		1.41
ß = 0.164 SE = 0.051 P <.0001			ß = -0.001 SE = 0.001 P = 0.5536	1.89

5.5. Discussion

This study investigated the impact that long stays in research facilities can have on the cumulative experience of laboratory rhesus macaques involved in neuroscience research. This was accomplished by measuring the subjects' changes in the frequency of the refined behavioural indicator of welfare *Inactive not alert*. The frequency of the *Inactive not alert* behaviour revealed high variability, with both increases and decreases over the years. However, the results show that on average, the frequency of *Inactive not alert* behaviour significantly increased over the years. In parallel, no signs of a significant increase were found in the combined frequency of both types of inactivity (*Inactive not alert* and *Inactive alert*). These results may suggest that the observed increases in *Inactive not alert* are not due to ageing and, thus, that the welfare of the subjects is deteriorating over time. This interpretation is further reinforced by subsequent analyses which showed that age was not the explanatory factor for this increase. However, this effect was further explored in the subsequent analysis in which the effect of ageing was investigated in combination with the other potential predictors. There is

no standardized frequency of *Inactive not alert* behaviour in the literature at which the long-lasting affective state of the subjects is considered to be compromised, as the emphasis is currently given to the relative change in the frequency over time. To benchmark the effect size, I compared the estimated value of the effect of time on the *Inactive not alert* frequency to the estimated value of a factor known to be detrimental to the welfare of the animals: single housing. Single housing was found to increase the frequency of *Inactive not alert by* 11% per year, compared to the 3% per year increase found over time in the investigated sample.

We aimed to explore the causes of the *Inactive not alert* behaviour variation. The impact of several factors that were hypothesised to induce such variation was tested.

The best predictor for the increase in the frequency of *Inactive not alert* behaviour was the involvement of the subjects in experiments. This predictor showed a significant impact on the frequency of the behaviour through all the models considered to best explain the increase of the behavioural indicator over time.

No *a priori* hypothesis of the effect of this predictor was made since it integrated multiple factors that could potentially have had opposing effects. The animals included in this study were involved in different experiments that required them to participate in behavioural tasks, be MRI scanned or be part of electrophysiological experiments. These experiments required transport from the home cages to the laboratory where the experiments take place, as well as head-restraint in the laboratory. All of these events could have a detrimental impact on the cumulative experience of the animals. The involvement in experiments can also provide social interaction with the researchers, in addition to control over the rewards they obtain when performing the various tasks, which could be beneficial for the animals. Finally, the transport to and from the experimental room could also offer the subjects the possibility to satisfy their curiosity and explore unknown parts of the facilities, inducing a positive impact on their affective state. The current results indicate a net negative impact of experiment involvement, suggesting that the potentially detrimental events that are integrated into this factor are not being fully compensated by the potentially enriching events.

Further discussion of these results is accomplished in Chapter 7.

Chapter 6: Local amount of grey matter in the hippocampus as an indicator of cumulative experience in laboratory rhesus macaques

6.1. Abstract

Hippocampal biomarkers are a promising approach to assessing cumulative experience in animals. They have been recently validated as sensitive and specific to the valence of the affective state of the animals, and they have been proven to integrate the impact of events on the affective state of the animals over time.

In this chapter, I used a macroscopic hippocampal biomarker to assess the cumulative experience of laboratory rhesus macaques. I investigated this in 15 adult rhesus macaques (10 males and 5 females) that I scanned using magnetic resonance imaging (MRI) every six months over up to 7 years. I used a voxel-based morphometry technique to extract the age effect and investigate the effect of factors suspected to impact negatively on the welfare of the subjects.

We found no signs of negative cumulative experience in the laboratory rhesus macaques investigated.

Limitations of the present study are discussed in the present chapter.

6.2. Introduction

As reviewed in section 1.4.5, hippocampal biomarkers are one promising approach to assessing cumulative experience in non-human primates (Poirier et al., 2019). This biomarker can be measured macroscopically via Magnetic Resonance Imaging (MRI), and microscopically via *post-mortem* analyses of the brain (see section 1.4.5 for details on the biomarkers). The microscopic level is more accessible as it does not require specialist equipment or training of the animals. However, it is invasive, and it only allows for between-subject analysis in cross-sectional studies. On the other hand, the macroscopic approach is less accessible but allows for the measurement of the biomarker longitudinally within subject. As currently there is no absolute value of the hippocampal biomarker that can be interpreted in terms of good or bad cumulative experience, the interpretation must be done relatively, assessing the changes in the cumulative experience within-subject, or between groups. From this standpoint, the macroscopic level of analysis appears to be a better candidate, as it allows the measurement of cumulative experience within subject, thus, requiring a smaller sample size, and the opportunity of assessing the effect of repeated procedures. Moreover, as it allows for the assessment of the cumulative experience during the lifetime of the animal, it grants the researcher the opportunity to make changes that can lead to the improvement of the welfare of that same animal. In this project, MRI was used as a non-invasive method to assess macroscopically the local amount of grey matter in the hippocampus of the animals.

6.3. Methods

6.3.1. Subjects

The subjects were selected from the total pool of subjects presented in Table 1. In addition to the inclusion criteria presented in section 2.2, for this study the animals were required to be trained and habituated to be scanned in the MRI, and to have a head-post implant that would ensure they are motionless (to avoid movement artefacts) during the procedure, and to have no metallic implants which are not compatible for the MRI use.

6.3.2. MRI image acquisition and processing

Two types of structural MRI images (T1- and T2-weighted) were acquired every six months for seven years, from Spring 2014 to Autumn 2021, at the Newcastle University facility. T1-weighted and T2-weighted images refer to the two most used MRI sequencies that differ in the amount of time between the pulse frequencies applied to the same slice and in the time between the delivery of the radio frequency pulse and the receipt of the echo signal. The two images provide different contrast images (i.e., Cerebrospinal fluid is dark on T1-weighted images and bright on T2-weighted images) and the use of both in the in the analysis can improve

the overall quality of the final image. All the MRI images were acquired using a Bruker vertical scanner with a 4.7 T field strength. The animals were awake and in an upright position during the acquisition of the scans.

The T1- and T2-weighted images were processed using the Voxel Base Morphometry (VBM) computational neuro-anatomic technique. This method allows computing the local amount of grey matter in each voxel (three-dimensional pixel) of an MRI scan (Kurth et al., 2015). MRI images were pre-processed with an in-house pipeline (AutoMacq) developed by my collaborator N. Kindred (Kindred et al., in prep.), using a combination of scripts from the Statistical Parametric Mapping software, version 12 (SPM12, https://www.fil.ion.ucl.ac.uk/spm/software/spm12), implemented in MATLAB (R2016A) and the FreeSurfer software (version 6.0.0). For each individual subject, the pre-processing consisted of the following main steps:

1) Manual reorientation of all the T1 and T2 images, matching the orientation and coordinate origin of the Salem and Logothetis macaque atlas (2012).
2) Serial longitudinal co-registration (i.e., method used to align multiple images to ensure the spatial correspondence of anatomy across different images) of the images acquired at different Time points, resulting in a subject-specific mean image, as well as time-point specific images where the amount of grey matter in each voxel was encoded relative to the value in the corresponding voxel of the mean image (Jacobian determinants). This step is required because the subjects differed slightly in their scanning position from one time point to the consecutive; this position difference was amplified in some occasions as a consequence of the subject's growth and additional head-implants.
3) Initial segmentation (differentiating the white matter, grey matter, cerebrospinal fluid and non-brain tissues in the images) of the mean image to automatically obtain a brain mask (binary image distinguishing brain from non-brain tissues).
4) Debiasing and refinement of the mask to ensure no brain tissue is excluded nor non-brain tissue included.
5) Final segmentation of the mean image, allowing for calculation of the mathematical transformation to co-register the subject-specific images (mean image and time-point specific ones) to the Salem and Logothetis atlas (2012).

6.3.3. Statistical analysis

The statistical analysis of the data was performed in SPM12 and was restricted to the hippocampus, using an explicit mask of the anatomical structure, bilaterally. First, a general

linear model was run in each subject dataset to estimate the slope of the age effect. Subject-specific images encoding the slope of the age effect in each voxel were then co-registered to the atlas, using the output of step 5 and smoothed using a 1mm Full-Width Half Maximum (FWHM) Gaussian kernel. Finally, a group-level General linear mixed model was run to assess whether the slope of the age effect of each subject could be explained by the number of factors the animal had been exposed to between the first and last scan (Table 4). This analysis was limited to the hippocampus, using an explicit mask of the bilateral anatomical structure. However, due to the unforeseen circumstances caused by the COVID-19 pandemic, the number of subjects included in this study was much lower than expected (original target sample size=25). This caused changes not only in the statistical power of the study but also in the intended statistical analyses. A smaller number of subjects increased the chances of overfitting the model planned to investigate the effect of the five factors (Table 4) on the grey matter volume of the hippocampus. To avoid overfitting, the effect of each predictor was investigated separately. As a result, five models were created (one for each predictor to be tested: Involvement in experiments, Scrapes, NSAID, Ketamine and Antibiotics). P values were adjusted using the familywise error (FWE) correction to account for the fact that one statistical test was performed per voxel (Mirman et al., 2018) and the α value was decreased to 0.01 using the Bonferroni correction to take into account that 5 models were run per voxel (Bonferroni, 1936).

6.4. Results

Following a visual examination of the MRI scans, two subjects were excluded, resulting in a dataset of 15 subjects (10 males and 5 females). The first subject was removed due to enlarged ventricles, and the second subject due to a visible large inflammation which resulted in an enlarged ventricle over time.

The analysis of the factors identified as potentially posing a detrimental impact on the cumulative experience on the NHPs revealed no significant effect on the hippocampal biomarker investigated (Table 7).

Table 7. Coordinates (position of voxel in the MRI image) and statistics of the voxel with the highest grey matter change caused by the factors suspected to affect the cumulative experience of the subjects.

Factor	**Coordinates**			**Peak level**	
	mm	mm	mm	P_{FWE}	P unc.

Involvement in experiments	14	17	-0.5	0.995	0.047
Ketamine	14.5	16.5	-0.5	0.984	0.028
Antibiotics	14.5	17	-0.5	0.962	0.019
NSAID	-8	14	1	0.254	0.001
Scrapes	-15.5	1.5	9.5	0.955	0.018

6.5. Discussion

In this chapter, I aimed to investigate the cumulative experience of laboratory rhesus macaques using a macroscopic hippocampal biomarker of cumulative experience: the local amount of grey matter in the hippocampus.

We did not find any evidence supporting the hypothesis that the age effect was influenced by the number of procedures the animals had been exposed to. However, there are certain limitations to consider.

The hippocampal biomarker has proven to be highly sensitive to changes in the affective states of the animals, with significant changes in the local amount of grey matter of rodents observed even just hours after exposure to mild stressors (Morais et al., 2017). However, it is crucial to consider that even though the biomarker is sensitive, different methods to measure this biomarker have different levels of precision. The acquisition and processing of MR images induce noise and artefacts which generate a measurement error that can hinder the sensitivity of detecting effects, particularly when those effects are small. This limitation can be overcome with a large sample size. In human MRI studies, the typical sample size is around 25 subjects, our original target sample size, and it has been argued that even such a size might lead to underpowered studies in some circumstances (Szucs & Ioannidis, 2020). However, as mentioned before, the Covid pandemic significantly reduced the number of animals available for this study (n=15), resulting in a lack of power. We, therefore, interpret our results as inconclusive.

Chapter 7: General discussion

At the beginning of this thesis, I outlined three main aims I planned on achieving. In the present chapter, I evaluate how the work presented throughout this thesis has tackled those aims. Afterwards, I put forward future research directions and recommendations that could expand my work and aid in the advancement of the field of captive primate welfare.

7.1. The cumulative experience of laboratory rhesus macaques

The first aim of this project was to assess the cumulative experience of laboratory macaques. To achieve that aim, I used a multidisciplinary set of cumulative experience indicators, and I investigated the change of those indicators in data covering up to seven years of the life of 15 to 30 laboratory rhesus macaques.

Four different indicators were used, of which, two showed evidence of negative cumulative experience in the laboratory macaques. These indicators were alopecia incidence and frequency of *Inactive not alert* behaviour (often termed as a depressive-like behaviour). Both the probability of developing alopecia and the frequency of *Inactive not alert* behaviour increased significantly over time. The study of *Inactive not alert* behaviour also indicated that the observed increase might be associated with negative welfare and not with ageing. This is evidenced by the lack of increase in general inactivity behaviours, which are known to be associated with ageing but not with a negative affective state. Furthermore, the additional analyses that investigated the effect of age in both the alopecia and inactivity studies failed to find any effect of age. The evidence leads to the conclusion that the observed increases in the indicators of a negative affective state are not caused by ageing (i.e., a welfare-wise neutral process), and might be associated with a deterioration of welfare over time.

The indicators that provided no conclusive evidence, in favour or against negative cumulative experience, were weight loss and the local amount of grey matter in the hippocampus. Weight gain was discarded due to the limitations in the study design. To understand the conflicting results between indicators, we must highlight the strength and limitations of each.

In this study, both body weight and alopecia were considered non-specific (changes in these indicators can be caused due to factors not related to the affective state of the animal) and non-sensitive (changes in the affective state might not always be reflected by changes in the indicator), yet they revealed different results. The first limitation of the use of weight as an indicator of cumulative experience in laboratory macaques is the inability to attribute weight gain to a change in the welfare state. At research facilities, each subject receives a regulated

quantity of food per day, appointed by the veterinary and technical staff. Animals have the control to eat less than the total amount of food they are offered. However, they do not have control over the maximum amount of food they can consume. Therefore, they do not have the ability to increase in weight freely. In conclusion, the lack of subject control over the maximum amount of consumable food prevents the reliable use of weight gain as an indicator of cumulative experience. Weight is also known to be regulated in many animal species according to their perceived risks (i.e., risk of starvation vs. costs of impaired locomotion arising from too much fat) (Halsey, 2018). The lack of current knowledge on the potential weight regulation strategies of laboratory rhesus macaques would hinder reliable interpretations of weight loss and gain as indicators of welfare. Another limitation when investigating changes in body weight over years is the non-linear shape of the indicator. For different life stages, different growing rates (changes in body weight over time) are expected, and that rate is predicted to vary per individual. On top of this, the fact that variability in both directions (overweight and underweight) could be indicative of a welfare impact, adds another layer of difficulty, since individual effects can be masked at the group level. In this regard, the use of alopecia incidence is an easier indicator of cumulative experience to interpret. The welfare impact of alopecia is expected to be linear and can only be interpreted in one direction, this is, an increase in alopecia incidence will be associated with a negative cumulative experience. However, unlike body weight, alopecia might be suffering from a flooring effect. The qualitative nature of the measure, in addition to the inability to encode the condition as better than the "perfect condition", might not enable the detection of subtle changes in the affective state of the animals unless they are in a negative valence state (cannot assess a change from a neutral to a positively valenced affective state).

As for the *Inactive not alert* behaviour, it was effectively refined as an indicator of cumulative experience and revealed a negative cumulative experience in the studied subjects. The refinement of this behaviour increased its sensitivity and specificity to the negatively valenced affective states of rhesus macaques. This was accomplished by optimising the minimal duration of displaying the behaviour and by distinguishing it from behaviours not associated with negative affective states (*Inactive not alert* with cage mate and *Inactive alert*).

Finally, the hippocampal indicator used in this project is known to be sensitive and specific to the negatively valenced affective state of the animals. However, as explained in section 6.5, the measurement error during the acquisition and processing of the MR images decreased the sensitivity of the approach significantly. This loss of sensitivity was not possible

to balance out as planned, due to the unpredictably small sample size. Therefore, the results obtained with this indicator are deemed inconclusive.

To better understand this combination of results, it is also important to highlight the size effect of the cumulative experience found. Using the effect of single housing (known to be detrimental to the welfare of NHPs) in the behavioural indicator frequency as a comparison, we found that the size of the change in cumulative experience is small. This finding concurs with the effect found in the study of alopecia incidence (Chen et al., 2010). This small effect size is consistent with the lack of negative cumulative experience found in the body weight and hippocampal indicators when we consider their limitations in terms of sensitivity and reliability.

7.1.1. Previous studies investigating the cumulative experience of laboratory rhesus macaques

A recent paper by Wegener and colleagues (2021) investigated the effect of laboratory procedures on rhesus macaques and reported that no indication of cumulative experience was found to be caused by laboratory procedures. The conclusion reached by Wegener and colleagues (2021) might seem to conflict with our results, however, to understand the joint results of both studies, we must direct our attention to the sensitivity and specificity of the indicators used. The Wegener (2021) study used both indirect measures of plasma cortisol levels with the neutrophile/lymphocyte ratio (N/L) as well as attributes of appearance and behaviour. As reviewed previously (section 1.4.3), the physiological measures traditionally considered as indicators of welfare, such as the hormonal indicator used in the Wegener (2021) study, have been shown to reflect the levels of arousal of the affective state, rather than its valence, and might not be sensitive for long-lasting affective states (Otovic & Hutchinson, 2015; Paul et al., 2005; Ralph & Tilbrook, 2016). Additionally, from the eleven attributes of behaviour and appearance used (coat, skin, orifices, nutritional status, mucosal membrane, behaviour at rest, behaviour in motion, solitary behaviour, social behaviour, behaviour in presence of observer and behaviour in presence of staff), three could be marginally comparable to the ones used in the present study: coat reflecting alopecia, nutritional status as weight and solitary behaviour as the behavioural indicator. These attributes were scored at four levels. In the nutritional status, both underweight and overweight were considered a welfare concern, and in the solitary behaviour, they included different categories of behaviour than are present in our ethogram. For example, apathy, which could be somewhat comparable to the *Inactive not alert* behaviour, is considered as less *disturbed healthiness* than the stereotypical behaviours. The change over time of the individual attributes is not reported; however, they report a small but significant increase in scores when comparing the animals before and after being involved in

experiments. Due to the grouping of the categories and the qualitative approach used in the Wegener study, the sensitivity and specificity are likely to have been lower than in the present study. Moreover, the use of human observers, known to alter the behaviour of the animals, might have caused an underestimation of the frequency of the investigated behaviours. Therefore, considering the small size of the effect found in our studied subjects, their results concur with our findings.

Similarly, Pickard (2013) investigated the presence of cumulative experience in several macaques across facilities and found no signs of cumulative experience. In this case, the only quantitative measure used was weight. In addition to the limitations of weight mentioned in the previous chapter (Chapter 4), this result is concordant with the results we obtained when using weight as an indicator of cumulative experience.

With the aim of identifying risk factors that could negatively impact the welfare of laboratory macaques, a series of studies have been investigating the effect of various husbandry procedures on the incidence of behaviours traditionally associated with low welfare status (Baker et al., 2012; Gottlieb et al., 2015; Lutz et al., 2003; Rommeck et al., 2009). These studies found that the probability of displaying abnormal behaviours (Lutz, 2013), self- injurious (Baker et al., 2012; Rommeck et al., 2009), stereotypic behaviours (Gottlieb et al., 2015) and anxiety-like behaviours (Baker et al., 2012) was significantly higher for animals that experienced single-housing at an early age, longer periods of single housing, were nursery reared and had a high number of blood draws. The results from these studies concur with the revealed negative cumulative experience caused by experimental procedures found in the present project. However, two main aspects that might hinder the compatibility between the two studies should be considered: (1) some of the behaviours found to be associated with the potentially detrimental factors might not be representing the long-term affective state of the animals (i.e., anxiety-like behaviours) or the valence of it (i.e., stereotypic behaviours); (2) due to the differences in legislation between the USA and UK, the animal facilities at the summarised studies differ significantly from the facility in the present project (Newcastle facility containing bigger cages, more enrichment), as does the early life experience of the subjects (Newcastle subjects being weaned at a later age and being always mother reared). Finally, a potential lack of the reliability in the analysis of these studies has also been highlighted due to the failure to replicate some of the study results (Baker et al., 2013).

In conclusion, previous studies revealed trends in the direction of the results found in the present study: a small but significant negative cumulative experience in laboratory rhesus macaques that seems to be driven by experimental involvement.

7.1.2. Factors influencing the cumulative experience

The second aim of this study was to identify the procedures which have a long-lasting effect on the welfare of the subjects, and the third and last aim was to assess the direction of this effect. In this section, I tackle both aims.

7.1.2.1. Involvement in experiments

Both indicators of cumulative experience that revealed an increasing negative net impact of experiences over the lifetime of the subjects (i.e., alopecia and *Inactive not alert* behaviour) concur on the factor that best explains that increase: involvement in experiments. This factor integrates several events that could explain its negative impact on the welfare of the macaques.

The first of the events suggested as a detrimental event for the welfare of laboratory macaques is fluid control (Prescott et al., 2010). The long-lasting physiological effect of this procedure has been investigated by Wegener and colleagues (2021) who found no indication of the negative impact of fluid control in the blood samples collected. However, to my knowledge, there is no study investigating the long-term psychological impact of fluid control protocols. Nevertheless, the study performed in Chapter 3 might aid in its understanding. When investigating the acute impact of fluid control on the welfare of the subjects, we found no acute effect of fluid control, no effect of consecutive days of fluid control (up to 5 days) and no signs of habituation nor sensitisation over the years. Therefore, considering the lack of acute impact of fluid control, even after several years of fluid restriction, and the absence of detrimental impact on the physical health of the animals, this procedure seems unlikely to be responsible for the negative cumulative experience we observed.

Transport to and from the home cage is another of the events integrated within the involvement in experiments factor. In this facility, the animals are trained with positive reinforcement to willingly move from the home cage to a mobile chair that allows the researcher to take the animal from the animal unit to the various laboratories. In this chair, their movement is limited but they can slightly move their head, as well as their limbs inside the chair, which allows them to observe the animals and environment around them. Individual preferences might play a role in the response of the animals when exposed to different environments for the first time, both offering them excitement and enrichment opportunities as well as causing them a negative stress response (Gottlieb & Capitanio, 2013). However, over time they are expected to habituate. Potential anxiety responses are also typically considered by the staff working with

these animals to ensure minimal impact on their welfare. Therefore, the net impact of this event over time is not suspected to be significant, at least at a group level.

Closely correlated with the transport, the event of social separation from the cage mates could also be having a negative impact on the welfare of the subjects. However, most social separation studies have been restricted to infant-mother separations, or adolescent macaques (Truelove et al., 2017). Therefore, even if this event is known to be highly detrimental, this negative impact is expected to be smaller if not null when it comes to the separation from non-related adult conspecifics (Mineka & Suomi, 1978). On the other hand, rhesus macaques are highly social animals, and the deprivation of visual and olfactory contact with other macaques could potentially be detrimental even if only for a few hours a day.

When animals are involved in experiments, they will inevitably spend more time with human beings. This event might constitute an enriching event cognitively and socially, in a contrast with the potential lack of stimulation available at their home cages. However, this event might also be perceived as a taxing experience. But, in the same way as the transport event, this effect is probably transient, with the animals habituating to potential detrimental impacts.

Most of the subjects that were involved in experiments in the Newcastle facility were head-restrained during the experimental tasks. Some of these tasks involved acute single and multi-unit electrophysiological recordings and brain imaging (MRI and fMRI), electroencephalography (EEG), transcranial ultrasound stimulation (TUS), behavioural tasks requiring eye-fixation, etc. These tasks required the subjects to be awake and relied on the lack of movement of the animal's head to reduce the artefacts in the data. Moreover, in the case of the former task (acute single neuron electrophysiological recordings), the movement of the head can also be a risk to the health of the animals through brain injuries (Isoda et al., 2005; Lanz et al., 2013) as well as the stability and quality of the recordings themselves. Therefore, the use of this method is favoured. However, head-restraint is suspected to have a detrimental impact on the welfare of macaques, as a result, the maximum duration of head-restraint is stated in the licenses required for the use of this procedure. A study by Clarke and colleagues (1988) investigated the response of rhesus macaques to physical restraint both by collecting cortisol samples and encoding behavioural responses. They found an increase in cortisol levels after the restraint compared to the pre-restrained levels and reported behavioural displays of struggle and increases in the duration of "passive-eyes closed" behaviour, which was considered an indicator of depressive-like behaviour. The physical restraint in the Clarke and colleagues (1988) study was significantly more severe than the head-restrain common in neuroscience research and the animals were not trained for it. Moreover, this study aimed to investigate the acute impact of

this event on the welfare of the animals. Therefore, further research is needed in order to discard the potential cumulative impact of this procedure.

During the experimental tasks, the subjects had control over the rewards (food or fluid) that they were given, in addition to being involved in cognitive tasks considered to be enriching. These two events are suspected to have a positive impact on the welfare of the rhesus macaques. Since the cumulative experience found through the present study was negative, the potential positive impact of these events might have been surpassed by other negative events.

Finally, the main additional events integrated into involvement in experiments will therefore be the experiments themselves. I briefly mentioned the experimental procedures the animals in this study were involved in (e.g., ephys, fMRI, EGG, TUS, eye-fixation). These experiments were not considered painful for the animals and most of them were non-invasive procedures routinely performed on humans. Moreover, the invasive aspect of those procedures was already controlled for in the other factors investigated in this study (e.g., Scrapes). Nevertheless, the individual cumulative impact of each of these procedures has not been, to my knowledge, investigated in rhesus macaques. Lastly, the sole event of being repeatedly involved in experiments might, in itself, already be inducive of a negative affective state independently of the nature of said experiment.

As observed, the factor *Involvement in experiments* is a complex one, and disentangling all the events integrated within is a challenge that could not be achieved solely in the present study. However, social separation from the cage mate and head restrain seem to be the best candidates responsible for the negative cumulative experience observed in our subjects.

7.1.2.2. Other potential factors suspected to impact cumulative experience

None of the other factors investigated in this project was revealed to have a statistically significant impact on the changes of the cumulative experience indicators used. However, it is not yet possible to discard completely the potential impact of these factors.

Three out of the four welfare indicators were investigated according to the six factors (Involvement in experiments, NSAID, Ketamine, Scrapes, Antibiotics and ageing) suspected to impact the welfare of the subjects: alopecia, *Inactive not alert* behaviour, and local amount of grey matter in the hippocampus.

During the statistical analysis of alopecia and *Inactive not alert* behaviour, five and four statistical models respectively were found to be equally as good as to explain the changes in the indicators over time. All of these models included involvement in experiments, which

reinforces the argument of this factor being the main, if not the only, factor inducing the negative cumulative experience. However, five other indicators were included in these models: Age, Ketamine, Antibiotics and NSAID. Ketamine and NSAID were both associated with a positive cumulative experience during both the analysis of alopecia incidence and *Inactive not alert* behaviour change. Therefore, it is unlikely that these two factors are drivers of the negative cumulative experience observed over time. Antibiotics was found to be associated with a negative cumulative experience in the analysis of *Inactive not alert* behaviour, and with a positive one in the analysis of alopecia incidence, rendering the results unreliable to interpret. In the specific case of age, as previously described, the results from the present study do not support it as the driving factor behind the negative cumulative experience.

Scrapes factor was discarded from the analyses of both *Inactive not alert* and alopecia indicators. Due to the predicted risks associated with this factor, the lack of observed effect of this factor was unforeseen. This lack of effect could be explained by the reduced variability of this factor when compared to the other factors investigated. Dura scape surgeries are not as common as ketamine sedations or treatment with NSAIDs or antibiotics in the Newcastle research facility. In addition, the lack of observed impact in the analysis of *Inactive not alert* behaviour should be taken with extra caution. Due to the high correlation between age and involvement in experiments, the decision was made to use the so-called "Early" group of subjects to avoid multicollinearity biases in the study of the factors responsible for the increase in behaviour over time. Although this decision allowed for the analysis of the data, it also included a few limitations. As previously explained (see section 2.2), the "Early" group of subjects was not as involved in experiments as the "Late" group. This difference implies that the variability of the factor Scrapes, which requires long-term involvement in experiments, was not as frequent as in the "Late" group. As a result, the interpretation of this factor might be inaccurate, not allowing to reject the hypothesis that this factor has no negative cumulative experience impact on the study subjects

Finally, even though none of the factors was revealed to have a significant impact on the age effect on the volume of grey matter change in the hippocampus, this might be due to a lack of sensitivity resulting from a small sample size.

In conclusion, while not significant, the impact of NSAID, Ketamine, Scrapes and Antibiotics cannot be completely discarded. Nonetheless, considering the relatively small effect size of the cumulative experience, and the compelling results suggesting involvement in experiments as the main driver of this negative effect, we do not predict a large impact of the other investigated factors.

7.2. Recommendations for the assessment of the cumulative experience of non-human primates

Throughout the duration of the present project, I have acquired knowledge on the assessment of the cumulative experience of laboratory rhesus macaques involved in neuroscience experiments. In this section, I will share some recommendations to aid the welfare assessment of these animals in other facilities.

7.2.1. Body weight

Body weight has been recently recommended as a welfare indicator for laboratory rhesus macaques (Prescott et al., 2022; Truelove et al., 2020). After careful consideration during the present project, I believe this indicator should be used with caution.

Body weight increases during the first years of the life of rhesus macaques as a biological mechanism of growth and stabilizes after 5-8 years of age. The ratios of body weight increase per year can be related to welfare. However, the natural growth rate will vary per individual, as does the exact age when it will stabilize. Individual differences are related to sex, genetic background (subspecies), pregnancy status, early life experiences, diet type, rank, living environment (outdoors vs. indoors), etc. Obtaining a reliable baseline growth curve from animals that share all these characteristics is, in most cases, highly improbable.

One way of dealing with this limitation is by using body fat percentage, or body weight/height ratio. As these refined measures do not tend to change within-subject over time (except for pregnant females), they ease the interpretation of the data, as well as set more clear limits for the exact values at which the welfare of the animal starts to be at risk. Of course, these measures will also vary due to the same factors affecting the absolute body weight measure. On top of the limitations that still need to be considered when using this indicator, the use of body fat or body weight ratios removes the main advantage of this indicator. Body weight is considered an easy, non-invasive, fast method of assessing welfare. If the animals need to be assessed based on body fat or being measured in height, they will most likely need to be anaesthetised or trained to obtain a reliable measure, which hinders these advantages.

Additionally, a more complex limitation of this indicator needs to be considered. Both weight gain and weight loss can be indicative of exposure to detrimental stressors for the welfare of the animals. This lack of unidirectionality in the indicator hinders the interpretation of the measures when used as raw body weight and/or as a refined measure. At an individual level, very clear ranges of body weight need to be defined to reliably assess if a weight loss is indicative of a positive cumulative experience (when the animal is within the optimal

individualised weight range) or indicative of a negative experience (when the animal is distancing from the optimal individualised weight range). When the analysis is done at a group level, the individual displays of negative cumulative experience through body weight (weight loss vs. weight gain) might be masked at the group level. There is no evidence suggesting that all subjects in the facility should follow a specific direction. Moreover, an individual might be reflecting cumulative experience variation through body weight in different directions as a response to the type of stressor or life stage. This limitation is evidenced by the diagnostic criteria for major depressive disorder in human beings, which includes both weight gain and weight loss as one of the main five criteria broadly used (Regier et al., 2013).

Finally, the intrinsically controlled environment of the research facilities does not allow for *ad libitum* consumption of food. The maximum amount of food will be controlled by the human staff. Therefore, interpretations of weight gain should be made with caution.

7.2.2. Alopecia incidence

During our study, we used a customised alopecia protocol to investigate the incidence of alopecia in our population of subjects. Comparative analysis between our custom protocol with other previously used protocols suggests the former might be more sensitive to changes in alopecia incidence, at least given the same circumstances.

In our study, we compared the Honess and colleagues (2005) protocol of scoring the alopecia level from 1 to 5 on the back of the animals, and our protocol which mixed the Honess and colleagues (2005) method with the Bellanca and colleagues (2014) protocol, which investigates the alopecia of the whole body of the animal (more detail in section 4.3.2). The incidence of alopecia was found to be higher when using the whole-body measure, than when restricting the analysis to the back. This result suggests the whole-body analysis of the alopecia is more sensitive. A few reasons for this increased sensitivity were hypothesised in section 4.3.2. We consider that depending on the question and characteristics of the study, the use of a whole-body score of alopecia should be considered to increase the sensitivity of the study and avoid potential biases.

7.2.3. Inactive not alert behaviour

After careful deliberation over the existing literature on behavioural indicators of welfare, we considered depressive-like behaviours to have two dimensions that are highly valuable for their use as an indicator of cumulative experience: (1) sensitivity to long-lasting affective states, (2) sensitivity to negatively-valenced affective states. In addition, a third dimension, which is not required to be used as an indicator of cumulative state, but should be

considered when interpreting the results, is (3) the sensitivity to affective states with low levels of arousal.

The sensitivity to long-lasting affective states of this behaviour is evidenced by studies such as the ones by Qin and colleagues (2015), where the frequency of this behaviour was only significantly increased after months of exposure to a stressor linked to human depression (i.e., shortened photoperiod). Additionally, the study performed by Perera and colleagues (2011) found that the frequency of the depressive-like behaviour was affected by treatment with anti-depressant drugs, known to only have long-term effects on the affective state, as opposed to an acute one. However, our results suggest that in addition to being sensitive to long-lasting affective states, this behaviour might also be sensitive to acute affective states. This is suggested by the variability in the frequency of the behavioural indicator of inactivity between sessions that are recorded within the same week or month. Factors affecting the acute affective state of the animals might be responsible for this variability.

The sensitivity to negatively valenced affective states has also been evidenced by studies that induced the display of this behaviour by exposing the animals to stressors linked to human depression such as social isolation or a shortened photoperiod (Harlow & Zimmermann, 1959; Hennessy et al., 2014, 2017; Li et al., 2013; Qin et al., 2015) and by the above-mentioned pharmacological validation using anti-depressant drugs (Perera et al., 2011).

Regarding the arousal level, this behaviour is associated with, the data is not conclusive. The inactive and hunched form of the behaviour suggests an association with a low-level arousal affective state. However, human studies show high comorbidity between disorders commonly considered as having a low level of arousal (i.e., major depressive disorder) and others with high-level arousal (i.e., generalized anxiety disorder) (Hirschfeld, 2001), which emphasises the complexity of delineating arousal levels according to their commonly associated behavioural forms. In addition, pain (which on many occasions might be associated with an affective state with a high level of arousal, due to inducing an accelerated heart rate) has also been found to be associated with some forms of inactivity behaviour in rhesus macaques (Polanco, 2021, not peer-reviewed). Therefore, the arousal level of depressive-like behaviour is currently unknown. However, from a welfare perspective, the arousal level of the behaviour is irrelevant, since the aim is to assess the valence of the affective state. In conclusion, depressive-like behaviours are suitable indicators of cumulative experience.

Nonetheless, however advantageous the use of this behaviour might be, its use as an indicator of a negative affective state has not yet been fully explored. One of the reasons behind

this is that the lack of a standardized definition and sampling method compromises its reliability and comparability as a welfare indicator.

In this study, we refined the use of this behaviour to increase their sensitivity and specificity to long-lasting, negatively valenced affective states, as well as to increase their comparability between studies. The behaviour was refined and made operational by testing different durations, as well as estimating the number of sessions required to assess reliably the frequency of the behaviour. In addition, the refined behaviour does not require information regarding baseline behaviour levels, nor is dependent on external stimuli. To improve the comparability, we termed this behaviour *Inactive not alert* and developed a standardized definition based on the previous terminology used in the literature.

We investigated the correlation between this *Inactive not alert* behaviour with *Inactive alert* behaviour to test the hypothesis that the *Inactive not alert* behaviour is not reflecting a non-welfare related inactive behaviour. We found no correlation between the behaviours, which suggests that indeed these two behaviours are reflecting different underlying factors and should be differentiated in future studies.

The observed high frequencies of *Inactive not alert* behaviour in the present study significantly surpass the expected ones, based on the testimonies of the staff at the facility (i.e., researchers, technicians and veterinarians) and the variability reported by previous studies (Polanco, 2021, not peer-reviewed). This is even more surprising when considering that the facility at which the author's investigated subjects are housed is highly enriched and follows some of the highest standards of animal welfare for laboratory rhesus macaques. The high welfare standards of this facility are further evidenced by the low incidence of alopecia when compared to other facilities (Honess et al., 2005), the lack of underweight animals and the almost virtual absence of display of abnormal behaviour within the subjects, which is not the norm in other facilities (Polanco, 2021, not peer-reviewed). This suggests two main concerns: (1) the presence of a human observer might have made the research community drastically underestimate the frequency of this behaviour (Hennessy et al., 2014) and, (2) the duration of the behaviour is critical to allow enough variability in the frequency of the behavioural indicator to analyse the behaviour reliably.

The refined definition of the behaviour developed in this study, which includes a specific and optimized duration, resolves the first of the concerns. Moreover, the wide use of the refined duration would also allow further comparability between studies.

Regarding the second concern, the use of remote video recordings would avoid the bias caused by human observation. The use of remote video recording has further benefits:

1. It might encourage larger sample sizes, as it does not require a researcher to be in person during the collection of the video data.
2. It allows for a better quality of results since multiple observers can encode the recording with no logistical limitations. In this way, inter-and intra-observed reliability can be assessed and guaranteed.
3. It allows for reproducibility studies between research centres which might aid to identify suboptimal methodological designs and, in turn, improve the quality of the studies.
4. It encourages collaborative studies by sharing video recordings between research groups. This also tackles the reduction aspect of the 3Rs, since the same recording can be used for different research questions without the need for additional animal subjects.
5. It would allow for automatic recognition of the behaviours using deep learning algorithms which would, in turn, increase the power of the studies by analysing high magnitudes of behavioural data pragmatically impossible to analyse with human observers.

The main potential limitations of the use of remote video recording in some research facilities are the required large digital storage space, the lack of facilities to record remotely in the animal units, and the risk of misuse of the video recordings by activists, which might discourage some research staff to integrate this approach in their facilities. However, rapid technological and digital security advancement in those areas is promising, envisaging a prompt solution.

Due to pragmatic decisions required for the encoding of the behaviour, the following limitation must be considered when interpreting the results: the *Inactive not alert* behaviour, as currently encoded, cannot be distinguished from sleep. In the present study, this limitation was controlled with the scheduling of the recording times, which were restricted to the early morning when the animals were not expected to sleep. It was also considered that if the animals were asleep during these unexpected times, this specific sleep behaviour might be representative of a failure to sleep during the night or of excessive sleep. Both of these events (i.e., insomnia and hypersomnia) are considered in humans as one of the main five criteria for the diagnosis of major depressive disorder (Regier et al., 2013) and could, thus, also be associated with a negative affective state in NHPs. However, if the behaviour is to be encoded during other times of the day, it is pivotal to consider the inability to distinguish between *Inactive not alert* and sleep.

As a result, I suggest the use of the *Inactive not alert* behavioural indicator when aiming to assess the long-lasting negative affective states in rhesus macaques.

7.2.4. Macroscopic hippocampal biomarker

Despite the multiple advantages of using the macroscopic measure of the local amount of hippocampal grey matter as an indicator of cumulative experience (see section 1.4.5), the technical side of this method might need further development if it aims to decrease measurement error and allow studies to have high sensitivity with a low number of subjects.

Currently, a collaborative approach between multiple research facilities sharing MRI data (e.g., PRIME-DE) might be the most feasible direction for using this hippocampal indicator. A collaboration would maximise the sensitivity without hindering the reduction of subjects supported by the 3Rs principles.

7.2.5. Further promising biomarkers

Novel promising techniques relevant to animal welfare science are steadily emerging from a wide range of scientific fields. One of these novel techniques is the application of epigenetic knowledge to assess detrimental impacts of the environment on the cumulative experience of non-human animals. When an individual is exposed to a stressor, it concatenates measurable epigenetic effects such as methylations (da Silva et al., 2022). The use of Epigenetics biomarkers to assess animal welfare in farmed animals is rapidly spreading. However, the study of these biomarkers has been limited to a few species such as pigs and chicken (da Silva et al., 2022; Hao, Cui, et al., 2016; Hao, Liu, et al., 2016; Pértille et al., 2017, 2020). The promising results (da Silva et al., 2022) highlight the value of investigating its potential use in laboratory NHPs.

7.3. Recommendations for the refinement of husbandry and experimental procedures to improve the welfare of laboratory rhesus macaques

The results obtained in this study aim to inform future decisions in the field of non-human primate welfare research as well as in the use of non-human primates as animal models in biomedical research.

Ensuring the best possible welfare status of the animal model is required from both an ethical and a scientific standpoint. Many of the studies requiring NHP models are publicly funded which reinforces the responsibility of easing the growing concern from the general public regarding their use by guaranteeing high welfare standards. This ethical requirement is also recognised in the legislation that regulates the use of these animals as models in research. From a scientific point of view, higher welfare of the animal model will decrease the noise in

the results obtained with it, decreasing, in turn, the number of trials (and/or subjects) required for such experiments (see section 1.1), once again, highlighting the importance of the welfare of the animals involved.

The present project revealed signs of cumulative experience associated with the subjects' involvement in experiments. These results suggest that the welfare of the animals is deteriorating over time and can have important implications for the research community and open a discussion of potential refinements in husbandry and experimental procedures to reduce this effect. The present study aspires to inform the decisions of future study designs that will investigate these potential refinements, in order to prioritize research projects and support data-driven research.

The first result to be considered for future decision-making is the factor found to be the main driver for this decrease in welfare over time: involvement in experiments. Previously I have outlined the potential singular events within this factor that might be responsible for this effect. Here I suggest potential refinements to said individual events, as well as future studies that may aid in understanding this negative impact.

An initial study that would further the understanding of the effect of experiments on the welfare of the animals is the assessment of the cumulative experience of laboratory rhesus macaques when they are not continuously involved in experiments. Such a study would shed light on the cumulative nature of this factor and assess the potential benefits of implementing no-experimentation periods for the animal models. These no-experimentation periods, during which the laboratory animals would not be involved in experiments, would aim to have a positive impact on the affective state of the animal, and restore the cumulative experience of the subjects to pre-experimentation periods. The present study does not include enough data during non-experimental periods that would allow a reliable analysis to answer this question.

Head-restraining is one of the events within the involvement in experiments that might be driving the negative cumulative experience. Within this event, two components must be considered: (1) the headpost implant and side effects caused by it (i.e., pain caused by undetected infections during head restraint), and (2) the physical restraint that might be a source of psychological distress. The potential detrimental impacts of the implant are already being investigated by working on refinement strategies that will decrease the incidence of infections (Lanz et al., 2013; Ortiz-Rios et al., 2018). An additional measure that might alleviate the detrimental impacts of this implant is its routine monitoring. This monitoring can be done using X-Ray imaging (as done when a complication is suspected) and utilizing the instances when

the animal is anaesthetised for other motives, such as, during the annual health monitoring. Regarding the potential impact of physical restraint, a study investigating multiple durations of experimental sessions that require this type of restraint might be informative.

Social separation is another single event suspected to be the driver of the negative cumulative experience. The animals are separated from their cage mates for transport to experimental rooms. The development of experiments that can be performed in the home cage while in the presence of a cage mate is a promising alternative to avoid social separation. These types of experiments are already in use in some research groups. Two main developments are needed to ensure the spread of this technology: (1) the advancement in equipment that can be used in the home cage without the presence of a researcher (i.e., cheaper strong experimental testing equipment), (2) face recognition technology that would enable multiple animals to be in the experimental area at the same time, while reliably detecting which subject is performing the experiment, and (3) the development of wireless chronically implanted data acquisition devices that would enable electrophysiological data acquisition in the home cage. The second point is pivotal since the experimental tasks need to be customised to the skills and training level of each subject.

Equally as relevant for the prioritisation of future welfare studies, I subsequently present the factors that failed to induce a negative impact on the welfare of the animals. These results will provide the information needed for a data-driven decision that allows a more proficient allocation of resources for future studies.

The first of these results involves the use of fluid control protocols. These protocols showed no signs of having a negative impact on the welfare of the animal models, not when enforced acutely, nor consecutively (with a maximum of five consecutive days), nor in an accelerating or decelerating manner (i.e., habituation and sensitisation). These results, aim to settle the current concern about these widely used protocols (Prescott et al., 2010) both for the scientific welfare community and/or the public.

A commonly considered detrimental event for the welfare of laboratory non-human primates is fights between cage mates. In the present study, we investigated the impact of these fights by using the treatment with NSAIDs and antibiotics as a proxy. Most of the fights are considered to have a negative impact causing injuries that are subsequently treated with anti-inflammatory drugs. The lack of effect found with the factor NSAID suggests that the impact of these fights is very small, if not null. Therefore, the current husbandry practices enforced by

the facility (separation of animals often involved in fights) seems to balance out or reduce the potential impact of this event.

7.4. Concluding remarks

The present work shows evidence supporting cumulative negative experiences in laboratory rhesus macaques. Through a multidisciplinary use of welfare indicators, we identified that the animal's involvement in experiments is likely the factor responsible for driving the cumulative negative effect. Our data do not support fluid control protocols to be the main event responsible to cause a negative impact on the cumulative experience of the animals. In addition, we discussed the potential individual events integrated within the involvement in experiments that might be responsible for the negative cumulative impact; and we put forward recommendations to tackle those events, as well as proposing future studies that might aid to disentangle them. Finally, informed by the obstacles encountered during the present work, I developed recommendations for the assessment of cumulative experience in laboratory macaques.

The results obtained in this study also highlighted the following: the methods used to assess welfare need to be highly sensitive to the affective state of interest. The area of animal welfare research has been developing fast in the last decades. The public pressure to ensure the welfare of non-human primates accelerated the research in this area and assisted significant improvement in laboratory animal welfare standards. As the overall welfare standards improve, we must develop methods that are sensitive to more subtle changes in the affective state of the animals. The present study accomplished this through different approaches (i.e., the alopecia scoring method was refined by increasing the area of study, and the *Inactive not alert* behaviour was refined by creating an operational definition informed by a variety of preliminary analyses). The presented work has illustrated the potential to improve existing approaches and develop highly sensitive indicators. These new indicators have been shown to be successful in identifying areas for improvement within laboratory non-human primate welfare.

This work aims to encourage refinement in experimental techniques and result in an improvement of the welfare of the laboratory rhesus macaques, used models of healthy humans in neuroscience research, across research centres. Finally, animal welfare research commonly prioritises the study and identification of factors having detrimental impact on the welfare of the animals. However, overlooking the positive dimension of the valence of animal's affective states might be hindering critical advancement in animal welfare science. Therefore, further research into potential enhancers of positive cumulative experience and the identification of

indicators sensitive and specific to positive affective states should also be considered in future studies.

References

Abbott, C. C., Jones, T., Lemke, N. T., Gallegos, P., McClintock, S. M., Mayer, A. R., Bustillo, J., & Calhoun, V. D. (2014). Hippocampal structural and functional changes associated with electroconvulsive therapy response. *Translational Psychiatry*, *4*(11), e483–e483. https://doi.org/10.1038/tp.2014.124

Animal Welfare Act. (2006). *Animal Welfare Act Animal Welfare Act* (p. Chapter 45).

Animals in Science Committee. (2017). *Animals in science committee code of practice and working protocol*. *November*, 1–34.

Ares, B. E. (2022). *Animal Welfare (Sentience) Bill*. *March*, 1–31.

Arnone, D., McIntosh, A. M., Ebmeier, K. P., Munafò, M. R., & Anderson, I. M. (2012). Magnetic resonance imaging studies in unipolar depression: Systematic review and meta-regression analyses. *European Neuropsychopharmacology*, *22*(1), 1–16. https://doi.org/https://doi.org/10.1016/j.euroneuro.2011.05.003

Arnone, D., Mckie, S., Elliott, R., Juhasz, G., Thomas, E. J., Downey, D., Williams, S., Deakin, J. F. W., & Anderson, I. M. (2013). State-dependent changes in hippocampal grey matter in depression. *Molecular Psychiatry*, *18*(12), 1265–1272. https://doi.org/10.1038/mp.2012.150

ASC. (2015). *Review of harm-benefit analysis in the use of animals in research.*

ASRU. (2015). *Advice note 05/2015: The Harm-Benefit Analysis Process – New Project Licence Applications.*

Aureli, F. (1992). Post-conflict behaviour among wild long-tailed macaques (Macaca fascicularis). *Behavioral Ecology and Sociobiology*, *31*(5), 329–337. https://doi.org/10.1007/BF00177773

Aureli, F., Das, M., & Veenema, H. C. (1997). Differential kinship effect on reconciliation in three species of macaques (Macaca fascicularis, M. fuscata, and M. sylvanus). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *111*(1), 91–99. https://doi.org/10.1037/0735-7036.111.1.91

Baker, H. J., Bradford, L. G., & Montes, L. F. (1971). Dermatophytosis due to Microsporum canis in a rhesus monkey. *Journal of the American Veterinary Medical Association*, *159*(11), 1607–1611.

Baker, K. C., & Aureli, F. (1997). Behavioural Indicators of Anxiety: An Emperical Test in Chimpanzees. *Behaviour*, *134*, 1031–1050.

Baker, K. C., Bloomsmith, M. A., Oettinger, B., Neu, K., Griffis, C., & Schoof, V. A. M. (2013). Comparing options for pair housing rhesus macaques using behavioral welfare measures. *American Journal of Primatology*, *76*(1), 30–42. https://doi.org/10.1002/ajp.22190

Baker, K. C., Bloomsmith, M. A., Oettinger, B., Neu, K., Griffis, C., & Schoof, V. A. M. (2014). Comparing options for pair housing rhesus macaques using behavioral welfare measures. *American Journal of Primatology*, *76*(1), 30–42. https://doi.org/10.1002/ajp.22190

Baker, K. C., Bloomsmith, M. A., Oettinger, B., Neu, K., Griffis, C., Schoof, V., & Maloney, M. (2012). Benefits of pair housing are consistent across a diverse population of rhesus macaques. *Applied Animal Behaviour Science*, *138*(1–2), 148–156. https://doi.org/10.1016/j.applanim.2011.09.010

Bar-Haim, Y. (2007). Threat-Related Attentional Bias in Anxious and Nonanxious Individuals: A Meta-Analytic Study. *Psychological Bulletin.*, *133*(1).

Barton, K. (2022). *Package ‘ MuMIn ’ Version 1.46.0. 1*.

Bates, D., Maechler, M., Bolker, B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *Package ‘lme4’ Version 1.1-28*.

Bateson, M., Emmerson, M., Ergün, G., Monaghan, P., & Nettle, D. (2015). Opposite effects of early-life competition and developmental telomere attrition on cognitive biases in juvenile European starlings. *PLoS ONE*, *10*(7), 1–23. https://doi.org/10.1371/journal.pone.0132602

Bateson, M., & Nettle, D. (2016). The telomere lengthening conundrum - it could be biology. *Aging Cell*, *16*. https://doi.org/10.1111/acel.12555

Bateson, M., & Poirier, C. (2019). Can biomarkers of biological age be used to assess cumulative lifetime experience? *Animal Welfare*, *28*(1), 41–56. https://doi.org/10.7120/09627286.28.1.041

Bateson, P. (2011). *Review of Research Using Non-Human Primates*.

Bayne, K., & Würbel, H. (2014). The impact of environmental enrichment on the outcome variability and scientific validity of laboratory animal studies. *Revue Scientifique et Technique (International Office of Epizootics)*, *33*(1), 273–280. https://doi.org/10.7892/boris.67436

Beisner, B. A., & Isbell, L. A. (2008). Ground substrate affects activity budgets and hair loss in outdoor captive groups of rhesus macaques (Macaca mulatta). *American Journal of Primatology*, *70*(12), 1160–1168. https://doi.org/10.1002/ajp.20615

Beisner, B. A., & Isbell, L. A. (2009). Factors influencing hair loss among female captive rhesus macaques (Macaca mulatta). *Applied Animal Behaviour Science*, *119*(1–2), 91–100. https://doi.org/10.1016/j.applanim.2009.03.016

Bellanca, R. U., Lee, G. H., Vogel, K., Ahrens, J., Kroeker, R., Thom, J. P., & Worlein, J. M. (2014). A simple alopecia scoring system for use in colony management of laboratory-housed primates. *Journal of Medical Primatology*, *43*(3), 153–161. https://doi.org/10.1111/jmp.12107

Bethell, E., Holmes, A., MacLarnon, A., & Semple, S. (2016). Emotion Evaluation and Response Slowing in a Non-Human Primate: New Directions for Cognitive Bias Measures of Animal Emotion? *Behavioral Sciences*, *6*(1), 2. https://doi.org/10.3390/bs6010002

Bethell, E. J., Holmes, A., MacLarnon, A., & Semple, S. (2012). Evidence That Emotion Mediates Social Attention in Rhesus Macaques. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0044387

Bleby, J. (1986). Animals (Scientific Procedures) Act 1986. *The Veterinary Record*, *119*(1), 22. https://doi.org/10.1136/vr.119.1.22-a

Blois-Heulin, C., & Jubin, R. (2004). Influence of the presence of seeds and litter on the behaviour of captive red-capped mangabeys Cercocebus torquatus torquatus. *Applied Animal Behaviour Science*, *85*(3–4), 349–362. https://doi.org/10.1016/j.applanim.2003.10.005

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche e Commericiali Di Firenze*, *8*, 3–62.

Bonini, L. (2019). Refinement techniques in non-human primate neuroscientific research. *Ann Ist Super Sanità*, *55*(4), 408–412. https://doi.org/10.4415/ANN_19_04_18

Brambell;, F. W. R., & Systems, T. C. to E. into the W. of A. kept under I. L. H. (1965). *Report of the Technical Committee to Enquire into the Welfare of Animals kept under Intensive Livestock Husbandry Systems*. London : Her Majesty's Stationery Office.

Browne, C. A., & Lucki, I. (2013). Antidepressant effects of ketamine: Mechanisms underlying fast-acting novel antidepressants. *Frontiers in Pharmacology*, *4 DEC*(December), 1–18. https://doi.org/10.3389/fphar.2013.00161

Bryant, C. E., Rupniak, N. M. J., & Iversen, S. D. (1988). Effects of Different Environmental Enrichment Devices on Cage Stereotypies and Autoaggression in Captive Cynomolgus Monkeys. *Journal of Medical Primatology*, *17*(5), 257–269. https://doi.org/10.1111/j.1600-0684.1988.tb00388.x

Camus, S. M. J., Blois-Heulin, C., Li, Q., Hausberger, M., & Bezard, E. (2013). Behavioural Profiles in Captive-Bred Cynomolgus Macaques: Towards Monkey Models of Mental Disorders? *PLoS ONE*, *8*(4), e62141. https://doi.org/10.1371/journal.pone.0062141

Camus, S. M. J., Rochais, C., Blois-Heulin, C., Li, Q., Hausberger, M., & Bezard, E. (2013). Birth Origin Differentially Affects Depressive-Like Behaviours: Are Captive-Born Cynomolgus Monkeys More Vulnerable to Depression than Their Wild-Born Counterparts? *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0067711

Camus, S. M. J., Rochais, C., Blois-Heulin, C., Li, Q., Hausberger, M., & Bezard, E. (2014). Depressive-like behavioral profiles in captive-bred single- and socially-housed rhesus and cynomolgus macaques: a species comparison. *Frontiers in Behavioral Neuroscience*, *8*(February), 1–15. https://doi.org/10.3389/fnbeh.2014.00047

Carlsson, H., Schapiro, S. J., Farah, I., & Hau, J. (2004). *Use of Primates in Research : A Global Overview*. *237*(May), 225–237. https://doi.org/10.1002/ajp.20054

Carpenter, C. R., & Schultz, A. H. (1940). A Field Study in Siam of the Behavior and Social Relations of the Gibbon (Hylobates Lar). *American Sociological Review*, *6*(3), 446. https://doi.org/10.2307/2086239

Cassidy, L. C., Hannibal, D. L., Semple, S., & McCowan, B. (2020). Improved behavioral indices of welfare in continuous compared to intermittent pair-housing in adult female rhesus macaques (Macaca mulatta). *American Journal of Primatology*, *82*(10). https://doi.org/10.1002/ajp.23189

Castles, D. L., Whiten, A., & Aureli, F. (1999). Social anxiety, relationships and self-directed behaviour among wild female olive baboons. *Animal Behaviour*, *58*(6), 1207–1215. https://doi.org/10.1006/anbe.1999.1250

Charbonneau, J. A., Amaral, D. G., & Bliss-Moreau, E. (2022). Social housing status impacts rhesus monkeys' affective responding in classic threat processing tasks. *Scientific Reports*, *12*(1), 4140. https://doi.org/10.1038/s41598-022-08077-4

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics: Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Chiaradia, I., & Lancaster, M. A. (2020). Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo. *Nature Neuroscience*, *23*(12), 1496–1508. https://doi.org/10.1038/s41593-020-00730-3

Cicchetti, D. v, & Feinstien, A. R. (1990). High Agreement but Low Kappa. *Journal of Clinical Epidemiology*, *43*(6), 551–585.

Clarke, A. S., Mason, W. A., & Moberg, G. P. (1988). Interspecific contrasts in responses of macaques to transport cage training. *Laboratory Animal Science*, *38*(3), 305–309.

Committee on Cruelty to Wild Animals. (1951). *Report*.

Contia, G., Hansman, C., Heckman, J. J., Novak, M. F. X., Ruggiero, A., & Suomi, S. J. (2012a). Primate evidence on the late health effects of early-life adversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8866–8871. https://doi.org/10.1073/pnas.1205340109

Contia, G., Hansman, C., Heckman, J. J., Novak, M. F. X., Ruggiero, A., & Suomi, S. J. (2012b). Primate evidence on the late health effects of early-life adversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8866–8871. https://doi.org/10.1073/pnas.1205340109

da Silva, A. N., Araujo, M. S., Pértille, F., & Zanella, A. J. (2022). How epigenetics can enhance pig welfare? *Animals*, *12*(1), 1–17. https://doi.org/10.3390/ani12010032

Darwin, C. (1872). *The expression of the emotions in man and animals*.

Das, M., Penke, Z., & van Hooff, J. A. R. A. M. (1998). Postconflict affiliation and stress-related behavior of long-tailed macaque aggressors. *International Journal of Primatology*, *19*(1), 53–71. https://doi.org/10.1023/A:1020354826422

Davenport, M. D., Tiefenbacher, S., Lutz, C. K., Novak, M. A., & Meyer, J. S. (2006). Analysis of endogenous cortisol concentrations in the hair of rhesus macaques. *General and Comparative Endocrinology*, *147*(3), 255–261. https://doi.org/10.1016/j.ygcen.2006.01.005

Dawkins, M. S. (1980). *Animal Suffering. The science of animal welfare*.

Dawkins, M. S. (2017). Animal welfare with and without consciousness. *Journal of Zoology*, *301*(1), 1–10. https://doi.org/10.1111/jzo.12434

de Waal, F. B. M. (1984). Coping with social tension: Sex differences in the effect of food provision to small rhesus monkey groups. *Animal Behaviour*, *32*(3), 765–773. https://doi.org/https://doi.org/10.1016/S0003-3472(84)80152-9

de Waal, F. B. M., & van Roosmalen, A. (1979). Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology*, *5*(1), 55–66. https://doi.org/10.1007/BF00302695

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

Dell'Anna, F., Schino, G., & Aureli, F. (2022). Anxiety in Geoffroy's spider monkeys (Ateles geoffroyi): Can scratching be used as an indicator? *American Journal of Primatology*, *n/a*(n/a), e23373. https://doi.org/https://doi.org/10.1002/ajp.23373

Departmental Committee on Experiments in Animals. (1965). *Report*.

Desimone, R., Olson, C., & Erickson, R. (1992). The Controlled Water Access Paradigm. *ILAR Journal*, *34*(3), 30–31. https://doi.org/10.1093/ilar.34.3.30

Directive 2010/63/EU of the european parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes., (2010).

Fairbanks, L. A., Jorgensen, M. J., Bailey, J. N., Breidenthal, S. E., Grzywa, R., & Laudenslager, M. L. (2011). Heritability and genetic correlation of hair cortisol in vervet

monkeys in low and higher stress environments. *Psychoneuroendocrinology*, *36*(8), 1201–1208. https://doi.org/10.1016/j.psyneuen.2011.02.013

Fanselow, M. S., & Dong, H.-W. (2010). Are the Dorsal and Ventral Hippocampus Functionally Distinct Structures? *Neuron*, *65*(1), 7–19. https://doi.org/https://doi.org/10.1016/j.neuron.2009.11.031

Feng, G. C. (2013). Factors affecting intercoder reliability: A Monte Carlo experiment. *Quality and Quantity*, *47*(5), 2959–2982. https://doi.org/10.1007/s11135-012-9745-9

Fox, John & Weisberg, S. (2011). An R Companion to Applied Regression. In *Thousand Oaks {CA} Sage* (p. Third edition). http://cran.r-project.org/web/packages/car/citation.html

Fraser, D. (2008). Understanding animal welfare. *Acta Veterinaria Scandinavica*, *50*(SUPPL. 1), 1–7. https://doi.org/10.1186/1751-0147-50-S1-S1

Fraser, D., Weary, D. M., Pajor, E. A., & Milligan, B. N. (1997). A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare*, *6*(3), 187–205.

Frodl, T., Jäger, M., Smajstrlova, I., Born, C., Bottlender, R., Palladino, T., Reiser, M., Möller, H.-J., & Meisenzahl, E. M. (2008). Effect of hippocampal and amygdala volumes on clinical outcomes in major depression: a 3-year prospective magnetic resonance imaging study. *Journal of Psychiatry & Neuroscience : JPN*, *33*(5), 423–430. https://pubmed.ncbi.nlm.nih.gov/18787661

Gottlieb, D. H., & Capitanio, J. P. (2013). Latent Variables Affecting Behavioral Response to the Human Intruder Test in Infant Rhesus Macaques (Macaca mulatta). *American Journal of Primatology*, *75*(4), 314–323. https://doi.org/10.1002/ajp.22107

Gottlieb, D. H., Maier, A., & Coleman, K. (2015). Evaluation of environmental and intrinsic factors that contribute to stereotypic behavior in captive rhesus macaques (Macaca mulatta). *Applied Animal Behaviour Science*, *171*, 184–191. https://doi.org/10.1016/j.applanim.2015.08.005

Gray, H., Bertrand, H., Mindus, C., Flecknell, P., Rowe, C., & Thiele, A. (2016). Physiological, behavioral, and scientific impact of different fluid control protocols in the rhesus macaque (Macaca mulatta). *ENeuro*, *3*(4), 1–15. https://doi.org/10.1523/ENEURO.0195-16.2016

Gray, H. E. (2016). *the Impact of Fluid Restriction Protocols on Rhesus Macaques (Macaca Mulatta) and Refinements To Their Use*. *September*.

Hadidian, J. (1980). Yawning in an Old World Monkey, Macaca nigra (Primates: Cercopithecidae). *Behaviour*, *75*(3/4), 133–147. http://www.jstor.org/stable/4534078

Hage, S. R., Ott, T., Eiselt, A. K., Jacob, S. N., & Nieder, A. (2014). Ethograms indicate stable well-being during prolonged training phases in rhesus monkeys used in neurophysiological research. *Laboratory Animals*, *48*(1), 82–87. https://doi.org/10.1177/0023677213514043

Halsey, L. G. (2018). Keeping Slim When Food Is Abundant: What Energy Mechanisms Could Be at Play? *Trends in Ecology & Evolution*, *33*(10), 745–753. https://doi.org/https://doi.org/10.1016/j.tree.2018.08.004

Hao, Y., Cui, Y., & Gu, X. (2016). Genome-wide DNA methylation profiles changes associated with constant heat stress in pigs as measured by bisulfite sequencing. *Scientific Reports*, *6*(May), 1–13. https://doi.org/10.1038/srep27507

Hao, Y., Liu, J. R., Zhang, Y., Yang, P. G., Feng, Y. J., Cui, Y. J., Yang, C. H., & Gu, X. H. (2016). The microRNA expression profile in porcine skeletal muscle is changed by constant heat stress. *Animal Genetics*, *47*(3), 365–369. https://doi.org/10.1111/age.12419

Harlow, & Zimmermann. (1959). Affectional in. *Science*, *130*(3373), 421–432.

Harrison, R. (1964). *Animal Machines*.

Hau, J., & Schapiro, S. (2006). Non-human primates in biomedical research. *Scandinavian Journal of Laboratory Animal Science*, *33*, 9–12.

Heagerty, A., Wales, R. A., Prongay, K., Gottlieb, D. H., & Coleman, K. (2017). Social hair pulling in captive rhesus macaques (Macaca mulatta). *American Journal of Primatology*, *79*(12), 1–11. https://doi.org/10.1002/ajp.22720

Henazi, S. P., & Barrett, L. (1999). The value of grooming to female primates. *Primates; Journal of Primatology*, *40*(1), 47–59. https://doi.org/10.1007/BF02557701

Hennessy, M. B., Chun, K., & Capitanio, J. P. (2017). Depressive-like behavior, its sensitization, social buffering, and altered cytokine responses in rhesus macaques moved from outdoor social groups to indoor housing. *Social Neuroscience*, *12*(1), 65–75. https://doi.org/10.1080/17470919.2016.1145595

Hennessy, M. B., McCowan, B., Jiang, J., & Capitanio, J. P. (2014). Depressive-like behavioral response of adult male rhesus monkeys during routine animal husbandry procedure.

Frontiers in Behavioral Neuroscience, *8*(September), 1–8. https://doi.org/10.3389/fnbeh.2014.00309

Hinde, R. A., & Rowell, T. E. (1962). Communication By Postures and Facial Expressions in the Rhesus Monkey (Macaca Mulatta). *Proceedings of the Zoological Society of London*, *138*(1), 1–21. https://doi.org/10.1111/j.1469-7998.1962.tb05684.x

Hirschfeld, R. M. A. (2001). The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*(6), 244–254. https://doi.org/10.4088/pcc.v03n0609

Home Office. (2020). *Annual Statistics of Scientific Procedures on Living Animals, Great Britain 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/327854/sp animals13.pdf*. 1–59.

Honess, P. E., Gimpel, J. L., Wolfensohn, S. E., & Mason, G. J. (2005). Alopecia scoring: The quantitative assessment of hair loss in captive macaques. *ATLA Alternatives to Laboratory Animals*, *33*(3), 193–206. https://doi.org/10.1177/026119290503300308

Howarth, E. R. I., Kemp, C., Thatcher, H. R., Szott, I. D., Farningham, D., Witham, C. L., Holmes, A., Semple, S., & Bethell, E. J. (2021). Developing and validating attention bias tools for assessing trait and state affect in animals: A worked example with Macaca mulatta. *Applied Animal Behaviour Science*, *234*, 105198. https://doi.org/https://doi.org/10.1016/j.applanim.2020.105198

Hudetz, J. A., & Pagel, P. S. (2010). Neuroprotection by Ketamine: A Review of the Experimental and Clinical Evidence. *Journal of Cardiothoracic and Vascular Anesthesia*, *24*(1), 131–142. https://doi.org/10.1053/j.jvca.2009.05.008

Ieraci, A., Madaio, A. I., Mallei, A., Lee, F. S., & Popoli, M. (2016). Brain-Derived Neurotrophic Factor Val66Met Human Polymorphism Impairs the Beneficial Exercise-Induced Neurobiological Changes in Mice. *Neuropsychopharmacology*, *41*(13), 3070–3079. https://doi.org/10.1038/npp.2016.120

Iredale, S. K., Nevill, C. H., & Lutz, C. K. (2010). The influence of observer presence on baboon (Papio spp.) and rhesus macaque (Macaca mulatta) behavior. *Applied Animal Behaviour Science*, *122*(1), 53–57. https://doi.org/10.1016/j.applanim.2009.11.002

Isbell, L. A. (1995). Seasonal and social correlates of changes in hair, skin, and scrotal condition in vervet monkeys (Cercopithecus aethiops) of Amboseli National Park, Kenya. *American Journal of Primatology.*, *36*(1), 61–70.

Isoda, M., Tsutsui, K. I., Katsuyama, N., Naganuma, T., Saito, N., Furusawa, Y., Mushiake, H., Taira, M., & Tanji, J. (2005). Design of a head fixation device for experiments in behaving monkeys. *Journal of Neuroscience Methods*, *141*(2), 277–282. https://doi.org/10.1016/j.jneumeth.2004.07.003

Jackowski, A., Perera, T. D., Abdallah, C. G., Garrido, G., Tang, C. Y., Martinez, J., Mathew, S. J., Gorman, J. M., Rosenblum, L. A., Smith, E. L. P., Dwork, A. J., Shungu, D. C., Kaffman, A., Gelernter, J., Coplan, J. D., & Kaufman, J. (2011). Early-life stress, corpus callosum development, hippocampal volumetrics, and anxious behavior in male nonhuman primates. *Psychiatry Research: Neuroimaging*, *192*(1), 37–44. https://doi.org/https://doi.org/10.1016/j.pscychresns.2010.11.006

Jarczok, M. N., Koenig, J., Shively, C. A., & Thayer, J. F. (2018). Behavioral depression is associated with increased vagally mediated heart rate variability in adult female cynomolgus monkeys (Macaca fascicularis). *International Journal of Psychophysiology*, *131*(October 2017), 139–143. https://doi.org/10.1016/j.ijpsycho.2017.11.004

Jennings, M., Prescott, M. J., Buchanan-Smith, H. M., Gamble, M. R., Gore, M., Hawkins, P., Hubrecht, R., Hudson, S., Jennings, M., Keeley, J. R., Morris, K., Morton, D. B., Owen, S., Pearce, P. C., Prescott, M. J., Robb, D., Rumble, R. J., Wolfensohn, S., & Buist, D. (2009). Refinements in husbandry, care and common procedures for non-human primates: Ninth report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Laboratory Animals*, *43 Suppl 1*, 1–47. https://doi.org/10.1258/la.2008.007143

Judge, P. G., Evans, D. W., Schroepfer, K. K., & Gross, A. C. (2011). Perseveration on a reversal-learning task correlates with rates of self-directed behavior in nonhuman primates. *Behavioural Brain Research*, *222*(1), 57–65. https://doi.org/10.1016/j.bbr.2011.03.016

Kemnitz, J. W., & Francken, G. A. (1986). Characteristics of spontaneous obesity in male rhesus monkeys. *Physiology and Behavior*, *38*(4), 477–483. https://doi.org/10.1016/0031-9384(86)90414-2

Kikusui, T., Winslow, J. T., & Mori, Y. (2006). Social buffering: relief from stress and anxiety. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2215–2228. https://doi.org/10.1098/rstb.2006.1941

Kim, J.-I., Lee, J. W., Lee, Y. A., Lee, D.-H., Han, N. S., Choi, Y.-K., Hwang, B. R., Kim, H. J., & Han, J. S. (2013). Sexual activity counteracts the suppressive effects of chronic stress on adult hippocampal neurogenesis and recognition memory. *Brain Research*, *1538*, 26–40. https://doi.org/https://doi.org/10.1016/j.brainres.2013.09.007

Koolschijn, P. C. M. P., van Haren, N. E. M., Lensvelt-Mulders, G. J. L. M., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Human Brain Mapping*, *30*(11), 3719–3735. https://doi.org/10.1002/hbm.20801

Kozorovitskiy, Y., Gross, C. C., Kopil, C., Battaglia, L., McBreen, M., Stranahan, A. M., & Gould, E. (2005). Experience induces structural and biochemical changes in the adult primate brain. *Proceedings of the National Academy of Sciences*, *102*(48), 17478–17482. https://doi.org/10.1073/pnas.0508817102

Kramer, J., Fahey, M., Santos, R., Carville, A., Wachtman, L., & Mansfield, K. (2010). Alopecia in Rhesus macaques correlates with immunophenotypic alterations in dermal inflammatory infiltrates consistent with hypersensitivity etiology. *Journal of Medical Primatology*, *39*(2), 112–122. https://doi.org/10.1111/j.1600-0684.2010.00402.x

Kroshko, J., Clubb, R., Harper, L., Mellor, E., Moehrenschlager, A., & Mason, G. (2016). Stereotypic route tracing in captive Carnivora is predicted by species-typical home range sizes and hunting styles. *Animal Behaviour*, *117*, 197–209. https://doi.org/10.1016/j.anbehav.2016.05.010

Kurth, F., Gaser, C., & Luders, E. (2015). A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (SPM). *Nature Protocols*, *10*(2), 293–304. https://doi.org/10.1038/nprot.2015.014

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lagarde, D., Laurent, J., Milhaud, C., Andre, E., Aubin, H. J., & Anton, G. (1990). Behavioral effects induced by beta CCE in free or restrained rhesus monkeys (Macaca mulatta).

Pharmacology Biochemistry and Behavior, *35*(3), 713–719. https://doi.org/https://doi.org/10.1016/0091-3057(90)90312-6

Lair, S., Crawshaw, G. J., Mehren, K. G., & Perrone, M. A. (1999). Diagnosis of hypothyroidism in a western lowland gorilla (Gorilla gorilla gorilla) using human thyroid-stimulating hormone assay. *Journal of Zoo and Wildlife Medicine : Official Publication of the American Association of Zoo Veterinarians*, *30*(4), 537–540.

Lanz, F., Lanz, X., Scherly, A., Moret, V., Gaillard, A., Gruner, P., Hoogewoud, H. M., Belhaj-Saif, A., Loquet, G., & Rouiller, E. M. (2013). Refined methodology for implantation of a head fixation device and chronic recording chambers in non-human primates. *Journal of Neuroscience Methods*, *219*(2), 262–270. https://doi.org/10.1016/j.jneumeth.2013.07.015

Li, X., Xu, F., Xie, L., Ji, Y., Cheng, K., Zhou, Q., Wang, T., Shively, C., Wu, Q., Gong, W., Fang, L., Zhan, Q., Melgiri, N. D., & Xie, P. (2013). Depression-Like Behavioral Phenotypes by Social and Social Plus Visual Isolation in the Adult Female Macaca fascicularis. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0073293

Li, Y., Zhu, X., Ju, S., Yan, J., Wang, D., Zhu, Y., & Zang, F. (2017). Detection of volume alterations in hippocampal subfields of rats under chronic unpredictable mild stress using 7T MRI: A follow-up study. *Journal of Magnetic Resonance Imaging*, *46*. https://doi.org/10.1002/jmri.25667

Line, S. W., Markowitz, H., Morgan, K. N., & Strong, S. (1989). Evaluation of attempts to enrich the environment of singly- caged nonhuman primates. *In: DRISCOLL, J. W. (Ed.): Animal Care and Use in Behavioral Research: Regulations, Issues and Applications Animal Welfare Information Center, National Agricultural Library, 103–117*.

Luchins, K. R., Baker, K. C., Gilbert, M. H., Blanchard, J. L., Liu, D. X., Myers, L., & Bohm, R. P. (2011). Application of the diagnostic evaluation for alopecia in traditional veterinary species to laboratory rhesus macaques (Macaca mulatta). *Journal of the American Association for Laboratory Animal Science*, *50*(6), 926–938.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lutz, C. K. (2021a). Effect of pregnancy and age on alopecia in adult female baboons (papio hamadryas spp). *Journal of the American Association for Laboratory Animal Science*, *60*(4), 484–488. https://doi.org/10.30802/AALAS-JAALAS-20-000156

Lutz, C. K. (2021b). Effect of pregnancy and age on alopecia in adult female baboons (papio hamadryas spp). *Journal of the American Association for Laboratory Animal Science*, *60*(4), 484–488. https://doi.org/10.30802/AALAS-JAALAS-20-000156

Lutz, C. K., Coleman, K., Worlein, J., & Novak, M. A. (2013). Hair loss and hair-pulling in rhesus macaques (Macaca mulatta). *Journal of the American Association for Laboratory Animal Science*, *52*(4), 454–457.

Lutz, C. K., Menard, M. T., Rosenberg, K., Meyer, J. S., Melinda, A., & Primate, N. (2019). *Alopecia in rhesus macaques (Macaca mulatta): Association with pregnancy and chronic stress*. *48*(4), 251–256. https://doi.org/10.1111/jmp.12419.Alopecia

Lutz, C., Well, A., & Novak, M. (2003). Stereotypic and self-injurious behavior in rhesus macaques: A survey and retrospective analysis of environment and early experience. *American Journal of Primatology*, *60*(1), 1–15. https://doi.org/10.1002/ajp.10075

Maestripieri, D. (1993). Maternal Anxiety in Rhesus Macaques (Macaca mulatta). *Ethology*, *95*(1), 19–31. https://doi.org/https://doi.org/10.1111/j.1439-0310.1993.tb00453.x

Maestripieri, D., Schino, G., Aureli, F., & Troisi, P. (1992). Maestripieri et al 1992 a modest proposal displacement activities as indicator of emotions in primates.pdf. *Animal Behaviour*, *44*(2), 967–979.

Majolo, B., Ventura, R., & Koyama, N. F. (2009). Anxiety level predicts post-conflict behaviour in wild japanese macaques (macaca fuscata yakui). *Ethology*, *115*(10), 986–995. https://doi.org/10.1111/j.1439-0310.2009.01685.x

Major, C. A., Kelly, B. J., Novak, M. A., Davenport, M. D., Stonemetz, K. M., & Meyer, J. S. (2009). The anxiogenic drug FG7142 increases self-injurious behavior in male rhesus monkeys (Macaca mulatta). *Life Sciences*, *85*(21–22), 753–758. https://doi.org/10.1016/j.lfs.2009.10.003

Maninger, N., Capitanio, J. P., Mason, W. A., Ruys, J. D., & Mendoza, S. P. (2010). Acute and chronic stress increase DHEAS concentrations in rhesus monkeys. *Psychoneuroendocrinology*, *35*(7), 1055–1062. https://doi.org/10.1016/j.psyneuen.2010.01.006

Marais, L., Daniels, W., Brand, L., Viljoen, F., Hugo, C., & Stein, D. J. (2006). Psychopharmacology of maternal separation anxiety in vervet monkeys. *Metabolic Brain Disease*, *21*(2–3), 201–210. https://doi.org/10.1007/s11011-006-9011-8

Mason, G., & Rushen, J. (2008). *Stereotypic Animal Behaviour. Fundamentals and Applications to Welfare*. CAB.

Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. *British Journal of Psychiatry*, *130*(1), 79–83. https://doi.org/10.1192/bjp.130.1.79

McCulloch, S. (2012). A Critique of FAWC's Five Freedoms as a Framework for the Analysis of Animal Welfare. *Journal of Agricultural and Environmental Ethics*, *26*. https://doi.org/10.1007/s10806-012-9434-7

McFarland, D. (1989). *Problems of Animal Behaviour* (L. S. & Technical, Ed.).

Mechling, H., & Netz, Y. (2009). Aging and inactivity—capitalizing on the protective effect of planned physical activity in old age. *European Review of Aging and Physical Activity*, *6*, 89–97. https://doi.org/10.1007/s11556-009-0052-y

Meishvili, N. v., & Chalyan, V. G. (2021). Body Weight and Anxiety in Cynomolgus Macaques (Macaca fascicularis) Males. *Bulletin of Experimental Biology and Medicine*, *170*(6), 802–804. https://doi.org/10.1007/s10517-021-05159-7

Mellor, D. J., & Beausoleil, N. J. (2015). Extending the "Five Domains" model for animal welfare assessment to incorporate positive welfare states. *Animal Welfare*, *24*(3), 241–253. https://doi.org/10.7120/09627286.24.3.241

Mellor, D. J., & Reid, C. S. W. (1994). Concepts of animal well-being and predicting the impact of procedures on experimental animals. *Improving the Well-Being of Animals in the Research Environment*, 3–18. http://org.uib.no/dyreavd/harm-benefit/Concepts of animal well-being and predicting.pdf

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Michopoulos, V., Reding, K. M., Wilson, M. E., & Toufexis, D. (2012). Social subordination impairs hypothalamic-pituitary-adrenal function in female rhesus monkeys. *Hormones and Behavior*, *62*(4), 389–399. https://doi.org/10.1016/j.yhbeh.2012.07.014

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin*, *133*(1), 25–45. https://doi.org/10.1037/0033-2909.133.1.25

Mineka, S., & Suomi, S. J. (1978). Social separation in monkeys. *Psychological Bulletin*, *85*(6), 1376–1400. https://doi.org/10.1037//0033-2909.85.6.1376

Mirman, D., Landrigan, J.-F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia*, *115*, 112–123. https://doi.org/10.1016/j.neuropsychologia.2017.08.025

Mitra, R., Sundlass, K., Parker, K. J., Schatzberg, A. F., & Lyons, D. M. (2006). Social stress-related behavior affects hippocampal cell proliferation in mice. *Physiology & Behavior*, *89*(2), 123–127. https://doi.org/https://doi.org/10.1016/j.physbeh.2006.05.047

Morais, M., Patrício, P., Mateus-Pinheiro, A., Alves, N. D., Machado-Santos, A. R., Correia, J. S., Pereira, J., Pinto, L., Sousa, N., & Bessa, J. M. (2017). The modulation of adult neuroplasticity is involved in the mood-improving actions of atypical antipsychotics in an animal model of depression. *Translational Psychiatry*, *7*(6), e1146–e1146. https://doi.org/10.1038/tp.2017.120

Moser, M.-B., & Moser, E. (1998). Moser MB, Moser EI. Functional differentiation in the hippocampus. Hippocampus 8: 608-619. *Hippocampus*, *8*, 608–619. https://doi.org/10.1002/(SICI)1098-1063(1998)8:6<608::AID-HIPO3>3.0.CO;2-7

National Institutes of Health. (2011). *NIH Guidelines for Care and Use of Animals for Experimental Procedures*.

NC3Rs. (n.d.). *The Macaque Website| Welfare assessment*. https://macaques.nc3rs.org.uk/welfare-assessment/

NC3Rs. (2017). NC3Rs Guidelines: Non-human primate accommodation, care and use. *National Centre of Replacement, Refinement, & Reduction of Animals in Research*, 1–52.

Nettle, D., Gadalla, S. M., Lai, T.-P., Susser, E., Bateson, M., & Aviv, A. (2021). Measurement of Telomere Length for Longitudinal Analysis: Implications of Assay Precision. *American Journal of Epidemiology*, *190*(7), 1406–1413. https://doi.org/10.1093/aje/kwab025

Novak, J., Bailoo, J. D., Melotti, L., & Würbel, H. (2016). Effect of Cage-Induced Stereotypies on Measures of Affective State and Recurrent Perseveration in CD-1 and C57BL/6 Mice. *PloS One*, *11*(5), e0153203. https://doi.org/10.1371/journal.pone.0153203

Novak, M. A., Hamel, A. F., Coleman, K., Lutz, C. K., Worlein, J., Menard, M., Ryan, A., Rosenberg, K., & Meyer, J. S. (2014). Hair loss and hypothalamic-pituitary-adrenocortical axis activity in captive rhesus macaques (Macaca mulatta). *Journal of the American Association for Laboratory Animal Science*, *53*(3), 261–266.

Novak, M. A., Hamel, A. F., Kelly, B. J., Dettmer, A. M., & Meyer, J. S. (2013). Stress, the HPA axis, and nonhuman primate well-being: A review. *Applied Animal Behaviour Science*, *143*(2), 135–149. https://doi.org/https://doi.org/10.1016/j.applanim.2012.10.012

Novak, M. A., & Kessler, M. J. (2009). Alopecia: Possible causes and treatments, particularly in captive nonhuman primates. *Comparative Medicine*, *59*(4), 320.

Novak, M. A., Menard, M. T., El-Mallah, S. N., Rosenberg, K., Lutz, C. K., Worlein, J., Coleman, K., & Meyer, J. S. (2017). Assessing significant (>30%) alopecia as a possible biomarker for stress in captive rhesus monkeys (Macaca mulatta). *American Journal of Primatology*, *79*(1), 1–8. https://doi.org/10.1002/ajp.22547

Orlans, F. B. (1991). Prolonged Water Deprivation : A Case Study in Decision Making by an IACUC. *Issues For Institutional Animal Care and Use Committees (LACUCs)*, *33*(3), 48–52. https://doi.org/10.1093/ilar.33.3.48.

Ortiz-Rios, M., Haag, M., Balezeau, F., Frey, S., Thiele, A., Murphy, K., & Schmid, M. C. (2018). Improved methods for MRI-compatible implants in nonhuman primates. *Journal of Neuroscience Methods*, *308*(August), 377–389. https://doi.org/10.1016/j.jneumeth.2018.09.013

Otovic, P., & Hutchinson, E. (2015). Limits to using HPA axis activity as an indication of animal welfare. *Altex*, *32*(1), 41–50. https://doi.org/10.14573/altex.1406161

Paul, E. S., Harding, E. J., & Mendl, M. (2005). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*, *29*(3), 469–491. https://doi.org/10.1016/j.neubiorev.2005.01.002

Paule, M. G., Li, M., Allen, R. R., Liu, F., Zou, X., Hotchkiss, C., Hanig, J. P., Patterson, T. A., Slikker, W., & Wang, C. (2011). Ketamine anesthesia during the first week of life can cause long-lasting cognitive deficits in rhesus monkeys. *Neurotoxicology and Teratology*, *33*(2), 220–230. https://doi.org/10.1016/j.ntt.2011.01.001

Pepper, G. v., Bateson, M., & Nettle, D. (2018). Telomeres as integrative markers of exposure to stress and adversity: A systematic review and meta-analysis. *Royal Society Open Science*, *5*(8). https://doi.org/10.1098/rsos.180744

Perera, T. D., Dwork, A. J., Keegan, K. A., Thirumangalakudi, L., Lipira, C. M., Joyce, N., Lange, C., Higley, J. D., Rosoklija, G., Hen, R., Sackeim, H. A., & Coplan, J. D. (2011). Necessity of hippocampal neurogenesis for the therapeutic action of antidepressants in adult Nonhuman primates. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0017600

Pértille, F., Brantsæter, M., Nordgreen, J., Coutinho, L. L., Janczak, A. M., Jensen, P., & Guerrero-Bosagna, C. (2017). DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. *Journal of Experimental Biology*, *220*(19), 3579–3587. https://doi.org/10.1242/jeb.157891

Pértille, F., Ibelli, A. M. G., Sharif, M. el, Poleti, M. D., Fröhlich, A. S., Rezaei, S., Ledur, M. C., Jensen, P., Guerrero-Bosagna, C., & Coutinho, L. L. (2020). Putative Epigenetic Biomarkers of Stress in Red Blood Cells of Chickens Reared Across Different Biomes. *Frontiers in Genetics*, *11*(November). https://doi.org/10.3389/fgene.2020.508809

Peterson, E. J., Worlein, J. M., Lee, G. H., Dettmer, A. M., Varner, E. K., & Novak, M. A. (2017). Rhesus macaques (Macaca mulatta) with self-injurious behavior show less behavioral anxiety during the human intruder test. *American Journal of Primatology*, *79*(1), 1–8. https://doi.org/10.1002/ajp.22569

Pfefferle, D., Plümer, S., Burchardt, L., Treue, S., & Gail, A. (2018a). Assessment of stress responses in rhesus macaques (Macaca mulatta) to daily routine procedures in system neuroscience based on salivary cortisol concentrations. *PLoS ONE*, *13*(1), 1–13. https://doi.org/10.1371/journal.pone.0190190

Pfefferle, D., Plümer, S., Burchardt, L., Treue, S., & Gail, A. (2018b). Assessment of stress responses in rhesus macaques (Macaca mulatta) to daily routine procedures in system neuroscience based on salivary cortisol concentrations. *PLoS ONE*, *13*(1), 1–13. https://doi.org/10.1371/journal.pone.0190190

Pickard, J. (2013). *LIFETIME EXPERIENCE IN NON- - HUMAN PRIMATES USED IN*. *77*.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & {R Development Core Team}. (2011). *\texttt{nlme}: Linear and Nonlinear Mixed Effects Models*. http://www.r-project.org/

Poirier, C., Bateson, M., Gualtieri, F., Armstrong, E. A., Laws, G. C., Boswell, T., & Smulders, T. v. (2019). Validation of hippocampal biomarkers of cumulative affective experience. *Neuroscience & Biobehavioral Reviews*, *101*, 113–121. https://doi.org/10.1016/J.NEUBIOREV.2019.03.024

Poirier, C., Oliver, C. J., Flecknell, P., & Bateson, M. (2019). Pacing behavior in laboratory macaques is an unreliable indicator of acute stress. *BioRxiv*, *April*, 1–8. https://doi.org/10.1038/s41598-019-43695-5

Polanco, A. (2021). Validating Markers of Cumulative Experience in Laboratory Rhesus Monkeys (Macaca mulatta). *A Thesis Presented to The University of Guelph In*.

Pomerantz, O., Meiri, S., & Terkel, J. (2013). Socio-ecological factors correlate with levels of stereotypic behavior in zoo-housed primates. *Behavioural Processes*, *98*(August), 85–91. https://doi.org/10.1016/j.beproc.2013.05.005

Pomerantz, O., Terkel, J., Suomi, S. J., & Paukner, A. (2012). Stereotypic head twirls, but not pacing, are related to a `pessimistic'-like judgment bias among captive tufted capuchins (Cebus apella). *Animal Cognition*, *15*(4), 689–698. https://doi.org/10.1007/s10071-012-0497-7

Poole, T. (1997). Happy animals make good science. *Laboratory Animals*, *31*(2), 116–124. https://doi.org/10.1258/002367797780600198

Prescott, M. J., Brown, V. J., Flecknell, P. A., Gaffan, D., Garrod, K., Lemon, R. N., Parker, A. J., Ryder, K., Schultz, W., Scott, L., Watson, J., & Whitfield, L. (2010). Refinement of the use of food and fluid control as motivational tools for macaques used in behavioural neuroscience research: Report of a Working Group of the NC3Rs. *Journal of Neuroscience Methods*, *193*(2), 167–188. https://doi.org/10.1016/j.jneumeth.2010.09.003

Prescott, M. J., Leach, M. C., & Truelove, M. A. (2022). *Harmonisation of welfare indicators for macaques and marmosets used or bred for research [version 1 ; peer review : 1 approved with reservations]*.

Qin, D., Chu, X., Feng, X., Li, Z., Yang, S., Lü, L., Yang, Q., Pan, L., Yin, Y., Li, J., Xu, L., Chen, L., & Hu, X. (2015). The first observation of seasonal affective disorder symptoms in Rhesus macaque. *Behavioural Brain Research*, *292*(32), 463–469. https://doi.org/10.1016/j.bbr.2015.07.005

Quigley, M. (2007). Non-human primates: the appropriate subjects of biomedical research? *Journal of Medical Ethics*, *33*(11), 655–658. https://doi.org/10.1136/jme.2007.020784

Ralph, C. R., & Tilbrook, A. J. (2016). Invited Review: The usefulness of measuring glucocorticoids for assessing animal welfare. *Journal of Animal Science*, *94*(2), 457–470. https://doi.org/10.2527/jas.2015-9645

Rand, C. S., & Macgregor, A. M. (1990). Morbidly obese patients' perceptions of social discrimination before and after surgery for obesity. *Southern Medical Journal*, *83*(12), 1390–1395. https://doi.org/10.1097/00007611-199012000-00006

Ratterree, M. S., & Baskin, G. B. (1992). Congenital hypotrichosis in a rhesus monkey. *Laboratory Animal Science*, *42*(4), 410–412.

Regier, D. A., Kuhl, E. A., & Kupfer, D. J. (2013). The DSM-5: Classification and criteria changes. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *12*(2), 92–98. https://doi.org/10.1002/wps.20050

Reinhardt, V. (1994a). Caged rhesus macaques voluntarily work for ordinary food. *Primates*, *35*(1), 95–98. https://doi.org/10.1007/BF02381490

Reinhardt, V. (1994b). Pair-housing rather than single-housing for laboratory rhesus macaques. *Journal of Medical Primatology*, *23*(8), 426–431. https://doi.org/10.1111/j.1600-0684.1994.tb00131.x

Reinhardt, V. (2008). *Taking Better Care of Monkeys and Apes: Refinement of Housing and Handling Practices for Caged Nonhuman Primates*. Animal Welfare Institute.

Reinhardt, V., Cowley, D., Eisele, S., & Scheffler, J. (1991). Avoiding undue cortisol responses to venipuncture in adult male rhesus macaques. In *Animal technology : journal of the Institute of Animal Technology: Vol. v. 42*.

Reinhardt, V., Reinhardt, A., & Houser, D. (1986). Hair pulling and eating in captive rhesus monkey troops. *Folia Primatologica; International Journal of Primatology*, *47*(2–3), 158–164. https://doi.org/10.1159/000156272

Reinhardt, V., Rodgers, J. C., Shayne, K., Schwartz, J., Bell, L., Vila, P. M., Kerwin, A., Rodgers, J. C., Schultz, P., (Genny), G. A.-K., Donnelly, M., Augusto, V., Moore, A., & Cawston, D. (2007). *A Discussion Relating to Animals in the Research La*.

Reite, M. L., Rhodes, J. M., Kavan, E., & Adey, W. R. (1965). Normal Sleep Patterns in Macaque Monkey. *Archives of Neurology*, *12 (2)*, 133–144. https://doi.org/10.1001/archneur.1965.00460260023003

Rommeck, I., Anderson, K., Heagerty, A., Cameron, A., & McCowan, B. (2009). Risk factors and remediation of self-injurious and self-abuse behavior in Rhesus Macaques. *Journal of Applied Animal Welfare Science*, *12*(1), 61–72. https://doi.org/10.1080/10888700802536798

Rossi, J. (2009). Nonhuman Primate Research: The Wrong Way to Understand Needs and Necessity. *The American Journal of Bioethics*, *9*(5), 21–23. https://doi.org/10.1080/15265160902788728

Saleem, K., & Logothetis, N. K. (2012). *A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates*. *2nd editio*.

Santarelli, L., Saxe, M., Gross, C., Surget, A., Battaglia, F., Dulawa, S., Weisstaub, N., Lee, J., Duman, R., Arancio, O., Belzung, C., & Hen, R. (2003). Requirement of Hippocampal Neurogenesis for the Behavioral Effects of Antidepressants. *Science*, *301*(5634), 805–809. https://doi.org/10.1126/science.1083328

(SCHEER), S. C. on H. E. and E. R. (2017). *Final Opinion on The need for non-human primates in biomedical research, production and testing of products and devices (update 2017)*.

Schino, G., Maestripierf, D., Scucchf, S., Giovanni Turillazzi, P., Schino, G., & Giovanni Turillazzi, P. (1990). Social Tension in Familiar and Unfamiliar Pairs of Long-Tailed Macaques. *Behaviour*, *113*(3–4), 264–272. https://doi.org/10.1163/156853990X00518

Schino, G., Perretta, G., Taglioni, A. M., Monaco, V., & Troisi, A. (1996). Primate displacement activities as an ethopharmacological model of anxiety. *Anxiety*, *2*(4), 186–191. https://doi.org/10.1002/(SICI)1522-7154(1996)2:4<186::AID-ANXI5>3.0.CO;2-M

Schino, G., Troisi, A., Perretta, G., & Monaco, V. (1991). Measuring anxiety in nonhuman primates: Effect of lorazepam on macaque scratching. *Pharmacology, Biochemistry and Behavior*, *38*(4), 889–891. https://doi.org/10.1016/0091-3057(91)90258-4

Schoenfeld, T. J., & Gould, E. (2012). Stress, stress hormones, and adult neurogenesis. *Experimental Neurology*, *233*(1), 12–21. https://doi.org/https://doi.org/10.1016/j.expneurol.2011.01.008

Shively, C. A., & Clarkson, T. B. (1987). Regional Obesity and Coronary Artery Atherosclerosis in Females: a Non-human Primate Model. *Acta Medica Scandinavica*, *222*(723 S), 71–78. https://doi.org/10.1111/j.0954-6820.1987.tb05930.x

Shively, C. A., & Day, S. M. (2015). Social inequalities in health in nonhuman primates. *Neurobiology of Stress*, *1*(1), 156–163. https://doi.org/10.1016/j.ynstr.2014.11.005

Shively, C. A., Laber-Laird, K., & Anton, R. F. (1997). Behavior and physiology of social stress and depression in female cynomolgus monkeys. *Biological Psychiatry*, *41*(8), 871–882. https://doi.org/10.1016/S0006-3223(96)00185-0

Shively, C. A., Register, T. C., Friedman, D. P., Morgan, T. M., Thompson, J., & Lanier, T. (2005). Social stress-associated depression in adult female cynomolgus monkeys (Macaca fascicularis). *Biological Psychology*, *69*(1 SPEC. ISS.), 67–84. https://doi.org/10.1016/j.biopsycho.2004.11.006

Shively, C. A., & Willard, S. L. (2012). Behavioral and neurobiological characteristics of social stress versus depression in nonhuman primates. *Experimental Neurology*, *233*(1), 87–94. https://doi.org/10.1016/j.expneurol.2011.09.026

Smith, J. J., Hadzic, V., Li, X., Liu, P., Day, T., Utter, A., Kim, B., Washington, I. M., Basso, M. A., & 1. (2006). Objective measures of health and well being in laboratory rhesus. *J Medical Primatology*, *35*, 388–396. https://doi.org/10.1111/j.1600-0684.2006.00188.x

Snyder, J. S., Glover, L. R., Sanzone, K. M., Kamhi, J. F., & Cameron, H. A. (2009). The effects of exercise and stress on the survival and maturation of adult-generated granule cells. *Hippocampus*, *19*(10), 898–906. https://doi.org/10.1002/hipo.20552

Steinmetz, H. W., Kaumanns, W., Dix, I., Heistermann, M., Fox, M., & Kaup, F. J. (2006). Coat condition, housing condition and measurement of faecal cortisol metabolites - A non-invasive study about alopecia in captive rhesus macaques (Macaca mulatta). *Journal of Medical Primatology*, *35*(1), 3–11. https://doi.org/10.1111/j.1600-0684.2005.00141.x

Steinmetz, H. W., Kaumanns, W., Neimeier, K.-A., & Kaup, F.-J. (2005). Dermatologic investigation of alopecia in rhesus macaques (Macaca mulatta). *Journal of Zoo and Wildlife Medicine : Official Publication of the American Association of Zoo Veterinarians*, *36*(2), 229–238. https://doi.org/10.1638/04-054.1

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience*, *15*(10), 655–669. https://doi.org/10.1038/nrn3785

Summerhoff, G. (1990). *Life, brain and consciousness.* North Holland: Amsterdam Trust.

Swencionis, C., Wylie-Rosett, J., M.A., M. R. L., Ginsberg, M., Christopher Cimino, Sylvia Wassertheil-Smoller, A. C., & Segal-Isaacson, C. J. (2013). Weight Change, Psychological Well-Being, and Vitality in Adults Participating in a Cognitive-Behavioral Weight Loss Program. *Health Psychol*, *32*(4), 439–446. https://doi.org/10.1037/a0029186.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13–21. https://doi.org/10.1007/s00265-010-1037-6

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*(July), 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Thornton, A. E., Cox, D. N., Whitfield, K., Fouladi, R. T., Thornton, A. E., Cox, D. N., Whitfield, K., Fouladi, R. T., & Columbia, B. (2008). *Journal of Clinical and Experimental Neuropsychology Cumulative concussion exposure in rugby players : Neurocognitive and symptomatic outcomes Cumulative concussion exposure in rugby players : Neurocognitive and symptomatic outcomes*. *November 2014*, 37–41. https://doi.org/10.1080/13803390701443662

Troisi, A. (2002). Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects. *Stress*, *5*(1), 47–54. https://doi.org/10.1080/102538902900012378

Truelove, M. A., Martin, A. L., Perlman, J. E., & Bloomsmith, M. A. (2017). Two methods of social separation for paired adolescent male rhesus macaques (Macaca mulatta). *Journal of the American Association for Laboratory Animal Science*, *56*(6), 729–734.

Truelove, M. A., Martin, J. E., Langford, F. M., & Leach, M. C. (2020). The identification of effective welfare indicators for laboratory-housed macaques using a Delphi consultation process. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-77437-9

UNO, H. (1986). The stumptailed macaque as a model for baldness: effects of minoxidil. *International Journal of Cosmetic Science*, *8*(2), 63–71. https://doi.org/10.1111/j.1467-2494.1986.tb00432.x

Wadden, T. A., & Bray, G. A. (2018). *Handbook of Obesity Treatment* (Second). The Guildford Press.

Wallen, K. (1996). Nature needs nurture: The interaction of hormonal and social influences on the development of behavioral sex differences in rhesus monkeys. *Hormones and Behavior*, *30*(4), 364–378. https://doi.org/10.1006/hbeh.1996.0042

Weatherall, S. D. (2006). *The use of non-human primates in research*.

Webster, J. (1995). *Animal Welfare: A Cool Eye Towards Eden*.

Wegener, D., Oh, D. Q. P., Lukaß, H., Böer, M., & Kreiter, A. K. (2021). Blood analysis of laboratory macaca mulatta used for neuroscience research: Investigation of long-term and cumulative effects of implants, fluid control, and laboratory procedures. *ENeuro*, *8*(5). https://doi.org/10.1523/ENEURO.0284-21.2021

Westlund, K. (2012). Can conditioned reinforcers and Variable-Ratio Schedules make food- and fluid control redundant? A comment on the NC3Rs Working Group's report. *Journal of Neuroscience Methods*, *204*(1), 202–205. https://doi.org/10.1016/j.jneumeth.2011.03.021

Willems, R. A. (2009). Regulatory issues regarding the use of food and water restriction in laboratory animals. *Lab Animal*, *38*(10), 325–328. https://doi.org/10.1038/laban1009-325

Wilson, M. E., Fisher, J., Fischer, A., Lee, V., Harris, R. B., & Bartness, T. J. (2008). Quantifying food intake in socially housed monkeys: Social status effects on caloric consumption. *Physiology and Behavior*, *94*(4), 586–594. https://doi.org/10.1016/j.physbeh.2008.03.019

Wolfensohn, S., & Honess, P. (2005). Handbook of Primate Husbandry and Welfare. *Handbook of Primate Husbandry and Welfare*, 1–168. https://doi.org/10.1002/9780470752951

World Health Organization. (2021). *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Xu, F., Wu, Q., Xie, L., Gong, W., Zhang, J., Zheng, P., Zhou, Q., Ji, Y., Wang, T., Li, X., Fang, L., Li, Q., Yang, D., Li, J., Melgiri, N. D., Shively, C., & Xie, P. (2015). Macaques

Exhibit a Naturally-Occurring Depression Similar to Humans. *Scientific Reports*, *5*, 1–10. https://doi.org/10.1038/srep09220

Yamada, H., Louie, K., & Glimcher, P. W. (2010). Controlled Water Intake: A Method for Objectively Evaluating Thirst and Hydration State in Monkeys by the Measurement of Blood Osmolality. *J Neurosci Methods*, *191*(1), 83–89. https://doi.org/10.1016/j.jneumeth.2010.06.011.Controlled

Young, E., & Korszun, A. (2010). Sex, trauma, stress hormones and depression. *Molecular Psychiatry*, *15*, 23. https://doi.org/10.1038/mp.2009.94

Zehr, J. L., van Meter, P. E., & Wallen, K. (2005). Factors regulating the timing of puberty onset in female rhesus monkeys (Macaca mulatta): Role of prenatal androgens, social rank, and adolescent body weight. *Biology of Reproduction*, *72*(5), 1087–1094. https://doi.org/10.1095/biolreprod.104.027755

Zuckerman, S. (1932). *The social life of monkeys and apes*. https://doi.org/10.4324/9781315009469

Appendix A

Permission to use images from rhesus macaques and NHP facility

Permission granted by Paul Dearden on 31st May 2022 to use Figure 3 and Figure 6 for the purpose of the present work (**Figure 28**).

Re: NHP unit and macaque photos

Paul Dearden <paul.dearden@newcastle.ac.uk>
Tue 31/05/2022 11:43

To: Janire Castellano Bueno (PGR) <J.Castellano-Bueno2@newcastle.ac.uk>;Michelle Clayton Wood <Michelle.ClaytonWood@newcastle.ac.uk>;Kathy Murphy <Kathy.Murphy@newcastle.ac.uk>

Dear Jainire,

Please accept this email as confirmation that you have approval to use the photographs in your thesis.

Regards

Paul

Paul Dearden

I am currently working between home and campus, you can call me on 07811101829

Business Manager of Operations
0191 2086188 / paul.dearden@ncl.ac.uk

Newcastle University
Comparative Biology Centre, Medical School,
Newcastle upon Tyne, NE2 4HH, UK
www.ncl.ac.uk
Visit the CBC Intranet for guidance, resources and support and keep up to date with current news on the CBC Researcher News page.

****If you require advice regarding urgent Home Office matters and you do not receive a response within 3 working days please call me****

Figure 28. Email by Paul Dearden on 31st May 2022 granting permission to use Figure 3 and Figure 6 for the purpose of the present work.