If I Could Talk to the Animals:

Measuring Subjective Animal Welfare

Heather Browning

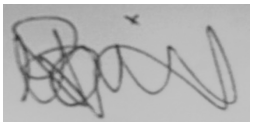
A thesis submitted for the degree of Doctor of Philosophy

The Australian National University

December 2019

Statement

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, all help received in preparing this thesis and all sources used have been duly acknowledged.



Heather Browning

December 2019

To Will, for his unending love and support throughout.

Acknowledgements

Writing a thesis is a huge undertaking, and one that would not have been possible without the help of so many people around me along the way. To begin with, greatest thanks to my primary supervisor Kim Sterelny, for consistently providing constructive feedback on my many drafts and for his unerring ability to find the philosophical core of any problem. Thanks to the other members of my supervisory panel, Seth Lazar and Rachael Brown, for feedback on drafts, career advice and support. Thanks also to previous panel members Ben Fraser, Ray Briggs and Clive Phillips for earlier discussions and literature recommendations.

I have adored being part of the ANU graduate student community, who are a constant source of inspiration and support. Being surrounded by likeminded people all in the same situation as myself helped indescribably in navigating this whole process. In particular our Kioloa workshops have been some of my best memories of my times here, for both the philosophical and social aspects.

I am grateful to the staff and visitors at the Global Priorities Institute at Oxford University, where I spent the summer of 2019, who opened up my mind and my work to myriad new ideas and showed me how productive collaborative academia can be. I am also grateful to David Mellor, Mark Budolfson and Ngaio Beausoleil for taking the time to talk through some of my work and for providing fresh perspectives.

Versions of these chapters have been presented at numerous conferences and workshops and benefitted greatly from the surrounding discussion, including the International Society for the History, Philosophy and Social Studies of Biology, Australasian Association of Philosophy, Sydney-ANU Philosophy of Biology workshops, Philosophy of Biology at Dolphin Beach and ANU's PhilSoc seminars.

On a more personal note, I would like to thank my family – that large collection of parents, sisters, partners and nephews - for their continuing support and for always believing that I could be Dr. Hez one day. To my brilliant partner and unofficial agent Walter Veit, for love and support and reminding me to stay passionate about philosophy. And to my husband Will, for always being there for me in all the ways I needed and who made all of this possible for me.

My doctoral research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Abstract

Animal welfare is a concept that plays a role within both our moral deliberations and the relevant areas of science. The study of animal welfare has impacts on decisions made by legislators, producers and consumers with regards to housing and treatment of animals. Our ethical deliberations in these domains need to consider our impact on animals, and the study of animal welfare provides the information that allows us to make informed decisions. This thesis focusses on taking a philosophical perspective to answer the question of how we can measure the welfare of animals.

Animal welfare science is an applied area of biology, aimed at measuring animal welfare. Although philosophy of animal ethics is common, philosophy focussing on animal welfare science is rare. Despite this lack, there are definitely many ways in which philosophical methods can be used to analyse the methodologies and concepts used in this science. One of the aims of the work in this thesis is to remedy this lack of attention in animal welfare. Animal welfare science is a strong emerging discipline, but there is the need for conceptual and methodological clarity and sophistication in this science if it is to play the relevant informative role for our practical and ethical decision-making. There is thus is a strong role here for philosophical analysis for this purpose.

The central aim of this thesis is to provide an account of how we can measure subjective animal welfare, addressing some of the potential problems that may arise in this particular scientific endeavour. The two questions I will be answering are: what is animal welfare, and how do we measure it? Part One of the thesis looks at the subjective concept of animal welfare and its applications. In it, I argue for a subjective welfare view - that animal welfare should be understood as the subjective experience of individuals over their lifetimes - and look at how the subjective welfare concept informs our ethical decision-making in two different cases in applied animal ethics. Part Two of the thesis looks more closely at the scientific role of welfare. Understanding welfare subjectively creates unique measurement problems, due to the necessarily private nature of mental states and here I address a few of these problems, including whether subjective experience is measurable, how we might validate indicators of hidden target variables such as welfare, how we can make welfare comparisons between individual animals and how we might compare or integrate the different types of experience that make up welfare. I end with a discussion of the implications of all these problems and solutions for the practice of welfare science, and indicate useful future directions for research.

Contents

1. Chapter One – Introduction

Animal welfare is a normative as well as a scientific concept – a 'bridging concept' between science and ethics (Fraser, Weary, Pajor, & Milligan, 1997). Welfare plays a role within both our moral deliberations and the relevant areas of science. Animal welfare scientists study welfare, both to learn more about animals – their evolution, minds, behaviour and physiology – and for its role in normative decision-making. The study of animal welfare has impacts on decisions made by legislators, producers and consumers with regards to housing and treatment of animals. Animal welfare considerations are important because there are many human activities that use or affect animals, often involving a large degree of harm. Over 70 billion animals are farmed annually for human food production[1] with around another 1-3 trillion fish caught per year[2]. Animals are used for biomedical and other types of research, estimated at over 25 million per year in the US alone[3]. They are also used for entertainment, in sports, circuses and zoos. Animals are kept as pets in our homes. Our actions also affect wild animals numbering in the trillions, both directly - through practices such as killing of pest animals - and indirectly, through our environmental impact. If we take animal welfare as a target of ethical concern, this then creates strong moral reason to evaluate the amount of harm caused to animals by humans, and act to reduce it. Our ethical deliberations in these domains need to consider our impact on animals, and the study of animal welfare provides the information that allows us to make informed decisions.

1.1. Three questions

There are three broad questions that make up the study of animal welfare. The first is: what sort of moral consideration should we give non-human animals? This is primarily a philosophical question, and one that has been well-explored within moral philosophy, by writers such as Peter Singer (1995) and Tom Regan (1983). The second question is: to which animals should we extend moral consideration? This is a question with both philosophical and scientific aspects – in first deciding which properties of animals will be morally salient, and then in determining which animals possess these properties. The final question, and that which will be the primary focus of this thesis project is: how do we measure the welfare of animals? This has largely been addressed as a scientific question, trying to find particular measures that

[1] https://www.worldanimalprotection.org.au/our-work/animals-farming-supporting-70-billion-animals
[2] http://fishcount.org.uk/fish-count-estimates-2
[3] https://www.humanesociety.org/resources/animals-used-biomedical-research-faq

represent the welfare of animals, but there is much room for philosophical work as well. For example: in clarifying what is meant by welfare, in examining the relationship between measures of welfare and the state itself, and in discussion of which aspects of welfare might be most important. Addressing these questions and issues helps strengthen the science of animal welfare, and has implications for its applications in animal ethics.

1.1.1. What is the moral status of animals?

The first question, and one that has been well-discussed in the animal ethics literature, regards the moral status of animals. That is, what is their moral status, and what sort of consideration should we give them in our ethical decision-making? Historically, within Western society, the treatment of animals was prevalently based on a religious ethic of human dominion. Animals were considered to have been placed on Earth by God explicitly for human use and so did not require moral consideration. Philosophical views tracked similar beliefs; e.g. with Aristotle arguing that as animals lacked reason, so too they lacked moral status (Aristotle, in Regan & Singer, 1976). This perspective shifted slightly in the mid 19th century with the emergence of Darwin's theory of evolution by natural selection, which undermined the foundational concept of a fundamental difference between humans and other animals. Around this time, philosopher Jeremy Bentham also made a case for moral consideration of animals with the famous line: "The question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?" (Bentham, 1879, p. 309). This led in the late 19th century to strong public opposition to the practice of vivisection (experiments performed on live animals) and the founding of the National Anti-Vivisection Society in Britain[4].

The contemporary case for moral treatment of animals gained widespread attention following the publication of Peter Singer's *Animal Liberation* (1975) and Tom Regan's *The Case for Animal Rights* (1983). Although they took different approaches to moral theory (Singer as a utilitarian and Regan a deontologist), both convincingly argued that if we assume that humans are worthy of moral consideration, and as many nonhuman animals share the features we consider morally relevant, then there is no reason not to extend this consideration to these animals. There have also been cases made for the moral consideration of animals in other areas of ethics, such as virtue ethics (Sandøe, Crisp, & Holtug, 1997; Hursthouse, 2011) and feminist 'ethics of care' (Donovan, 2003; Gruen, 2011). The literature on animal ethics is vast and detailed (see e.g. overviews in Beauchamp & Frey, 2011; Gruen, 2011), and one that

[4] http://www.navs.org.uk/about_us/24/0/299/

I cannot hope to do real justice to here in a thesis primarily situated within philosophy of science, rather than ethics. There is still considerable debate as to the degree of moral status we should accord to animals, and how we should frame our interactions with them. Here I do not presuppose or argue for or any particular view within animal ethics. Throughout a range of views, as well as within public opinion, it is generally accepted that animals possess at least some moral standing and, all else being equal, it is important to consider their welfare. This is particularly true regarding captive animals, where humans are entirely responsible for the conditions for and quality of lives of the animals in their care. If we allow animals even some moral consideration, this is sufficient to motivate what follows regarding the importance of measuring and ensuring welfare.

1.1.2. Which animals are moral subjects?

It is not hugely controversial that animals possess some moral status and that their interests and welfare should be given consideration. This then leads to the second question, which is: to which animals we should extend moral consideration? This will depend on which features of animals we take to ground their moral status. As I will argue for in Chapter Two, here I will take the relevant feature to be sentience – the ability to experience positive and negative mental states (i.e. affect) and thus possess a welfare that can be benefitted or harmed. This is the view of 'sentientism' about moral status: "you have moral status, i.e. you are a subject of moral concern, if and only if you are sentient, i.e. if and only if you are capable of phenomenally consciously experiencing pleasure or pain" (Sebo, 2018, p. 4). Sentience research then gives us the targets of moral concern - those animals which possess welfare - and forms the basis for the practice of welfare science. This will be true, regardless of what the nature of sentience is found to be.

The presence of sentience is an empirical question – whether or not certain animals possess subjective mental states depends on their particular characteristics and properties, not on our own ideas or opinions. Although there may be a 'grey area' of consciousness between non-sentient and fully sentient organisms, as with the transition between life and non-life (Ginsburg & Jablonka, 2019, p. 456), what determines this are features of the world. However, there is still plenty of debate on how exactly scientists might test for sentience and what given test results actually mean, as will be discussed below. The traditional Cartesian perspective is that animals completely lack awareness; their apparent reactions to pain are merely automatic, unaccompanied by a mental state that we could describe as suffering (Descartes, in Regan & Singer, 1976), but this view has largely been rejected today. It is now commonly accepted that

at least some animals are sentient, as embodied by the Cambridge Declaration on Consciousness in 2012 – "the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Non-human animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates" (Low et al., 2012, p. 2). We accept that at least some non-human animals are sentient; the empirical question is then which animals these are and how far the boundaries of sentience extend.

A large part of the work in animal sentience is in attempting to identify which animals are sentient – in particular, which are capable of experiencing pain, as the avoidance of pain is one of the most basic conditions for welfare. This requires determining a set of criteria that are required to establish the presence of pain perception, as opposed to just nociception, which is the non-conscious detection of tissue damage. Several such sets have been proposed (e.g. Sneddon, Elwood, Adamo, & Leach, 2014), but it is still controversial which criteria should appear in them. There is a role here for the interplay between animal welfare science and animal sentience research in determining suitable criteria, as I will discuss in Chapter Eight.

Research into animal sentience is constantly changing the boundaries of where we think sentience lies. Most sentience research has been focussed on vertebrates, and evidence now supports the conclusion that all vertebrates are sentient, due to similarity in nervous system and brain structure (Proctor, 2012). The sentience of fish has been the subject of more recent debate. On the one hand, opponents argue that fish could not be sentient, as they lack the brain structures and connectivity thought to be responsible for creating conscious experience in mammals (Key, 2016). On the other hand, the behavioural evidence strongly supports the presence of fish sentience. Fish will avoid noxious stimuli, they will change their behaviour and become less responsive to novel stimuli when injured, and will show behavioural changes when under the effects of analgesics, which act in other animals to suppress the conscious experience of pain (Proctor, 2012). There is also a plausible proposed neural mechanism for processing pain experience, albeit a different one than in mammals (Sneddon et al., 2014) and the general consensus is now on the side of fish sentience.

Data on invertebrates are less clear. Jones (2013) goes carefully through the current evidence to conclude that even some insects and crustaceans must feel pain, and the presence of innate analgesics in some snails and earthworms may also be suggestive of their capacity for pain. Broom (2016) considers that there is sufficient evidence for sentience in all vertebrates, cephalopods and decapod crustaceans, and thinks that cognitive ability in stomatopod crustaceans (mantis shrimp), spiders and some insects, including bees and ants, makes them

good candidates for investigation. Ginsburg & Jablonka (2019, p. 351) attribute sentience to vertebrates, cephalopods, arthropods (insects and crustaceans) and possibly some annelids (worms). Where behavioural evidence for sentience is observed in invertebrates, it has often been dismissed as automatic reflex, as the physical setup of their nervous systems is so different (Proctor, 2012). However, if sentience depends on the functional properties of the neural system rather than a specific architecture, then there is the possibility of the same function being multiply realisable, using different morphological structures. Where there is behavioural evidence to suggest otherwise, the absence of the accepted physical substrates should not be sufficient for dismissal. What is needed is assessment across a range of physiological and behavioural criteria, and insects meet many of those proposed thus far (Sneddon et al., 2014). The evidence is certainly suggestive enough to encourage further exploration. As currently only a small fraction of extant species and even taxonomic groups have been studied to determine sentience, the boundaries are still unclear.

Studies in animal sentience inform animal welfare science as to the appropriate targets for study – once an animal species has been identified as sentient, the conditions for its welfare are then an open question. The conceptual and methodological underpinnings of sentience research and welfare science will also inform each other in important ways, as I will detail in Chapter Eight.

1.1.3. How do we measure animal welfare?

Once we've accepted that animal welfare is morally important, and that there are some animals that possess the relevant capacities for welfare, the final question is then how we can measure their welfare and investigate under what conditions the lives of these animals will be improved or worsened. There has historically been scepticism about our ability to gain information about mental states, which we are unable to access directly, though this view is uncommon today. The behaviouristic tradition of the early and mid 20th century relied solely on descriptions of environment and behaviour – stimulus and response – to describe animals (Panksepp, 2005). Mental states were not considered appropriate targets for study, and instead focus was shifted to external, observable behavioural variables. It is true that studying sentience is difficult, as it relates to subjective mental experiences, which are necessarily private and cannot be directly related to particular anatomy or physiology. Even with developments in neuroscience linking brain activity to specific emotions, we still cannot gain complete certainty about the mental states of others (Proctor, 2012). Particularly since we cannot communicate directly with animals through speech, we are entirely reliant on indirect indicators of sentience.

However, this does not mean we cannot be confident in the knowledge we can gain through these means. There has been widespread opposition to this sort of scepticism: "that one cannot oneself experience the feelings of another person or animal does not mean that one has no access to the other's feelings at all" (Wemelsfelder, 1997, p. 9). Many areas of science – including human psychology - similarly rely on indirect access to their targets, but are still able to progress successfully. Animal sentience or welfare research should not require a higher standard: "whilst other areas of science will often make do with imperfect data, animal sentience is required to buck the trend and provide unequivocal proof" (Proctor, Carder, & Cornish, 2013, p. 883). We must of course be cautious about collection and use of our data under these conditions. Jones (2013) argues that "since our epistemic access to the mental lives of animals is even more limited than access to each other's minds, we must be cautious about cognitive attributions, and selective about the kinds of evidence for such attributions we have at our disposal" (2013, p. 3). We should thus be looking for "robust and valid evidence of animal sentience" (Proctor, 2012, p. 630). Although some prominent animal welfare researchers (e.g. Dawkins, 2008) still contend that subjective experience is 'untestable', as we cannot determine whether observed indicators are necessarily linked to corresponding subjective states, we should not require such certainty. Though we may never have direct access to the contents of mental states, we can refine our background theory and methods to obtain increasingly accurate data that is best explained by the presence of subjective experience (a claim I will develop further in Chapter Five). Through careful observation and experimentation, we may still hope to gain solid understanding of animal mental states and welfare.

I will argue further for this position – that subjective mental states are measurable – in Chapters Two and Four, as well as discussing in Chapter Five how we might find valid indicators of animal mental states. These problems mean that we need extra caution in collecting and interpreting data, but not that we should abandon research of this type entirely. Animal welfare is measurable, and its measurement is then the subject of animal welfare science.

1.2. Animal welfare science

Animal welfare science is an applied area of biology, looking at how we can measure the welfare of animals. It draws on work in a range of biological sciences, including behavioural ecology, evolutionary biology, neuroscience, animal behaviour, genetics and cognitive science (Dawkins, 2006). Where we are concerned about animals and their welfare – such as when working with animals, using their products, or just as a result of caring about them - it is obviously critical to understand how to measure welfare; how it is that we can know when they are doing better or worse. As I will argue in Chapter Two, animal welfare should be understood as consisting in the subjective experience of animals over their lifetime, and thus animal welfare science is at its core, a scientific inquiry into the contents of animal experience. The private nature of subjective experience raises particular issues for measurement, and these will be addressed throughout this thesis. Even for those who may not accept this view of welfare, the problems discussed are still relevant; they will arise for any view which accepts that subjective experience comprises at least part of the welfare of animals.

Animal welfare science aims at measuring the welfare of animals. Scientists use different indicators, such as changes in behaviour or of physiological variables such as heartrate, to measure the changes in welfare under different conditions. So, we might want to test whether chickens are less stressed in the company of conspecifics, and could look for the presence of stress hormones to find out. Or we might want to see whether a mouse prefers woodchip flooring to sawdust flooring and use a choice test to see which one it chooses. Or we might want to determine whether a high level of visitor contact is detrimental to a bear in a zoo, through measuring levels of stereotypic behaviour. All these are examples of the practice of animal welfare science.

Animal welfare science is important because without objective measurement, all we can do is rely on our best guesses as to what animals might want, or what is good for them, and these guesses can often be wrong. For example, many zoo visitors prefer monkeys to be displayed in naturalistic open-air enclosures, thinking them better for welfare (Kagan & Veasey, 2010). But the monkeys feel differently – cage-style exhibits often provide more climbing surfaces and opportunities for activity, promoting good welfare (Browning & Maple, 2019). Similarly, chickens being caught up for transport can be harvested from their cages by hand, or by machines. Most people assume human handling would be preferable for welfare, but changes in heart rate and fear behaviour indicate the machine handling is better for the chickens (Duncan, Slee, Kettlewell, Berry, & Carlisle, 1986). Although one might think that animals would minimise the amount of work they need to do to receive food, preference tests show that

in many cases, animals will choose to work for food over obtaining it for free (a phenomenon known as 'contrafreeloading') (Osborne, 1977). These sorts of counterintuitive results are not uncommon, and it is for this reason we need an objective science of animal welfare to gather information about what is good or bad for welfare, from the point of view of the animals themselves. We need to find and use valid and reliable external measures of the internal experiences of animals, ones that are not influenced by the subjective values and opinions of experimenters.

Animal welfare science is a relatively new discipline, having come into prominence in the 1970s and 1980s and continuing to grow and develop, tracking a general increasing public interest in animal welfare concerns (Lawrence, 2008). Its origin followed the publication of Ruth Harrison's *Animal Machines* in 1964 (Harrison, 1964), detailing the conditions of animal housing within intensive agriculture, and the subsequent Brambell report on animal welfare (Brambell, 1965) commissioned by the British government. This report highlighted the production practices of most concern, listed what they considered to be the most important welfare provisions, and outlined the need for scientific research into animal welfare. This led to an increase in applied animal behavioural studies, looking to develop methods for seeing the world from the 'animal perspective' (Lawrence, 2008). The publication of *Animal Suffering* by Marian Stamp Dawkins in 1980 (Dawkins, 1980), critically assessing the different methods of measuring welfare, was an important landmark in the development of this new science.

Historically, there has been a primary focus on animal pain and suffering (see e.g. Dawkins, 1980). Welfare science has looked to measure whether (or how much) animals are suffering, and under what conditions this occurs. Almost all studies in welfare science focussed on measures of suffering, such as behavioural and physiological signs of fear and stress. This tied to an idea that absence of negative states was sufficient for good animal welfare (Mellor & Stafford, 2008). It is true that the primary goal of animal welfare interventions is usually to prevent suffering, due to its greater urgency and impact. However, there is increasing recognition that while the prevention of suffering is necessary for welfare, it is not sufficient. As well as removing these negative experiences, there is a need to focus on providing positive experiences (Yeates & Main, 2008; Sandøe & Jensen, 2011). Once suffering is removed, and animals are existing at a neutral baseline, it is not the case that there is nothing more that could be done. Welfare is more than just the absence of suffering – it is also the presence of positive states, such as pleasure or satisfaction. For those working with animals in situations such as zoos, in which (if all is going well) there should be limited suffering, there is still reason to be interested in how to improve welfare. In considering animal welfare and the question of what

animals want, we should move beyond the mere prevention of suffering to include the promotion of positive states. There is then space here to find measures of positive welfare – signs of improved welfare above this neutral baseline. Modern animal welfare science works on measuring both positive and negative welfare states.

1.3. The role of philosophy in animal welfare science

Animal welfare science is not a science that has been the focus of much philosophical attention. Although philosophy of animal ethics is common, philosophy focussing on animal welfare science is rare. By contrast, the measurement of human wellbeing is the subject of huge amounts of work within psychology, economics and philosophy. Despite this lack, there are definitely many ways in which philosophical methods can be used to analyse the methodologies and concepts used in this science. One of the aims of the work in this thesis is to remedy this lack of attention in animal welfare.

To make the scope of this work clearer, with regards to the role of philosophical analysis, we can draw a comparison to another scientific discipline which is similar in many ways and has been the recent focus of philosophical study. This is the science of conservation biology. Conservation biology is centrally the science of how to measure and preserve biodiversity. There are many similarities between the two disciplines:

- Both aim at finding measures of a central target concept: conservation biology focuses on biodiversity, and animal welfare science on welfare, but both contain similar issues in defining and using a central concept
- Both require the use of surrogate measures to get at the target: conservation biology uses many surrogate measures - such as species richness or levels of endemism – to represent biodiversity in a region, while animal welfare science uses indicators such as physiology or behaviour to represent the underlying state of welfare in an animal
- Both are value-laden: both sciences are aiming to measure and promote something that is considered to be of value and the concepts must therefore do normative as well as scientific work
- Both are multi-disciplinary: both sciences draw on the work of many disciplines within the sciences and humanities
- Both have a practical and a theoretical component: both sciences work in a theoretical manner - determining which are the best measures for the particular

target; and then in a practical manner – using these measures to make judgements about particular situations

- Both must often work with partial and incomplete data: because of the importance of action in conservation or welfare protection, these disciplines cannot wait for complete data to become available before making decisions
- Both require consideration of trade-offs: in considering promoting the central value, both disciplines must consider trade-offs in use of resources to maximise outcomes, as well as trade-offs with other competing values (e.g. human interests)
- Both raise issues of commensurability: both sciences must try to compare measured values between quite different entities – welfare in distantly related species, or biodiversity in dissimilar ecosystems

The goals of philosophical work in animal welfare science are thus similar to those in conservation biology – to elucidate the central concept, to analyse how surrogate measures might map onto this concept, and to examine what areas of research might best lead to the desired outcomes for the science. Similarly too for health science, another discipline which has a descriptive and normative component, requires the use of proxy measures, and involves calculating trade-offs between competing values.

As discussed in Section 1.2, animal welfare science is a strong emerging discipline, but there is the need for conceptual and methodological clarity and sophistication in this science if it is to play the relevant informative role for our practical and ethical decision-making. There is thus is a strong role here for philosophical analysis for this purpose.

1.4. Overview of the thesis

The central aim of this thesis is to provide an account of how we can measure subjective animal welfare, addressing some of the potential problems that may arise in this particular scientific endeavour. The two questions I will be answering are: what is animal welfare, and how do we measure it?

Part One of the thesis looks at the subjective concept of animal welfare and its applications. There are multiple ways of understanding animal welfare, and what it consists in. The most common views are subjective experience, physical functioning, teleology (naturalness) and preferences. In Chapter Two, I argue for the subjective welfare view; that animal welfare should be understood as the subjective experience of individuals over their lifetimes. A welfare concept needs to play both scientific and normative roles, and I will describe how the subjective

welfare concept meets the requirements of being normatively significant, fundamental and measurable. I address some potential objections to understanding welfare subjectively, particularly those coming from the literature on human wellbeing. I then go on to examine the other candidate welfare concepts and argue that none of them succeed; in fact they are only relevant to welfare insofar as they impact on subjective experience.

The way we understand welfare will have effects on our science and our ethics. In Chapter Three, I will look at how the subjective welfare concept informs our ethical decision-making in two different cases in applied animal ethics. The first is the practice of management euthanasia used in zoos: where otherwise healthy individuals are killed when surplus to the requirements of breeding programs, in order to create space for other animals. The second is in de-extinction programs: in which cloning, genetic engineering or back-breeding techniques are used to re-create extinct species. In both of these cases, I discuss the potential welfare harms - as understood subjectively - and how we might look at trading these off with other competing values to make decisions on the acceptability of these practices.

Part Two of the thesis looks more closely at the scientific role of welfare. Understanding welfare subjectively creates unique measurement problems, due to the necessarily private nature of mental states. As discussed above, this shouldn't make us think that measurement of welfare is impossible, but should give us reason to be cautious about our methods and of how we interpret our results. I begin in Chapter Four by addressing the question of whether subjective experience is measurable – not in the general sense of detecting subjective states as discussed above, but whether it is an appropriately quantifiable target for scientific inquiry. I will examine welfare through the lens of measurement theory and argue that welfare is meaningfully quantifiable and we can do so using a ratio scale.

One problem arising from measurement of subjective welfare is in how we validate our measurement indicators. We cannot directly access the mental states of others, and so must use indirect proxy indicators for our measurement. For such indictors to be valid, it must be the case that they are measuring the intended target state, and not some other object. Indicators can usually be validated through looking for correlation between target and indicator under controlled manipulations, but where we cannot access the target state directly, this is not possible. In Chapter Five, I address this issue and outline a procedure, using robustness analysis, that allows us to validate the indicators of hidden target variables such as welfare.

Another problem in measurement of welfare is how we can make welfare comparisons between individuals. Often, we will need to make such comparisons – such as when deciding on fair allocation of resources – but as we don't necessarily have reason to think that the welfare

capacities of different individuals are the same, we can't make direct comparisons. In Chapter Six, I describe the similarity assumptions we need to make in order to make such comparisons, and the conditions under which they are likely to be justified. I also discuss what we might do in cases where the assumptions fail to hold.

Taking animal welfare to consist in the subjective mental states of animals also poses an additional problem, as there are a large variety of such states. It is not obvious that these are of the same kind, such that we could compare them in making decisions about trade-offs between different positive and negative experiences, or combine them all into a single metric of welfare. In Chapter Seven I argue that we have good reasons to think that there is a 'common currency' underlying different mental states, that will allow us to compare or combine these states as needed for assessing welfare or making management decisions. I will also describe the procedures by which we might go about determining the weightings of different mental states in their contribution to overall welfare.

Finally, in the Conclusion I will discuss the implications of all these problems and solutions for the practice of welfare science and indicate useful future directions for research. In particular, continuing animal sentience research - examining the evolution and functioning of sentient experience - will greatly inform and improve animal welfare science.

PART ONE – the Subjective Welfare Concept and Its Applications

2. Chapter Two – Animal Welfare is Subjective Welfare

2.1. Introduction

As discussed in Chapter One, the goal of animal welfare science is to measure the state of welfare in an animal, and how it changes under different conditions, using a variety of behavioural and physiological indicators. This science therefore requires use of a meaningful welfare concept, describing what it is we're trying to measure – what comprises the *state* of welfare itself. In this chapter I argue that subjective welfare – the experience of positive and negative mental states by the animal - should be understood as the primary state of animal welfare, and that this concept fulfils the requirements for a concept of welfare. I will also describe the common competing concepts - a tripartite welfare concept, under which welfare consists in feeling good (subjective welfare), functioning well (physical welfare) and living naturally (teleological welfare), and a preference-based concept under which welfare consists in meeting animal preferences – and show how the other proposed components (physical, teleological and preferences) can be collapsed onto subjective welfare. These components may form an important part of the conditions required for the realisation of good welfare, but do not themselves comprise the state of welfare.

To clarify some terminology: I will use the term 'state' of welfare to refer to the base state that we are interested in. The state of welfare is that underlying property that constitutes welfare; that thing which is affected by changes in conditions and in virtue of which we say that welfare is increasing or decreasing. This is what will be captured by the subjective welfare concept I will describe. This state is typically understood to be continuous, on a continuum from good to poor welfare. The properties of welfare and its measurement will be described in more detail in Chapter Four. Next there are the *conditions* for welfare – those things that will serve to increase or decrease welfare, such as, say, food quality or social companionship. These are factors which are instrumentally important for welfare, as they impact on the state itself, but do not themselves determine welfare (Kagan et al., 2015). Welfare science is concerned with finding out what these are for different species. Finally, there are the *indicators* of welfare, those behavioural and physiological measures that are used to determine whether an animal is experiencing good or poor welfare (see e.g. Kagan & Veasey, 2010). For example, we might imagine investigating a pig in a sow stall – we could look at measures of stereotypic behaviour and blood cortisol levels, in comparison with a pig on straw bedding. The behavioural measures

and hormone sampling would be the indicators of welfare. The housing (sow stall or straw bed) would be the conditions for welfare. What it is that's being improved or worsened under these conditions would be what comprises the state of welfare. These distinctions are laid out in Table 2.1.

	Definition	**Examples**
Welfare state	Property that constitutes welfare	Subjective experience, physical health, preferences
Welfare conditions	Things that make life better or worse	Housing type, diet, social interactions
Welfare indicators	Measures used to determine changes in welfare	Heart rate, choice tests, blood cortisol, vocalisation

Table 2.1: Distinctions in welfare terminology

In order to practice this science, it is plainly important to have a meaningful concept of welfare. We need to be clear on exactly what it is we mean when we speak of welfare – what it is that comprises this state in an animal. Without knowing exactly what it is that we're trying to measure, we can't hope to assess which conditions matter, or which indicators will be the most accurate.

This also has applications for those who interact with animals, such as those who keep captive animals. As a large part of their job should be ensuring the greatest welfare of their charges, it is of central importance that there is a clear understanding of what welfare is, to prevent spending time and resources on providing conditions that may appear to increase welfare without actually doing so. Think, for example, of the zoo manager who may decide to house all primates on 'island' style exhibits, rather than aviary-style cages, in the thought that it is more naturalistic and creates a sense of freedom. Instead, the arboreal animals are denied the cage walls and roof they could have otherwise enjoyed using as locomotory surfaces (see Browning & Maple, 2019). It is thus important that a conception of welfare be tied to the science and practice of animal husbandry and welfare.

2.2. The case for subjective welfare

The concept of welfare plays two roles – a normative role and a scientific role. It is therefore important that the concept we use is sufficient for both of these roles. The overall goal of the science and practice of animal welfare is to determine how to maximise welfare, an end which is considered morally important. Thus, in order to fill the normative role, welfare must be something that is valuable, something that matters morally. Fraser et al. (1997) describe welfare as a 'bridging concept', linking science and ethics. This requires a concept that is "amenable to scientific study" (Fraser et al., 1997, p. 188) but also reflects the underlying ethical concerns people attach to welfare.

The scientific role of welfare is primarily as a target for investigation – a state we want to measure and to understand – and therefore in order to fill the scientific role, it must be something that is measurable. If we take welfare to be the appropriate target for investigation, a central concept in itself rather than a property or proxy of some other state, it must also be something which is fundamental. As well as playing a normative role in our moral deliberations, welfare can play an explanatory and predictive role in biological sciences. As I will discuss in Chapter Seven, welfare (understood as the combination of negative and positive experiences) may also play a role in determining how animals make trade-offs and decide on particular actions and so understanding this type of integrated subjective experience will help us understand animal behaviour.

Webster (2005, p. 2) also differentiates between the scientific and the moral roles that the concept of welfare needs to play. In this work, he suggests that the concept will vary depending on the different application – with scientists taking something like Broom's (1986) concept of coping within an environment, considering physical and mental states, physiological and behavioural needs, while those concerned with ethics focus more upon the subjective experience, and the fact that animal feelings matter *to the animals*. It is clear that there are different roles for the concept, and these may have different requirements. However, in the end, we want these to refer to the same entity, and so require a single concept to work in both areas. If we were to adopt a different concept for the moral role than for the scientific role, we would then lose the ability to use the findings from welfare science in our moral deliberations, without constructing an additional framework to map one concept to the other. As informing moral decision-making is one of the primary functions of animal welfare science, using different concepts would be detrimental. We need a concept that captures the basic state that is of scientific interest, and morally relevant.

The conception of welfare I will be arguing for is a subjective concept. Subjective welfare theory takes welfare to consist in the subjective mental states of the animal, often also called 'affects'[5] (Mellor, 2012). These states include (among others) hunger, breathlessness, pain, fear, curiosity and joy. Positive mental states add to welfare, negative mental states subtract from it. This is similar to the hedonist position in human welfare (Crisp, 2017), though not necessarily committed to pleasure as the only positive mental state. A note on terminology here – throughout the thesis, I will be using terms such as 'pleasure', 'enjoyment', 'happiness' and 'satisfaction'. These are meant only as general terms, as stand-ins for the suite of positive mental states, whatever they may be, and not as any type of commitment to what the nature of these states may actually be. As per Duncan (2009), "the strong negative feelings are often lumped together as 'suffering' and the positive feelings as 'pleasure'" (2009, p. 2); though I continue the practice here, it does not imply a strong commitment to the grouping of these states. In Chapter Seven, I will take a closer look at the types of positive and negative experiences that comprise welfare, and how they may relate to one another and interact to create the overall state of welfare, taking up the issue of commensurability.

Under a subjective conception, welfare consists in the experience of an animal over its lifetime[6]. This then requires an animal be sentient – i.e. capable of some form of subjective experience – in order to have welfare. Some writers have argued that this is a form of 'sentientism', where non-sentient forms of life (or even non-life) are excluded. However, this is not meant to imply a strong claim about what matters morally. Although, as I will argue, welfare is of moral importance, this does not necessarily mean it is the only thing that is. Non-sentient organisms may have a 'good' in a weaker sense that could still be considered in moral decision-making. But sentience is morally important - though perhaps not the only type of moral claim, it is a unique one (Donaldson & Kymlicka, 2011).

The subjective conception of welfare is common throughout the animal welfare literature. Singer (1995) considers sentience – the ability to experience pleasure and pain – as grounding animal interests, which he takes as central to welfare. Tom Regan (1983) describes animal welfare as "the experiential quality of their life, as considered over time" (1983, p. 96), where this 'experiential quality' refers to life from the point of view of the animal; their subjective

[5] Some preference-based accounts of human welfare are referred to as 'subjective' as they are grounded in (subjective) preferences. Here I use subjective welfare to refer to the narrower 'hedonistic' view and will consider preferences separately in Section 2.4.3.

[6] There are additional considerations as to whether it is simply the total number of experiences that matter for welfare, or also their distribution throughout the lifetime. These 'shape of life' concerns are important, but will not be considered here; it is sufficient for this work to take some function of lifetime subjective experience as determining welfare.

experience[7]. Grandin (2009) focuses her discussion on animal welfare on providing for the emotional needs of animals: "I believe the best way to create good living conditions for any animal . . . is to base animal welfare programs on the core emotion systems in the brain" (2009, p. 3). This is strongly based in a subjective view of welfare, in which what really matters are the mental states underlying particular behaviours or husbandry conditions (emotions in this work being taken as conscious experiences). Maple and Perdue (2013) use the concept of 'wellness', defined as "a balance of mind, body, and spirit that results in an overall feeling of well-being" (2013, p. 49) – importantly here it is the overall 'feeling of well-being', or subjective experience, that is central. Recently, work by David Mellor (e.g. Mellor, 2016) has argued for a 'Five Domains' model of welfare, where the four physical domains (nutrition, environment, health and behaviour) matter only in relation to their impact on the fifth domain, of subjective experience. In this model, subjective experience is considered to be the central state comprising welfare, with the other four domains acting as conditions affecting this. Subjective welfare is a commonly used conception of welfare and, I will argue, is the best way of understanding animal welfare, as it is normatively significant, fundamental, and measurable. This debate mirrors one in human wellbeing, where this subjective or 'hedonic' view of welfare is typically rejected. In Section 2.3.2 I will discuss the reasons why, and argue that the reasons we have for accepting or rejecting such a concept in the human case will not necessarily also apply in the animal case. The goal of this chapter is to establish the subjective welfare concept for animals and thus I take no strong stand on whether the same considerations apply in the human case.

2.2.1. Subjective welfare is normatively significant

Firstly, let's look at the normative role for welfare. Welfare is morally important. "Any conception of animal welfare inherently involves values because it pertains to what is better or worse for animals" (Fraser et al., 1997, p. 188). It is a central target for many moral theories (Crisp, 2017). Many of our ethical deliberations revolve around ways to increase or ensure welfare, or wellbeing. Utilitarian ethical theories take welfare as the primary target of moral behaviour, with the aim of performing acts which maximise welfare (Sinnott-Armstrong, 2015). Deontological ethical theories focus primarily on rights (Alexander & Moore, 2016), but particularly within the animal realm, these are often aimed in service of protecting interests

[7] Though Regan would take capacity of welfare of this type as insufficient for moral consideration; an animal must also possess additional mental capacities that make it what he calls the 'subject of a life'.

or welfare (Cochrane, 2012; Regan, 1983). Welfarist theories take welfare to be the only thing that is morally important (Crisp, 2017). As welfare plays a central role in many moral theories, the concept we use to fill this role must therefore contain something we consider valuable. As I will now show, subjective experience succeeds in doing so.

It is highly intuitive that sentience, or subjective experience, is morally important. This is suggested by the fact that much of the work on why we should care about animals - why we should consider them morally important - focuses on their subjective experience. Sentience is a primary target for most utilitarian ethical theories, tracing back to Bentham (1879), who endorsed a 'sentiency criterion' for utilitarianism (Frey, 2011), "to an extent that this criterion has become virtually identified with utilitarian approaches to the moral status of animals" (Beauchamp, 2011a, p. 9). This has been reinforced by more modern work by Peter Singer (1995), who also identifies sentience as the ground for moral consideration. Many other views in animal ethics also take sentience, or some form of subjective experience as necessary (if not sufficient) for moral standing. Tom Regan (1983) in his work on animal rights focusses on being the 'subject of a life' as grounding rights, which requires sentience alongside additional capacities such as agency and a sense of identity over time. Other work on rights grounds rights in possession of interests, which arise from the capacity to experience pleasure or suffering (sentience) (e.g. Beauchamp, 2011b; Cochrane, 2012; Gruen, 2011). Korsgaard's Kantian-based animal ethics focusses on possession of a 'natural good' - grounded in sentience, or having an awareness that can make things good or bad for an individual - as creating grounds for duties towards animals (Korsgaard, 2011). Even virtue ethicists can see recognition of the sentience of nonhuman animals as giving rise to virtues such as compassion and respect in our interactions with them: "any of the vices listed above and the virtues opposed to them may be manifested in relation to our treatment of and attitudes to sentient animals" (Hursthouse, 2011, p. 124).

This focus on the moral importance of subjective experience also comes across in much of the work in animal welfare science. It is emphasised in the early work of Dawkins (1980, 1988, 1990, 1998) – that animal welfare concerns the subjective experiences of animals, in particular their suffering. Fraser (1999) states that "the study of animal welfare is at least partly an attempt to understand the animal's own perceptions of its quality of life" (1999, p. 183). The general claim is that it is the first-person experience of their own good which makes sentience morally significant: "It is the fact that sentient beings care about how their lives go that generates a distinctive moral claim on us" (Donaldson & Kymlicka, 2011, p. 33). In some cases, this seems intended as a stronger claim about self-awareness, but for the most part rests simply on the

capacity to experience. Singer (1995) similarly argues that "a stone does not have interests because it cannot suffer. Nothing that we can do to it could possibly make any difference to its welfare" (1995, p. 8). Almost all the major contributions to the literature emphasise the moral relevance of subjective experience, reinforcing the intuitive pull of this claim: that subjective experience is valuable.

As well as being composed of something of normative worth, we want our welfare concept to identify the bearers of this worth. As welfare is morally important, it follows that if a creature is one such that it can experience greater or lesser welfare, then this is the sort of creature we should care about. If welfare matters morally, then so do the bearers of welfare. Beings capable of experiencing welfare are those which should form a part of our moral deliberations[8]. In defining the concept of welfare and what it consists in, we will also be making some ruling on what types of beings invoke this sort of care. Subjective welfare allows us to make this distinction. As per Jeremy Bentham's famous quote: "The question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?" (Bentham, 1879, p. 309). It is the capacity for pleasure and suffering, as experienced subjectively, that provides cause for moral concern and delineates those we should give consideration to from those we shouldn't.

Our moral consideration of those beings we attribute with subjective experience certainly seems to differ from the consideration of those beings or objects that are thought to lack it. Sentience provides an experience of good-for-the-animal, as opposed to merely 'good of its kind', which we may see in plants, or human artifacts. "They are aware of how they feel *and it matters to them*" (Webster, 1994, p. 249). It is the states of pleasure or suffering that make animal welfare important – why most people would consider it problematic to pull the ear off a cat but not the branch off a tree. If then, the reason we care about animal welfare at all is because it matters *to the animal*, then it seems logical to say that it is this experience that is constitutive of welfare. Subjective experience is necessary for moral consideration and thus subjective welfare is normatively significant. This of course still leaves open the possibility that other things may also be normatively significant, and in Section 2.4 I will identify other possible contenders and argue that they are only instrumentally so, in virtue of their effects on subjective experience, and thus subjective experience is also sufficient for animal welfare.

[8] Though, as mentioned earlier, may not compose the entire moral community; we may have other reasons to value other organisms or objects, but this will usually be a weaker consideration.

2.2.2. Subjective welfare is fundamental

There are a couple of ways in which we can consider subjective experience to be fundamental to welfare, such that facts about welfare are grounded in facts about subjective experience. The first is that it is a necessary component of welfare – it does not seem possible to characterise welfare without it. As previously discussed, almost all accounts of animal welfare include subjective experience. Even the alternative accounts discussed later in the chapter, such as preference satisfaction, typically include subjective experience as a part of the account, alongside other properties. Subjective experiencing is a necessary prerequisite for welfare considerations. None of the other views typically grant welfare status to beings or objects without sentience, even if they take welfare to consist in something more than merely the sentient experience itself. A plant may do better or worse in some way under particular conditions, but it does not seem to most that it has a *welfare* that is being harmed. Although it is certainly true that it may in some sense do better or worse, this is not considered a welfare concern. "Although plants, bacteria, viruses and cells in cultures are alive and may be said to have needs, there is no reason to believe they have interests. That is, there is not a shred of evidence that these things have any awareness or consciousness and consequently we cannot say that the fulfillment and thwarting of these needs 'matters' to them anymore than getting oil matters to a car" (Rollin, 2006, p. 104). A plant or single-celled organism can function more or less well under particular conditions, can even act to promote its own survival, but this does not seem enough to grant it welfare. There needs to be something more, and this something more is subjective experience. It is a fundamental necessary component of welfare. Even for those who might nevertheless wish to deny this strong claim, it is still the case that where an animal has the capacity for subjective experience, the content of this experience is central to its welfare.

The second is that it is intrinsically important to welfare. That is, it is not a concept that can be understood as important due to its instrumental effects on some other component (or set of components) of welfare; it cannot be collapsed onto any other state. There is no further 'something else' that subjective experience influences in order to create welfare change. I will demonstrate this in Section 2.4 and show that the reverse is actually true - subjective experience underlies all the other components that can be thought of as important to welfare. These can all be understood as instrumentally important through their effects on subjective experience and thus can be collapsed onto subjective welfare, establishing that it is the more fundamental concept.

Finally, it seems that subjective experience could also play a fundamental psychological role. Webster (2005) describes psychological processes in animals as the reception, categorisation and interpretation of stimuli. The animal experiences something, perceives it, then categorises the experience according to its conceptual framework and ends with an interpretation of the experience that is either a positive or a negative emotion. This final stage is subjective experience, which then forms the basis for learning, motivation and behaviour. Ginsburg and Jablonka (2019) similarly propose a model in which animals encounter different stimuli, and the effects of their interactions with these will cause feedback in the form of subjective experience and a 'felt value' which will then help guide future motivation and action. In Chapter Seven, I will look at how this categorisation and compiling of experience can then function to help animals consider trade-offs between competing demands and thus make decisions for action, so that other psychological processes flow from this more fundamental state. Similarly, work on cognitive bias demonstrates how mood - an overall emotional state integrating a variety of affects and roughly equivalent to welfare experience at a time - can influence behaviour and decision-making (e.g. Baciadonna & McElligott, 2015; Clegg, 2018; Mendl, Burman, Parker, & Paul, 2009). Subjective experience is thus likely to be a fundamental psychological process, and one that can play a causal and explanatory role in other processes and behaviours.

2.2.3. Subjective welfare is measurable

Welfare is the central target of welfare science, which aims to measure changes in welfare under different conditions. It is thus crucial for this scientific role that welfare is something measurable. Although not all concepts used in science may refer to measurable states, measurement is certainly the aim of animal welfare science and thus measurability matters here. Perhaps the biggest potential problem with a subjective conception of welfare, is thus whether it can ever be of any practical use as a scientific concept in this sense. As subjective states are necessarily private, how is it that we might ever measure them, to know anything about the welfare of other animals?

In essence, this concern is a version of the problem of other minds. That is, how it is we might know whether creatures other than ourselves (including other human beings) have minds at all, as opposed to being complex non-conscious 'machines', and how we might gain access to the contents of their experience. This is a problem in philosophy even for knowing the minds of other humans, where we can use language to attempt to communicate. For animals that cannot communicate in this way, it is even more difficult to ascertain their private mental states.

For many decades, the tradition of behaviourism in psychology, starting with Watson (1913) held that we could not investigate conscious states; that in fact it was unscientific to ask such questions. We could only ever have information on what animals *do*, never what they think or feel. However, within animal studies, this tradition has become less popular and most scientists accept "that some bodily states and behaviour can be used as reasonably reliable guides to what a human or other animal is experiencing" (Dawkins, 1980, p. 11). If we accept that these sorts of tests are accessing subjective experience, then it is measurable. Indeed, unless we think subjective experience is epiphenomenal, having no causal impact on the world, then there must be behavioural and physiological effects of mental states, which we can then detect and can form the basis for measurement (see Chapter Four for discussion of why subjective welfare is also a measurable entity).

The arguments for assuming the presence of other minds take the form of arguments from analogy and parsimony – we assume others have minds because of their similarities to ourselves, and because we have no reason to assume ourselves to be the privileged unique. These same arguments can apply in the case of animals. Although they are different from us in many ways, there are behavioural, physiological and evolutionary similarities that would suggest that they also possess minds. While the sceptical worry may be correct that we can never be certain that we have accurately gauged the mental state of an animal, there certainly seem to be methods by which we can make a close estimate, using their behaviour and physiology. Tests such as preference tests appear to give powerful information as to the mental states of animals; at the very least to reveal what they find rewarding or aversive.

There are then of course further issues to do with the measurement of welfare, which I will address in the second part of the thesis, such as how accurately we can measure it, what type of measurement scale we can use, and whether we can compare welfare experience across time or across individuals and between different types of mental states. These concerns are important ones and will be discussed in detail in later chapters. For now, it is sufficient to establish that, at least in principle, we can measure subjective states through their detectable effects.

2.3. Objections to subjective welfare

2.3.1. Torpid Tigers

I have shown that we have good reason to think that subjective experience is constitutive of the state of welfare, and that it is adequate for both the normative and scientific roles the concept must fulfil. What, then, are the potential problems with it? The primary objection to a subjective welfare account is that it will end up being too narrow, requiring other components to fully capture welfare (as per the tripartite conception that will be discussed further on). That is, we might have an animal that perfectly meets the subjective criterion for welfare, but we would still not want to say it is experiencing ideal welfare, without adding some other conditions.

Take this animal, the 'torpid tiger'[9]. Tigers can be challenging for zoo managers, as they are wide-ranging carnivores in the wild, and in captivity can become frustrated if unable to perform roaming, hunting and killing behaviours (Szokalski, Litchfield, & Foster, 2012). This frequently manifests in pacing, and tigers are often seen moving up and down a single fence of their exhibit on what is clearly a well-worn path. Now we imagine that one zoo manager, eager to combat this obvious welfare issue in tigers, begins a breeding program. They select only the quietest tigers, those that seem to prefer sleeping to roaming or hunting. After a few generations, they have created the 'torpid tiger'. This animal shows no desire to hunt or kill, as evidenced by its lack of interest in enrichment items designed to channel these behaviours. It does not pace, instead choosing to rest and sleep throughout its days, rousing only to eat when necessary[10]. It seems this animal has great subjective welfare – it has what it wants and is happy all the time. And yet, our instinct is that there is something wrong with this picture. There is something lost in the 'tigerness' of this animal. It does not have ideal welfare[11]. This is similar to the 'adaptive preferences' criticism of subjective welfare concepts in the human case, which will be discussed in Section 2.3.2.

There are two lines of response to this case. The first is to refute that there is any real problem with welfare, claiming instead that the problem simply lies within our own expectations. The second is to identify the lack not within the tiger itself, but in where we set our baseline – by what standards we judge the 'best' level of subjective welfare.

---

[9] This example and discussion also appears in Browning, H. (2019a). The natural behaviour debate: Two conceptions of animal welfare. *Journal of Applied Animal Welfare Science*: 1-13.

[10] A similar real-world example is presented by Sandøe et al. (2014), with blind hens who do not engage in feather-pecking or cannibalistic behaviour when kept in intensive housing systems and thus may have increased welfare.

[11] A similar challenge is sometimes presented, with the example of whether it would be considered beneficial to animals' welfare to consistently drug them into a state of happiness, and I think roughly the same set of replies presented here also applies in that case.

In the first instance, it seems entirely possible to refute that there is a welfare problem here at all. Perhaps our intuitions are just incorrect. Rather than our feeling of 'wrongness' reflecting any welfare problem with our torpid tiger, it simply reflects our own biases - as Rollin (2006) puts it, "a queasiness that is at its root aesthetic" (2006, p. 128). We are conditioned to seeing tigers in particular ways, to enjoying certain features of them. When we see a tiger that lacks these features, we are disappointed. This certainly seems to be the case with the multitude of zoo visitors who constantly express their dismay at seeing sleeping animals, apparently unaware that most animals – particularly big cats – also spend the majority of their time in the wild sleeping. The problem lies not with the welfare of our perfectly content tiger, but simply in our own categorisations of what animals should be like. Indeed, it seems that the burden of proof may rest on the shoulders of those who feel troubled, to justify why it is that the perceived problem is one of poor tiger welfare, rather than, say, a human concern with aesthetics, or the ethics of manipulating tiger lives.

This response may still be unsatisfying. It does not get to the heart of our feeling that there is a problem in welfare itself – that this animal is not experiencing welfare as high as it should be. In response, we need to examine where we set our baseline for welfare. When we are saying an animal has good or poor welfare, we need to be explicit as to what this is in comparison to. Rather than setting the comparison point at zero, so that any animal in a positive state is doing well, we could instead look at the maximum possible for the animal. It is not enough to simply say that an animal is perfectly content within itself – that it doesn't know any differently. What we really want is a comparison between the experience of this animal in its current situation, and how it could be in its best possible situation. Although the torpid tiger is not experiencing any suffering, and is content with its days of sleeping, there may be a range of positive mental states it is lacking - those associated with achieving the goals of hunting or killing. Even if the tiger would not choose these activities, it does not follow that it would gain no subjective benefit from them – the activities we select are not always those which bring us the greatest pleasure (see discussion on 'liking' vs 'wanting' in Section 2.4.3). This is not to say that these pleasures are in some way of a more valuable type than others, but that our intuitions are that they are simply *more* pleasurable. If we simplified welfare into something like 'happiness units', it could turn out that the number of attainable units for a torpid tiger are not as many as it could otherwise have had, had the situation been different (see Chapter Six for discussion on making such welfare comparisons between individuals). The tiger is then said to have reduced welfare, not because it is suffering, but because it is not in the best possible state it could be in.

Even if this tiger is experiencing its best possible welfare in terms of the maximum happiness it can obtain, we may think we have harmed its welfare by creating it such that it can only obtain this reduced level of happiness. There is a welfare problem for our torpid tiger if its experience of subjective welfare is lower than that of a traditional tiger that has its needs met. A content torpid tiger may still have better welfare than a frustrated traditional tiger. But it does not have welfare as high as a content traditional tiger and has thus in this way been deprived. Though it may be true that where conditions cannot be improved for the animal in terms of providing additional behavioural opportunities, breeding such that it does not experience frustration at this lack could actually be better for welfare in these cases (see also discussion in Sandøe et al., 2014). I take it that in most cases, it is possible (even if costly) to improve conditions and this will then still be the better option for improving welfare. The choice between a frustrated traditional tiger and a content torpid tiger is thus a false one, as there is another superior alternative potentially available (Sandøe et al., 2014). We have reduced the tiger's welfare by breeding it to be such that it is not capable of experiencing a fuller range of positive states. It is this that leads us to see a welfare problem with our torpid tiger. This is, of course, not to say that it is therefore impermissible to create such a tiger or that we are obligated to only create tigers with the best possible welfare. This relationship between permissibility of an action and potential welfare harm will be discussed in Chapter Three.

2.3.2. The human case

One potentially strong objection to the subjective conception of welfare, is that it is one that is widely rejected in the human case. The most common conceptions of human wellbeing are a preference-satisfaction account (that welfare consists in the satisfaction of preferences) and an objective list theory (that welfare consists in meeting a set of objective criteria set out as necessary for wellbeing), as well as hybrids of these views. Given that the subjective conception of welfare is so unpopular in the human case, we might think that this counts against it in the animal case. Why should we think that welfare in humans is an entirely different kind of thing to that of other animals?

I have two responses to this objection. The first is that I am far from convinced by the arguments against subjective welfare in the human case. The second is that there do seem to be important differences between humans and other animals, that could ground such a difference in welfare. The reasons for rejecting a subjective account of human welfare do not then apply to non-human animals.

In debates about the nature of human welfare, the subjective conception is often rejected. The primary objection against this account is that it fails to capture some objective goods we think are important for welfare (Crisp, 2017). The classic thought experiment supposed to illustrate the insufficiency of a subjective view of welfare is that of Nozick's 'experience machine' (Nozick, 1974). In this experiment, Nozick asks us to imagine we are presented with the offer to be hooked up to an experience machine. The machine will provide us with a virtual reality experience that is entirely pleasant to us, in which all those things that we wish for are provided. Further, we will have no recollection of having been made this offer, and will experience this as though it were our genuine life. Despite this, the intuition of most people is that it would be undesirable to take up such an offer. This is thought to demonstrate that there is more to welfare than simply positive experiences and thus refutes the 'subjective' conception of welfare (Crisp, 2017).

Firstly, there is good reason to be suspicious that our intuitions are reliable when thinking about this case. People presented with the case cannot imagine it correctly – given that our current selves know about the machine, we are unable to properly imagine the future experience of not knowing that we are in it, once we enter, and thus we still take on the feeling of some harm that we would actually not experience. Weijers (2014) found that when the case was presented in different ways, people's intuitions about the acceptability of the experience machine changed dramatically. Primarily, when the case was presented as discovering you were *already in* an experience machine, and the offer was to unplug and live a slightly worse, but more authentic life, far fewer people chose to leave the machine. This suggests other psychological mechanisms, such as a status-quo bias, are at play when considering the thought experiment, and lowers our confidence in our intuitions as a reliable guide in this case.

Further, even if we were to accept that the intuitions were reliable, I think the conclusion drawn from them is mistaken, and turns on what we think represent *conditions for* welfare, rather than the state of welfare itself. That is, the subjective account of welfare still handles our intuitions with regards to the experience machine, if we allow that pure hedonistic pleasures may not be the only things that bring us positive mental states – that aspects such as our feeling of autonomy may also serve such a role. Then it becomes clear that even if we consider our welfare to be compromised in using the experience machine (and it is not clear that it is not instead a clash between welfare and other values)[12], this is not necessarily because we have an

[12] DeGrazia (1996) would seem to agree with this judgement [about the experience machine], that perhaps in this case we are relying on intuitions about goods that are not welfare: "Suppose . . . internalism is correct: your well-

objective conception of welfare, but could also be because we consider autonomy to be an important component of our own welfare, perhaps more important than pure pleasure. If this is the case, then the knowledge of (or belief about) our own autonomy therefore simply brings us a *subjective* benefit rather than being a good in and of itself. Autonomy would thus be valuable because we like it. As Crisp (2006) points out, lists of 'objective goods' such as these only count as counterexamples if they are examples of goods that are "both widely accepted as a contributor to well-being, and never enjoyed" (2006, p. 637). As the examples offered – autonomy, achievement – are things which we tend to enjoy the feeling of possessing or attaining, they do not serve to weaken the case for a subjective conception of welfare. Indeed, it seems odd to consider someone's welfare to have been compromised if they have no strong feelings towards their future autonomy and have joyfully accepted the offer, now contentedly living the life of the experience machine. This argument then does not seem to rule out subjective welfare in this richer sense.

Even if one were to reject this and insist that there are some things that have a direct benefit to human welfare, outside of their subjective value to an individual, it still does not seem that this must affect our decision in the animal case. This is because, although the list of those components considered important for human welfare in an objective sense may vary, very few of these apply to the animal case. For example, it is difficult to imagine what autonomy may mean to an animal in any deep sense, or authenticity. "Given our historical and moral emphasis on reason and autonomy as nonnegotiable ultimate goods for humans, we believe in holding on to them, come what may . . . in the case of animals, however, there are no . . . higher priority values, like freedom and reason lurking in the background" (Rollin, 2006, pp. 126–127). If, say, it turns out that great apes or other animal groups, do have understanding of such ideals, then perhaps we would have reason to accord their preferences such primacy, but for most animals this is unlikely to be an issue. Although this does not rule out the use of an objective list for animals, perhaps containing other goods, there would need to be a separate case made for this, as the arguments made in the human case are insufficient. This then does not give us reason to reject the subjective conception of welfare.

The other type of welfare concept preferred in the human case is preference-satisfaction. In Section 2.4.3, I will discuss why this account of welfare can be collapsed onto a subjective

being is affected only if your body or mind is. This still allows the possibility that external goods matter morally, even if they do not benefit the person to whom they are external . . . suggesting that considerations of well-being do not exhaust what is morally important. Maybe not all wrongs are harms." (1996, p. 224).

account. However, even if we do not accept this as relevant to the human case, this seems to be because of the type of rich preferences described, which require cognitive and cultural sophistication beyond the capacities of most animals, resting on a deep belief-desire psychology and a strong sense of self and agency persisting over time, that many animals do not possess.

Another primary line of argument against the subjective welfare concept (as well as preference satisfaction accounts) is that of adaptive preferences (e.g. Khader, 2011). Like the torpid tiger, this type of thought experiment is supposed to capture the problems with subjective accounts of welfare in the human case. For example, we may have a woman living in the slums of a poor country, with very little food or opportunities for self-expression. However, due to her limited experience of the world, she does not wish for more. She is completely contented with her lifestyle, experiencing only positive mental states – therefore having good subjective welfare. By contrast, we have a woman living in a wealthy country, in a comfortable middle-class house. She has all her material needs well met, and as a result is able to find the time for her creative pursuits. Again, she feels only positive mental states and has good subjective welfare. However, we surely do not wish to say that both these women have the same level of welfare. As this example is similar to that of the torpid tiger, my responses to it are of the same form.

Firstly, I can say that the difference here is not one of welfare. If both women really do have the exact same level of subjective welfare in terms of their positive mental states, then I am happy to concede they have the same welfare overall. I would suggest that our intuitions otherwise lie outside of considerations of welfare. Instead, we may be worried about something like injustice, or instead be imagining ourselves in this situation, where we can't conceive of ourselves having truly positive mental states as a result.

Secondly, it seems unlikely that two women in these situations really *do* have the same level of subjective welfare. The poor and hungry woman is likely to experience negative effects of physical deprivation, feelings of hunger or illness that will impact her mental state. A subjective mental state is more than just lack of wishing for something different, it will contain feedback from physical functioning etc. Even if this is not the case, if she does not experience any negative mental states, this does not automatically mean that she has the same level of subjective welfare as the second woman. As in the torpid tiger case, this is a question of baselines. Subjective welfare does not necessarily reach a maximum simply because of the absence of negative mental states. There is still the possibility for a greater number or richness of positive states. It may differ between individuals as to what creates the fullest level of

positive subjective states, but we can imagine for humans that things such as creative self-expression may rank highly. Therefore, someone lacking this may have a lower level of subjective welfare than someone who has it, even if its absence is not noticed or causing frustration. Our baseline for comparison shouldn't be a neutral state of merely lacking negative feelings, it should be the fullest experience of positive feelings. If, given all this, we still wish to say that the first woman contains the same level of subjective welfare as the second woman, I don't see any reason to say there is a problem with her welfare.

Working through these examples, it seems that there is nothing in the case of the parallel debate on human welfare that rules out the 'subjective' conception of welfare for the animal case and since there are positive reasons to accept it, as detailed above, we should do so. This may mean that we need to accept different concepts of welfare for humans and for animals, based on differing capacities and interests[13]. I suspect that we have reason to prefer using the same concept for both animal and human welfare - and for the reasons outlined in this section that it is probably a subjective welfare concept in both cases - but it is not my project here to draw conclusions about human welfare and I will not defend this view. My aim has been simply to show that the parallel debate in the human wellbeing literature does not provide reason to reject a conception of subjective welfare for animals.

2.4. Competing accounts

I have provided here a positive account for why we should consider subjective welfare to be a good candidate for composing the state of welfare. I will now show how other proposed accounts are insufficient for the task. They are actually describing conditions for welfare that can be collapsed onto subjective experience, so that where these factors are important for welfare, they are only instrumentally so in virtue of their effects on subjective experience. I will do so by showing that for any proposed welfare component C:

- Cases in which C impacts welfare are cases in which it impacts subjective experience
- Cases in which C does not impact subjective experience are not relevant to welfare concerns
- We cannot trade off a decrease in subjective welfare for an increase in C
- We can trade off a decrease in C for an increase in subjective welfare

[13] This may lead to some animals, such as great apes, possessing the 'human'-type welfare, and possibly some so-called 'marginal' humans possessing the 'animal'-type welfare, but in terms of providing a concept for use in animal welfare science and ethics, this will not be a problem for most cases.

When these points are all true, it will be the case that although C may be an important condition for welfare, it is not an intrinsic component, and subjective welfare is the more fundamental state composing welfare.

There are two primary alternative accounts of animal welfare: a tripartite account which considers welfare to consist in multiple components – subjective, physical and teleological; and a preference-based account which considers welfare to consist in meeting the preferences of animals.

The first account – the tripartite – is probably the most common. Under this framework, welfare consists in feeling good (subjective welfare), functioning well (physical welfare) and living naturally (what I call teleological welfare) (e.g. Fraser, 1999; Webster, 2005). This is similar to the 'objective list' theories of human wellbeing, where these three components form the list of conditions that must be met for good welfare. We have already discussed subjective welfare. Physical welfare refers to the physical functioning of an animal, its bodily health and comfort. Teleological welfare refers to the 'naturalness' of an animal, how closely its behaviour and lifestyle match that in which it has evolved to live. This tripartite theory might seem like an appealing alternative as it recognises the importance of subjective experience, but also includes other things we might think are central for welfare. However, there are potential problems in using a tripartite concept like this – in adjudicating conflicts between the different components, and that the components themselves are not actually necessary for welfare.

As acknowledged by Maple and Perdue (2013), the components of each of these may differ and can thus create conflicts (2013, p. 26). For example, a concrete and tiled cage that allows for proper cleaning and disinfecting will provide maximally for physical welfare in terms of health, but will strongly detract from psychological welfare due to lack of stimulation, and cannot be considered at all natural. By contrast, addition of natural pieces of cage furniture, such as logs, will increase naturalness and provide increased mental stimulation, but may harbour disease-causing organisms. Allowing the animal free choice of diet would greatly increase subjective experience (at least in the short term, as I will discuss later on), but will have negative physical effects in terms of weight gain and malnutrition.

In cases such as these, it does not seem that there is a good way to adjudicate between these competing demands. There is not a common currency by which we can compare changes in each of the different components, or combine them into a single welfare score[14]. If each of these factors is seen as equally primary in welfare, then there is no reason that one should win

[14] See Chapter Seven for discussion of a similar possible worry for different types of mental states.

out over another. Using this framework seems like it could lead to problematic conclusions. For instance, we might imagine an animal kept in full health in a highly natural enclosure, but finding this a source of constant distress. We surely would not want to say that this animal has better welfare than a happy animal in an enriched unnatural enclosure, even if the 'overall' measure is higher. Our intuition in these cases points us towards giving greater weight to the subjective experience of the animal within the environment.

Dawkins (2004) provides an example to illustrate how calculations of this type can fail, a case in which different components contradict one another and subjective welfare should be considered as the most important. When hens are given access to an 'enriched' environment (containing materials such as litter for dustbathing), they will preferentially choose this environment. However, the hens also show a rise in cortisol levels (indicating stress) and decrease in shell thickness of the eggs they lay. Considered as a list, two measures (cortisol and shell thickness) have indicated poor welfare, while only one has supported positive welfare (choice). We may then conclude that the environment is not beneficial to the animal, but Dawkins argues that this would be a mistake: "What the birds wanted is thus not just another measure of welfare, but a necessary piece of evidence that gives valence and meaning to the more physiological measures" (2004, p. 54). The preferences of the birds is the indicator most closely tied to their subjective experience, and though still fallible (animals – and people – are not always good at knowing what is best for them), it is considered to be the one more central to indicating welfare (in Section 2.4.3 I will show why it is the link to mental states, and not simply preferences themselves, that matters here).

In contrast to a list-based account, one in which only a single item constitutes welfare – in this case, subjective experience – has the benefit of allowing us to make calculations of overall welfare without needing to search for some additional common currency with which to consider trade-offs. Of course, this still leaves open the issue of considering trade-offs and comparisons within the category of subjective experience, and this will be addressed in Chapter Seven. When considering whether some change will be beneficial to the welfare of the animal, we need only ask, 'Will it improve the overall subjective experience of life by the animal?' If the answer is yes, it is a worthwhile welfare improvement; if the answer is no, then even if the improvement seems more natural or perhaps enhances physical health, an alternative should be considered. As I will show, it will often be the case that improvements to physical health or naturalness will lead to increased welfare through their effects on subjective experience, but where they do not, they do not then have any further welfare value.

2.4.1. Physical welfare

First consider physical welfare. It is undeniable that physical functioning is important for welfare. It's extremely unlikely that an animal that is unwell, or in discomfort, is going to have good welfare. However, we can account for this as an effect of the unpleasant subjective experience of these conditions. Cases in which physical health impacts welfare are cases in which it impacts subjective experience, and where it does not do so, we don't think there is a welfare concern. Illness and injury are almost always associated with strong negative subjective states, such as pain: "Usually when an animal is ill, it will also feel ill, so that taking care of its mental state (i.e., how it feels) will automatically take care of its physical health. However, there may be cases in which the animal is not in full physical health, but feels all right, and we conclude that its welfare is all right" (Duncan & Petherick, 1991, p. 5018). However, when the subjective experience associated with physical health is not negative, it does not seem that welfare is compromised. To motivate this, imagine an animal with impaired physical functioning that doesn't seem to have any associated subjective welfare lack. Animals are notoriously good at coping with physical injuries and deformities. I once worked with a lemur who had lost an arm in an accident, and would quite happily tripod around his habitat. He was able to do almost everything a normal lemur could do, and it thus seems odd to say that his welfare was compromised by his physical lack. Similarly, if an animal has an infection, but we are able to give it medication to relieve all symptoms so that it subjectively feels well, it doesn't seem that we want to say it is experiencing poor welfare, as we watch it move about and enjoying life as it did before.

We might want to deny this still true in the case of a physical ailment which creates no subjective suffering but shortens life – this is impaired physical functioning with no subjective component, but in a case where we want to say welfare is impacted. However, a subjective concept can account for the welfare problem in premature death, as this is removing the possibility of future positive experiences. As described earlier, welfare consists in the subjective experience of an animal over its lifetime. An animal which dies young, will have lower lifetime welfare than it would have had if it had lived a longer life. This issue, of the relation of length of life to welfare, will be discussed further in Chapter Three. These examples all show that cases in which we think physical health is impacting welfare are actually cases in which we are concerned about the impact on subjective experience, and in the absence of such an impact, welfare is not harmed.

Let us consider the cases of trading off one for the other. If both physical and subjective welfare are equal components of welfare, then it seems we could decrease one while increasing

the other, leaving total welfare the same. However, this is not the case. Many zoos used to try this – to keep animals in sterile concrete cages to help avoid disease, while causing the animal psychological harm in the forms of boredom and frustration. This practice has been largely discontinued, primarily because the harm to subjective welfare was considered to outweigh any benefit to physical health.

What of the reverse – a case in which we decrease physical welfare without also decreasing subjective welfare? As described in the cases above, for a physical lack without corresponding distress, or a subclinical or managed disease, we do not see a welfare problem: "An animal with a tumour it cannot feel does not have a welfare problem, even if it does have a health problem" (G. Mason & Mendl, 1993, p. 302). Performing surgery under anaesthesia is another example – it is extremely physically invasive, but as the animal is not conscious, it is not considered harmed by the procedure. We do not consider it a problem to decrease physical welfare if subjective welfare is not also impacted.

One example raised by opponents to the subjective account is one in which subjective pleasure is increased through allowing an animal to eat large amounts of junk food, with a corresponding decrease in physical health as the animal becomes fat and sick over time. This is intended to show that we can't only consider subjective experience. However, as a fat and sick animal is typically going to have many associated negative mental states, such as sluggishness and frustration from lack of mobility, it is not easy to separate the two components. If there is decreased lifespan (as there almost certainly would be), then the welfare impact of this can be accounted for through lack of future positive experiences, as discussed above. If it truly is the case that the sum of the gustatory pleasures experienced by the animal outweigh the negative feelings associated with obesity and illness, and those future positive experiences that may be excluded, then in these cases it may just be true that the animal is experiencing good welfare, despite possible intuitions to the contrary.

Although there may be difficulties in establishing how to compare measurements of subjective welfare over time, this is not necessarily just a problem for the subjective welfare conception – proponents of a physical welfare concept are also wanting to retain subjective welfare, and comparing physical changes to subjective ones may be even more intractable. The issue of comparing subjective experience over time will be discussed further in Chapter Six.

Physical functioning is likely to be highly important for welfare in most cases, but this is as a *condition* for subjective experience, rather than primary in itself. It may also function as a useful indicator – where we think physical functioning will be impacting on subjective states, we can use measures of health to track subjective welfare. However, physical health itself is

not a necessary part of the state of welfare and it will matter for welfare only inasmuch as it impacts on subjective experience.

2.4.2. Teleological welfare

I turn now to a more controversial position – what I have called teleological welfare; whether naturalness forms a central part of welfare. Perhaps we might aim to rescue physical functioning in cases described above, by pointing out that these animals may be prevented by their physical lacks from behaving in natural ways, and so this is where welfare is harmed. However, again there is no strong reason to think this is the case. There does not seem to be any necessary link between naturalness and welfare, whether we take natural to mean functioning according to an evolved design (or *telos*) or simply the conditions the animal experiences in the wild (see Browning, 2019a for a more complete discussion). There are many highly natural states, that seem to be quite bad for the welfare of the animals experiencing them – think starvation, predation, conspecific aggression. And perhaps many unnatural states that appear to enhance the welfare of animals – such as great apes using electronic devices for mental stimulation (Perdue, Clay, Gaalema, Maple, & Stoinski, 2012), or animals receiving veterinary interventions such as parasite removal.

Natural behaviour is important for welfare where it impacts subjective experience, and its performance will often correlate with increased subjective welfare. Animals have evolved with psychological reward systems that make many natural behaviours feel good to them (though some evolved natural behaviours such as predator avoidance will have associated negative affects, and we do not usually want to say that these are associated with good welfare). In trying to find optimal conditions for animal welfare, it will often be the case that the natural environment and behavioural repertoire will be the best starting points. For example, Dawkins (1989) analysed the time budgets of red junglefowl (*Gallus gallus*) as a model for domestic fowl (*Gallus domesticus*). She found that even when fed regularly, the junglefowl still spent a lot of time in foraging activities, which suggested they might be important, and flagged this as a starting point for further testing on domestic fowl. Špinka (2006) also argues that natural behaviour can be a useful criterion for animal welfare as it may be the easiest way of meeting the animal's desired ends, provide positive emotions and have longer-term effects on health and welfare that may not be assessed. For example, mink that are allowed the opportunity to play in water will later show more play behaviour in their cage (Špinka, 2006). Veasey et al. (1996) suggest that, all other things being equal, it is probably better for a captive animal to spend its free time performing natural behaviours, and a captive animal performing all relevant

wild behaviours is probably (but not definitely) less likely to be suffering than one that is not. However, as gestured to earlier, in cases where natural behaviour is not linked to subjective states (or where it is linked to negative states), naturalness is not doing any additional work for welfare.

Even those who seemingly advocate a teleological view of welfare appear to base it on subjective experience. Rollin (2006) has been one of the primary advocates of the concept of *telos* in determining animal welfare. However, it seems that underlying his view is a subjective concept of welfare. He is at pains to stress that only sentient animals are capable of having a *telos* of moral importance, as it must be such that it *matters to the animal*. He considers it "necessary that we can say sensibly of the animal that it is aware of its struggle to live its life, that the fulfilling or thwarting of its needs matter to it" (2006, p. 100). He further goes on to say that it is the suffering caused by frustration of these interests that gives it moral importance: "The animal must be capable of being aware that the thwarting of the need is a state to be avoided, something undesirable" (2006, p. 102). Although he rejects a simple pleasure/pain framework, he includes other possible negative mental states, such as frustration, boredom, anxiety etc. In this sense, then, *telos* can simply be viewed as a useful shorthand, a way of getting at what is likely to matter to the animal.

Another writer who emphasises a teleological view is Nussbaum (2004, 2011). She argues for the importance of a 'dignified existence' for animals, which entails pursuing their own interests to flourish naturally: "It is good for that being to flourish as the kind of thing it is" (2004, p. 306). She calls this view neo-Aristotelian in its implicit reference to telos: "the Aristotelian idea that each creature has a characteristic set of capabilities, or capacities for functioning, distinctive of that species, and that those more rudimentary capacities need support from the material and social environment if the animal is to flourish in its characteristic way" (2011, p. 228). From this she builds a list of objective 'entitlements' animals should possess.

On the face of it, she appears to wish to differentiate herself from a subjective view of welfare, arguing that "there seem to be valuable things in an animal's life other than pleasure, such as free movement and physical achievement" (2004, p. 304), valuing these things even if they are not the source of positive experience to the animal. However, she also believes that not all natural functions will be equally valuable: "there is waste and tragedy when a living creature has the innate, or 'basic', capability for *some functions that are evaluated as important and good*, but never gets the opportunity to perform these functions" (2004, p. 305, italics added). We must thus somehow evaluate *which* of these we wish to preserve and here she does seem to fall back on subjective concerns, arguing that we should try to avoid the "pain of

frustration" (2004, p. 311) for animals denied particular opportunities. So again, we might see this work as a teleological view in practice, but grounded in a subjective conception. That is, teleological considerations may inform our decisions about what is likely to matter for welfare, but this is true in virtue of its effects on subjective state.

We don't allow a trade-off to increase naturalness at the expense of subjective welfare. Allowing animals to suffer large parasite burdens, or fight one another to the death, while closely tied to 'natural' states or evolved functions, are considered unacceptable practices and decrease subjective welfare. Conversely, decreasing the naturalness of an animal's life in order to increase its subjective wellbeing does not seem problematic – come back to our apes on iPads, or monkeys in cages as opposed to more naturalistic open-air enclosures.

The underlying assumption to this claim for the importance of naturalness in welfare appears to be that the wild state represents the best possible welfare for animals. This is a version of the 'appeal to nature' fallacy: the mistaken belief that because something is natural, it must be good. It is a clearly dubious assumption, as there are many examples of animals suffering in the wild. Individuals are often physically injured, malnourished, stricken with disease and exposed to unfavourable environmental conditions. As mentioned, evolved behaviours such as predator-avoidance will have associated negative mental states. In a behavioural sense, it is not the case that wild animals are free to perform all their natural behaviours. Animals suffering from illness or injury will clearly have a more limited behavioural repertoire. Additionally, many subordinate animals will be denied the opportunity to access particular food resources, or mates, through competition with conspecifics, or threat from predators. Animals in the wild are far less free than is commonly assumed.

As shown here, there is no necessary connection between naturalness and welfare. There is no reason to think that the conditions experienced by wild animals have any link to welfare, and though many evolved functions are associated with positive experience, where they are not, we do not think they are beneficial to welfare simply in virtue of their naturalness. Although natural behaviours will often increase welfare, this is because of their impact on subjective experience and not because of their intrinsic value. Natural behaviour may sometimes be a condition for welfare, as many natural behaviours are pleasurable, and the performance of natural behaviours may help indicate good welfare, but teleological welfare is not itself a central component of welfare.

2.4.3. Preference-based welfare

As described earlier, one popular account of welfare in the human case is that welfare consists in the satisfaction of preferences. That is, for your life to be going well is for your preferences to be satisfied, and for it to be going poorly is for your preferences to be frustrated. Some writers (e.g. Dawkins, 2003) apply a similar account to animals, grounding welfare in the choices they make. Under this view, if animals have what they want (providing it doesn't impact negatively on physical health), then they have good welfare.

However, the preference account can also be collapsed back onto subjective welfare. That is, preference-satisfaction is only important because satisfaction of preferences produces a strong feeling of subjective wellbeing, while their frustration is associated with negative feelings. In the human case, this move is resisted. After all, humans might want things which don't feel good, and these preferences are still considered important for their wellbeing. For animals though, without the attendant desires for autonomy or authenticity, preference-satisfaction does not seem to consist in anything more than subjective experience. Without higher-order preferences, it is difficult to imagine exactly what preferences would be if not just the positive association with one state of affairs and the negative association with its frustration. Because this will often be a very strong link, it will almost always be the case that satisfying preferences will produce (at least short-term) increases in subjective welfare, while denying preferences will decrease welfare due to the associated frustration.

The cases in which we think preference-satisfaction impacts welfare are just cases in which these subjective states occur. If an animal had no real affective response to the satisfaction of particular preferences – was in some sense subjectively neutral towards it – this does not seem a concern for welfare. Further to this, some preferences may unexpectedly create negative mental states. In these cases we would want to say that welfare had been harmed, even though preferences were satisfied (controlling for the subjective effects of preference satisfaction or frustration). In the human case, some would want to say that having your preferences thwarted is detrimental to your welfare even if you never know about it, such as if your children die after your own death. This objection is unconvincing in the animal case, largely because animals are unlikely to hold preferences for objects or states of affairs outside their immediate experience. Although preferences and mental states are closely linked, it is the impact on subjective experience that makes preference-satisfaction important.

This line of argument could be pursued in reverse – to argue that where preferences and mental states overlap, it is instead preferences that are primary. DeGrazia (1996) takes this view – accepting the link between mental states and desires, he defines pleasure as coming

from an experience desired for the way it feels, and therefore increasing pleasure just is meeting these desires (1996, p. 124). Subjective welfare in his view therefore collapses onto preference satisfaction. If preferences and subjective welfare were identical, as in this view, then there would be no reason to prefer one to the other – they are the same thing. Something would only be preferred to the degree that it was liked by the individual. Those who advocate this view do so because they believe that preferences are an accurate guide to liking (Camerer, 2006). However, it is not true that they are identical, as these processes are separable. The 'wanting' and 'liking' systems in the brain, (roughly analogous with preferences and subjective pleasures respectively) are not the same thing as one another (Berridge, 1996).

'Wanting' systems, underlying preferences, are mediated through dopamine (Berridge, 1996). These cause a particular object or experience to become more attractive, and increases motivation to seek it out (Berridge & Robinson, 1995). 'Liking' systems, which create positive subjective experiences, are mediated through opioids (Berridge, 1996). These systems are found through different parts of the brain and can be activated separately – so that an animal can like something they don't want, or want something they don't like (Berridge, 1996). This has been established through studies that use different indicators for each system – using behavioural measures such as consumption or level of work for wanting, and change in facial expressions for liking. Manipulations of dopamine can increase the behaviours indicative of appetite (wanting) without showing any effect on facial expressions indicating palatability (liking) (Berridge, 1996). This separation may be familiar to anyone who has experienced compulsive or addictive behaviour they did not enjoy – this is activation of the 'wanting' system without the 'liking' system. In most cases, activation of the 'liking' system will cause activation of the 'wanting' system: "wanting may be conceived as the imprinting of what was once liked and disliked" (Spruijt, van den Bos, & Pijlman, 2001, p. 160). The reverse is not true – things can be wanted without being liked (Berridge, 2009).

Liking, or affect, is also likely to be the more psychologically fundamental process. Ginsburg and Jablonka (2019) provide a detailed description of the probable evolution of and functioning of animal consciousness. Under their model, animals start by encountering various stimuli and receiving feedback from their interactions with these, in the form of positively or negatively valenced subjective experiences. This feedback then guides learning and memory, which guides future motivation and action. Preferences here are thus the output of a valuation system: "the neural states are motivating because they ascribe value to actions and percepts based on the overall (systemic) homeostatic state of the body in the world" (Ginsburg & Jablonka, 2019, p. 367). If this is the case, then something is preferred because it is judged

valuable, not the other way around (see further discussion on the role of affect and motivation in Chapter Seven).

These two systems – liking and wanting - are often closely linked, but can come apart. In these cases, we need to determine which it is that matters to welfare – the preferences or the subjective experience. Thinking about examples in which this occurs can help us judge which we think is more important, and these lead us to prefer subjective experience. Compulsive behaviours that are strongly 'wanted' but occur without enjoyment - such as gambling, or repetitive stereotypic behaviours in animals - would not be considered to increase wellbeing; while experiences that were not desired but were enjoyable, such as an unexpected encounter, do increase wellbeing. Impact on subjective experience is what matters.

This conclusion is further strengthened through consideration of acceptable trade-offs: whether we would consider it acceptable to decrease subjective welfare in order to satisfy more preferences, or to frustrate some preferences in order to increase subjective welfare. Intuitively, the latter is acceptable, while the former is not. There are many preferences held by animals, the satisfaction of which could decrease their subjective welfare, either now or in the future. Animals will not always understand the choices that they are making, and they are unlikely to have the ability to make good decisions for their future welfare (after all, even humans often do poorly in this regard). For example, most animals have a strong desire for fatty and sugary foods, consumption of which can lead to poor health (and, as discussed, poor subjective welfare). Addiction to drugs will lead to a strong desire for their consumption, which is detrimental to health and often to psychological wellbeing. Many animals will have desires for aggressive behaviour when faced with competitors, but acting on these desires can lead to stress and injury. If we were to allow animals to have all the things they desired, then under the preference-satisfaction account they would have good welfare (saving the effects of current preference-satisfaction on the ability to satisfy future preferences), but these examples illustrate why we would not think this to always be the case.

For the same reason, it seems acceptable to make the trade-off in the other direction – to allow a decrease in preference satisfaction where it will create an overall increase in subjective welfare. We would be happy to deny an animal something it wants, if we know it will actually gain greater pleasure from some alternative state of affairs. It is not the case that all things that will bring positive mental states are those things which an animal will directly choose, as they will not be fully aware of the potential outcomes of all their choices. As Nussbaum (2004) points out, animals may do well with conditions they do not know to prefer: "what the [preference satisfaction] view cannot consider is all the deprivation of valuable life activity that

they do not feel" (2004, p. 305). They may not have the necessary information to prefer these activities, but would enjoy them if they experienced them. We may thus have cases where we frustrate preferences, in favour of creating pleasure, and in these cases the intuition is that we have not harmed the animal but instead increased its welfare.

In the human case, this objection is dealt with through appeal to a modified set of preferences. These can be higher-order preferences, or preferences about our preferences, such as a preference not to gamble excessively, regardless of our strong desires at the time. These preferences are usually grounded in our own sense of self, and our overall desires for the features of our life (Griffin, 1986). They require an understanding of the self, and one's continuation through time. As most animals are unlikely to hold higher-order preferences of this type - lacking this sort of psychological cohesion and self-reflection - this move won't work to overcome the objection in the animal case[15].

The second way the preferences may be modified is through constructing some sort of idealised preference set an individual might hold if they had a greater knowledge and understanding of what is good and bad for them. There are two problems with this response. Firstly, it is unclear what the criteria would be for building such an idealised set of preferences. If it is just those things which would make us happiest, were we to know about them, then this is just the subjective experience view I have been arguing for. If it is based on increasing some other set of important goods, this starts to look like an objective list theory, and my previous discussion of this view applies. Secondly, and most importantly here, even if this would work for humans, it is difficult to imagine what an ideally rational version of a parrot looks like, or what it would want. By the time we constructed a set of preferences that some ideal parrot would hold, taking into account their values, and adding understanding of current and future consequences of actions, this seems to be far from the abilities of any actual parrot. If we were able to work according to our ideal parrot, this animal would be so vastly different from an actual parrot that its preferences may not be a good guide to what our real parrots would want, meaning many of their actual preferences would be frustrated.

There is still a role for preference satisfaction in understanding welfare. Preference satisfaction is likely to be an extremely important condition for welfare, in that satisfaction of preferences will usually lead to increased welfare, due to the attendant positive subjective experience. Preferences have evolved to track those things which are important for the survival

[15] It may turn out that some animals - great apes for instance - do possess these cognitive abilities, and so perhaps for them some version of preference satisfaction may work as a welfare concept, if it can overcome the other problems mentioned.

and reproduction of animals, which can often be associated with positive experience (Spruijt et al., 2001) (though, as discussed in the section on teleology, these links can come apart for captive animals). Over an animal's lifetime, preferences develop in accordance with those things they have found rewarding or aversive via associative learning (Berridge, 1996), and thus often track well with subjective welfare. Frustration of preferences and motivated behaviours can often create negative affect (Gygax, 2017), and so meeting preferences will also increase subjective welfare in this way. Preferences are also likely to be a very useful indicator for welfare, with the preferences of an animal showing us what they find to be positive or negative and how strongly they are motivated to obtain or avoid certain conditions (Kirkden & Pajor, 2006). However, it is not the case that these themselves will be central to what welfare is – preferences are only important to welfare where they are associated with subjective experience.

2.5. Conclusion

I have argued here that a subjective account of welfare – that animal welfare consists in the subjective experience of an animal over its lifetime - best captures what is important about animal welfare, in both the normative and scientific roles. I have shown that this account is preferable to the most commonly used competing accounts – the tripartite account in which animal welfare consists in subjective, physical and teleological welfare; and the preference-based account – and that these accounts can be collapsed onto the more fundamental component, subjective welfare.

If we accept this view, it has implications for the practice of welfare science, some of which will be examined in the second part of this thesis. In particular, it means we should narrow the focus to indicators that will map onto the subjective experience of an animal, rather than those that might currently only measure physical states or amount of natural behaviour. This does not necessarily lead to a large-scale revision within welfare science. Many of the measures used will already be doing this. It does mean that we have a means of adjudicating when different indicators might be telling us different things – we should go with what ties most closely to subjective experience. Finally, we should still keep physical, teleological and preference-related considerations in our toolkit. These conditions will frequently map closely onto welfare and are useful both as indicators and conditions for welfare. However, we should only use them while keeping in mind that they matter insofar as they impact subjective experience.

3. Chapter Three – Applications in applied animal ethics

Accepting the subjective welfare concept has implications for animal welfare science, and how we measure welfare, as will be addressed in Part Two of the thesis. However, it also has implications for applied ethics. As described in the previous chapter, animal welfare is a normative as well as a scientific concept. I argued for a subjective conception of welfare in part because it is morally significant. In this chapter I will look at some of the specific implications that holding this subjective welfare concept will have for problems in applied animal ethics and how welfare of this type can interact with other values in decision-making. Here I will address two cases which have not previously been much discussed in this area – management euthanasia in zoos, and de-extinction programs.

3.1. Management Euthanasia and Animal Welfare[16]

3.1.1. Introduction

In February 2014, Copenhagen Zoo became the subject of a media frenzy when they euthanized a young male giraffe, subsequently using his body for a public autopsy and eventually as food for the carnivores. This was controversial because the euthanasia took place not due to illness, but because he was surplus to requirements. Since then, similar incidents have followed, with similar responses (e.g. Nicholls, 2018; Parker, 2017), and much debate both for and against the practice. Those against argued that it was 'wrong and disturbing' to kill a healthy animal and use the body in such a way (Maple, 2014); while those in favour responded by pointing out that the killing was done humanely so the giraffe did not suffer, and that the limited resources of zoos created these difficult decisions in maintaining viable breeding populations (Rincon, 2014). This debate is not new in the zoo industry, with the problem of management euthanasia, or 'culling', having been discussed for decades (e.g. Lacy, 1991; Lindburg, 1991). In this chapter I will look at both sides of the discussion – coming from the animal 'rights' and 'welfare' positions respectively – before describing an alternative way of seeing the welfare position that might speak against the practice, and looking at some of the conditions under which it might be considered acceptable.

[16] A version of this chapter section has been published as: Browning, H. (2018a). No room at the zoo: Management euthanasia and animal welfare. *Journal of Agricultural and Environmental Ethics*. **31**: 483-498.

Within zoos, the practice described above – the killing of otherwise healthy surplus animals – is referred to as 'management euthanasia'. Euthanasia in general refers to humane or painless killing (most commonly through lethal injection performed under anaesthesia); it is a term based in ancient Greek that roughly translates to 'good death'. Euthanasia is typically associated with the practice of ending the life of an individual that is terminally ill or in chronic pain, so that the choice to end life can be considered an act of mercy or kindness and is also a 'good' death in this way[17]. What differentiates management euthanasia from this usual practice is that the animals involved are otherwise healthy. The decision is not made from consideration of their expected quality of life but instead, the animals euthanized are those considered surplus to the requirements of the institution: that is, those animals that are not on the overall management plan and which the institution lacks the resources to support. The practice of management euthanasia, though not often made public, is relatively common, with estimates that European zoos within the European Association of Zoos and Aquaria (EAZA) euthanize between three and five thousand animals each year (Barnes, 2014). There are suggestions that this is also occurring within American zoos, though numbers are not available (Parker, 2017).

Surplus animals can arise from a variety of causes. The simplest is uncontrolled breeding. If animals are allowed to breed without restriction, very soon the population will grow beyond a size that any particular institution has the resources to support. This sort of practice is rare in zoos, where populations are carefully managed for genetics and demographics. However, even in carefully controlled breeding programs, surplus can arise. In polygynous species, which are common in, for example, primates and hoofstock, an equal sex ratio at birth will lead to a surplus of adult males in the population as only a few are required for breeding. Maintenance of genetic diversity will require careful breeding of only those individuals which have under-represented genetics, and so any animals from already well-represented lines will be surplus to breeding requirements. So too for post-reproductive animals, those which have already made their breeding contribution to the next generation. Creation of viable self-sustaining captive populations requires careful use of all available spaces to house genetically and demographically valuable breeding animals, and using spaces to house surplus animals can threaten the viability of such programs (Penfold, Powell, Traylor-Holzer, & Asa, 2014). Powell & Ardaiolo (2016) list the common reasons for surplus – large litters, uncontrolled breeding,

[17] Some writers, e.g. Regan (1983) consider that the second requirement - the death being in the interests of the individual - is also essential for a practice to be considered euthanasia rather than killing. Here I simply follow the common usage within animal industries of 'euthanasia' as referring to the manner of killing rather than its intention.

unexpectedly high offspring number or survival, requirement for ongoing breeding to preserve fertility, sex ratio and presence of post-reproductive individuals. It is inevitable that even the most carefully managed breeding programs will create some surplus animals, which the institution must then manage in some way.

Often, opponents to management euthanasia cite alternatives that zoos should be using instead of culling animals. There has been a lot of writing done on these potential alternatives and their benefits and drawbacks (e.g. Asa, 2016; Glatston, 1998; Lacy, 1991; Lindburg, 1991; Lindburg & Lindburg, 1995; Penfold et al., 2014) and I will only summarise them here, to show that they are not always viable. For the purposes of the rest of this chapter, I will assume that when we are talking about management euthanasia, it is for situations in which there are no good alternatives available. The first alternative management strategy is to try and prevent or minimise the creation of surplus in the first place. This involves both careful strategic planning on which animals to breed and when (Hutchins, Willis, & Wiese, 1995), and use of contraception to prevent unplanned breeding. Neither of these methods is perfect. As discussed above, even well-managed programs will create some surplus, and contraception options can often have negative physiological and behavioural effects (Asa, 2016; Glatston, 1998; Penfold et al., 2014). Animals (particularly females) kept on contraception for too long can have difficulties in breeding in the future, threatening the long-term viability of breeding programs (Penfold et al., 2014), and can be at risk for health problems such as cancers. Prevention of breeding, particularly through separation of the sexes, can lead to behavioural problems and the potential for decreased welfare through lack of opportunities to perform beneficial breeding behaviours (Penfold et al., 2014).

There are then, obvious problems with preventing the creation of surplus animals, both in lack of effectiveness and undesirable side-effects. Other alternatives are aimed at other ways of managing these surplus animals once they do exist – housing within the institution, dispersal to other institutions, and release to the wild. Housing within the institution is the strategy endorsed by Kagan et al. (2015), who contend that "high-quality environments must be provided for animals removed from plans, 'retired' from breeding programs, or removed for other reasons" (2015, s3). Housing within the institution is usually possible, but as resources are limited, doing so will necessarily take resources away from other animals – taking up space that might be used for more valuable breeding animals, or resources that could be used to improve the housing and husbandry of other animals in the zoo. I will turn later to examination of these sorts of trade-offs, but suffice to say for now, that no zoo can continue to house all surplus animals over time without large potential costs to breeding programs and the welfare

of other animals. Dispersal to other institutions holds similar problems. Within any region, accredited zoos are managed as a whole, with spaces allocated throughout the region for particular breeding programs. This means that although in the short term, other institutions may have space to take on surplus, eventually the same problems that arise for a single institution will arise for the region, as all zoos reach carrying capacity. Dispersal to non-accredited institutions is problematic as they often will not meet suitable welfare standards for the animals. Lindburg (1991) has suggested creation of large-scale 'holding facilities' to which a number of zoos could contribute, and use to house their surplus. Here though, similar resource problems will arise, as resources are always scarce and those used for these facilities will necessarily be taken from that which could be used for management of other zoo animals within breeding programs. Release to the wild is not a viable option for most species. Even the most carefully-managed release programs are not often successful (Harrington et al., 2013). Captive-born animals do extremely poorly in the wild unless they undergo extensive training for release, which requires resources that are likely to be unavailable, and is usually unpleasant for the animals. For most species, there is also no suitable habitat available for release, that isn't under threat or already at carrying capacity and so fierce competition, fighting and predation are highly probable, and will lead to decreased welfare. Release of most animals would be condemning them to a much slower and more unpleasant death than that of management euthanasia.

In some cases, these potential alternatives can work to reduce creation of surplus, or manage surplus animals where they are created, but there will be many cases in which they are not possible or appropriate. Although almost all zoos will aim to use alternative options where possible, their use always requires trade-offs in other areas, such as decreased resources to put towards other animals, or breeding programs. There will still be situations where, all things considered, management euthanasia may be one of the better available options. The question of interest here should then be, when (if ever) is management euthanasia permissible, and under what conditions? The aim of this chapter is not to definitively answer this question, as the answer is likely to be highly context-sensitive and reliant on the values at play in particular institutions. Instead, I aim to discuss some of the considerations that are likely to play a role in forming an answer – considerations of the rights and welfare of the particular animal, as well as other potential competing values that exist within zoos and animal management.

3.1.2. Two sides – the rights and welfare views

Opinions about the practice of management euthanasia tend to differ depending on the underlying ethical framework at play. There are two major frameworks within animal ethics – the rights view and the welfare view[18] (Lindburg, 1999). Although these two views agree about many of the issues to do with our treatment of animals, they differ in the underlying motivations and ideology. They also form the basis for the two sides of the debate on management euthanasia, with most of the arguments against the practice being grounded in the rights view, and the arguments in favour coming out of the welfare view. Here I will briefly describe each of these views, and what they say about the practice of management euthanasia.

The rights view sees our responsibilities towards animals as being grounded in their rights as sentient beings. As animals are individuals with their own lives, thoughts and feelings, this creates a moral duty in others not to interfere with these. One of the early developers of this view is Tom Regan (1983), who sees the morally relevant feature of animals as being that they are "subject[s] of a life", with their own individual set of beliefs and desires, their own wellbeing, which ground their individual rights. Rights can include welfare rights - those things which prevent physical or emotional harm - as well as additional rights, such as to "some form of protection of their lives and liberty[19], irrespective of the impact on their welfare" (Gray, 2017, p. 91). This view comes out of deontological ethical theories, in which other individuals should be treated as 'ends in themselves' rather than 'means' towards our own goals, and it is not generally considered acceptable to infringe on the rights of an individual for some greater overall benefit (Alexander & Moore, 2016). In respecting these rights, animal rights advocates typically oppose any use of animals for human ends. One prominent organisation which operates within this framework is PETA (People for the Ethical Treatment of Animals), who consistently fight for the abolition of human use of animals (Lindburg, 1999).

The arguments against management euthanasia typically come out of the animal rights view. It seems clear that within this framework, if we should not use animals for human ends, then we definitely should not end their lives for human reasons, such as the allocation of resources. To do so would violate the right to life. This is one of the most fundamental rights as without life, it is impossible to enjoy other rights. Although there are situations in which rights can be

[18] There are other potential ethical frameworks through which we can view our treatment of animals, such as virtue ethics or feminist ethics of care (see e.g. Beauchamp & Frey, 2011; Gruen, 2011), but these are by far the two most common and as they are the two which frame this particular debate, they will be the only two considered here.

[19] Liberty not necessarily here implying complete freedom from captivity, but at least some measure of choice and control for animals within their environments; though many in the animal rights camp would take captivity itself as a violation of rights.

overridden when the stakes are sufficiently high (such as protecting stronger or more important rights), this is not the case for management euthanasia. The potential benefits in terms of increasing the welfare of other animals, or of increasing the success of breeding programs, are not of the sort that could be considered to override the right to life of the euthanized animal. Although there may be competing rights and values at play in the protection of biodiversity (as will be discussed in Section 3.1.4), these are not generally considered by rights theorists to be of sufficient strength to override rights to life. Rights are typically accorded to subjects, which the earth and its ecosystems are not. Under a rights view, the practice of euthanasia is only acceptable when it is aimed to directly benefit the animal itself, through ending suffering. This is what Regan (1983) terms 'preference-respecting' euthanasia; where in the case of animals, were the animal able to voice an opinion, it would choose this option for itself – typically in situations such as chronic or terminal illness. Because of this, it is considered morally permissible. Management euthanasia, by contrast, is a decision to kill a healthy animal whose preference would be to go on living, and for this reason is morally unacceptable. In the context of the question 'when, if ever, is management euthanasia permissible, and under what conditions', we can see that the answer from a rights perspective is therefore going to be 'never'.

In contrast to the rights view, the welfare view sees the moral status of animals as grounded in their experience as sentient beings, those capable of experiencing pleasure and suffering. Our responsibilities towards them are then those of providing the best possible welfare to make their lives go as well as possible. Welfare here is usually understood in terms of the subjective welfare concept I argued for in Chapter Two – as subjective experience, over the lifetime of the animal. That is, an animal's total welfare is equivalent to something like the weighted total of positive and negative mental states experienced (see e.g. Phillips, 2009). It is important to consider lifetime welfare, as experiences can have differential impact depending on life context (Houe et al., 2011). Those experiences causing negative mental states, such as fear or pain, will decrease welfare, while those experiences causing positive mental states, such as joy or satisfaction, would increase welfare.

Our treatment of animals should then aim to maximize welfare in terms of the quality of this experience. In practice, this is most often understood as the prevention of cruelty, so that no (or minimal) suffering will be inflicted on animals throughout their lives. Importantly, differing from the rights view, this does not exclude the human use of animals in areas such as science, farming, or recreation, as long as the animals are not harmed in so doing, or if the harm is outweighed by a greater overall benefit (though, as I will argue, as the harms of death are

greater than they may first appear, the benefit will also have to be larger). The welfare view emerges from a utilitarian moral framework, in which an action is judged on the overall outcome for all those affected (in this case, in terms of welfare) (Lindburg, 1999). This means that it is acceptable to perform actions which compromise the welfare of an individual as long as there is some larger overall benefit. One major proponent of this view has been Peter Singer (e.g. Singer, 1995), who considers the prevention of suffering (and, to a lesser extent, the promotion of pleasure) as the ultimate moral good. The values of the welfare view can be seen in the actions of organisations like the RSPCA, which operate to protect the welfare of animals; focusing on ensuring their humane use rather than eliminating their use.

Due to these differences in the underlying moral framework, the welfare view then approaches the subject of management euthanasia quite differently than the rights view. In general, defenders of the practice, particularly those within zoological institutions, have been situated within the welfare view. Under the traditional welfare view (one I will argue against in the section that follows), management euthanasia is typically not considered to be problematic, as it does not harm the welfare of the animal. This is because, as long as the animal has previously been well-cared for, with predominately positive experiences, and the euthanasia is competently performed, the animal will not experience any suffering and thus welfare is not compromised. As, under the welfare perspective, the quality of subjective experience is what matters, a painless death does not create a welfare problem.

This perspective, that death is not a welfare issue, is relatively common within the animal welfare view (Jensen, 2017; Yeates, 2010). Jensen (2017) considers this to be because 'welfare' is typically considered to mean 'welfare at a time'. Yeates (2010) attributes it as a side-effect of a subjective conception of welfare, under which it is assumed that those things which matter to welfare are only those which an animal can subjectively experience. "In order for something to be good or bad for an animal, it has to be *experienced* as good or bad; and in order for this to be possible, the animal must be alive and conscious at the time … It follows that since there is no experience of the state of being dead, the concept of welfare does not apply to that state" (Jensen, 2017, p. 616). As death is necessarily the absence of such experiences, it can neither harm nor benefit the animal, and the only concerns for welfare are the circumstances surrounding the death. In cases of management euthanasia, so long as the sum of experiences for the animal's life has been positive (presuming it has been well cared for) and the act of euthanasia itself caused no suffering, the animal has had good welfare. Several writers defending the practice of management euthanasia use this line of argument: "culled individuals do not experience reduced welfare" (Powell & Ardaiolo, 2016, p. 197); "culled animals do not

experience reduced welfare compared to living animals, unless they are culled inhumanely" (Penfold et al., 2014, p. 25). As advocates of this view think that management euthanasia does not compromise welfare, it can therefore be an acceptable tool to use when necessary; and given the underlying utilitarian framework, it can be justified whenever there is an overall benefit arising from its use (though these calculations must also take into account other possible negative effects, such as harm to public opinion regarding the zoo, or to the personal feelings of the keepers involved, as will be discussed further on).

Under a traditional welfare view then, the answer to our question 'when (if ever) is management euthanasia permissible, and under what conditions?' would be, 'when there is some overall benefit to be had'. As the harm to the euthanized animal is minimal, in this view this would mean that (other factors taken into account), in principle the practice may be permissible for even quite small benefits. However, it is still the case that many within the welfare view are also opposed to the practice. Although this is sometimes attributed to our own feelings and attachments to the animals (Lacy, 1995), in the next section I will describe another possibility through an extension to the usual welfare view, which considers a greater harm to the animal from management euthanasia. This then changes the types of conditions under which we may find it acceptable; and I argue that we would need more compelling reasons in justification than previously thought, to outweigh the welfare harm.

3.1.3. Extending the welfare view

The view that death is not a welfare issue, is not satisfactory to many "who consider animal welfare to be an appropriate basis for decision-making in animal ethics but also consider that an animal's death is ethically significant" (Yeates, 2010, p. 229). Even under the welfare view, there is often a sense that killing is not harmless to welfare. "There are many who work at least partly within the animal welfare tradition who may consider the killing of a healthy animal to seem, at least in some cases, morally undesirable" (Yeates, 2010, p. 230). This seems true in the management euthanasia debate – although defenders of the practice are typically operating under the welfare view, there are those within the view who opposed it (e.g. Maple, 2014). To understand why, we need to expand the welfare view to allow for the welfare harm of death; which can be done through understanding that welfare is more than just 'welfare-at-a-time' as described above, but should also include exclusion of positive states (Yeates, 2010) and lifetime welfare (Jensen, 2017), as I argued for in Chapter Two.

As described above, the aim of the welfare position is to maximise the welfare of captive animals, where welfare is understood as the subjective experience of an animal over its lifetime.

This tends to focus on prevention of suffering, with the apparent underlying assumption that the maximal state of welfare is one in which there is no suffering. However, if we consider the nature of subjective experience, it should be clear that this can vary along both sides of the spectrum – into negative experience (suffering) and positive experience (pleasure) – and can change in magnitude along both these lines. Maximising welfare then does not just involve the prevention of suffering, but also the promotion of positive experiences. When we are considering whether the welfare of an animal has been negatively compromised, it is not enough to simply look at whether it has dropped below a neutral baseline level into the negative. As argued for the 'torpid tiger' example in Chapter Two, we should instead set our comparative baseline higher – at some optimal positive state of flourishing, and look at how we may be failing to reach this. Under this view, we can compromise welfare not just through the infliction of suffering, but through the failure to provide positive opportunities. Something like this view is discussed by Regan (1983) who differentiates between 'harms' to welfare – those actions which directly create negative experiences – and 'deprivations' – those actions (or lack of) which deprive an animal of opportunity for positive experiences. Here then, we can see a welfare problem arising for animals that are given insufficient opportunities for achieving positive welfare states, even when they do not experience any suffering.

It is important to keep in mind the second part of the welfare definition discussed earlier, that is subjective experience *over a lifetime*. This means welfare can be measured as something like the sum total of experiences over time (e.g. Phillips, 2009, pp. 8–9), rather than some sort of average of overall quality of experience[20]. All other things being equal, an animal with a longer lifespan is likely to have better welfare than one with a shorter life, as this life will contain more positive experiences. Jensen (2017) argues that something can be bad for the welfare of an animal if it makes the animal worse off than it would have been under some other possible scenario. In this case, longevity is an important consideration as a longer life will typically be better than a shorter one – "to say that death is bad for this person means that she would have had a better life, had she continued to live rather than die at this time" (Jensen, 2017, p. 617). In fact, he argues that really all our welfare assessments of animals are based on comparisons with some other possible state – we want to know whether an animal is better or worse off in their current state than they may be under some proposed intervention, and this can apply also to premature death. The early termination of life is a harm to welfare through

[20] For the latter view, see e.g. Penfold et al. (2014): "Welfare reflects a combination of positive and negative mental, physical, and emotional states that are co-dependent and vary over time. Longevity of an animal does not translate into "better welfare", as welfare is not a cumulative characteristic for the individual" (2014, p. 25).

the *removal of future opportunities for positive welfare experiences*. This perspective also explains why medical euthanasia is typically not considered a welfare problem – because for these animals, the future is not expected to hold many, if any, opportunities for positive experiences and instead is expected to be overwhelmingly negative. In this case, we are actually benefitting welfare through the early termination of life and the reduction of negative experiences.

Yeates (2010) points out that most of our welfare evaluations involve a comparison between different states (for example, in terms of their duration and severity) and that any state can only be considered good or bad in comparison to another. Since we routinely compare states when one or the other is not present, it is no problem to compare the presence of states with their absence (such as through death). Indeed, if we are not able to make such comparisons, it is difficult to make sense of the practice of humane euthanasia for veterinary reasons – this involves a comparison between the presence of the negative states that a sick or dying animal may experience, and the absence of these states that will occur with its death. "Non-existence means that all and any states of the animal that could otherwise have been present are actually absent … the overall welfare of an existent animal can therefore be compared to the absence of that overall state, i.e., to its non-existence" (Yeates 2010, 236). These calculations can apply to the value of death in cases where we expect the potential future of an animal to be largely positive or negative – "if the presence of a life would have positive value overall then death is a harm; if it would have negative value overall then death is a benefit" (Yeates 2010, 237). This then gives us a "a prima facie responsibility for an agent not to kill an animal that would otherwise have a life worth living" (Yeates 2010, 239).

Death then can be considered a welfare issue "insofar as it leads to the exclusion of relevant positive states" (Yeates 2010, 229) or "when the animal is deprived of good experiences or other good things in life" (Jensen 2017, 618). This gives us a position within the welfare view which speaks against the practice of management euthanasia. Where an animal might otherwise have been expected to have a life consisting of largely positive states, management euthanasia harms the welfare of that animal by depriving it of those future states. As there is some (potentially quite large) harm to the welfare of the animal through euthanasia in these cases, it will require a quite large benefit in trade-off to make the practice permissible. Although the answer to our question 'when (if ever) is management euthanasia permissible, and under what conditions?' may remain 'when there is some overall benefit to be had', under the extended welfare view, the magnitude of that benefit must be much larger to justify the practice.

3.1.4. Considerations and trade-offs

I have shown that even under a welfare view, management euthanasia may be more problematic than has been thought, as it compromises welfare through the deprivation of possible future positive experiences. All other things being equal, an animal with a long life will have higher welfare than an animal with a shorter life and the practice of management euthanasia then creates a welfare deprivation, which will require stronger justification in order to be acceptable. In answering the question of when management euthanasia is permissible, and under what conditions, it is necessary to also consider the other positive and negative outcomes that may arise from the practice, either in a specific instance, or in general. "In addition to the complexity of the welfare evaluation on its own, a decision whether or not to kill an animal will be complicated further by other external factors, such as productivity, economics, and health of other animals … these other factors might entail that, in some cases, the killing of an animal might be justified" (Yeates 2010, 238). We will then be able to do some sort of weighing of different factors to determine the best course of action in each case. In this section, I will look at some of the possible factors that may need to be considered in making such calculations – both under a monist perspective (considering only a single value – that of welfare maximization), and a pluralist perspective (considering multiple competing values). It is not my intention here to try and provide some sort of strict weighting that could be used in making these calculations, but only to draw attention to the potential considerations that must be kept in mind when making such decisions. In the end, decisions will depend on the specific circumstances and values for a particular situation or institution.

3.1.4.1. Monist (welfare)

Norton (1995) draws a distinction between a monist system of value, under which we only consider one value, and a pluralist system, under which we must adjudicate between multiple competing values. Under both of these systems, there are potential circumstances which create additional considerations in deliberating on management euthanasia. I will look first at a monist view, under which maximization of welfare is the primary concern. We have already considered how the act of management euthanasia can decrease the welfare of the euthanized animal, in depriving it of the opportunity for future pleasures. Here, I will look at some potential welfare harms that may occur as a result of failing to implement management euthanasia, that may then weigh in favour of the practice under these conditions. These relate both to the expected welfare of the euthanized animal, and welfare of other animals in the collection. There is certainly reason to think that there will be situations in which the expected welfare of surplus

animals over time will not be overwhelmingly positive. Additionally, there will be situations in which the welfare of other animals will be negatively affected to such a degree that the overall welfare of the group may be decreased through prevention of management euthanasia.

As described earlier, management euthanasia is usually considered because of a lack of viable alternatives. That is, that the other possible options for the animal are not good ones, and may not lead to higher welfare in the long term. Additionally, those in the position of making management decisions about specific animals are rarely those who control overall resource availability, further constraining available options. Consider keeping the animal within the institution. Leaving the animals in their existing enclosures, with the current social group, is often not beneficial. This can lead to overcrowding, potential disease transmission and social disturbance, such as fights between individuals. This was the case for Marius the giraffe, who was reportedly experiencing aggression from his father (Parker 2017). All of these are going to cause suffering and lower welfare. The most common alternative option is off-display housing, keeping the animals in another enclosure somewhere away from the public areas of the zoo. These enclosures are usually smaller and less well-furnished than the display exhibits, because resources tend to focus on those areas used by the public. They are also not often designed with particular species in mind, as they may be used for a variety of animals as needs arise, and so will not meet the needs of the animals as well as specialised display exhibits do. Animals housed in these enclosures will usually have decreased human contact, which is something they often find positive, and may also be housed away from conspecifics. All these factors can cause distress and decrease welfare.

The options for rehoming outside the zoo are also likely to be inadequate. As described, sending the animals to other accredited institutions is rarely possible, as they will have usually already bred or acquired animals to meet their capacity. Instead, dispersal options are often limited to smaller non-accredited institutions, where welfare standards cannot be guaranteed. The lack of funding in such institutions means they are unlikely to meet the needs of the animals as well as they should. Release to the wild is also unlikely to be successful, and as described earlier, is highly likely to compromise the welfare of the animals involved.

These points show that there are many cases in which management euthanasia may be considered the best option for the animal, as rather than removing opportunities for future pleasure, it is removing the likelihood of future suffering. In these cases, it begins to seem closer to a case of preference-respecting euthanasia, where were the animal given the ability to voice a preference, it may prefer death to a life of ongoing deprivation. In addition to this, we should also consider welfare effects on other animals in the zoo. There will be cases where

keeping the animal alive may cause a decrease in welfare in many other animals, decreasing overall welfare of the group. Within a large institution, although maximising the welfare of each animal is important, there needs to be a balance in which the total welfare of the animals across the zoo is as high as possible[21]. There will be some cases where management euthanasia is the option that will best achieve this.

There are two ways in which the presence of surplus animals is likely to decrease the welfare of other animals – the first is directly, through their immediate presence, and the second is indirectly, through diverting resources. As described above, holding too many animals in an enclosure can result in crowding, illness and social stresses such as aggression, which can affect all the animals in the group. There may also be a loss of breeding opportunities for the other animals. Zoos will breed to fill the space available, and surplus animals taking that space will restrict the breeding of others. Breeding creates many opportunities for positive welfare, in courtship, mating and parent-offspring bonds; opportunities that will be lost if breeding is prevented. "Breeding is a fundamental motivator of all animals. The life cycle of breeding, birthing and raising young is an engaging and satisfying behaviour for many animals" (Gray 2017, 80-81). The prevention of these behaviours could be described as an "arguable unkindness" (Parker 2017). However, there are those who doubt the real benefit of these opportunities: "The absence of breeding opportunity does not meet the definition of suffering or poor welfare. In nature, socially dominant animals do most of the breeding. Subordinate males are often found at the periphery of female herds led by mature matriarchs. This is true for hippos, elephants, buffalos and many other species. Do weaker males enjoy a lesser quality of life in nature? The case can be made that the mere opportunity to compete for access to females is a life-enriching experience" (Maple 2014, para. 6). These animals will typically take the chance to breed when available, and may even work for it, but this does not necessarily mean they suffer in the absence of it. More work may be needed to quantify the level of welfare cost experienced by animals prevented from breeding. In addition to this, maintaining additional surplus animals will use resources that could be used for increasing the welfare of animals more central to the larger aims of the collection. Resources such as money, and keeper time, can be used to improve the exhibits and husbandry of other animals. If the resulting resource deficit is large, it can result in a decrease in overall welfare. It may sometimes be the

[21] This is not necessarily the only consideration in welfare calculations – for example we might want to add something like: no animal should have their welfare drop below a certain baseline even if it maximises welfare overall.

case that the euthanasia of a single animal may be a means of increasing overall welfare of animals across the institution.

3.1.4.2. Pluralist (other values)

Even if all we value is animal welfare, I have shown that there may be reasons in favour of management euthanasia in particular cases. However, there are also many other values that may come into play when making decisions of this type. Maximising the welfare of the animals in their care is a huge part of the aim of a zoological institution. However, there are also many other things a zoo may wish to achieve, which will not always be in line with absolute maximization of welfare. "Moral pluralism is the view that we value many things in different ways, and that these differing values are sometimes in conflict. Further, these values may be incommensurate, so that they cannot be weighed in a common metric" (Norton 1995, 104). We may often need to make decisions within complex webs of competing values.

The first and probably most important value active in zoos is that of conservation and breeding programs, including the conservation value of public education and connection with wildlife (Lukas & Ross, 2005; Pearson et al., 2014; Powell & Bullock, 2014). In general, zoos no longer take animals from the wild and so must breed genetically compatible animals with high precision in order to maintain a healthy gene pool in captivity and ensure the future of the captive populations. The existence of surplus animals can get in the way of this, as described above, when they prevent ongoing breeding. Spaces that are allocated to housing surplus animals will be directly detrimental to the breeding program. Additionally, where these programs are necessary for direct conservation, such as release, the problem is of even greater importance. If the existence of a surplus animal interferes with an effective conservation program, this could be a strong consideration in favour of management euthanasia.

This clash of values, between the welfare of individual animals and the preservation of species or environments, is found throughout conservation biology. (Norton 1995). On the one hand, we value individual animals and seek to maximise their welfare. On the other, we value flourishing ecosystems and the continued existence of species. Often, concentrating on one of these ends requires sacrifice in the other. Norton (1995) argues that our responsibilities will differ depending on context. In the wild context, outside of human interference, the value of wildness and ecosystem preservation is dominant, and we are willing to compromise individual welfare to achieve this. In the domestic context, where we have taken animals into our care, our responsibilities towards them individually become more important. The issue with zoos is they seem to straddle both contexts – the animals have certainly been taken into human care,

with the attendant responsibilities, but also are there for the purposes of supporting conservation of their wild conspecifics, with some sacrifices possibly necessary to achieve this goal. There is unlikely to be a simple answer as to which of these values should take precedence in the zoo context, as it will depend on many other factors, such as the level of threat the wild population is under and the type of action the zoo is taking to assist (e.g. breeding for release, fundraising, public education). However, there will be some cases where it seems the demands of conservation will outweigh the welfare deprivations of management euthanasia.

Another important goal of zoos is public education and engagement. This is in service of conservation objectives, but in modern zoos, raising awareness on the plight of endangered species, and leading action to help preserve them, are probably more important than direct conservation action through breeding programs. Effective global conservation relies on the concern and action of the wider public, and zoos have a unique role in inspiring care for the natural world. In this case, it seems likely that management euthanasia will most often harm this outcome. Part of the effectiveness of such an approach is public engagement with individual animals; developing an emotional bond with them and then transferring that care to conservation efforts. The rising popularity of personal animal encounters in zoos seems to attest to this. Management euthanasia can harm that bond, decreasing perceived value of individual animals in favour of the group. There are different potential roles for zoo animals - "An animal can be a city's shared pet, or it can be a quasi-agricultural team member whose work is to be seen and to breed and, perhaps, to die young." (Parker 2017, para 26). Although some have advocated for this focus on groups, it does not seem that the public engage in the same way with large groups. If management euthanasia harms public engagement with zoo animals and thus harms conservation goals, this is a reason against its use. The huge public outcry against Copenhagen Zoo, including death threats (Parker 2017) seems to support this. Similarly, when management euthanasia causes public outcry, these harms to zoo reputation can decrease attendance, which as well as decreasing opportunities for education, will further decrease available resources for remaining animals. Although some of the effects of public impact can be reduced through education about zoo population management and the reasons behind decisions to euthanize, this is unlikely to improve problems with individual engagement, especially with children, who are often considered the most important target audience.

There is the problem of managing the conflicting narratives of zoos as caring sanctuaries for animals versus the clinical nature of euthanasia and dissection as presented at Copenhagen. Zoo animal welfare expert Terry Maple described the Marius incident as "a huge public-relations blunder" (quoted in Parker 2017, para. 71) and "counterintuitive to the mission of the

zoo community globally" (Maple 2014, para. 3); claiming that the negative effects this had on the zoo visitors and supporters served to undermine the good work zoos are doing in conservation, and overall zoo credibility, not just for Copenhagen but around the world. "It seems as though the public (and especially critics of zoos) do not judge each zoo as an individual institution, but more as a part of a larger zoo community. A crisis kicked off by one zoo could affect other zoos as well" (Schäfer 2015, 179-80). If zoos are to continue with the practice, it seems important that they are able to communicate it in such a way as to keep the public on side.

One final value to consider is that of human emotions. Management euthanasia can be difficult for people, who grieve the loss of particular animals. This is particularly true for keepers, whose job relies on their bond with their animals and their commitment to maintaining and improving the lives of these animals. Euthanasia of favoured animals is going to be upsetting to keepers and this emotional distress, as well as being in itself a negative, can also prevent them from bonding and doing their jobs as well in the future. Lacy (1991, 1995) argues that these sentiments from keepers are actually currently the primary motivating factor against management euthanasia; that even where arguments based on animal rights or welfare are put forward, these are actually a justification to protect our own feelings. Powell & Ardaiolo (2016) surveyed keepers and managers on their reaction to particular management euthanasia scenarios and found that keepers were more likely than managers to disapprove of management euthanasia, particularly with animals they are more prone to bond with, such as primates. Maple (2014) points out that "the bond between zoo animals, zoo managers and zoo patrons is based on mutual emotional ties between humans and animals that often originate in childhood. Zoo animals are valuable ambassadors between the wild and human world rather than a commodity displayed for the amusement of humans" (para. 11). This emotional impact is also an important consideration that may weigh against euthanasia in many cases.

3.1.5. Conclusion

Management euthanasia, the practice of euthanizing healthy surplus animals, is controversial. Traditionally, those arguments against the practice have come from the animal rights camp, who see it as a violation of the rights of the animal involved. Arguments in favour come from the animal welfare perspective, who argue that as the animal does not suffer, there is no harm in the practice and it is justified by its potential benefits. I have argued that an expansion of the welfare view, encompassing longevity and opportunities for positive welfare, give stronger considerations against management euthanasia, which then require greater

benefits to justify its use. I have also presented some of the other considerations that may play a role in making decisions about management euthanasia; from both the perspective of maximising welfare, and in consideration of other values, such as conservation. It has not been my aim here to make a definitive stand about the acceptability of management euthanasia, but to point out that in making such decisions, the welfare of the animal should be given more weight than is perhaps usually considered, when only suffering is taken into account. In the end, each institution will need to make decisions for itself, based on the overall context of the zoo and the particular circumstances surrounding each individual animal (Lacy 1991).

3.2. De-extinction and animal welfare[22]

3.2.1. Introduction

De-extinction is the process through which extinct species can be brought back into existence. There is some disagreement as to whether this actually results in the resurrection of a species or just something closely resembling the species (see discussion on this point in Section 3.2.4). For the welfare concerns discussed in this chapter, however, the distinction does not play a strong role. The process is usually undertaken with the aim to reintroduce species to the wild and restore ecosystems (Shapiro, 2017). It is controversial, with debates tending to focus on scientific viability, or the ethical issues accompanying such a project, with animal welfare concerns mentioned only briefly if at all. However, as pointed out by Kasperbauer (2017), there is good reason to think that the welfare of the animals involved will be poor. In this chapter, I will expand on the potential types of welfare harm that de-extinction programs can cause. In particular, when thinking of welfare subjectively, this gives increased reason to focus on the potential forms of psychological suffering. This welfare harm should be an important consideration when making decisions on de-extinction projects, and in the final part of the chapter I will look at the potential benefits of such projects and argue that in most cases they are insufficient to outweigh the potential welfare harm as it stands. With further development of the technology and careful selection of appropriate species as de-extinction candidates, these problems may be overcome.

There are three methods through which de-extinction can be achieved: selective breeding, cloning and genetic engineering (Cohen, 2014; Shapiro, 2015). Each of these has different benefits and drawbacks and each is useful for different cases. The first of these methods is

[22] A version of this chapter section has been published as: Browning, H. (2018b). Won't somebody please think of the mammoths? De-extinction and animal welfare. *Journal of Agricultural and Environmental Ethics*. **31**: 785-803.

selective breeding, or back-breeding. In this process, current relatives are selectively bred for those characteristics that defined the extinct species. For example, in order to re-create mammoths, elephants could be successively bred for their larger and hairier variations, until something closely resembling a mammoth is created. There are currently a few programs using back-breeding to attempt to recreate extinct species - projects attempting to bring back the quagga are selectively breeding zebras (Cohen, 2014), and to bring back Auroch through selective breeding of modern-day cattle (Shapiro, 2017). The process is limited by the availability of sufficiently similar relatives that are still capable of expressing the desired traits.

The second method is somatic cell nuclear transfer (SCNT), or cloning. This requires the nucleus of a cell taken from a recovered member of the species to be implanted into an egg cell of a related surrogate species. This creates a zygote genetically identical to the donor animal of the target extinct species (barring mitochondrial DNA). The zygote is then gestated and birthed by the surrogate animal. A famous example of the use of this process was in the creation of 'Dolly' the sheep. This has also been attempted for de-extinction purposes, in the cloning of the Pyrenean ibex, or Bucardo, where a clone was made of the last living individual; though this clone did not survive long after birth (Cohen, 2014). There has reportedly been some success in using this technique for creating new embryos of the extinct gastric-brooding frog (Cohen, 2014). This process is only possible where entire cells of the extinct species are available (which is only the case for very recent extinctions) and where appropriate surrogate species can be identified (Shapiro, 2017).

The third de-extinction method is genetic engineering. Here, DNA is recovered from preserved specimens of the extinct species. As it is almost never entirely intact, it is spliced with the DNA of a related species to create the closest possible genetic match to the original target. The spliced genetic material is then used to create a zygote to be gestated and birthed by an extant surrogate animal. Most current work on de-extincting mammoths is using this method, as the remaining mammoth DNA exists only in fragments. It is also the primary method in use for de-extinction of the passenger pigeon (Cohen, 2014). This is what Shapiro (2017) considers the "most likely route to de-extinction" (2017, p. 4) as it only requires fragmentary DNA from the target species, which can then be expanded into a whole genome. It is limited primarily through availability of preserved DNA, which rules out long-extinct species, and like the cloning techniques, requires the use of closely related living species both for use of their DNA for gene editing, and as surrogate mothers (Shapiro, 2017).

All of these methods for de-extinction are being used in current projects on various species, and each presents potential animal welfare problems, which I will detail in Section 3.2.2. In

Section 3.2.3 I will argue for the moral status of these animals within our deliberations. This is not merely hypothetical – de-extinction projects are happening now, and the welfare concerns for the animals thus created should be the focus of attention when evaluating these projects.

Before turning to the issues de-extinction raises for animal welfare, I will briefly note some of the other issues that have been raised for de-extinction projects. In the past few years, de-extinction has become the centre of increasing scientific and philosophical focus, with a number of books and papers published on the scientific feasibility of and ethical issues surrounding such projects. The first of these is the question of whether or not the animals created through these methods count as members of the original species, due to differences in genotype, phenotype, history and development (Shapiro, 2017), and whether this matters. Another set of issues relate to scientific feasibility; whether the breeding and cloning methods can even work, and which animals are most likely to be successful if reintroduced (K. E. Jones, 2014). Finally, there are the other ethical concerns of de-extinction projects. For example, whether de-extinction programs harm other forms of conservation, as the resources spent on de-extinction programs could result in decreased support for conservation of extant species (Bennett et al., 2017) and the discussion of de-extinction may "give the impression that extinction is reversible" and will therefore "diminish the gravity of the human annihilation of species" (Campagna, Guevara, & Le Boeuf, 2017, p. 48). Another criticism of the process is that it is 'unnatural' - both in terms of the technologies that are used to recreate the species, and in terms of the nature of the species thus created (C. Mason, 2017). All these questions have already been well-explored and will not be re-examined further here. Instead I will now turn to the issues of animal welfare arising from de-extinction projects, which in contrast have not yet been given much attention in the literature.

3.2.2. Welfare issues

Until recently, what has been absent in discussions of de-extinction is an exploration of the issues relating to the welfare of the animals created through these projects. Although mentioned briefly in many of the papers discussing ethical issues, animal welfare concerns are typically given only a few lines. These usually indicate that these issues are important but should not be difficult to work out, as they are the same sorts of issues that show up in other projects involving scientific research and species reintroduction. For example, Cohen (2014) brings up harm to animals as a potential source of negative utility in considerations of de-extinction projects but concludes "there is no reason to think de-extinction will cause a large animal welfare problem" (2014, p. 175). Sandler (2014) considers animal welfare concerns as a potential ethical reason

against de-extinction projects but concludes "compared with the number of animals already used in research … conservation cloning does not pose a special or very large animal welfare problem" (2014, p. 358). Although he acknowledges that research should take care to minimise suffering caused, he thinks "the animal welfare concerns it raises do not significantly differentiate it from many other research and conservation practices involving animals" (2014, p. 358). Rohwer & Marris (2018) point out that "creating a *mammoth* is morally *permissible* ... provided that suffering is minimal" (2018, p. 2, italics in original), and go on to describe some of the concerns with cloning, surrogacy and rearing; though again, follow Sandler in dismissing these as not much different from those in other similar conservation practices. Greely (2017) describes the potential welfare problems associated with cloning technology but considers that "the risks of de-extinction are not substantively different from those associated with gene editing" (2017, p. 34). Friese & Marris (2014) briefly describe some of the welfare concerns in the creation and rearing of de-extinct animals, and argue that "questions regarding animal care need to be understood as a crucial part of de-extinction experimentation, rather than downstream concerns" (2014, p. 2), however they conclude that these issues can be addressed through "a social science approach based upon the current realities of cloning, genetic engineering, back breeding, and species preservation today" (2014, p. 3).

Only Kasperbauer (2017) has really emphasised the importance of animal welfare considerations in making decisions about de-extinction programs, labelling it as "the most critical challenge for de-extinction" (2017, p. 1). He argues that "the current state of de-extinction technology provides good reasons to think the lives of de-extinct animals will indeed be full of suffering" (2017, p. 6) due to problems with cloning technology and reintroduction (though only the latter are applicable to de-extinction through back-breeding), and briefly outlines some of the ways in which these potential harms could come about. Here I will examine in detail exactly how and why these situations are likely to be harmful to animal welfare.

Although it is true that many of the welfare issues are continuous with those affecting other areas of animal research and conservation, I argue that de-extinction creates special issues for animal welfare that need to be considered when evaluating such projects. Though the number of animals used in de-extinction projects is far smaller than in other human uses of animals we currently permit, this does not give us the justification to carry out any harmful actions within this range, as will be discussed at the end of Section 3.2.2.1. Welfare issues can affect the de-extinct animals, other animals used in the process (e.g. surrogate mothers) and the wild animals that will be impacted through reintroduction (Cohen, 2014). In particular, there are welfare

issues surrounding the cloning procedures when these are used, and in the process of captive rearing and reintroduction, beyond those usually faced by zoos or conservation bodies doing this sort of work. This is in large part due to the lack of knowledge about these species and their requirements.

3.2.2.1. Welfare issues with cloning

In the first instance, there are issues surrounding the cloning technology used in both SCNT and genetic engineering. These will not be a problem for back-breeding projects, though this method will face a few problems of its own. So far, the use of cloning has been problematic for animal welfare, with cloned animals showing rapid aging, ongoing health problems and premature death. "Cloned animals suffer from impaired health, including placental abnormalities, foetal overgrowth, prolonged gestation, stillbirth, hypoxia, respiratory failure and circulatory problems, malfunctions in the urogenital tract, malformations in the liver and brain, immune dysfunction, anaemia, and bacterial and viral infections" (Gamborg, 2014, p. 6). Fiester (2005) outlines the different ways in which cloning procedures can have negative impacts on animal welfare – through the suffering inherent in the cloning procedure, gestational problems with surrogates, ongoing health of cloned animals and the future suffering cloned animals might endure through research, housing etc. These procedures are associated with miscarriage, stillbirth, early death, genetic abnormality and chronic disease. As the success rates (in terms of live birth) for even the most effective programs are only 5-12%, this creates a lot of excess donor procedures and surrogate pregnancies. As cloned foetuses show a higher than average birth weight, caesarean deliveries are also often necessary. Those offspring that are delivered alive show huge mortality rates, due to conditions such as developmental abnormalities and lung, heart and liver problems. The US Humane Society has advocated for a ban on these procedures due to the high incidence of welfare problems.

Take the famous Dolly, the first ever successfully cloned individual. Dolly was plagued with health problems, such as arthritis and lung disease, and died at six years, only around half the normal life span of a regular sheep of her kind (Williams, 2003). These sorts of problems only increase when using the technology to create and gestate extinct animals in close relatives rather than conspecifics, with low success rates and high levels of health problems and abnormalities in both the surrogates and foetuses in interspecific procedures (Sandler, 2014). A Pyrenean ibex cloned from the last individual, died of a lung defect within minutes of birth (Cohen, 2014).

Surrogacy can give rise to the problem of maternal-foetal incompatibility, which can be problematic to both the surrogate mother and the gestating foetus (Fiester, 2005). Similarly, there can be birthing complications when the target animal is larger or differently shaped than the surrogate, as would be the case with - for example - elephants carrying mammoth babies. Surgical delivery is the likely option in these cases, but surgery on an elephant is difficult and the chance of complications during surgery or recovery is high. There are also the chances of maternal rejection of the unusual offspring, creating potential social isolation. The lactation of the surrogate mother may not be appropriate for the offspring, creating nutritional and health problems. Shapiro (2017) points out that we need close relatives of the extinct species in order for the process to be successful, and these may not often be available; the less closely related the surrogate species, the higher the chances of problems arising.

Back-breeding can run into similar problems. If selecting for larger or somewhat different individuals within the population, we again have a risk of gestational complications (though lower than in the surrogate cases) and maternal rejection of unusual offspring. Back-breeding will also usually use a very small founder population, and so creates significant risk of inbreeding and the associated health issues (Shapiro, 2017).

Although these issues may be reduced with further research into the technology - looking for where the problems in development are occurring and repairing them - this further research will require the production of animals fated to suffer these physical and psychological problems. Some of these problems may be the same as those facing other animals created through these methods for research or agriculture, but some will be unique to de-extinction, particularly due to the requirement to use other species as surrogates. Additionally, the defence that these problems occur in other areas of science (the ‘accepted practice standard’) is not a strong one. The fact that one set of practices matches another provides no real justification if the first set of practices is also ethically problematic (Fiester, 2005), and the set of benefits and justifications for each will differ. Any project which uses the technology should thus be independently assessing the potential harms. These other applications are deemed acceptable in large part because of the perception of gains in other areas, so this response can only apply where there is sufficient justification of the benefits, to outweigh the potential suffering caused. This trade-off will be examined in Section 3.2.4.

3.2.2.2. Welfare issues with captive rearing

As discussed, there are potentially serious welfare problems with the use of cloning technologies as they currently stand. However, refinement of the procedures can possibly reduce or remove most of these problems over time (at least those involved with cloning itself – issues of maternal/foetal incompatibility seem potentially more serious). Of greater concern, and far less obviously surmountable, are the issues surrounding the rearing and release of de-extinct animals. Shapiro (2015) points out that "from an animal welfare perspective, the captive breeding stage is likely to be one of the most challenging steps of de-extinction" (2015, p. 195). These are in large part practical issues as to the feasibility of such practices, but as their failure harms the welfare of the animals under consideration, they are also strong welfare concerns. Again, these have largely been dismissed as the same issues that surround any breeding and reintroduction programs, something zoos have been addressing for many years (e.g. Beck, 1995), but this does not mean they do not need to be addressed independently for these projects. Additionally, there is good reason to think that de-extinction programs are going to have a unique set of challenges arising from lack of knowledge and lost ecological conditions, as well as a highly atypical social environment.

Breeding and rearing any animal in captivity requires a set of husbandry standards in order to succeed. These include detailing the recommended diet for the animal, appropriate housing conditions (such as temperature, shelter, access to water), social conditions, behavioural requirements, known health issues and how to treat them, among other information. For most captive animals, collection and collation of such information has taken decades, drawing on knowledge of the living conditions and habits of wild counterparts, or closely related species, and much trial-and-error on the animals in captivity – often resulting in poor health and short lifespans in the early members of captive populations. As an example, historically reindeer were notoriously difficult to keep in captivity, consistently suffering ill health and dying young. Eventually it was found that in the wild, their diet included a large amount of lichen, which provided essential nutrients (Steen, 1968). Addition of these to the captive diet fixed many of the problems previously encountered. Without the ability to check this in the wild population, this problem could not have been fixed and in the meantime would have led to ongoing suffering.

If it has been this difficult to create husbandry standards for the animals we have held in captivity for many decades, sometimes even centuries, with access to research on their wild relatives, it will be much harder to do so for animals for which we have no such information. In some cases, we might be able to use modern relatives as a starting-point: for example,

quagga are likely (though not certain) to have similar requirements to zebra. In other cases, the species may have gone extinct recently enough that we still have access to some relevant information. For example, the thylacine - which has been gone less than a century and was frequently held in captivity prior to this - is a species for which we are likely to still have ecological and husbandry data. For other species, this will be much more difficult. Take the de-extinction flagship, the mammoth. We have no good reason to believe their diet, habits or environment will at all closely resemble that of modern-day elephants. They are a vastly different species, which lived in a vastly different environment. Paleontological evidence is scarce (though we can go some way with information from preserved stomach contents and coproliths), and given the slow production rate of large animals like these, trial-and error, even if considered ethically acceptable for research purposes, is impractical. In pointing out that de-extinction would give us the ability to study and learn about extinct animals, Rohwer & Marris (2018) also demonstrate that there is a lot we don't know about species like mammoths – "how long do they nurse? What time of year do *mammoths* mate? How intelligent are *mammoths* as compared to elephants?" (2018, p. 7, italics in original). Seddon et al. (2014) list some of the types of knowledge we need for successful rearing and reintroduction – "knowledge of former distributions, social structure and behavior, diet, reproduction, parental care and growth, interspecific interactions, and biotic and abiotic habitat requirements is required" (2014, p. 143). Although they are somewhat confident that "valuable clues may be obtained from the biology and ecology of extant species that may be nearest living relatives or otherwise occupying a similar ecological niche" (2014, p. 143), for species that have gone extinct long before human observation and record-keeping, this is likely to be a much more difficult project than this suggests.

Even if we are able to determine what the appropriate conditions should be, there may also be large problems with providing them. Take the mammoth example again. If mammoths are anything like their elephant relatives, they live in large social groups of mixed age and sex. However, in the early stages of de-extinction projects all we will have are numerous juveniles. These may get some of their required social contact with elephant surrogates, but elephants are unlikely to have the required behavioural repertoire and social 'vocabulary' to match their mammoth companions. Provision of appropriate environmental features, and required dietary items could also prove intractable if these are no longer available.

3.2.2.3. Welfare issues with reintroduction[23]

Above, I have discussed the problems of housing the de-extinct animals in captivity. Just raising them to an age that they are suitable for release may prove to be impossible, and this is a huge welfare concern – the animals are likely to be malnourished and in poor health, with potential psychological and behavioural deficits. But in most of these programs, the animals are also destined to be released back to the wild (Shapiro, 2017)[24], and as such require rigorous behavioural conditioning for this process. Seddon et al. (2014) argue that "de-extinction is a conservation translocation issue" (2014, p. 140) and thus will require the same considerations and hold the same risks. Jørgensen (2013) summarises the IUCN guidelines for reintroduction – "background studies to allow identification of the species' habitat requirements, identification of lessons learned from prior reintroduction projects of similar species, evaluation of potential sites within the former range of the species, selection of appropriately diverse genetic stock, and an assessment of the socioeconomic context of the project" (2013, p. 719). It should be immediately clear that for de-extinct species, we will not have the means of obtaining most of these answers, as indicated in the previous section. Harrington et al. (2013) provide a detailed flow-chart of welfare considerations in reintroduction projects, demonstrating the huge number of considerations at play and knowledge required for such projects to go ahead successfully. A lot of these issues will be particularly salient for de-extinction projects, as the required data will not be available. Without this information, we have little chance of successful reintroductions, and this will lead to decreased welfare for reintroduced animals.

There are a number of potential animal welfare issues present in all reintroduction projects. These include mortality, disease, post-release stress and human conflict. Almost a quarter of reintroductions have mortality over 50%, due mainly to predation, traffic and other human effects, and disease and starvation (Harrington et al., 2013). Beck (1995) found that only 11% of the reintroduction programs he studied were successful in creating self-sustaining wild populations. Our lack of knowledge about de-extinct species and their ecology means these numbers are likely to be even worse for de-extinction projects. Harrington et al. (2013) looked at ways of measuring and monitoring welfare of released animals (e.g. health, condition,

[23] Though the use of the term 'reintroduction' here and throughout the chapter implies the released animals would be of the same species as the extinct ones, this is not meant necessarily as an endorsement of this view, but is rather following the common usage in the literature. The welfare issues discussed are the same regardless of whether or not the releases can be classified as reintroductions.

[24] Not all de-extinction projects aim at releasing animals back to the wild (Sandler 2014), and those which simply aim to create animals to hold in captivity for research or exhibition will not face this set of welfare problems, though the others, particularly those in Section 3.2.2.2, will still apply.

behaviour) and the types of supportive actions that may be used to improve welfare, such as health screening, pre-conditioning and provision of artificial food and shelter. These again all require knowledge of the species in order to succeed. Harrington et al. (2013) also noted that captive-bred animals were more likely to fail to cope after release than wild-caught animals. This is a problem for de-extinction as all animals in these projects will be captive-bred.

There are also concerns for the welfare of other wild animals that will come into contact with the released species, through habitat alteration, resource competition, predation or aggression (Seddon et al., 2014). This may be managed to some degree by careful choice of species for de-extinction (for example, large slow-breeding animals that can be more easily tracked) (Seddon et al., 2014), but lack of knowledge about the species will still make potential impacts hard to predict.

Even the preparatory training procedures for release can also be detrimental to welfare, as has been discussed for captive breeding and release programs of extant species (Beck, 1995). They require training and conditioning animals to tolerate the reduced conditions of the wild – lack of shelter, exposure to parasites, lack of food, avoidance of predators, social interactions with unfamiliar conspecifics etc. All of these conditions require a reduction in welfare as compared to the captive environment. Beck (1995) looks at some ways of improving post-release success and welfare – presence of a wild-born 'teacher' for natural behaviours, post-release support and monitoring, and careful study of which of the stressful pre-release conditions are actually required to help survival and flourishing. All of these, save post-release monitoring, will be extremely difficult to provide in a de-extinction context. For example, depending on our knowledge of the species, we are unlikely to know what sorts of wild conditions these animals need to be acclimatised to and will not have access to suitable model animals to act as 'teachers'.

In particular, training the necessary behavioural repertoire for a species that has never been observed in the wild seems a potentially insurmountable task. "It is unclear how emergent social behaviours would survive the de-extinction process" (K. E. Jones, 2014, p. 21). The ways in which these animals interact with their environment, find and extract food, find and make shelter, and interact with one another and other wild animals, are all unknown. Turner (2017) points out that most of the candidate species for de-extinction programs are large, charismatic vertebrates, such as mammoths or passenger pigeons, and these are precisely the sorts of animals for which the concerns are likely to be most pronounced, due to their behavioural complexity. At best, we may be able to perform 'soft release', in which animals

are monitored and cared for as they adapt to life in the wild, but this will still be a time of stress; and would require substantial additional resources.

Getting the preparation wrong before releasing an animal can have disastrous consequences, as has been seen in countless reintroduction projects performed with animals that have wild counterparts to study. In fact, the high rate of failure of such projects has led to them generally falling out of favour as conservation initiatives. Where animals cannot be properly prepared, when released they will suffer and are likely to die, with no ecological benefit. This is, of course, a huge welfare concern. The IUCN regulations for reintroductions state "the welfare of animals for release is of paramount concern through all these stages" (IUCN, 1998, p. 9). Reintroduction programs inevitably cause suffering, even for extant and well-studied species, and these problems will be worse for species with less suitable habitat and about which we have much less information. There may be some species for which these concerns are less strong, such as thylacines: more recently extinct, which have plenty of suitable habitat and some recorded knowledge. At the very least, it will require very careful assessment of potential candidates, to minimise these concerns (Seddon et al., 2014).

3.2.2.4. Measuring welfare

A final concern is that we don't have any good way of measuring what sorts of welfare harms are occurring in these projects. As I will describe in the next section, it is possible that the welfare harms described could be offset by benefits in other areas. But even if we are able to develop a framework that allows us to determine how much welfare harm is acceptable for gains in other areas, it is not at all clear that we can get a strong sense of the level of harm that is occurring. Making a decision about an action based on its harms and benefits requires at least a basic approximation of the degree of these harms and benefits, and this may be extremely difficult for the welfare of de-extinct species. Measurement of welfare requires using physiological and behavioural indicators that are usually specific to the species, and calibrated through testing of other individuals. Even where the indicators themselves may be similar across a large taxonomic range (e.g. cortisol measurements), the levels of significance (i.e. what measurements indicate problematic welfare) will not be. Our lack of knowledge of the normal behavioural and physiological parameters for these animals mean that we can make only very rough (and potentially anthropomorphised) judgements as to the welfare of the animals we are creating. As I will discuss in Chapter Six, there is the potential to use similarities between closely related species, but this is quite limited. There is also not a large pool of individuals that we can test to develop such indicators. Use of welfare indicators from other

species may suffice, but may also be greatly misleading if the species differ in the particular behaviours or physiological markers used. This may then have a large impact on the trade-off matrices we are considering.

3.2.3. The moral status of de-extinct animals

It is clear that there are many strong potential concerns for the welfare of the animals used in and created by de-extinction projects. Perhaps, though, we might think that these are not problematic as the animals do not fall into the right category to require our moral concern. The degree to which we should be concerned about the welfare of reintroduced de-extinct animals mirrors the discussion about de-domestication – the 'rewilding' of domesticated animals (see e.g. Gamborg, Gremmen, Christiansen, & Sandøe, 2010). There is a tension here about whether the animals should be considered as wild animals, or domestic animals, as these categories carry with them different ethical and legal implications. Domestic animals tend to be considered at the level of the individual animal, with welfare considerations in the forefront, while for wildlife the consideration is at the level of species or population (Gamborg et al., 2010) and it is generally considered acceptable to compromise animal welfare somewhat if there is an overall species-level or conservation benefit. Norton (1995) argues that wild animals, for the most part, do not need to fall within the human moral sphere, and that in fact because we value their wildness, we choose not to interfere in their lives: "It is not this *content* of animal experience but the *context* in which we encounter it that determines the strength and type of our obligations" (1995, p. 106, italics in original). The level of our interference in the lives of animals determines our responsibilities towards them. Captive-bred exotic animals, neither wild nor domestic, fall somewhere between these boundaries.

A difference between usual considerations of management of wild animals and of de-extinct animals is that we are not just dealing with animals as 'moral patients'; the additional fact that we have *created* them places on us extra duties of care (Cohen, 2014), similar to the responsibilities we have towards pets, or zoo animals. Gamborg et al. (2010) also stress the difference between animals which humans have been directly involved in creating or rearing, for which we should assume responsibility, as opposed to those we have not: "Because humans are responsible for the very existence of domestic animals … and because the latter often render the relevant animals dependent and vulnerable in ways wild animals are not" (2010, p. 72). As de-extinct animals are created by us, often for our own ends, and spend at least the early part of their life in our care, their welfare should be our concern.

3.2.4. Weighing up potential benefits

I have outlined some of the ways in which de-extinction efforts are likely to be harmful to the welfare of the animals involved. These are big problems, in some cases possibly insurmountable, and it is almost inevitable that these programs will result in animal suffering. Even if there are some positive experiences in the lives of the de-extinct animals, it seems likely that these will be far outweighed by the physical and psychological problems described above and most de-extinct animals would not have what would be considered 'lives worth living'. However, this does not necessarily mean we should not engage in such programs at all, as there are other potential benefits to weigh against the welfare harms.

Animal welfare should be a strong ethical consideration in any project that impacts it and most authors in the area agree that animal welfare is an important concern. Cohen (2014) claims that "beyond a certain level and probability of harm de-extinction may cause, we should refrain on moral grounds from performing it, despite sacrificing greater utility" (2014, p. 175). Sandler (2014) takes a milder approach, concluding that "while animal welfare concerns must be addressed, they do not justify abandoning deep de-extinction" (2014, p. 358). Kasperbauer (2017) concludes that de-extinction is "still permissible … but only if it can overcome the challenges I identify" (2017, p. 2), which particularly refers to the animal welfare cost and takes the strong view that "the ethical permissibility of de-extinction projects would be limited by their ability to ensure that the individuals brought back would not have lives full of suffering" (2017, p. 7). Rohwer & Marris (2018) similarly argue that de-extinction would be permissible "if and only if suffering is minimal" (2018, p. 1). But as important as welfare considerations are, they are not the only considerations in play. There are many potential values which will be positively or negatively affected, such as environmental and human values, and these should be considered against one another. This means that when considering the ethical permissibility of the de-extinction program, we must look at the potential benefits and how these might weigh against the welfare harms.

This sits within a larger dialogue about under which conditions it may be acceptable to cause harm to animals for some other benefit. It is beyond the scope of this thesis to assess this question in any meaningful way. I will assume that the extreme positions – that it is never okay to cause harm to any animal unless it is to benefit that animal itself, or that there is no problem in harming animals in pursuit of any human gains – are the least plausible, and that there will be at least some conditions under which we consider such harms acceptable. We then need to establish what the real gains of such projects will be, as well as the level of harm that will be occurring (as mentioned, not necessarily an easy task), and make some attempt at weighing

these concerns. Norton (1995) notes that there is unlikely to be a single moral measure on which we can make such decisions. Instead we should be moral pluralists, with differing values in competition, and between which we must adjudicate – "we value many things in different ways, and these differing values are sometimes in conflict" (1995, p. 104). He continues that "we have an obligation to minimize the suffering of individual animals in some situations and that we have obligations to emphasize species protection in other situations. The problem is to explain coherently and effectively how to tell the difference between these situations" (1995, p. 104). Diehm (2017) points out that the 'individualistic' ethic used in animal welfare concerns will not be the only important value in conservation considerations and "the broader conversation about de-extinction is likely to take place on terms substantially more holistic" (2017, p. 26), taking into account species value as well as individual. Where there is a sufficient gain of some other sort, we might accept the welfare harms of these projects.

The potential gains of de-extinction projects fall into four categories. These are: ecological – the improved quality of ecosystems with restoration of keystone species; aesthetic – human preference for the presence of such species; restitutive – that we are in some sense righting the wrongs we have committed in sending such species extinct; and scientific – leading to advancement of knowledge and technology. Several authors have analysed these potential benefits. Cohen (2014) and Sandler (2014) provide in depth analyses of all of these and both conclude that none of these provide sufficient justification for such a project. More recently, Rohwer & Marris (2018) assess potential benefits and conclude that human benefits are the most likely justification, but cannot overrule animal welfare concerns. Sandler (2014) argues that "deep de-extinction does not address any pressing ecological or social problems, and it does not make up for past harms or wrongs. As a result, there is not a very strong ethical case (let alone an ethical imperative) for reviving long extinct species or developing the capacity for doing so … taking on significant costs and risks or funnelling scare resources to pursue it is not justified … deep de-extinction is in many respects a luxury. It is fine to pursue it if people want, so long as it does not interfere with or compromise ethically important things" (2014, p. 359). Greely (2017) describes the proposed benefits as "vague and insubstantial" (2017, p. 35) in comparison to other potential uses of resources to solve environmental and human health issues.

Aesthetic and restitutive gains are the weakest of the proposed potential benefits (cf Lean, 2019 for detailed discussion of these points). Aesthetic gains are based in the value we place on the resurrection of the species, the sense of 'wonder' or 'awe' that in itself would hold intrinsic value (Cohen, 2014), however these are not a type of value that can outweigh the

suffering caused through de-extinction programs (Kasperbauer, 2017). These may then lead to increased emotional and commercial investment in conservation efforts, but this would then only be an instrumental benefit with the primary gain still falling into the ecological category (see Section 3.2.4.1). Restitutive benefits take de-extinction to be in some sense a matter of justice – something we 'owe' to the species we have driven extinct. However, the underlying assumption that species are the kinds of things which are able to hold such claims is unconvincing (Rohwer & Marris, 2018) and even if we could make sense of our duties towards extinct species, de-extinction may not be a way of discharging them as de-extinct animals may not even belong to the same species as the extinct species – Shapiro (2017) describes them as "proxies, not copies" (2017, p. 5). If this is right, and we do not have the same species, it is difficult to justify that de-extinction has benefitted or provided justice to the extinct species. Ecological and scientific benefits are those most likely to justify de-extinction programs, and will be discussed in the following sections. Here I will run through these proposed benefits, and the objections raised against them, to assess whether they are likely to be sufficiently great to outweigh the potential welfare harms; concluding that it is unlikely that any will be sufficient as things currently stand.

3.2.4.1. Ecological benefits

The strongest justification for de-extinction is ecological: that it can help improve the environment through restoring ecosystems. This is what Kasperbauer (2017) refers to as the 'instrumental' value of de-extinct species. It is a commonly held view in conservation biology that we have an obligation to sustain natural processes, and this obligation will offset some animal welfare harms (e.g. Norton, 1995). There are two different strands to this justification and the replies to it – whether we should be aiming at ecosystem restoration at all, and whether de-extinction is the best process to achieve this.

On the first point, it is not obvious that ecosystem restoration is the right target for conservation ecology. Ecosystems are dynamic, constantly changing, and there may be no principled way for choosing some historic state of the ecosystem as the one we should aim at restoring (Davis 2000). Under this view, there is no objective standard of 'ecosystem health' that we can aim at, and all these efforts would merely be based on an arbitrary judgement of the ideal state of an ecosystem from a human point of view (Rohwer & Marris, 2018). Instead of ecosystem restoration, we might instead think of the benefits of restoring lost ecosystem functions and supporting biodiversity, however in most cases these are likely to be equally well served (at lower cost) by other interventions (Lean, 2019).

Even if we were to accept the goal of ecosystem restoration, and were able to set an ideal target state, we don't know enough about ecology to predict whether our actions in this regard may be successful. Cohen (2014) concludes: "Although our analysis supported the essential and actual possibility of de-extinction's ecological benefit, probable changes to species' environment since extinction and the resultant risks reintroduction may pause [sic] to ecosystemic integrity will likely make the overall ecological value of de-extinction quite uncertain in most cases" (2014, p. 169). Single-species de-extinctions may be ineffective in restoration, as ecosystems require interactive networks of species and the target species would thus likely "need to be brought back with a cluster of other species" (Kasperbauer, 2017, p. 5). Most of the species are unlikely to thrive in the wild without assistance (hence their previous extinctions). De-extinct species may fail to provide the intended ecological functions, instead merely serving as "functionally ineffectual eco-zombies" (McCauley, Hardesty-Moore, Halpern, & Young, 2017, p. 1004), as ecosystems can change rapidly after extinction and the functional niche may not remain. Robert et al. (2017) are similarly concerned about possibility for success, due to problems of limited genetic variability and ecological divergence of the species from the ecosystem.

In terms of a conservation 'last resort' or safety net, de-extinction projects are likely to be of limited benefit, as they will not address the causes of species decline, and are probably not the best use of resources in this area (Sandler, 2014). There are strong reasons to think de-extinction projects are unlikely to succeed in restoring lost target ecosystems. At the very least, this justification is only as strong as the likely success of the de-extinction project in restoring the target ecosystem, which relies on a deep understanding of the ecology of the species, the availability of appropriate habitat, removal of the original causes of extinction and the role of the particular species within the ecosystem (Kasperbauer, 2017; Seddon et al., 2014).

De-extinction projects may also have some benefit in raising public interest and support for wider conservation projects, in the same way as 'flagship' species (Lean, 2019). Interest in de-extinction projects may raise funds and awareness for other conservation projects, and habitat protected for de-extinct species will also serve as general ecosystem preserves. However, as de-extinction programs are highly costly – both financially and, as I have demonstrated, in terms of welfare - the case would need to be strong that the degree to which these benefits attain (in comparison to those that come from other traditional programs) would be sufficient to outweigh these additional costs.

3.2.4.2. Scientific benefits

The second strongest benefit of de-extinction programs is the benefit of the science itself – the value in advancing our scientific knowledge and creating new technology. Sandler (2014) considers this to be the primary benefit of de-extinction programs. Similarly, Rohwer & Marris (2018) promote anthropocentric benefits, including scientific knowledge, as the primary goods of de-extinction projects. These projects could push forwards scientific knowledge in terms of the techniques and processes used, as well as the ability to study and understand the de-extinct animals themselves, and the subsequent ecosystem changes (Rohwer & Marris, 2018). These human benefits of knowledge accumulation are not strong ethical reasons, unless there is a case for some other tangible downstream impact of the knowledge gained, "therefore legitimate ecological, political, animal welfare, legal, or human health concerns associated with a de-extinction (and reintroduction) must be thoroughly addressed for it to be ethically acceptable" (Sandler, 2014, p. 354). There is a stronger case where research and understanding could provide more direct benefits to humans, such as improvements in medical research – for instance, suggestions that research into de-extinct gastric brooding frogs could improve understanding of infertility in humans (Zimmer, 2013). Concrete benefits to human lives could be weighed against animal welfare concerns in the same way that current medical testing does, but would require a convincing case that the benefits are likely and of a degree to outweigh welfare harms.

There are possibly other benefits for the technologies currently developed for de-extinction. They could be used in conservation projects for extant species, such as genetically engineering species to tolerate new environmental conditions caused by climate change (Kasperbauer, 2017) or the 'genetic rescue' of endangered species with low genetic diversity (Rohwer & Marris, 2018). "The scientific knowledge and progress that will likely occur also has a great potential to help currently endangered and threatened species" (Rohwer & Marris, 2018, p. 8). As these would be improving the quality of life for currently existing animals, there would be an obvious benefit to individual welfare that may offset other welfare problems. However, this would only provide a reason to develop the techniques in these other contexts, not for de-extinction itself. If the scientific processes are themselves valuable, they can be developed more directly for the projects in which they would be beneficial.

3.2.4.3. Creating future animals

One more potential argument in favour of de-extinction projects of this kind is that they may give rise to many future animals, who will have good lives. Kasperbauer (2017) quotes Brand – "if you can bring bucardos back, then how many would get to live that would not have gotten to live?" (in Kasperbauer, 2017, p. 6). This future benefit might then compensate for the current suffering caused. There are two parts of this argument – the presumption that future animals may actually have good lives, and that if they do then this will outweigh present suffering. In regards to the first claim, it is not clear that the future animals will have sufficiently good lives, due to many of the problems described earlier for rearing and reintroducing animals. "At the very least, they need to present evidence that that lives of future individuals will be good enough to justify the suffering of the first individuals brought into existence. If none of these lives are worth living, then de-extinction is clearly impermissible" (Kasperbauer, 2017, p. 7).

The second claim is a controversial one – it is not generally accepted that the potential future lives of others is a moral good, and certainly not one that outweighs current suffering. To paraphrase Narveson (1973), we want to make people happy, not make happy people. While we may have obligations not to bring into existence individuals who will have lives of suffering, we have no such mirroring obligation to bring into existence individuals who will have lives worth living (McMahan, 2002, p. 300). This means that the future good lives of other animals could never outweigh the suffering of the initial animals. "Many ethicists would be reluctant to accept that the possible existence of future animal lives could justify intense suffering for the first individuals" (Kasperbauer, 2017, p. 6). Kasperbauer (2017) concludes that the justification for creation of future lives could only work if the lives of the first animals are not full of suffering - "at the very least … the initial individuals could be guaranteed a certain level of wellbeing – in common parlance, a 'life worth living'" (2017, p. 6) and this seems unlikely, for the reasons discussed in Section 3.2.2.

3.2.5. Conclusion

I have shown here that none of the proposed benefits of de-extinction programs appear sufficient to outweigh the cost in terms of animal welfare, at least not as it currently stands. This is without taking into account other potential costs, such as economic costs of the research, opportunity costs in terms of other conservation projects that may have instead been funded, risks of harm to existing ecosystems and human populations from release of new species and the potential decrease in urgency of conservation efforts if extinction is seen as reversible (Camacho, 2015). These additional costs give even more weight to considerations against these

projects. Sandler (2017) points out that the way in which we consider these trade-offs will depend a lot on our starting point: "is the presumption that a de-extinction effort ought to be permitted to go forward unless there are compelling reasons, such as those that would emerge from a conservation cost–benefit analysis, against doing so? Or is the presumption that a de-extinction effort ought not to be permitted unless there are compelling reasons, such as those that would emerge from a conservation cost–benefit analysis, in favour of doing so?" (2017, p. 2). Which of these starting points we take will influence how strong the reasons for or against need to be in order to be decisive. The strong evidence for welfare harms gives us a presumption against de-extinction and thus would require compelling reasons in favour in order to outweigh these costs – reasons which we do not currently seem to have.

What this means is that we should at least wait to begin. These projects are, for the most part, not time-sensitive. The targeted animals will not become more extinct the longer we wait. Giving some more time would allow for improvements in the technology that may help reduce these welfare harms. Though the course of making these improvements might still require the use of animals that would be harmed, the number of animals could be smaller, and the larger-scale de-extinction projects could then take place in future with reduced suffering. For more recently extinct species, such as the thylacine, the problem is more pressing, as we may want to bring the species back before the ecosystem changes too much to support them. The likelihood of significant welfare problems, and the lack of strong justification for the projects, suggests that if such projects should go ahead at all, careful attention needs to be paid to the selection of candidate species in order to minimise the risks of suffering, and maximise benefits.

'Shallow' extinctions such as thylacines may be far better candidates for de-extinction projects than 'deeper' extinctions, such as mammoths. For the latter, our lack of knowledge, and changes in ecology, are likely to lead to greater welfare problems, as well as less chance of successful projects. Rohwer and Marris (2018) support this conclusion: "certainly, we believe that the case for bringing back very recently extinct animals is much stronger. Where their habitats and ecological interactions are still available, their return can be justified in the same way as a reintroduction of a locally extinct species" (2018, p. 12). For the projects to have the strongest benefit, and greatest potential to outweigh welfare concerns, these should be species which have a high chance of successful reintroduction, and those which are likely to pay the largest role in restoration of damaged ecosystems. For the lowest welfare impact, these should be species which can be more easily bred (most likely those with extant relatives), and those for which our knowledge of their requirements for rearing, husbandry and reintroduction

are good. Only in these sorts of cases, where we have sufficient information and well-chosen candidate species, with a high chance of success, are de-extinction projects likely to be permissible.

3.3. Conclusion

Animal welfare is, in part, a normative concept. The field of applied animal ethics analyses the different ways in which we use or impact animals; considering the moral permissibility of particular actions and our duties towards animals in different situations. Here, I have looked at two different issues in animal ethics: management euthanasia in zoos and de-extinction projects looking to re-create extinct species. Using a subjective concept of welfare influences how we think of the potential harms and benefits of practices like these. Whenever our actions impact animal welfare, our moral deliberations should take this into account. Using these case studies, I have shown how welfare considerations might interact with other values in our decision-making. I have not aimed to give definitive answers as to the permissibility (or not) of the practices described, but instead aimed to highlight the range of considerations we need to keep in mind when assessing them. One additional point raised throughout these discussions has been the importance of accurate measurement of welfare in order to inform our decision-making regarding necessary trade-offs and comparisons; whether regarding the quality of life lost for euthanized animals compared with the gain to other animals within the institution, or in quantifying the degree of harm experienced by animals in de-extinction programs. Without accurate measurement, we risk taking the wrong paths when faced with such choices. We thus require accurate quantitative measures of welfare as part of our moral decision-making. It is to this matter – the measurement of animal welfare – that I now turn.

PART TWO – Measurement of Subjective Welfare

4. Chapter Four – Measuring Welfare

4.1. Introduction

In Chapter Two, I argued that we should understand animal welfare subjectively; as consisting in the subjective experience of an animal over its lifetime, then went on to examine how this concept could influence the way we think about some cases in applied animal ethics. However, as well as being a normative concept, welfare is a scientific concept. In this chapter, and those that follow, I will shift to looking at the measurement of this subjective experience of animals, and addressing some of the potential problems with making such measurements. For what follows, it is not essential to accept the arguments of Chapter Two. Measurement of subjective welfare is still important, even if one does not take it as the sole constituent of animal welfare. Almost all conceptions of animal welfare agree that subjective experience is a necessary component of welfare, and if this is the case, then its measurement will be an important part of animal welfare science. As discussed in Chapters One and Two, I am not here addressing the 'problem of other minds'. I established there that we have good reason to think that animals experience conscious mental states, and that we can gain information about them. The other option is to assume that subjective states in animals are epiphenomenal - that is, that they have no effects on the animal itself or on the world. If we reject this, and instead accept that conscious experience has some effects, then we should also accept that - at least in principle - subjective welfare is detectable. This means it is accessible to measurement, though still insufficient to determine whether it has the required properties to be a measurable entity. In this chapter I will address this second concern.

There are two different questions that arise when we are looking at the measurement of welfare. The first is a theoretical question – is welfare a measurable entity? And the second is practical – can we actually measure it in practice, and how might we do so? In this chapter I will address the first question, looking at whether, in principle, subjective welfare is the sort of thing that can be measured; that is, "what quantitative statements about well-being can we give sense to?" (Griffin, 1986, p. 94). The chapters that follow will then look at some of the other issues that come up in the practice of trying to measure welfare. In particular, when thinking about the measurement of welfare, one might have concerns about the commensurability of the different types of subjective experience and whether they can combine into a single, cohesively measurable state – this problem will be addressed in Chapter Seven. Here, I will examine the question of whether in general subjective experience (for now considered as a whole) is a

measurable attribute; using some basic measurement theory, I will look at what is required for an attribute to be considered measurable, and show how subjective experience meets these criteria.

Measurement, at its most basic, is simply "the assignment of numerals to objects or events according to rule – any rule" (Stevens, 1959, p. 19). We already have in place a system of mathematics that we understand – we are able to apprehend the relationships between numbers, and have well-defined rules for the sorts of transformations we can perform on them, and how to interpret the results. When we measure something, we are mapping these mathematical structures onto that attribute, in the hope that we can then use the mathematical rules to better understand the properties of whatever it is we are measuring and how it relates to other instances of the same kind. In determining an attribute to be measurable, we are "showing that a system of empirical relations and operations is isomorphic with a certain system of numerical operations and relations" (Griffin, 1986, p. 95). We use basic rules or axioms to provide the necessary structure for ordering measures and determining the relationships between them. For example, "axioms of order ensure that the order imposed on objects by the assignment of numbers is the same order attained in actual observation or measurement" (Hosch, 2011, p. 227) and "axioms of difference govern the measuring of intervals" (Hosch, 2011, p. 227).

This mapping will always be somewhat imperfect, but so long as we have reason to believe that the empirical properties of what we are measuring are in some way mirrored by the mathematical properties of the assigned numerals, then we can hope to gain understanding of the system of interest. What is important to keep in mind is the relationship between the measured attribute and the measurement system applied. Sarle (1997) gives a clear example of the use of measurement and the relationship between the measured attribute and the assigned numbers:

> Suppose we have a collection of straight sticks of various sizes and we assign a number to each stick by measuring its length using a ruler. If the number assigned to one stick is greater than the number assigned to another stick, we can conclude that the first stick is longer than the second. Thus, a relationship among the numbers (greater than) corresponds to a relationship among the sticks (longer than). If we lay two sticks end-to-end in a straight line and measure their combined length, then the number we assign to the concatenated sticks will equal the sum of the numbers assigned to the individual sticks (within measurement error). Thus another relationship among the numbers (addition) corresponds to a relationship among the sticks (concatenation). These relationships among the sticks must be empirically verified for the measurements to be valid. (Sarle, 1997).

Because these mappings are imperfect, we need to look at which particular features might be relevant in any case and use the mathematical transformations accordingly. This gives us

different scales of measurement, which are different ways of assigning the numbers to the measured attribute. S.S. Stevens (1951) laid out four primary kinds of scale we might use for measurement – nominal, ordinal, interval and ratio. In Section 4.2, I will go over what each of these entails, and how we might apply them to subjective welfare. The important thing is that we match the right type of scale to the attribute or object being measured. It needs to be the case that the types of transformations that can be performed on the mathematical elements of the scale mirror those that can be performed on the measured attribute. What scale we use will inform which statistics we can use in analysis.

When choosing a measurement scale, it is also important to keep in mind our goals for measurement. Griffin (1986) notes that measurement is flexible enough that basically any attribute you can imagine will be measurable on some scale or other – "What we want to know is not just, Is well-being measurable? There are many different scales of measurement, and it would be astonishing if well-being were not measurable on at least one of the less demanding ones" (1986, p. 75). As we will see, the simpler scales of measurement are quite permissive, such that anything which can be categorised could be considered measurable in this basic form. This means the open question of "is welfare measurable?" is not necessarily an interesting one. What we are really interested in is whether welfare is measurable in such a way that it will be useful for our purposes – whether our capabilities fulfil our needs (e.g. to enable us to compare and contrast; to evaluate the results of interventions). This requires an idea of what our needs are – how we are going to use these measurements and what we will therefore require of them. Do our powers of measurement match the demands of application? As I go on to look at the different types of scales and their potential applicability to subjective experience, I will also comment on how appropriate these are to the goals of animal welfare science, as described below.

As discussed, animal welfare science is interested in the measurement of welfare. Most commonly, measurement of animal welfare will take the form of trying to determine whether some particular type of intervention – such as type of environment, or husbandry procedure – will increase or decrease the welfare of the animals that experience it[25]. This then often only requires a very basic type of measurement – all we need to determine is whether some outcome is higher or lower than a baseline, or some other outcome. In other cases we may have multiple

[25] As we are ultimately concerned with lifetime welfare (as discussed in Chapter Two), there are additional complexities in calculating lifetime welfare from a set of synchronic measures such as are typically used in welfare science. I take here an assumption that there is some such function that could do the job, though there is insufficient space to work out the specifics of how that might look in practice.

treatments which we have to compare, but we are still really just looking for a basic ranking – is it better for the welfare of these animals that we provide environment A, or environment B? In this sense, the degree to which welfare is affected is not really important, and so all we would need would be a basic ordinal scale, as will be discussed below (though there is still the additional issue of comparing interventions which may improve welfare in some respects and not others; this will be addressed in Chapter Seven). However, there are other cases in which we would require more information. It is not always enough to simply determine which of two (or more) environments is the best one; we will sometimes need to know by how much. Additionally, when making practical decisions about implementing changes to animal housing and husbandry, we will need to consider potential trade-offs, which will require more information about the degree of effect of particular treatments, and how they might compare to one another. This sort of decision-making requires quantitative comparisons of magnitude. We may also want to make welfare comparisons across different species, such as when deciding on best distribution of resources; this raises the problem of intersubjective commensurability that will be addressed in Chapter Six. The applications of measurement in welfare science and the types of measurement scale required for these are summarised in Table 4.1 below, and will be discussed throughout Section 4.2. When considering the needs or goals of measurement of animal welfare, it seems that, though in many cases a simple ranking procedure will be sufficient, when trying to consider trade-offs or comparisons we will need some more sophisticated system of measuring degree or magnitude of subjective welfare effects. In the following sections, I will look at whether these types of measurement might, in principle, be possible for subjective welfare.

Welfare application	**Measurement scale**
Comparing housing types	Ordinal
Looking for individual welfare changes over time	Ordinal
Comparing welfare between individuals	Ordinal
Comparing the effect of husbandry interventions	Ordinal
Aggregating individual welfare to assess groups	Quantitative (Interval or Ratio)
Calculating management trade-offs	Quantitative (Interval or Ratio)
Determining best resource distribution	Quantitative (Interval or Ratio)

Table 4.1: Types of measurement scale required for welfare applications

4.2. Measurement scales and application to welfare

As mentioned in Section 4.1, to measure some attribute is really just to assign numerical values to different values of the attribute, according to some rule. There are many different such rules that can be used, and here I will look at some of the more common ones, and how they might be applied to subjective welfare. The five primary types of scale we might use for measurement are nominal, ordinal, interval, ratio and absolute. Each scale has an associated set of permissible transformations that may be used; those which "preserve the relevant relationships of the measurement process" (Sarle, 1997). When choosing a type of scale on which an attribute might be measured, it is important to find the right match, such that the sorts of transformations that can be performed on the mathematical elements of the scale mirror those which can be performed on the measured attribute. Relating measurement to the attribute requires assumptions about the nature of the attribute and its properties (Sarle, 1997). In this section, I will outline the features of the different types of scales, and for each one, examine whether subjective welfare could have the required features to be measured on that scale.

4.2.1. Nominal

The first and most basic scale of measurement is the nominal scale. This is simply the application of labels to different categories – these can be numbers, but can also be words, or symbols. In a nominal scale, the mathematical relationships between the labels have no meaning, apart from the relationship of identity. All that matters are the number of categories and how many members they contain. Two items can be assigned the same label where they have the same value of the attribute (Sarle, 1997). This scale relies only on the relationship of equality – in determining whether some object is equal to another (same category) or unequal (different categories). The permissible transformations are one-to-one and many-to-one transformations (Sarle, 1997). That is, that each item within a category could be shifted to a new category label, preserving their distinctness, or items from multiple categories could be shifted into a new category, grouping them. For example, we might consider two categories of measurement for swans – black and white. We could label these categories '1' and '2', where the numbers represent only so far as the uniqueness of each number mirrors the distinctness of the groupings – the numerical properties of the labels have no bearing. We would observe our group of swans and assign them to either category, depending on their feather colour. Swans within the same category would count as equals, while those in different categories as unequal. We could perform a one-to-one transformation on these categories, where all members of one category (say, the black swans) are shifted to a new category: essentially a 'relabelling', where

‘1’ becomes ‘3’, or ‘A’. We could also perform a one-to-many transformation in which we create a new category of simply ‘swans’ (perhaps labelled ‘4’), and move all the members of ‘1’ and ‘2’ to this new set.

Subjective animal welfare could in theory be measured on a nominal scale. We could create different welfare categories, based on welfare scores – such as group ‘A’, ‘B’, ‘C’ - and place individuals within them, depending on their measure. Any animals placed within the same category would be taken to have the same welfare value (equality). These categories could be transformed through relabelling (e.g. A = 1, B = 2, C = 3), where equality relations between individuals within a category still hold, or through grouping (e.g. new category D = what was previously in B & C) where new equality relations are created.

This is, however, going to be of extremely limited use, as we cannot even make judgements as to the relative value of the categories (i.e. that category A represents higher welfare than category C); to do so would be to convert our measurements to an ordinal scale. Without saying anything about the relative empirical values of the categories (i.e. whether they represent higher or lower measured scores), they might still have some limited use if we state that we prefer membership of one category (say, A) to another (C). We could then assess whether more animals are in our preferred category (based on the relative membership of the two categories), or whether this number changes over time. In analysing a particular husbandry situation such as a farm, we could then have rough comparisons between different farms or the same farm over time. As this would be an additional judgement of normative value as opposed to the empirical values of the welfare scores used in measurement, this would not technically be a conversion to an ordinal scale. Returning to our earlier example, it would be equivalent to saying that we prefer to have more black swans than white, and then analysing data accordingly. We are not saying anything about the *measurement* of blackness or whiteness, but drawing further conclusions as to how we might prefer to use the data. However, given that in practice, our preferred categories will always be those of higher welfare over lower welfare, then this will push us to use of an ordinal scale, to which I will now turn.

4.2.2. Ordinal

The second type of measurement scale is ordinal. An ordinal scale allows us to rank objects in order, regarding some particular property on which an attribute is simply ranked as ‘higher’ or ‘lower’ than another within a list, and these are then assigned corresponding values. These values must reflect the ordering relationships between the members – e.g. those with higher levels of the attribute should be assigned higher numbers. It is important to note here that the

values themselves mean little – as long as they represent the ordering of the attribute, they could be 1,2,3 or just as easily be -14, 1.4, 3000 (Griffin, 1986). The differences between the values do not represent anything about the differences between the attributes. Building such an ordinal list simply consists of comparisons between members of a set. An example of this sort of scale is 'scratch' tests for rock hardness, in which rocks are ordered according to which other rocks they are able to scratch, or be scratched by. This type of scale still cannot give us many useful statistical analyses, but we are able at least to determine relations of greater than or less than. The permissible transformations are monotone increasing transformations, in which the numerical (or other) labels may change, so long as they retain the ordering relation between items (Sarle, 1997).

There are also several possible types of ordering we can use – Griffin (1986, pp. 96–98) describes strong ordering, weak ordering, partial ordering and vague ordering. Strong ordering is the most demanding and has the properties of reflexivity, transitivity, completeness and anti-symmetry. Reflexivity means that all objects stand in equality relations to themselves – in this case, meaning that they sit in the same location in the ordering (have the same assigned value). Transitivity means that, for example, if A > B and B > C, then A > C. Completeness means that all attributes are assigned a value and can be placed somewhere on the scale – for any pair of objects, either they are equal or one is greater than the other. Anti-symmetry means that no two items can hold the same value and sit in the same place on the scale. For a weak ordering, we remove anti-symmetry so that now two objects can be equal. For a partial ordering we remove completeness, so that some objects may not have a place on the scale relative to one another.

Vague ordering is a type of partial ordering, in which we lose completeness due to the presence of rough equalities. In rough equality, some item A might be roughly equal to B, which means it is neither greater than, less than or equal to B. Rough equality also removes transitivity – so where A is roughly equal to B and B better than C, A may not be better than C. Why then can't we just treat rough equality like strict equality and assume that A = B? According to Griffin, the reason is that if we add something to A, to create A+ (where A+ > A), unlike with strict equality, we still could not claim that A+ > B, or that A+ < B or A+ = B; we might still just have A+ roughly equal to B. "The trouble with rough equality is that it makes the strict ranking statements that it infects neither definitely true nor definitely false" (Griffin, 1986, p. 96). In these situations it will not be the case that A > B or A < B or A = B, as it is for most scales, and this is what gives us vague ordering. Unlike in a weak ordering, here where A roughly equals B, B roughly equals C, we might sometimes decide A roughly equals C and

sometimes not. This gives us what he calls a 'partially transitive weak ordering'. This can occur either "because it is hard to discriminate the differences in value that are really there or that there are no fine differences really there to discriminate" (Griffin, 1986, p. 80). In practice, we are much more likely to face the former situation than the latter (or not even know which we are in), as even when we might think there is some fact of the matter as to which of two items is greater, we will often lack the capacity to detect small differences.

The existence of rough equalities may not often matter much to the application of ordinal scales. In fact, Griffin claims that "it is uncommon for rough equality to matter to prudential deliberation. Where we have rough equality we treat the items, when it comes to choice, simply as equals … It is a mistake to conclude from the fact that rough equality crops up fairly often that the difference between strict and rough equality matters fairly often" (Griffin, 1986, p. 97). So for practical purposes, when using an ordinal scale, we can probably treat rough equals as strict equals and get the same results. When making decisions regarding welfare, these will come out roughly right for our purposes. In particular, it is rare that much will turn on decisions in which there is no detectable differences between alternatives – either both would be acceptable, or there will be some alternative use of resources that will provide a much larger measurable difference.

An ordinal scale is definitely appropriate for measurement of subjective welfare. The quality of subjective experience can be better or worse along the welfare scale, and particular instances can be ranked against one another (in Chapter Six I will address the question of whether and how we can make comparisons between individuals). It seems fairly intuitive that we are able to assess, at the very least, whether some welfare situation is better or worse than some other (barring the issues of integrating multiple types of experience that will be addressed in Chapter Seven). It may be the case that in many circumstances it is difficult to differentiate between closely-matched cases, but this is more a practical difficulty in application than a true problem with the scale. Although it may be difficult at times to judge the exact ranking of particular states of welfare, it is not the case that such states are unrankable. We shouldn't have much trouble deciding that a bear rolling around in a pool with its companions is in a better state than a malnourished lion pacing on concrete. It may be difficult to decide in more closely matched cases whether a lion hunting prey or a bear catching fish is experiencing better welfare, but this is more likely to be a result of measurement imprecision rather than a case of strict incommensurability (to which I will turn in Section 4.3). At least in principle then, subjective experience should be measurable on an ordinal scale. There may of course be cases in which making such comparisons is extremely difficult, as described for vague orderings above – we

might lack the measurement precision to determine differences in value, or may have the type of entities for which such differences do not exist. The first is simply a practical issue about the reach of our measurement practices. The second is potentially more difficult, and leads to Griffin positing rough equality, as described above, where "the roughness is not in our understanding but ineradicably in the values themselves" (1986, p. 81).

Due to this possibility, Griffin (1986) thinks vague ordering is the scale most likely to be applicable to measurement of wellbeing. He takes rough equality as inherent to measurements of wellbeing. However, Griffin has taken a preference-based approach to understanding wellbeing, in which there may actually be no strong fact of the matter about whether some preference truly outweighs another. For a subjective account of welfare, we might instead say that there is some fact of the matter about whether or not A > B, but we just don't have sufficiently precise measurements to tell. In these cases we could treat the items as roughly equal, in the knowledge that future measurement may allow us to be more precise and alter the ordering. In these cases too, as with vague ordering, we should not expect exact transitivity, as if we take two items (A & B) as roughly equal only because of imprecise measurement, it could easily be the case that some third item (C) may fall within the bounds of error for A but not for B and thus not be roughly equivalent to both. In cases where we have sufficiently precise measurements to determine the differences between closely matched items, we should be able to construct a weak ordering, as the items will show the properties of reflexivity, transitivity, and completeness. We should not expect a strong ordering, as it will be possible for some items to occupy the same position on the scale, and thus will not be anti-symmetric.

Using an ordinal scale allows us to determine whether one animal's welfare is better or worse than another's, or whether an animal's welfare has improved or decreased over time. This fits with the goals of much of animal welfare science, in trying to determine whether some particular type of intervention – such as type of environment, or husbandry procedure – will increase or decrease the welfare of the animals that experience it. With this scale, we are able to discriminate both changes across valence (shifting welfare from poor to good or vice versa) and within valence (e.g. shifting from more to less poor). We can determine whether some particular outcome is higher or lower than some comparison case or another outcome. As our aims should always be to shift welfare up on the scale, having an ordering such as this will meet many of our needs.

Although an ordinal scale will provide us with useful information for many applications, it will still not be sufficient in all cases. Although trying to determine which interventions will improve welfare is probably the most common application, it is not the sole use of such

measurement. We will often be making decisions in which it is important to consider potential trade-offs, which then require more quantitative information about the comparative magnitudes of different effects. Resources for animal management are always limited and we might, for example, want to decide whether to increase the size of an otter exhibit to give them more space to play, or to add a pond to a tiger exhibit to allow them to swim. This sort of decision-making requires quantitative comparisons – to determine whether the welfare benefit of more space to the otters is greater or less than the welfare benefit of a pond for the tigers. Leaving aside for now the issue of cross-species comparisons (which will be addressed in Chapter Six), we are still left with the task of comparing magnitudes. This comparison of magnitudes will be required in single-individual cases as well, such as if we were deciding whether the amount of suffering a kangaroo might experience from undergoing a vasectomy will be offset by how much pleasure it might gain from the ability to then live and mate freely with female companions.

Ordinal scales also do not allow for aggregation of data, such as averaging the welfare of an animal over many points in time, or finding the average welfare level of a group of animals under investigation. We may be able to rank each animal into a category (say – high, medium, low) and against one another, but cannot say whether, for instance, an equal number of 'high' and 'low' scoring individuals would produce the same average as all 'mediums' would, as we don't know if the difference between high and medium is the same as the difference between high and low. This could make a difference, for example, when trying to compare different farming systems in which one may contain a lot of both 'high' and 'low' rankings, while another contains a lot of 'mediums' – we would have no way of deciding between them.

Although an ordinal scale is appropriate for measuring welfare and will meet our goals in many cases, in other situations where we need to make trade-offs or comparisons, we will need to measure magnitude of welfare effects. For this we will need a quantitative scale, such as an interval or ratio scale.

4.2.3. Interval

We may often require more information than just knowing the relative ordering or rankings of a set. There can be important reasons to be interested in the size of the gaps between the items. A more informative type of measurement in this regard is an interval scale. Unlike an ordinal scale, the interval scale carries information about the size of the gaps between items in the ordered set – the difference between adjacent items. The differences between the assigned values are representative of the differences between the measured attribute – so the difference

between 1 and 2 is of the same magnitude as between 2 and 3. The exact scale selected, though, is arbitrary. An example of an interval scale is a temperature scale such as Celsius or Fahrenheit. Here we can see the arbitrary selection of scale – although both Celsius and Fahrenheit have equal intervals between their values, they differ from one another in the size of these intervals and in the origin point.

Use of a quantitative scale gives us the possible rate of substitution between two values – we can now talk meaningfully about trade-offs between items with equivalent units. The permissible transformations are affine transformations, of the form t(m) = c * m + d (where m is the measure, t the transformed value, and c and d constants) (Sarle, 1997). This will preserve the relative distance between points. This can be seen in the transformation from Celsius to Fahrenheit: C = 1.8 * F +32. It also allows us to use more sophisticated statistical analyses to compare different states. One worry with use of numerical scales such as these is that the numerical representations may then be more specific than the information they are being used to represent – there may be a false precision. We thus need to be sure that the measured attribute is quantifiable in the right way.

Are interval scales, then, applicable to subjective welfare? Intuitively, it seems like they are. It makes sense to think of subjective experience as something that varies along a linear scale – that it can move up and down by increments (see Chapter Seven for further discussion of how this would work for an integrated set of experiences). "If well-being were a simple state of mind that occurred in smoothly changing intensities, then we could at least hope to develop a powerful cardinal scale, and one that is without discontinuities or incommensurabilities" (Griffin, 1986, p. 75). Although Griffin disagreed that this was possible, because he took the nature of wellbeing to be informed desire, on the subjective welfare approach this seems plausible. The fact that we are able to conceive of rates of substitution, or trade-offs, in welfare, tells us there must be meaningful intervals between items on the scale.

However, application of an interval scale is a much more difficult prospect than an ordinal scale – now we are no longer just trying to rank particular states as better or worse than others, but we are trying to determine *by how much* they are so. This requires quantification of the underlying subjective states – to transform subjective experience into some sort of measurable units, such that we could say something like, "Giving this animal diet X increases its welfare by three units, while diet Y only increases welfare by 1 unit." For any individual, these units can be determined as increments on a scale from maximum to minimum experienced welfare, as discovered through measurement of indicators.

For practical purposes, we should be able to approximate such a carving through the types of quantifiable proxy measurements we use, such as physiological and behavioural variables. For many of these, it seems that they can be quantified and compared – blood concentrations of relevant hormones, frequency of particular behaviours, changes in heart rate or body temperature. When we measure welfare, we take data about measured indicators of welfare and then convert these to a welfare score. If we can assume that these proxies stand in the appropriate quantitative relationship to the central measure of subjective experience, then we will be able to give at least a rough cardinality to measurements of welfare. We need reason to think that the transformation of these data matches the relationship between the indicator and the welfare experience in order to create a useful scale. We want to avoid over-precisification in our final scores, based solely on the pattern of the indicator response. This does not mean that welfare is not quantifiable, but that we need to take care to ensure we understand the relationship between welfare experience and indicator response.

This may not always be easy to determine, as the relationship between measured indicators and underlying welfare state may not be proportional, or linear. For instance, in order to determine the welfare impact of human presence on an animal, we might measure the flight distance. However, we cannot say that the welfare difference between two animals which flee at a distance of 6m and 8m have the same difference in welfare experience as between an animal which flees at 2m and one that allows the experimenter to touch it (0m). Even though the intervals are the same, we would probably want to say that the first two animals have fairly similar fear, while the second pair are quite different (Botreau, Bracke, et al., 2007). Or, when measuring pacing behaviour as indicative of frustration, we would not necessarily think that twice as much pacing represents twice as much frustration. We need to be sure we have the right indicators to represent welfare, and use of multiple indicators will help to build welfare scores. Mapping the response profile for an indicator over a range of interventions will also help with the creation of the right transformation function. Where we have multiple indicators that show similar relations to welfare response, we can be more confident that our measured variables are capturing the underlying mathematical relations of welfare response (see further discussion in Chapter Five on validation of indicators).

Use of an interval scale would allow us to make important decisions in animal management, as described above. Knowing the intervals between items in our ordering allows calculation of rates of substitution and trade-offs for management decisions.

4.2.4. Ratio

Although interval scales can give us good quantitative information, the limiting feature of these scales is that they do not have a true 'zero' point and so we cannot say anything about the ratios of different values. For example, it would be meaningless to say that it is twice as hot on a day that measures 40°C than one that measures 20°C, because on the Fahrenheit scale the temperatures would read and 68° and 104°. The same numerical relationship will not hold, despite the same underlying relationship between the measured attribute of temperature: "the relationship 'twice-as' applies only to the numbers, not the attribute being measured (temperature)" (Sarle, 1997). As the assigned numbers are largely arbitrary, they do not allow for meaningful analyses like these.

A ratio scale, by contrast, is set around a non-arbitrary origin point and thus can allow these transformations. On a ratio scale, the ratios between the assigned values (doubles, triples etc.) will then stand in for corresponding ratios between the attribute being measured. This allows comparisons such as 'twice as much as'. The Kelvin temperature scale, beginning at absolute zero, is an example of such a scale, as are standard scales of length measurement (where the origin is 0cm, or no length at all). Permissible transformations are linear transformations of the form t(m) = c * m (where c is a constant). This represents the fact that though the origin is fixed, the units of measurement are still arbitrary (Sarle, 1997). An example of this transformation is in measurement between centimetres and inches, where length(cm) = 2.54 * length(in).

If we allow that animal welfare can be quantified, it seems to lend itself to a ratio scale just as well as to an interval scale. That is, there appears to be a natural zero point which gives sense to the ratio transformations[26]. That zero point would simply be the point of neutral welfare – neither positive nor negative. This may arise because of an absence of either positive or negative states, or due to an equal balance between the two. We can then measure deviations from this set point in either the positive or negative direction – certain interventions may increase welfare up above the neutral point (positive welfare), while others will decrease welfare below the zero point (negative welfare, or suffering). A ratio scale is a good representation of welfare, and allows all the same applications in making comparisons and trade-offs as an interval scale. In Chapter 7, I will describe some of the specific measurement

[26] There are additional interesting questions about the zero point – such as its exact nature, and how it would be measured - but these are beyond the scope of this thesis to explore.

methods and frameworks used for animal welfare, which will further illustrate the usefulness of a ratio scale.

4.2.5. Absolute

The final, and most demanding type of scale is an absolute scale. Here, the properties of the assigned numbers are identical to the properties of the measured attribute. The only permissible transformations are identity transformations, which always return the same value for an item (Sarle, 1997). Examples of absolute measures include counting of objects and measurements of probability. An absolute scale is unlikely to apply to welfare, as we have no reason to think that there are non-arbitrary units of measurement along the welfare scale.

4.3. Incommensurability

Incommensurability between two items, or attributes, occurs when there is no way of placing items on the same measurement scale and thus no way of making comparisons between them. Think of trying to answer a question like "is this potato heavier than this banana is yellow?" There is no answer; the question itself is meaningless. The two values are incommensurable – there is just no common scale on which both items could be ranked. If incommensurable, "two items cannot be compared quantitatively at all; the one is neither greater than, nor less than, and not equal to the other" (Griffin, 1986, p. 79). Chang (2015) terms this incomparability - the inability to place two items on an ordinal scale, such that one will be greater than, less than, or equal to the other and takes incommensurability to be a weaker notion, of inability place two items on the same cardinal scale. Whichever specific terminology we use, the concern is still the same - that we might be unable to compare items.

Although welfare itself is measurable, we might still have the problem of incommensurability between its items. If this were the case - such that we could not make comparisons between different animals, or types of welfare experience - then its measurement may not be of much use for the required applications. In this section I will argue that this is not the case; arguments I will follow up on in more detail in Chapters Six and Seven. Some of these questions are moral rather than scientific – as to how much moral weight we should place on different animals or on the states of pleasure and suffering – but the measured level of impact on subjective welfare will also be relevant.

Griffin (1986) points out that an appeal to incommensurability is unconvincing in almost all cases – it seems that for almost any two imagined variables, there will be some point at which we think a large enough amount of one outranks a tiny amount of the other. If we allow this,

then there is at least some minimal sense of comparability. We only think there is a problem in cases where the values start to converge and we cannot decide how to rank them. This however does not mean that they are 'unrankable', just that we may not be able to discriminate or decide. The notion of 'rough equality', introduced in Section 4.2.2, can help with these sorts of cases. Here the reason two cases may be difficult to compare is because there are no real differences present to discriminate between. Griffin (1986) speculates that this sort of rough equality may actually be quite common.

At first pass, it might seem like the problems of intersubjective comparison are examples of incommensurability – that we simply cannot decide whether the subjective experience of one animal is better or worse than another, because the fundamental differences in their subjective experiences may not track one another. However, it seems we have at least some minimal ordinal comparability – we can easily think of situations in which the poor state of one animal is clearly representative of lower welfare than the positive state of another, where it would seem crazy to insist that there is no way of judging their relative rank. The presence of a common 'zero point' of welfare experience also implies comparability of at least a basic sort – we can at least compare two individuals as to where they sit in relation to this point. It may often be difficult to judge, but this is not the same thing. I will discuss this issue further in Chapter Six, and describe a way in which we may be able to make intersubjective comparisons of welfare.

One type of incommensurability is trumping, which "allows comparability, but with one value outranking the others as strongly as possible. It takes the form: any amount of A, no matter how small, is more valuable than any amount of B, no matter how large." (Griffin, 1986, p. 83). This means that some feature A will always outweigh some other feature B, regardless of its magnitude. This might be how we'd imagine the divide between positive and negative welfare to sit – perhaps negative welfare is just so important in its impact, that inflicting some degree of negative welfare will never be outweighed by some increase in positive welfare. This is the view of negative utilitarians – that it is only the negative welfare that counts (Sinnott-Armstrong, 2015). This also seems unconvincing, because again we can imagine cases in which a tiny loss could be offset by a massive gain and it seems odd to deny that these cases would benefit overall welfare. In fact, Griffin (1986) points out that there is unlikely to be trumping in many cases at all, as there aren't many things we wouldn't sacrifice a tiny bit of in order to gain a very large amount of something else. The reason we may think that there are some things we wouldn't sacrifice may instead just be an effect of our inability to comprehend the type of massive gain that would be required to offset some losses.

Another type of incommensurability is pluralism – the idea that there are irreducibly many values, without any sort of common scale on which to rank them. This might be the sort of incommensurability that would occur if we were to accept something like the tripartite conception of welfare, with three different but equally important values contributing to welfare. One of the advantages of an account of subjective welfare is we do not have to worry about the potential problems arising from trying to compare different types of welfare – for example, trying to decide how a particular level of natural behaviour compares to some amount of subjective wellbeing. Although these different components may not be incommensurable, there would certainly be difficulty in developing a framework to compare their measurements in such a way as to build up a meaningful picture of overall welfare – would, for instance, an animal with low physical welfare but high subjective welfare be better or worse off than an animal with low subjective welfare but a high level of natural behaviour? Trying to decide on the weightings of the various components could be, in practice, almost prohibitively difficult. But, as I have argued, we should understand welfare as actually just consisting in the single target of subjective experience. This means that in determining the welfare of an animal, we only need to consider how this one attribute varies, which should give us monism.

It is only if we accept that individual subjective experiences are so different, or positive and negative welfare are so different, that they count as distinct values, will this problem arise for the subjective welfare account. Although Griffin argues "if the denial of reduction is just the denial that there is a single mental state running through all the things that we rank in terms of which we rank them (the denial of a crude mental state account, or of hedonism), then, it can be agreed by everyone, utilitarians included. In one sense, different kinds of pleasure, or pleasure on the one hand and pain on the other, are incomparable: namely, there is no deeper unitary mental state in terms of which they can be compared" (Griffin, 1986, p. 89), there seems no reason to think that this is true – that there is not some sort of underlying subjective quality of life onto which individual pleasures and pains will map (see Chapter Seven). If we assume that these are not different in kind between different individuals we also do not have this problem for the intersubjective case (see Chapter Six). So long as we are able to work out some sort of trade-off between these two different scales, we must still have some underlying scale ('overall welfare') to which we are referring, and this is all we need.

Overall, there is no reason to think that we have incommensurability with measures of subjective welfare, either between individuals or types of mental states, and I will go on in Chapters Six and Seven to show how these comparisons might be made.

4.4. Aggregation

One set of problems we come up against in measurement of animal welfare is issues in how to aggregate or combine measures into a metric of welfare. So far, I have taken welfare to be a fairly simple function of the total number of positive and negative subjective experiences for an animal over its lifetime. But welfare may not be a simple additive function of different positive and negative states (Sandøe, Corr, Lund, & Forkman, 2019). For example, in order to suffer, there may be a threshold level of intensity, duration and combination of experiences below which one is not too bothered, and above which one suffers. The combined effects of different experiences may be greater than the individual contribution of each. The presence of multiple aversive stimuli that can create conflict in an animal will add an additional degree of stress which further reduces welfare (G. Mason & Mendl, 1993). This is not a problem of combining affects per se, just one that means the way in which we do so may be more complicated than simply a linear sum (or weighted sum). Use of whole-animal measures to compare the relative value of different experiences, alone and in combination, will help to determine this (see Chapter Seven).

There is also a 'counting' problem, in weighing the value of different experiences against their simple measure. For example, it is unlikely to be the case that an animal experiencing 1 unit of pain over 10 days will be equivalent to one experiencing 10 units of pain over 1 day (Sandøe et al., 2019). This can possibly be overcome in two ways. The first is to look at the decision-making process of the animal itself – whether it will work equally hard to avoid one day of extreme pain than 10 days of mild pain, for instance. By looking at these measures we can build equivalences or trade-offs between different intensities and durations. The second way is to build this information into the scores themselves. This would include answering questions (beyond the scope of this thesis) about why something should count as '1 unit' of pain, as compared to '10 units'; what it is about this experience that is 1/10 of the other. "Assessing humaneness is complex, not least because it involves comparing durations and intensities of suffering, and making such judgements as "is extreme breathlessness worse than nausea?" and "is a few hours of intense pain better or worse than several days of milder distress?"" (G. Mason & Littin, 2003, p. 21). It seems the value the animal places on these experiences, and their relative weightings with regards to trade-offs, should be part of what determines the scale.

Aggregation of welfare states and measures into a single welfare metric is complicated, and I will return to the issue in Chapter Seven. This should not be taken to say anything about the measurability of welfare however. Although it may be difficult in practice to combine

measurements, and to determine the relative weighting and impact of different experiences, this is merely a problem of what we know about their relation to welfare; not in itself a reason to think that welfare is not a measurable attribute. As I have already shown, we have good reason to be confident that (at least in theory), we can measure welfare using both ordinal and ratio scales, depending on our requirements for application.

4.5. Conclusion

For an attribute to be measurable, it must be the case that we can assign numbers according to a rule, such that the properties of and relations between the numbers mirror the empirical properties of the attribute. There are a number of common measurement scales that can be applied – nominal, ordinal, interval, ratio and absolute. Importantly, the scale chosen must both match the properties of the measured attribute and play the appropriate role for the use of the measurements.

I have argued in Chapter Two for use of the subjective welfare concept, in part because it is measurable. In this chapter, I have shown that, at minimum, we can rank welfare states on an ordinal scale, which will allow us to make many of the required judgements about the comparative value of interventions. When making decisions regarding trade-offs, we will need to use a quantitative (interval or ratio) scale to compare relative magnitudes. The natural zero point for welfare experience (neutral welfare) means it can be measured on a ratio scale for these applications. There is the possibility of incommensurability between different welfare measures, such as between individuals, and of different types of mental states, which will be further explored in Chapters Six and Seven.

Although there may be further concerns about exactly how we bring together measured items to create a total welfare score (an issue that will also be further addressed in Chapter Seven), this does not give us reason to think welfare is not measurable, but just that there is further work to be done in relating simple measures to the underlying state we are measuring. Welfare is a measurable entity, though its measurement is not straightforward. In the chapters that follow, I will look at some of the problems that arise in measuring subjective welfare, beginning with the problem of validating the observable indicators used for measurement.

5. Chapter Five – Validating Welfare Indicators

5.1. Introduction

I have established that animal welfare is, at least in principle, a measurable entity. In Chapter Two I introduced the indicators of welfare used for measurement in welfare science - those behavioural and physiological measures that are used to determine whether an animal is experiencing good or poor welfare. It is important when using indicators such as these that they are valid – that is, that they actually measure the intended target state. In this chapter, I will discuss in more detail the different types of indicators, and what is required for validation. I will end by proposing a four-step process, using robustness analysis, to validate indicators of welfare, or any similarly 'hidden' scientific target.

In animal welfare science, we aim to measure the welfare of animals under different conditions. This is similar to many other areas of science that are interested in the measurement of particular target entities – objects or states of the world. These entities are measured to look for their values, or to look for changes under differing conditions. The targets can be measured directly, or through the use of indicators. Some examples of direct measurement are measures of length, weight, counting (e.g. number of animals), and behavioural observations. Surrogate measurement occurs when for some reason we can't or don't want to measure the target directly and involves the use of surrogate measures, or indicators. Indicators correlate with the target of measurement, and changes in the target will be reflected by changes in the indicator. An example of measurement by indicator might be something like using a thermometer to measure temperature – we are not measuring temperature directly but instead are observing its effect on the expansion of a liquid. I have created three categories for targets that may not be measured directly, but instead through use of indicators – the target may be a construct, may be difficult to measure, or may be inaccessible to measurement (hidden).

In the first case, the target state may not exist except as a construct or composite of several other states. Health is an example of this sort of system. It is not possible to directly measure 'health' because no such entity really exists. Instead, health refers to something like the absence of malfunction in any one of a number of physiological systems, and measurement of the functioning of these systems (or some target subset thereof) will function as an indicator of the total state of health. Socioeconomic status is another example, where items such as household income, neighbourhood of residence and level of education may combine to form the total

state. Again, direct measurement of 'socioeconomic status' is not possible simply because no such thing exists independently of the set of items that compose it. Intelligence is another potential example, where intelligence might be just some sort of composite of particular types of reasoning skills and knowledge. Measurement of a construct is then some function of measurement of the constituents, weighted to produce an overall score.

In other cases, some targets may be difficult to measure directly and it is simply cheaper or easier to use an indicator measure than one that is a direct measure of the target. An example of this is biodiversity. If biodiversity is considered as something like the number or diversity of species present in a particular area, it is usually not practical to conduct a thorough survey to determine this. Instead, for expedience, something like species abundance or diversity, or variety of vegetation cover may instead be used as a surrogate measure (Lindenmayer & Likens, 2011; Sarkar, 2002). Another example comes up in medical research. As clinical trials often show a significant time lag between the intervention and the outcome of interest, having a surrogate measure that shows up earlier can help speed up the research. Some examples are bone mineral content as a surrogate for incidence of fractures, and blood cholesterol levels as a surrogate for cardiovascular outcomes (Gøtzsche, Liberati, Torri, & Rossetti, 1996).

Finally, we have target states that it is not possible to measure, simply because we do not have direct access to them – they are 'hidden'. These are sometimes referred to as 'latent variables', particularly within psychological models, and cannot be directly observed (Markus & Borsboom, 2013). This can be a result of current limitations in our knowledge or technology, or a feature of the target state. One example of this is temperature, as mentioned earlier – we cannot directly access the kinetic energy of molecules, but instead can measure the resultant expansion or contraction of particular chemicals. Much of the work in the historical sciences, such as palaeontology, by necessity relies on indicators since we cannot have direct access to information about, say, the ecology of the stegosaurus.

Measurement of animal welfare is of this third type – measurement of a hidden target. This is because we can't directly access the mental states that make up welfare for external or objective measurement. In this case we have to rely entirely on indicator measures such as changes in behaviour or physiology. It might seem that welfare could be of the first type – a composite target. This is because welfare consists of the mental states of animals, which are a large and heterogeneous set including, for example, hunger, pain, comfort and curiosity. In Chapter Seven I will argue that these mental states can all be integrated into a single state of welfare through use of a 'common currency' and can be measured in its entirety. Further, even if we consider welfare to be a composite target, the mental states which compose it are still

themselves hidden targets and thus the problem of validation that I will describe will still hold. I will note here that this may not always be the case. With increasing understanding of the neuroscience underlying different mental states (Ledger & Mellor, 2018), we could make a case that were to accept that mental states are identical to neural states then direct measurement of brain activity through neuroimaging is in fact direct measurement of these mental states. If this were true, then welfare would no longer be considered a hidden target, but merely a difficult one. In this case, we could use neuroimaging as a method of validating other indicators of welfare (as will be described in Section 5.3 for difficult targets), that may be cheaper or easier to use in most contexts. As this work is still new, I will not argue further for this claim, but merely flag it as potentially important for future understanding of validating welfare indicators. For now, we can treat measurement of welfare as measurement of a hidden target, which presents a special problem for validation of the indicators used.

When selecting an indicator measure to stand in for some target state, it is important that the indicator is an appropriate one. There are a number of features of indicators that make them more or less appropriate for particular measurement tasks, such as precision (how fine-grained the readings are), lack of bias, ease of use, sensitivity and reliability (the same result always given in the same circumstances). Some take reliability to be the same thing as validity, but I wish to keep the ideas separate here, following Markus & Borsboom (2013). Some measures could be valid in the sense I use, while not being reliable (e.g. if it measures with low precision). For this chapter, the feature I am interested in is validity - that is, our level of certainty that a particular indicator is accurately tracking the target state in question; that the indicator changes when the target changes, and does not change in the absence of a change in the target. "Validity refers to the extent to which the test or instrument measures what it is intended to measure" (Bringmann & Eronen, 2016, p. 28). Where the other features of an indicator may be tracking their quality, validity seems to track whether a particular measure is really an indicator at all.

5.2. Causal and effect indicators

For something to function as an indicator, it must be the case that it reliably correlates/covaries with the underlying state that it is standing in for, and this requires a causal relationship with the target state. Correlation on its own is insufficient – many different and unrelated factors may correlate under particular test conditions (Borsboom, Mellenbergh, & van Heerden, 2004). Borsboom et al. (2004) argue that validity is grounded in causation – a test is valid when there is a causal link between the target and the indicator. Otherwise, measurement is just not taking place. Having a causal relationship between indicators and the

target also gives some other epistemic benefits (Markus & Borsboom, 2013) – it gives us increased reason to believe the correlation will persist under change (holding fixed relevant background conditions), and reason to have confidence in our predictions.

Sidestepping as much as possible the literature on the relation between correlation and causation, we can generally assume that when there is a reliable correlation between two variables A and B, it is either because A causes B, B causes A or there is a common cause for both A and B: "a covariance structure model implies potential nonzero covariances among measured variables if (a) there is a correlational, direct, or indirect path between the measured variables or (b) the measured variables share a common source variable or correlated source variables" (MacCallum & Browne, 1993, p. 539). When we are looking for indicator measures, they will stand in one of these three relationships with the target state – they will either be a cause of the target, an effect of the target, or a mutual effect of a common cause. These three categories of indicators will have different features, both mathematically and pragmatically. Here I will focus on the first two categories, the 'causal' and 'effect' indicators[27], as the 'common cause' type are likely to be much less common. Animal welfare science commonly uses both causal and effect indicators.

Bollen & Lennox (1991) differentiate between "indicators that influence, and those that are influenced by, latent variables" (1991, p. 305) – causal and effect indicators. Effect indicators are those that stand causally 'downstream' from the target state. Changes in the indicator are a result of changes in the target. They can be characterised by an equation such as $Y_i = \lambda_{i1}\eta_1 + \varepsilon_i$, where Y is the indicator, η is the target state, ε is the level of measurement error and λ is a coefficient representing the level of effect of the target on the indicator (Figure 5.1). These indicators are thus determined by the underlying state we want to measure.

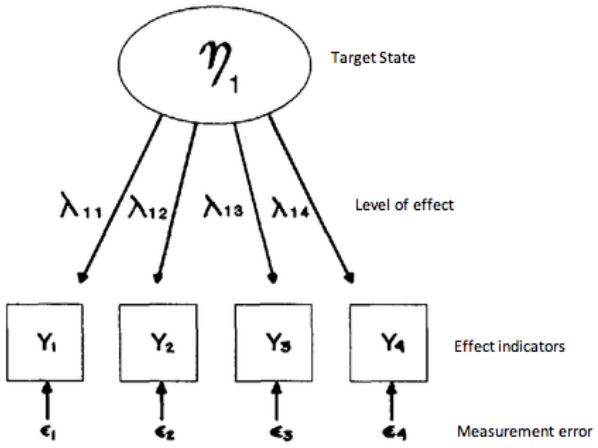


Figure 5.1: Path diagram of effect indicators (modified from Bollen & Lennox, 1991, p. 306)

[27] Sometimes referred to as 'formative' and 'reflective' models (Markus & Borsboom, 2013).

Effect indicators stand downstream from the target state; they are effects of this state. These measures will covary with the target state because changes in the target will cause changes in the measures. There are many examples of the use of effect indicators in science. In medicine, an effect indicator might be white blood cell count – it is an indicator of infection, since white blood cells increase as a result of the presence of foreign micro-organisms. In animal welfare science these are what I referred to in Chapter Two simply as 'indicators', and are often referred to in the animal welfare literature as 'animal-based' measures (e.g. Botreau, Bracke, et al., 2007) or 'output' measures (e.g. Kagan et al., 2015). They are physiological and behavioural indicators that are used to measure changes in welfare, where it is assumed that a change in the indicator reflects a change in the underlying subjective experience. Examples include measurements of blood cortisol levels or approach and withdrawal behaviour towards a particular stimulus. These indicators change as welfare changes.

By contrast, causal indicators stand causally 'upstream' from the target state, where changes in the indicator are a *cause* of changes in the target. Causal indicators are characterised by a more complex equation of the form $\eta_1 = \gamma_{11}\chi_1 + \gamma_{12}\chi_2 + \ldots + \gamma_{1n}\chi_n + \zeta_1$, where χ is an indicator, η is the target state, γ is a coefficient representing the level of effect of each indicator on the target and ζ is a variable representing error or additional causal factors (Figure 5.2). The crucial difference here is that the indicators are *determining* the target variable rather than *determined by* it. Although both types of indicators will correlate with the target state, with effect indicators we are observing the effects of an underlying state, while with causal indicators we are observing the causes of that state.

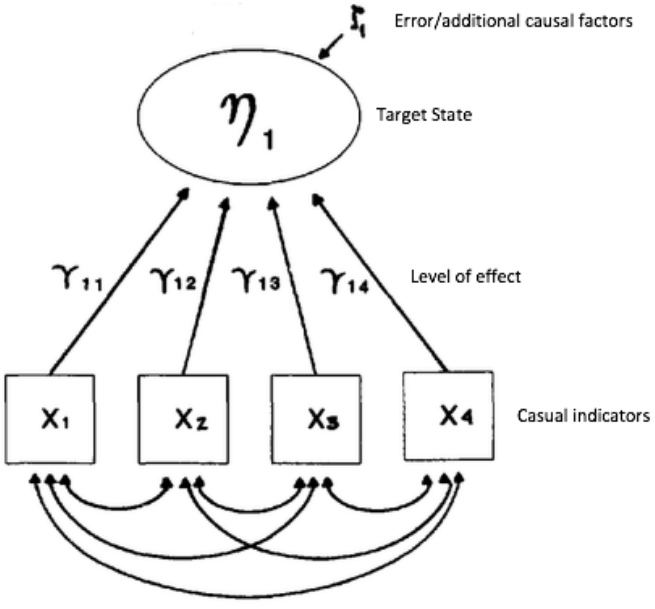


Figure 5.2: Path diagram of causal indicators (modified from Bollen & Lennox, 1991, p. 306)

Causal indicators stand upstream from the target state; they are themselves causes of this state. They covary with it, as changes in the indicator will create changes in the target. Some authors (e.g. Markus & Borsboom, 2013) don't wish to admit causal indicators as true measures at all, as they define measurement as occurring in a single causal direction. In their view, causal indicators predict the target rather than measure it. For my purposes, as long as there is a reliable correlation between the indicator and the target, with an underlying causal relationship, this is sufficient to serve as an indicator - even if not a true 'measure' in this formal sense - as knowledge of the value of the indicator tells us about the value of the target. An example of the use of causal indicators in ecology is the use of rainfall measures to estimate biodiversity, as the level of rainfall will affect the type and abundance of species in an area. In the case of animal welfare, these form what I have called 'conditions' for welfare (see Chapter Two), sometimes also known as 'provisions' for welfare; those things that will cause changes in the subjective states that compose welfare. These are also referred to in the animal welfare literature as 'environment-based' indicators (e.g. Botreau, Bracke, et al., 2007) or 'input' measures (e.g. Kagan et al., 2015). We can measure changes in the conditions for welfare in order to infer changes in the state of welfare. For example, some of the conditions for welfare will be presence of adequate food and water, freedom from disease and adequate mental stimulation. Direct measurement of these conditions can serve as proxies for measurement of welfare itself. These types of indicators are commonly used in animal welfare assessments. Several different measurement frameworks are used to assess the welfare of animals under particular husbandry conditions, such as the Welfare Quality® (Botreau, Veissier, & Perny, 2009) or Five Domains (Mellor, 2016) frameworks, which will be discussed in more detail in Chapter Seven. These sorts of measures are less commonly used in animal welfare science itself.

The third set of indicators would be those that covary with the target state due to the presence of a common cause. For example, we might think that the visual presence of lightning could be used as an indicator for thunder, as both will correlate as a result of the common cause of an electrical discharge. There do not seem to be any examples of this sort of measure being used in animal welfare science, but it is the sort of thing that could potentially be used here, or in other areas that use proxy measures, such as in conservation biology. For example, using population levels of a particular species as a surrogate for overall biodiversity is likely to be of this type, as the environmental conditions that affect biodiversity will also affect the numbers

of the surrogate species, and thus they will covary – an example of such a common cause in this case could be water availability.

There is a difference in how we validate causal and effect indicators. Bollen & Lennox (1991) look at some of the common 'guidelines' in use for the selection of and validation of different indicators. Importantly, the types of procedures that can validate indicators will differ between causal and effect indicators, particularly when considering those that rely on measure of correlation between different indicators. The conventional wisdom has been that indicators that are positively correlated with the same concept, should be positively correlated with one another. Additionally, there has been disagreement about what level of correlation should be considered ideal. They looked at these claims mathematically and found that for effect indicators, they should always be positively correlated (that is, a negative or zero correlation says they are not measuring the same thing) and will be best off when the correlation is as high as possible (as it is a direct reflection of the correlation between each indicator and the target). For example, when considering indicators of animal welfare, we would expect to find a correlation between the change in blood cortisol levels and stress-related behaviour as they are both effects of the common cause of welfare.

By contrast, for causal indicators there is no reason to expect any correlation between indicators, as they work independently, and they thus will not correlate with one another. There is no reason to think that two common causes of a state will covary. For example, when considering animal welfare, there is no reason to think that the availability of food and water would necessarily have any relationship with access to a social group. Importantly, this means that while effect indicators can be, in part, validated through measures of correlation with one another, causal indicators can only be validated through embedding in a model which also contains effect indicators; a point which will be further explored in Section 5.5.

The second claim they examined regarded whether it was necessary for validation to select a variety of indicators. The claim is that selection of a diversity of indicators will 'capture different facets' of the target, and thus use of these different indicators is necessary for complete and valid measurement of the target. They found that this will be true only for causal indicators. In the case of effect indicators there is no reason to require use of diverse indicators, as removal of particular indicators would have no significant impact on the measurement of the target variable. If any single effect indicator is providing a measure of the target variable, the magnitude of change in this indicator will be representative of the magnitude of change in the target, regardless of how many other effect indicators are also used. For cases where this is not true, it is because we have a multi-dimensional concept (a construct or composite target)

which can then be broken down and analysed in terms of each individual facet. Take as an example the composite target of health. Because this is composed of a large number of different components, such as cardiac functioning and immune response, no single measure will be sufficient to capture the entire target, and many must be used in conjunction. But in typical single-target cases, while having multiple effect indicators will help in reliability (as failures in any one indicator will not ruin the results), and each indicator should be correlated with the target, which 'facets' they each measure will not matter.

However, this is not the case for causal indicators. When using causal indicators, it is important that all relevant factors are included in the model. The removal of even one causal factor will have a strong impact on the measurement of the variable, as they are all necessarily contributing to changes in the target. Consider measurement of welfare: if we were to try to measure welfare through causal indicators (welfare conditions), we might include things like stocking density, food availability and social interactions. We would then measure the level of all these variables to determine welfare level. However, if we left out an important contributing condition, such as presence of injury, our results would be inaccurate. We might look at an animal with lots of food and a soft place to sleep, concluding it has good welfare, but have failed to take into account the strong negative effect of pain. Only by including all causal indicators will we get an accurate measure of the target.

The important points to come out of this are as follows:

- Effect indicators can be, in part, validated through measures of correlation with one another
- Causal indicators can only be validated through embedding in a model which also contains effect indicators
- Causal indicators must all be measured to give a reliable measure of the target

It may seem here that this all speaks against the use of causal indicators at all, which many authors seem to agree on (e.g. Markus & Borsboom, 2013). However, in many cases (particularly within animal welfare assessment), the causal indicators are easier to see than the effect indicators and can be used for quick large-scale assessments that effect indicators would be impractical for. For example, trying to do behavioural and physiological assessments on even a small sample of the animals on a farm is going to take far longer than looking for the causal husbandry variables which will impact all the animals and drawing conclusions based on these.

As causal and effect indicators stand in different relations to the target state, they are each going to have their own unique features and drawbacks. Each of the indicator types have different features and what is important is that we accurately identify which type of indicator is operating, in order to use it correctly. As I will detail in Section 5.5, correctly identifying which type of indicator we are using will be crucial for the process of validation.

5.3. Validation

The validity of a test or measure refers to whether or not it is really measuring what it purports to – whether the observed data are actually tracking the intended phenomenon. Validation of indicators is thus testing to ensure that the indicators are tracking the right target state – that the values and changes in indicators are correlating with changes in the target. In particular, we need to establish that one of the types of causal relationships discussed above holds between the indicator and the target. The process of validation will vary depending on what type of target we are talking about and in this section, I will apply some of the discussion of validation to the different categories of targets I introduced earlier. I will show that for hidden targets such as welfare, there is a particular problem for validating the indicators.

In some cases, we may use the presence of adequate predictions as a form of validation (Markus & Borsboom, 2013). The idea being that, if we are able to make such predictions from measurements of the indicators, this gives us reason to think that we are measuring the correct target. The success of the predictions is best explained by the validity of the indicators. Similarly, Bringman & Eronen (2016) suggest that the success of theories that are built using the measurements will add to our confidence in the validity of the measurements. When using the measures, we work with the assumption that the measures are valid, and if the theory is successful, in terms of explanatory and predictive power, this supports the assumption. It is very unlikely we will have accurate predictions based on invalid measures. “What increases confidence in the validity of measurements is the success of the theories that are based on them, and what justifies the success of those theories is their explanatory and predictive power. Testing the latter need not involve the same types of measurements whose validity is in question” (Bringmann & Eronen, 2016, p. 36). There are two ways in which predictions might be seen as a form of validation. The first is that if, using our measurements, we are able to make further predictions (e.g. that given a certain measurement of physiological variable x, we should see behaviour y), then this gives us confidence our measures are valid. This method is not necessarily strong, as there can be other explanations for the success of predictions. Although it may form one strand of evidence (and may be part of a robustness analysis, as

described further on), it does not seem sufficient to stand alone for validation. The second is that the predictions about the measures themselves – the inputs and outputs of our causal model about animal welfare – are accurate; so that targeted interventions on input give the expected outputs. This would give us confidence in the content of the model, and is a similar process to that I will propose in Section 5.5.

For 'composite' targets, there is not really a unique problem of validation. As the target state is simply an aggregate of the measured indicators, it is going to be true by definition that the indicators are measuring the target. There may be separate problems of deciding which features to include within the composite, but this not a validation issue. All that may be required is some modelling to determine the relative weights of the contribution of different indicators to the target and to decide on which aggregation function to use. For example, think of a simple case – the composite 'bachelor'. Bachelorhood is not a natural property, but rather a construct of sex (male) and marital status (single). We do not need to validate to know whether measuring sex and marital status will measure the composite; it must be true. In some cases, our indicators may be indicators of the components of the composite, rather than themselves being components (e.g. using blood pressure as an indicator of cardiac function, which is a component of the composite 'health'). In these cases, we would need to ensure that these were themselves valid indicators. Here, validation of the indicators would proceed according to one of the other two categories ('hidden' or 'difficult'), depending on the particular example.

For 'difficult' targets, validation is relatively straightforward though direct measurement of the target. It is necessary to establish a causal link between the target and the indicator. Borsboom et al. (2004) describe the process of validation as requiring the establishment of a reliable correlation, and providing a theoretical explanation for the causal pathway between the target attribute and the measurement outcomes. This involves first determining the causal direction (whether we have a causal or an effect indicator). This can be done through using theory - embedding within a theoretical framework that explains the causal connections between the target and the indicators (Bringmann & Eronen, 2016; Lindenmayer & Likens, 2011) – or through testing to look for timing and direction of effect.

The second step is establishing a reliable correlation by measuring both the target and the indicator under a range of conditions and (preferably) interventions, where particular conditions will be deliberately varied to alter the target, and the indicator will be checked to ensure it tracks these changes. Where interventions or manipulations are not possible, we can try to use 'natural' experiments; using the results of natural change or randomness (Markus & Borsboom, 2013). If we see a reliable correlation between the target and indictor under a range

of conditions, we have good reason to think that there is a valid causal connection. What we require is correlation over a range of interventions (Markus & Borsboom, 2013).

We thus have a change in a condition (either induced experimentally or tracked naturally), which causes a change in the target state (which we can track through direct measurement), and which then also causes a change in the indicator (which we track through measurement of the indicator) (Figure 5.3). When we observe such correlation between measures of the target and measures of the indicator, we can take the indicator to be valid. The testing may be difficult in practice due to the nature of the target, but does not provide any conceptual difficulties. Once a single indicator has been validated, we can either validate further indicators by also testing against the target, or through correlation with other known indicators (for effect indicators). This process is similar for the one I will go on to describe for hidden targets, but in that case, the central change in the target cannot be measured for comparison and so must be validated another way.

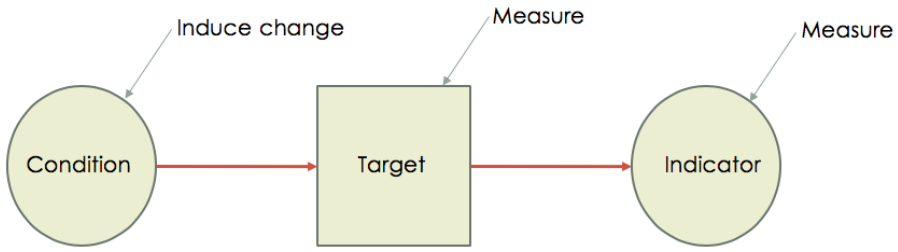


Figure 5.3: Validation of difficult targets

Hidden targets provide a special problem for validation. In this case, we are still assuming there is some real-world state that we are trying to map onto. However, it is completely inaccessible and so we cannot use the above strategy to validate our indicators. It is impossible for us to get correlational data between the target and the indicator, because the target cannot be measured. All we can get is data about changes in the various indicators – we are missing the central link used in validation of difficult targets. Our tests can involve varying particular conditions and looking for changes in the indicators, but we as we cannot get direct measures of the target, we are unable to show the necessary correlations to validate these indicators directly (Figure 5.4).

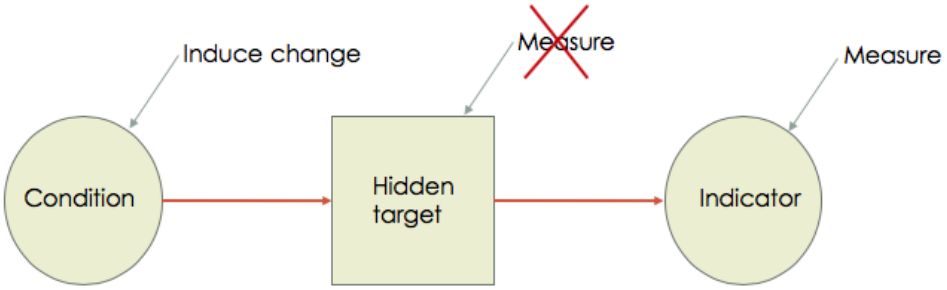


Figure 5.4: Validation problem for hidden targets

Although we can get some degree of validation through correlation between indicators (Markus & Borsboom, 2013), this still requires initial validation of at least one indicator. We cannot start to validate one indicator against another until we are fairly certain the first indicator is itself valid, otherwise we could end up quite far from our target. Here though, there is no starting point at which we can connect an indicator to the target. Schikore & Coko (2013) point out that in these cases, "a set of background assumptions is needed to describe how the unobservable entities bring about the experimental outcomes" (2013, p. 297). Duncan & Fraser similarly argue that "compared to the simple, empirical study of processes that can be observed directly, developing an understanding of unobservable processes involves additional logical steps and assumptions" (1997, p. 23). We are making assumptions about the causal link between the target and indicators, but the problem arises in justifying or testing these assumptions without access to the target. In the following sections I will outline how robustness analysis can help resolve this problem and serve as a test of the assumptions.

5.4. Robustness

Robustness is a concept used in much philosophy of science, and applied in many different contexts. In a general sense, robustness is the property of being "invariant under a multiplicity of independent processes" (Soler, 2014, p. 203). Something is robust where it stays the same, despite changes to the conditions surrounding it. Robustness can give us increased confidence in our measurements and predictions. There are multiple types of robustness discussed within the literature, that correspond to different types of entities and processes. Here I will outline three, along the lines laid out by Calcott (2011). What they all have in common is that there is "one thing [that] remains stable, despite changes to something else that, in principle, could affect it" (Calcott, 2011, p. 284).

The first type of robustness is robustness of models (also known as robust theorems, or derivational robustness). This is probably the type of robustness most commonly discussed in

the current philosophical literature. Robust theorems are those models whose derivation can be supported in multiple ways, using different assumptions. To rule out the possibility that the results of the models are simply an artefact of the idealisations made within the models, multiple models are constructed with different idealisations to look for some common structure or result. The stability of this common structure gives reason to think that it could represent something in the real world, as opposed to just the model.

Robust phenomena are those phenomena which are present under a variety of different (actual or possible) conditions. This will usually be the result of some mechanism which will function across a range of interventions, as opposed to the result of contingent processes.

Measurement robustness (also known as robust detection, triangulation, or multiple lines of evidence) is a form of robustness where multiple, independent means of detection are used to support a claim about the world. It is this type of robustness that is of interest here. In Calcott's terminology, we have an 'R-source' which is the thing in the world which we are aiming to detect, multiple (independent) 'R-variants' which are the methods of detection and the 'R-target' which is the data that these detection methods converge on. When these variants all give the same result (stable target) then we have good reason to think we have the right R source (Figure 5.5). Increasing the number of variants decreases the chance that there are coincidentally multiple sources providing our results.

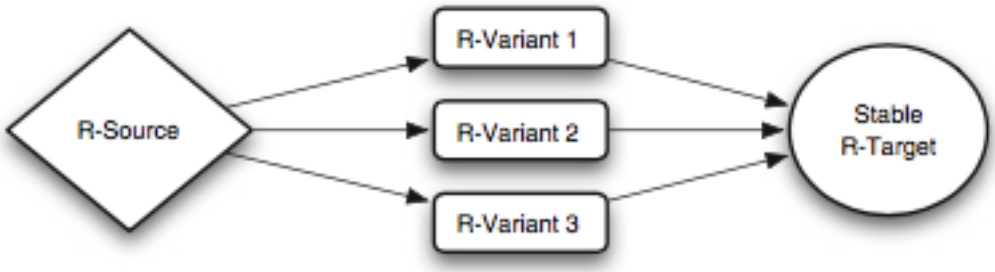


Figure 5.5: An example of robust detection (from Calcott, 2011, p. 285)

In the case of animal welfare, I take it that the R-source would be the subjective welfare of the animal, the R-variants the indicators we are using to try and measure it, and the R-target the resultant conclusions drawn about changes in welfare. As I will discuss in Section 5.5, if multiple independent indicators give us the same conclusions, this gives us reason to believe that they are really tracking changes in welfare.

Robustness analysis helps improve our confidence in particular results – distinguishing 'noise' from 'signal' and "that which is regarded as ontologically and epistemologically

trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting" (Wimsatt, 2012, p. 63). The different conditions under which robustness can hold could be, for example, different sensory modalities, different experimental procedures, different assumptions or different models. Wimsatt (2012) sees robustness analysis as having four stages, and something counts as robust if it comes out as invariant on this analysis.

- To analyse a *variety* of *independent* derivation, identification, or measurement processes.
- To look for and analyse things which are *invariant* over or *identical* in the conclusions or results of these processes.
- To determine the *scope* of the processes across which they are invariant and the *conditions* on which their invariance depends.
- To analyse and explain any relevant *failures of invariance*. (Wimsatt, 2012, p. 62)

For measurement robustness, this would look like using a variety of measurement processes for the same target, looking for the overlap in conclusions and determining the conditions under which this measurement relationship does and does not hold.

Wimsatt (2012) justifies the use of robustness analysis by looking at the impact or errors in different types of reasoning. He describes the traditional scientific method, which aims to establish a small number of fundamental axioms and derive the rest from these. Because there is some small chance of error in any operation in the chain of derivation, long serial chains of reasoning like this will have a much higher chance of error overall. In a serial chain of reasoning, any one step could fail and that will cause a failed result. Small errors in each step multiply, so the more steps there are, the greater the chance of and impact of errors. In Wimsatt's words, "fallible thinkers should avoid long serial chains of thinking" (Wimsatt, 2007, p. 50). If we rely on only one method of measurement, any error in that method will infect our results.

By contrast, a 'parallel' or 'network' setup for reasoning will help each strand reinforce the others, as the chance of error in each one has less chance of impacting the final conclusion and this will decrease further with the addition of more lines of evidence. The more steps there are, the more chance of success in the result: "With independent alternative ways of deriving a result, the result is always surer than its weakest derivation" (Wimsatt, 2012, p. 66). A similar result is found with the application of Condorcet's jury theorem. This showed that "if each member of a jury has an equal and independent chance better than random, but worse than perfect, of making a correct judgment on whether a defendant is guilty (or on some other factual

proposition), the majority of jurors is more likely to be correct than each individual juror, and the probability of a correct majority judgment approaches 1 as the jury size increases" (List, 2013, s 1.1).

We should be more confident in more robust results because of a 'no miracles' explanation – it would be a miracle if a variety of independent tests produced the same erroneous result, so the explanation that they are providing an accurate result is more likely (Soler, 2014). This is particularly important in the sort of tests of hidden targets that are under consideration here: "When targets of triangulation that are not directly accessible are at stake, one needs to resort to often highly complex theories to establish the nature of those phenomena and events in multiple experimental settings" (Schickore & Coko, 2013, p. 304). If we run a preference test on an animal, and also measure blood cortisol levels, both showing high welfare; we would think it strange to conclude that they were likely to be both erring in the same direction when welfare was actually poor, rather than that they were producing a correct result.

It is clearly important to have independent lines of evidence to avoid making the same errors across them all, but achieving true independence may not be as easy as it seems. Many new techniques are calibrated to existing techniques and so are not independent in this sense (Stegenga, 2012). If all our tests are validated simply through correlation with other valid measures (see Section 5.5.4), then the 'serial chains of reasoning' problem will recur where any errors in the first will 'infect' the others. All tests will have similarities and differences along different dimensions, and it will not always be easy to delineate which are the relevant ones for independence (Soler, 2014). For the animal welfare case, as I will outline, we will take independence as requiring use of different background theories and assumptions, but it is not always clear what background theories are in use at a time and what features are shared. We must identify which are the *relevant* ones that need to differ (Stegenga, 2012), as I will advocate in Section 5.5. Additionally, we will not always be able to identify exactly which assumptions underlie each test, and there may be important shared assumptions which weaken independence.

Coko & Shickore (2012) point out that it is difficult to get good robustness in available methods: "it is evident that the epistemic ideal of detection by multiple independent means in the Wimsattian sense is never fully realized in actual experimental practice" (2012, p. 682). In real science, there are usually a limited number of methods available. These are often not fully reliable, and not fully independent. They may also give somewhat divergent results. In assessing the independence of different tests, we need to look at where the possible sources of error may arise – in technique and in background assumptions - and vary so as to reduce overlap

in these. We can think of robustness as a matter of degree – a result is more robust where it is derived by processes that are more reliable, more independent and give greater congruence of results.

Other than failures of independence, there are other ways in which robustness analysis may fail. Schickore & Coko (2013) point out that in practice, robustness techniques may be problematic to implement. "Obviously, in real (scientific) life, multiple means of determination are not always available. Scientists do not always have an array of different instruments and techniques at their disposal, which they can use to investigate an unobservable phenomenon of interest … Even when different kinds of instruments and techniques *are* available, it is likely that the investigators' epistemic situation is less-than-ideal – only a couple of techniques might be available; the techniques might not be completely independent of one another; the different means of determination may be limited in their reach etc." (2013, p. 302). This is particularly true in animal welfare science, where limits on resources can strongly constrain which tests are feasible to perform in particular settings – on-farm assessments of multiple animals, for instance, will not usually allow for individual blood sampling or behaviour profiling. Stegenga (2012) similarly points out that we might not always have access to multiple independent modes of gathering evidence and that even with robust evidence, the hypothesis may be false. It may be a case of 'the best we can do' rather than ideal practice; over time the sciences can improve through trying to overcome these problems as best they can.

When validating indicators, robustness gives us reason to be confident in the link between the indicator and the target state, and that our measured results are representing real changes in our target. As an example, Schickore & Coko (2013) describe the derivation of Avogadro's number (a number representing the number of particles within a standardised unit of a substance) as a means of the use of robustness analysis to determine a hidden target. There is no direct way of measuring this target due to the unobservable size of the constituent particles. Instead, thirteen different experiments were used as indirect measurements. The experiments required a set of background assumptions to "describe how the unobservable entities bring about the experimental outcomes" (Schickore & Coko, 2013, p. 297). As each experiment used different assumptions, they served as mutual tests of these assumptions, as well as of the target. The large number of independent lines of evidence pointing to the same answer despite different background assumptions gave stronger confidence in the results. In the next section, I will detail a similar process for using robustness to validate indicators of hidden targets.

5.5. Application to the validation problem

As described above, there is a validation problem for hidden targets such as welfare: as we cannot access the target, we have no means of directly establishing a correlation between the target and the indicators. Instead, we must make some assumptions about the relationship between the target and indicators, and these assumptions may not be justified. The only information we have access to regards the variation in causal and effect indicators, and we only have assumptions about their link to the target state. It is important then that we find means to justify these assumptions. Here I will describe a process by which assumptions can be justified through theoretical plausibility, and tested through the collection of multiple independent lines of evidence that support the assumptions made – robustness analysis. The procedure I propose for the validation of indicators follows a general four-step framework:

1. Make a (plausible) starting ***assumption*** relating a causal or effect indicator to the target
2. Test for correlated variation in an indicator ***of the other type***
3. Repeat tests for the indicator using ***different assumptions*** to give robust results
4. Use validated indicators as starting point to test others

Below, I will detail what is involved in each of these steps, how they fit into the process and how they will help with the problem of validating hidden indicators.

5.5.1. Make a (plausible) starting **assumption**

The first step in validating an indicator of a hidden target is to make a (plausible) starting assumption relating a causal or effect indicator to the target. An assumption of this type is necessary, as we cannot in the beginning have any knowledge about the relationships between the target and its indicators. Even in the 'difficult' case, as described earlier, we must still begin with a similar assumption. The difference in this case is that we are not then immediately going on to test this assumption as we would with the difficult targets, but are instead using the assumption as a base to test other hypothesised target-indicator relationships.

The assumption made will be of the form 'I (indicator) is causally related to T (target)'. 'I' in this case may either be a causal or an effect indicator. We can start by selecting a causal indicator we think is directly related to the target, such that any changes we make to the causal state will cause a change in the hidden target state. Or we can select an effect indicator we think is related, such that any changes in the target state will cause a change in the effect indicator. Whichever indicator we are making the assumption about, we can call the 'set' indicator (Figure 5.6). In any particular test, we will hold this assumption fixed, using it as a basis to test

other indicators (as described in Step 2); but overall we gain support for this assumption through use of different tests (as described in Step 3).

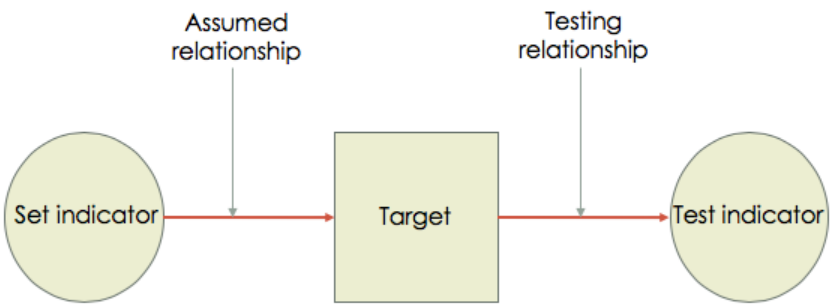


Figure 5.6: Set and test indicators

An example of this type of technique can be seen in the testing of cognitive bias measurements. The details of cognitive bias testing will be discussed in Chapter Seven, but it involves looking for an optimistic or pessimistic judgement bias in animals. When trying to validate the link between the results of these tests and welfare, researchers attempt to induce a change in welfare through a change in conditions, relying on an assumption about the link between a particular condition (such as barren or enriched housing) and welfare (Clegg, 2018). The use of this assumption for testing is in line with the first step described here.

The idea of setting an initial assumption has been occasionally mentioned in the animal welfare literature. For example, Duncan & Fraser (1997) state that we have to "postulate that these unobservable phenomena [the feelings and emotions of animals] have certain properties, are affected by certain influences, and in turn have certain effects that we can observe" (1997, p. 23). It is also often implicitly used in testing or validating indicators, such as in the example described above. However, the need for and use of such assumptions is rarely discussed or given explicit justification. In particular, the further link between this step and the robustness analysis described in Step 3 has not been explored.

One important feature of this step is that we want the starting assumption to be plausible. This means that we must have some good reason to think the assumption is true, or at least justified, independently of the results of these tests. Plausibility of this type is usually achieved through embedding within an accepted theoretical framework: one that can give a description or explanation of the assumed causal relationship between the target and the indicator. If the theoretical framework is a well-accepted and well-supported one, we have good support for the

plausibility of assumptions that fit within it. This is a role for existing data and accepted theory in the relevant area (Markus & Borsboom, 2013). In animal welfare science, this will be sentience research. The relevant theoretical framework is scientific understanding of the neurophysiology of mental experience, as well as the mechanisms that underlie processing of causal indicators and expression of effect indicators (Beausoleil & Mellor, 2017), and this link will be discussed further in Chapter Eight.

One method sometimes used in animal welfare science for validating indicators is through analogy with humans (G. Mason & Mendl, 1993). Here we find particular indicators of welfare changes in humans (such as heart rate increases, or changes in blood cortisol) and make the assumption that similar changes seen in animals represent the same sorts of welfare experience. This method is only as strong as the assumption of analogy between humans and animals. I will discuss such cross-species comparisons of welfare in Chapter Six, but this assumption is only likely to hold in cases where we think there is relevant similarity in the underlying mechanisms linking the indicator to welfare change.

5.5.2. Test for correlated variation in an indicator **of the other type**

After setting a starting assumption, the second step in validating hidden-target indicators is to test for correlated variation in an indicator of the other type. This means we measure changes in the 'set' indicator and then look for correlated variation in the indicator we are interested in testing – the 'test' indicator (as per Figure 5.6, above). As described in Section 5.3, in a standard 'difficult' target case, we are able to directly test for correlation between the indicator and the target. In the hidden case, as this is not possible, we must instead test for correlation between the test indicator and an indicator we are assuming stands in another position in the causal pathway. If we are assuming that variation in the set indicator reflects variation in the target state, then correlation between the set and test indicators should directly reflect correlation between the target and test indicator. This gives us good reason to think that *given the truth of our starting assumption* then the test indicator is a valid indicator of the target. This may seem like a large caveat, if we don't have strong reason to believe in the starting assumption. Our reasons for this will derive partially from the plausibility described in Step 1, and also through the robustness testing that will be described in Step 3.

If the set indicator is a causal indicator and the test indicator an effect, then these tests will ideally take the form of deliberate manipulations on the set indicator, looking to induce associated variation in the test indicator, which stands causally 'downstream'. For example, for tests in animal welfare we can make changes to food availability, or provision of environmental

features or even pharmacological interventions, using drugs known (or assumed) to cause changes in welfare-relevant mental states; then look for measured changes in effect indicators such as heart rate or behaviour. In some cases, where direct manipulation is not possible (for example, some ecological measures), we can approximate the same sorts of tests by collecting data in a variety of situations under which the set indicator varies naturally, and look for correlated variation in the test indicator. If the test indicator shows variation alongside the manipulations of the set indicator, this can be presumed to be a result of the changes in the set indicator (e.g. food availability) causing changes in the target (welfare), which then cause changes in the test indicator (heart rate) (in the next section I will discuss how we can rule out other potential explanations of observed correlation).

If the set indicator is an effect indicator, the tests will be of roughly the same form, but the inferences taken from them will be different. We cannot simply reverse the tests, as the causal direction runs the other way and manipulations on the effect indicators will not necessarily have any corresponding changes in the causal indicators. We could not, for instance, induce a change in heart rate and take this to tell us anything about the environmental conditions for an animal. Instead, as described above, we would still carry out manipulations (or tests under natural variation) of the causal indicator and look for correlated changes in the effect indicator. However, what would change would be our interpretation of the results. Given that in this case the effect indicator is the set indicator, then when we see correlated variation between indicators this would count as validation of the causal indicator, as it is our test indicator.

An example of this process of validating a causal indicator might be investigating whether type of handling correlates with welfare changes in sheep. Here, our test indicator is the causal indicator of handling type. In this case, we would set up tests of different types of handling (human vs. machine) and then as our set indicator use effect indicators such as heart rate changes (either previously validated, or with an assumed relationship to welfare, as per 5.5.1) to measure whether a change in welfare is taking place. If a correlation is found between handling type and changes in heart rate, this helps validate the causal indicator; telling us that handling type is a valid causal indicator for sheep welfare. Where we have a cause affecting a target, which in turn affects the indicator, this time the causal link between the target (welfare) and the effect indicator (heart rate) is based on an assumption, which can then be used to test and validate the link between the causal indicator (handling) and the target (welfare).

As discussed in Section 5.2, it is important that these tests are done with an indicator of the other type than the set indicator – that is, if the set indicator is causal than the test indicator should be effect, and vice versa – as they stand in different positions in the causal pathway.

Particularly, recall that while effect indicators can be, in part, validated through measures of correlation with one another, causal indicators *can only be validated through embedding in a model which also contains effect indicators*. This means that testing of causal indicators can only ever be done using effect indicators, as in the example of sheep handling above.

The reverse is not always true. Effect indicators can be validated through testing for correlation with one another. For instance, we could validate the effect indicator of heart rate against that of changes in ear position, by looking for correlations between the two. However, this will only really work when using an effect indicator which is already known to be valid (see Step 4 for more on this). In this stage of assumption-based testing, if both the set and the test indicators are effect indicators, an additional assumption will be required for testing. Although effect indicators will correlate, this is due to them being effects of a common cause (the target) rather than a direct causal link. That means that direct intervention on an effect indicator will not necessarily show a change in other effect indicators. Correlated variation will only occur through interventions on the common cause target state, which requires the use of causal indicators. If these causal indicators are not already validated (in which case, we are again at Step 4), then we are making an *additional* assumption about the relationship between causal indicator and target, that will weaken our tests. Thus, all testing at this stage should be of indicators of the other type to that used in the assumption.

This can be seen through the following example. If we were looking to validate heart rate against ear position, we would first make an assumption about the relationship between ear position and welfare (as per Step 1), making this the set indicator. We would then look for correlated variation in heart rate, making this the test indicator. However, in order to induce change in these indicators, we could not simply intervene on our set indicator, ear position. We could manually change the position of an animal's ears without expecting to see any correlated variation in other indicators of welfare (unless the animal were to become annoyed by our interventions). We could only expect correlated variation if we were to induce change through intervening on the common cause of both indicators, i.e. welfare. Any intervention on welfare will necessary go through one of our causal indicators – one of the conditions for welfare, such as provision of social companions. But by introducing this causal indicator to the testing process, we must either already know that this indicator is valid (which takes us to Step 4), or we must make an additional assumption regarding the link between social companions and welfare. Having two assumptions within the one test (about both the link between social companions and welfare, and ear position and welfare) will weaken our confidence in our results and in the validity of the indicators.

In some cases, we may not get the expected results in these tests. For instance, when setting a causal indicator we think would be a condition for increased welfare (such as providing companions for a social animal), we might then find that the indicators go the opposite way, perhaps telling us that they don't enjoy this condition, despite our initial assumption that they would. Here we need to decide when we would abandon our initial assumption, or when we would reject the indicators. In these cases, we would need the robustness analysis described below – we can test both the assumption and the indicator under different conditions to see which one holds and which does not.

5.5.3. Repeat tests for the indicator using **different assumptions** to give robust results

As flagged earlier, there is a weakness so far with the described procedure. That is, that our confidence in our results is only as strong as the starting assumption we have made. This is the role of the third step - to increase our confidence in the results, and thus in the validity of our test indicator, through use of multiple tests, each using different starting assumptions. "A variety of sources of evidence that test an assumption in different ways will generally offer stronger validity evidence than a single line of evidence that tests the assumption in just one way" (Markus & Borsboom, 2013, p. 587). This is robustness analysis, as described in Section 5.4. Animal welfare science often uses a similar process for validation as the one I have outlined so far – to subject animals to a presumed stressor, measure the corresponding effects and then take these to be valid indicators of stress that can then be used to test for stress under other circumstances (G. Mason & Mendl, 1993). However, what this process misses is the explicit identification of the assumption and subsequent repetition of the tests in order to test this initial assumption and build robust results.

As discussed in Section 5.4, a key feature of this sort of analysis is that these lines of evidence are independent (or at least, as much as they can be). "If there are several independent ways of measuring something, this increases our confidence in the measurements" (Bringmann & Eronen, 2016, p. 29). There is a lot of discussion about what independence means in this context, but the general characterisation is one which defines independence in terms of chance of the same types of error occurring. That is, that the differences between the types of tests tries as much as possible to minimise the overlap in the same type of error, so errors are independent and robustness helps build our confidence in the result as described earlier. In this case, what is most important is that the tests rely on *independent background assumptions*. Although all tests will share at least some assumptions, here what matters is that "any *problematic* or *unconfirmed* assumptions should not be shared by the different ways of access" (Eronen, 2015,

p. 3969). All assumptions may be problematic or unconfirmed to some degree, but it is the most problematic background assumptions - those which we are uncertain about - that we primarily need to account for.

In the case of the procedure for validating indicators of hidden targets, assumptions will be problematic when they are assumptions about an unmeasurable causal relationship between the set indicators and a hidden target state. If we repeat the tests using different background assumptions, it means that the collective results do not rely on any one assumption in the way that a single test would: "A variety of sources of evidence that test an assumption in different ways will generally offer stronger validity evidence than a single line of evidence that tests the assumption in just one way" (Markus & Borsboom, 2013, p. 587). In the single test case, as described in Step 2, our confidence in the validity of the tested indicator is only as strong as our confidence in the truth of our starting assumption. If the assumption fails, so too do our results. However, when we have a set of results relying on a set of different starting assumptions, as described earlier, it becomes increasingly unlikely that that the observed results are due to some mistake or failure in these starting assumptions, and instead we have good reason to think they reflect a real valid causal link between the target and the test indicator.

These assumptions should differ in that they use different set indicators, while still testing a single test indicator. For example, we might test an effect indicator of animal welfare first by using a set causal indicator of food quality, which we assume to have an effect on welfare, and then by using the causal indicator of access to social companions. As these two types of causes are different from one another, and the mechanism by which each is thought to affect the target state are different, we would have sufficiently independent assumptions to give robust results. If the tested effect indicator showed the right kind of variation in both cases, we would have good reason to think it is a valid indicator of welfare.

As mentioned in Step 2, one of the problems in running these sorts of tests is ruling out alternative explanations for the observed correlation between set indicator and test indicator. Testing 'difficult' targets just involves inducing a change in the condition and comparing resultant measures of the target and the indicator for correlations. The process is transparent, in that we can see that changes in the indicator are a result of changes in the target state. This is not true for hidden targets, where the process is opaque and although we can vary particular conditions and look for changes in the indicators, we are unable to validate these directly to the target as we cannot get direct measures of the target.

The problem is that the results of our tests on hidden targets (covariation between changes in the conditions and changes in the indicator) could be a result of one of three different scenarios:

- The change in condition is directly causing a change in the indicator rather than doing so indirectly through changes in the target. That is, the indicator, rather than serving as a measure of the hidden target, is instead serving as a measure of the condition itself (Figure 5.7). This may be because the target in this case does not really exist, or because the indicator is just not related to it.

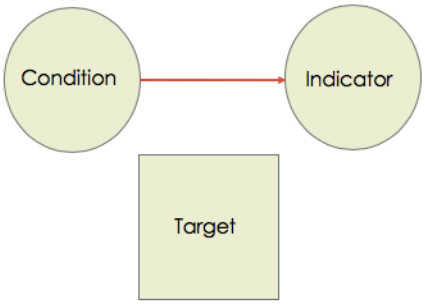


Figure 5.7: Case 1- Indicator measuring condition directly

- The indicator is actually measuring some other target that is affected by the condition. Here we have the condition causing changes in some other target, that the indicator is then tracking (Figure 5.8). Again, this may be because the intended target is absent, or simply because the indicator is the wrong one for the target in this case.

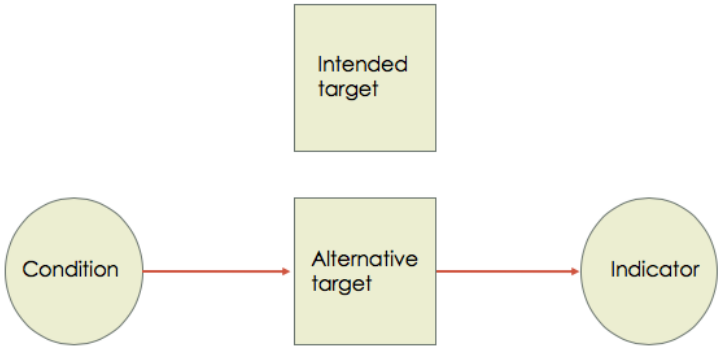


Figure 5.8: Case 2 - Indicator measuring an alternative target

- The indicator is tracking the intended target, and observed changes in the indicator are a result of changes in the target under varied conditions (Figure 5.9). This is the case we are hoping stands when validating indicators.

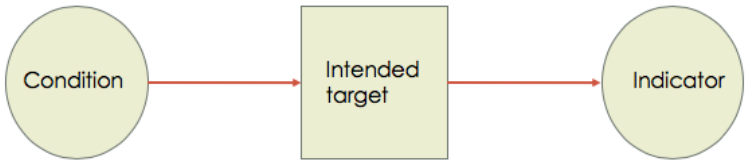


Figure 5.9: Case 3 - Indicator measuring the intended target

What is important is that we have a method of distinguishing between these possible scenarios. This is where robustness analysis can assist. In the first instance, this requires use of accepted theory to create a likely model of the relationship between a set of conditions and the target, and the target and some indicators. We can then use different conditions to strengthen our testing and create more robust results.

Take the first scenario – that the indicator is measuring the intended condition. This becomes much less likely if we repeat the testing under changes in some other condition that we have reason to think impacts on the target. It may be the case that condition A directly affects the indicator, but if we then bring in condition B, which is quite different from condition A, it becomes less likely that observed changes in the indicator still reflect both these conditions directly rather than reflecting the effect of the conditions on the hidden target. Figure 5.10 shows this line of reasoning employed by Miller (cited in Fraser, 2009), who argued that the most parsimonious explanation for the links between several different causal and effect indicators will often be use of an intervening variable, rather than multiple independent stimulus-response links. The more lines of evidence we add, in terms of different conditions, the less likely the direct relationship is the explanation. This also links back to the earlier discussion of predictions as a form of validation – under the first model we could not make predictions about changes in the effect indicators, except through changes in the listed causal indicators. In the second model, we could introduce some new causal factor we think would influence thirst (such as an increase in ambient temperature) and expect to see changes in the effect indicators. Where such change is observed, we have further reason to believe this explanation.

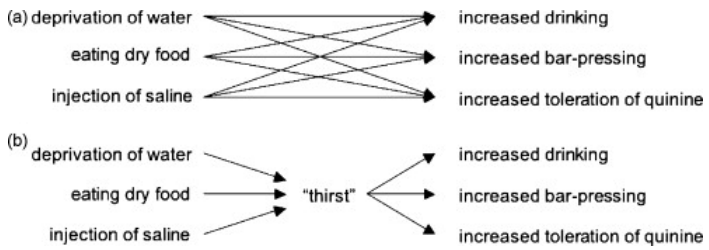


Figure 5.10: Use of an intervening variable to simplify explanation of links between indicators (from Fraser, 2009, p. 114)

In the second scenario, a similar approach also works. We run tests under a range of conditions and look for changes in the indicator. If changes in the different conditions all result in changes in the indicator, we have good reason to think that they are working through some intermediate state rather than directly on the indicator. But even then we still might not have reason to think that the intermediate state is our intended target. Perhaps there are other potential states that could serve this role just as well. We cannot immediately distinguish between the second and the third cases. It is true that there is not necessarily going to be any set of tests that will definitively differentiate between these scenarios. However, there are two ways we may strengthen our belief that we are measuring our intended target.

The first is through targeted testing – putting in place tests which are specifically designed to rule between the effects of the intended target and some other hypothesised state that could be doing the work (Schupbach, 2016; Soler, 2014). Schupbach (2016) even defines independence of tests in these terms - "each newly cited detection in such cases [of] RA [robustness analysis] differs from previous means of detection if it is capable of ruling out another class of competing potential explanations of the result left standing by these previous means" (Schupbach, 2016, p. 13). This will not give us the result that that intended target is the definite cause of our results, but can rule out the most probable alternatives. For example, say we observe a change in availability of favoured foods correlating with a change in indicators of preference behaviour. We could have two alternative explanations – that the change is occurring through a change in welfare experience, or that it is instead mediated through some other non-conscious reflex mechanism. If we had a way of intervening that would affect experience but not non-conscious processing (e.g. use of mood-altering drugs), then this would give us reason to prefer the former explanation.

The second way of strengthening is the use of theory. If we can tell a plausible causal story about the mechanisms through which changes in the conditions affect the target state and go on to affect the indicator, and this causal story fits well with our best established theories in the area – especially if it can be independently tested - then we have reason to prefer this interpretation over others. Think again of the example above. We could use our current best theories about the relationship between subjective experience and behavioural motivation (e.g. Ginsburg & Jablonka, 2019) to lend support to the welfare interpretation. This gives us a way of deciding on the third explanation over the second. If the relationship between the indicator and the target holds under a range of conditions, then we can be confident in our assumptions and accept the indicator as valid.

5.5.4. Use validated indicators as starting point to test others

Once we have used the three steps above to validate an indicator, we can repeat for as many indicators as we wish to. However, we can also make the process simpler by using the validated indicators to test others ("concurrent validity" - Botreau, Bracke, et al., 2007). The validated indicator would then take the place of the set indicator used in the starting assumption. We can use validated causal indicators as starting points to test effect indicators, and validated effect indicators to test casual indicators. Correlation between a validated indicator and a test indicator tells us they are likely to be mapping onto the same target state, and thus that the test indicator is also valid. Additionally, because of the correlation between effect indicators, as discussed above, we can also use effect indicators to test one another. Although this will still require assumptions for causal indicators (or use of validated causal indicators), correlation with other validated effect indicators is a strong additional line of evidential support.

As an example, Panksepp (2005) suggests that we could use results from human tests to validate behavioural indicators of welfare in other animals. The suggestion is that we could take neurochemical agents known through self-report to cause changes in emotional states in humans (e.g. increasing or decreasing joy or sadness). Taking the assumptions that self-report is a reliable enough guide to human experience, and that neurochemical agents are likely to act the same way in other similar brains (i.e. containing similar relevant neural pathways), we can take these causal indicators as valid and then use manipulations in these to test for correlated changes in the effect indicators, such as playful behaviour or vocalisations. Where correlated changes are seen, we have good reason to think that these indicators are valid for the changes in welfare.

Because this method does not rely on starting assumptions, but on established validated indicators, it therefore doesn't require the third step of multiple testing for robustness. Our confidence in the validated indicator gives us confidence in the results of the tests. However, in many cases it will still be valuable to run multiple tests. Although the initial testing process may give us confidence in the validity of our tested indicators, it does not give us certainty, particularly as the sensitivity and scope of the indicator response may vary. And any mistakes in that process would be amplified if these are then used as the basis for testing others – recall Wimsatt's 'chain of reasoning'. Running multiple independent tests, using different assumptions or other (independently) validated indicators, gives us increased confidence that there are no such mistakes causing an impact on our results, and thus is still a useful step in testing. Our higher level of confidence in previously validated indicators, as compared to confidence in assumptions about the set indicators might be reflected in the need for fewer

lines of testing than we would need initially, but it would usually be advisable to have more than one.

5.6. An example of the process

The framework I've developed here gives us a method for validating welfare indicators. Although something like this method, or components of it, are often used in animal welfare science, the process has not yet been made explicit. An example of the use of a method like this can be found in Briefer et al. (2015) in their work in developing indicators to measure both the valence (positive/negative) and strength of welfare in goats, in which they used multiple tests to increase confidence in the results. Here I'll detail how this study fits into my framework, showing its application and benefit.

1. *Make a (plausible) starting **assumption** relating a causal or effect indicator to the target*

The initial assumptions related causal indicators to goat welfare. Goats were placed under differing conditions that were assumed to have positive or negative effects on welfare – access to food or social groups versus being unable to access food and experiencing social isolation. These assumptions – that, for example, access to food improves welfare, and seeing but being unable to access it causes reduced welfare – seem fairly plausible and are based on expert knowledge of the animals.

2. *Test for correlated variation in an indicator **of the other type***

While varying the causal indicators, the goats were then assessed for changes in various potential effect indicators such as ear position, type of vocalization and change in heart rate. This is a case of a causal (set) indicator affecting the target, which in turn affects the effect (test) indicator. Those effects that varied reliably with the changes in conditions were supported as valid indicators.

3. *Repeat tests for the indicator using **different assumptions** to give robust results*

This experiment used two different set indicators for testing – food availability and social contact - each with its own independent assumption relating these to welfare. This meant that even if one assumption was erroneous, it was unlikely that both would be. It is far less likely that the observed effect indicators were indicating some other factor; in the framework, welfare is the most likely link between the food and social conditions.

4. *Use validated indicators as starting point to test others*

Although not used in this experiment, the indicators that were tested and validated here could form the basis of future testing of both causal and other effect indicators.

5.7. Conclusion

If we take animal welfare to consist in the subjective mental states of animals, we have a measurement problem as these states cannot be measured directly; instead we must use indirect indicator measurements such as behaviour or physiology. The measurement indicators used must be valid ones – that is, it must be the case that the indicators are measuring the intended target rather than some other target (or nothing at all). This requires a causal relationship between the target and the indicators. This causal relationship can go in either direction – the indicators can either be causes or effects of the target state. These two types of indicators need to be tested against one another for validation. Here I have proposed a four-step approach for validating these indicators, which requires making some assumptions about the causal links between the target and the indicators, and testing these assumptions using multiple independent lines of evidence to increase our confidence in them via robustness analysis. Indicators showing a reliable correlation throughout testing can then be taken to be valid measures of the target state. This framework could also apply in other cases of measurement of 'hidden' target variables, in particular other properties of animal minds, such as consciousness or cognitive abilities.

6. Chapter Six – Intersubjective Welfare Comparisons

6.1. Introduction

As already discussed, animal welfare science is the scientific study of the welfare of animals. Scientists use different indicators, such as changes in behaviour or of physiological variables such as heartrate, to measure the changes in welfare under different conditions. I have argued that we should understand welfare as consisting in the subjective experience of animals – their positive and negative mental states. The private nature of this subjective experience raises particular issues for measurement. In Chapter Five I looked at how we can validate indicators of welfare, as a hidden target state. Here I will examine another such issue: whether we can justify making intersubjective comparisons of welfare – that is, comparing the welfare of different individuals.

When measuring the welfare of any individual animal, we are quantifying its welfare – in some sense defining the 'units' of welfare for that individual. We use indicators to gain information about the states of subjective welfare – both their valence (whether they are positive or negative) and their intensity (how strongly the animal feels them). We can also gather information about how the indicators reflect welfare (i.e. to what degree the indicators change when welfare changes), how strongly welfare changes under different conditions, and the boundaries of maximum and minimum welfare for that individual, both in terms of under what conditions this will occur and how strongly the indicators will reflect this.

We are able to measure subjective states of an individual with (hopefully) some degree of accuracy. We can certainly provide an ordinal listing of states of welfare, in which we rank some as higher than, lower than, or the same as, others. For example, if looking at whether sheep prefer handling by humans or machines during shearing, we could measure indicators like their behavioural responses to each situation (such as approach/avoidance), ear positions and changes in heart rate to determine which situation provides better welfare. Based on these indicators, we could then rank the conditions accordingly. Additionally, we may think that we can even, to some degree, quantify the magnitude of differences in welfare through the magnitude of difference in the measured indicators. If, for example, we see twice as much cortisol in blood sampling, we might infer that that animal is twice as stressed, though these sorts of inferences must be taken with caution. We can create a rough ordering of sets of conditions for an individual, from those that create the best welfare down to those that create

the worst and understand how the measured indicators reflect these states. Though this is somewhat straightforward for any individual animal, it becomes much more complex if we want to compare this information between different animals. We may run into the problem of incommensurability, as discussed in Section 4.3. In this chapter, I will look at this problem of intersubjective comparison, and describe one way to overcome it. In Section 6.2, I will describe the different types of situations under which we might want to make comparisons and the types of comparisons they require. In Section 6.3 I will use a worked example to outline the problem of intersubjective comparison in more detail. I will then go on to provide my solution to this problem in Section 6.4, looking too at some potential problems, and finally in Section 6.5 look at how the solution applies to the different types of comparisons we might want and what we can do in cases where it will not apply.

6.2. Types of animal welfare comparison

There are three primary areas in which we might want to make intersubjective comparisons of animal welfare. The first is in making management decisions for the life of any particular animal; in weighing the trade-offs that might be required for making their lives go well overall. There are often times when we might want to inflict a negative experience in order to enable a positive experience, and need to compare their intensity to ensure the trade-off is justified. For example, whether we should put a young animal through a painful medical procedure in order to prevent health problems in later life, or cause frustration through denial of a favourite food type that could cause future obesity. These comparisons will be *intra*subjective, and will require comparative information about the degree of harm and benefit of these different actions for a single individual.

The second area is in using animals for experimental welfare science. In animal welfare science, the welfare of animals is studied under different conditions. This requires taking groups of animals and placing them under conditions such as different feeding regimes, environmental parameters or social groupings. Measurement of behavioural and physiological indicators is then used to draw conclusions about the effects of these conditions on the welfare of the animals. Importantly here, the tests are performed on small groups of animals, with results that are assumed to be relevant to other members of the species. Often, the different experimental conditions will be performed on different groups, and the results from each group compared. Here we have two ways in which intersubjective comparisons are necessary – in making comparisons between experimental groups and in extrapolating results to other

members of the species; both of which will typically occur within species. These comparisons will usually be *inter*subjective, but *intra*-species.

Finally, there are comparisons made for institutional (or individual) decision-making about the distribution of resources between animals in order to achieve the best overall outcomes. In most situations, there are limitations to resources, such as money and husbandry time, and these sorts of decisions about distribution under scarcity will require making trade-offs between provision of benefits to different individuals or groups of animals; with some animals necessarily being privileged over others. For example, consider a zoo manager trying to decide whether to use part of the zoo's budget to provide a lion with underfloor heating in its den, or a group of lungfish with new exhibit furniture Decisions of this type require intersubjective comparison of the potential harms and benefits the different animals may experience, and comparisons will frequently be inter-species.

All three of these applications require meaningful comparisons of welfare in order to make the requisite inferences and decisions. If such comparisons cannot be made, this will impact many areas of welfare science and animal husbandry. However, making such comparisons does seem problematic. It is not immediately obvious that we could be justified in thinking that different animals have experiences of welfare that are comparable in this way. In the next section, I will explore why this is the case.

6.3. The comparison problem

In order to make comparisons of welfare between different individuals, we must have a common scale onto which we can map their 'units' of welfare. That is, to be able to say something about how many measured units of welfare for one animal are equivalent to a number of units of welfare for another. Think of temperature measurement – although we have different temperature scales (Fahrenheit and Celsius) we are easily able to convert units in one into units of the other. If there is no such common underlying scale for welfare, or we cannot gain the required information to do the necessary conversions, then we will not be able to make intersubjective comparisons of welfare.

On the surface, it seems we have good reason to be sceptical of this possibility. It is entirely plausible that different individuals could experience vastly different levels of welfare, and that they do not reflect these differences in measurable indicators. We see versions of this in real-world situations – it seems, for example, that people can vary quite a lot with respect to pain thresholds and the degree to which they express pain reactions, and this can make it very difficult to compare pain experience between individuals. It may be the case that some animals

have reduced affect – that their highs are not particularly high nor their lows particularly low. The intensity of all their experiences may be small. Some animals, by contrast, might be capable of reaching far higher heights and far deeper lows – their intensity is just greater overall. If it is possible (which it seems to be) that such individuals could exist without showing different indicator responses, then as the underlying subjective states are private and inaccessible, we might never know whether or when they occur, and this undermines our ability to trust such comparisons. Making intersubjective comparisons of welfare will then require further justification. This problem has been widely discussed in the literature on human welfare, and the approaches taken there will be discussed in detail in Section 6.4.1. Here I will make the problem clearer through use of an example, and discuss the types of comparison the problem may apply to, before turning to my proposed solution.

6.3.1. An example

The problem of intersubjective comparison can be illustrated with this example. I used to work with two otters – Sneezy and Paddy. Imagine that each are given some yabbies, and their behavioural and physiological responses measured – say, the amount of vocalisation, and changes in heart rate. We see that Paddy shows a higher level of response on all measured indicators than Sneezy does. What conclusions can we draw from this? There are a few options:

- Paddy enjoys receiving yabbies more than Sneezy does and her indicators reflect this
- Paddy and Sneezy enjoy receiving yabbies the same amount, but Paddy is more inclined to 'display' her pleasure in measured indicators than Sneezy is
- Paddy enjoys receiving yabbies *less* than Sneezy does, but her reactions are much higher

There is also a possibility that Paddy actually dislikes the yabbies, and her indicators are instead demonstrating the strength of this dislike rather than enjoyment. Here, I will introduce the distinction between valence and intensity. Indicators such as heart rate and level of vocalisation may tell us how strong the reaction is (intensity), while there are other indicators, such as behaviour and type of vocalisation, that tell us whether the reaction is positive or negative (valence). Here, I will take it as fixed that the otters have the same indicators of valence – that is, that they will show the same signs of positive and negative experience. This is plausible, due to their shared evolutionary history and development. Here then, we are only

interested in whether the measured levels of response intensity correspond to the same intensity of experience.

The problem is, that there does not seem to be any evidence here that gives us reason to prefer any of these alternatives over the others. There are multiple conclusions, all equally consistent with the observations and there is not any reason to consider one of the possibilities more likely than the others. This is not just hypothetical – within-species differences in individual behavioural and physiological responses to positive and negative stimuli are common (e.g. Boccia, Laudenslager, & Reite, 1995; Izzo, Bashaw, & Campbell, 2011; Manteca & Deag, 1994), and it is difficult in these cases to determine whether or not results imply a welfare difference.

We cannot escape the problem using tests for the strength of preferences, as they will hold the same problem as other indicators. We could look to see whether Paddy will work harder for her yabbies than Sneezy would, but the results we see only give us comparative information for each individual – how much they value yabbies vs. work – and can't be used to make intersubjective comparisons without assumptions such as that they both find work equally unpleasant. Sneezy may work less hard, but that won't tell us whether he values yabbies less, or dislikes work more. He may merely be lazy, and less motivated in general to try and receive his favourite things, despite enjoying them just as much. Again, the information we have gives us no reason to prefer one of these options to the others, and there does not seem to be any further information we could gather that could give us such reason. Similarly, repeating the tests with a larger sample size would not give us further insight. Even if we had a test in which we saw 100 otters showing a higher response and 100 showing a lower response, we would still not have the required information in place to determine whether the higher responses were a result of greater enjoyment of yabbies, or of differences in reactivity.

The problem arises because we have two dimensions along which the animals can vary from one another. The first is in the degree or intensity of their subjective response to stimuli: something like their 'capacity' for pleasure or suffering. This includes both the level of their maximum and minimum welfare (scope) – how high their highs and how low their lows – as well as their usual response to stimuli – whether it might be overall more positive or negative than others. The second dimension is the relationship between the level of subjective response and the change in the measured indicators. Some animals may be highly reactive, showing large changes in their measured indicators to only small increases or decreases in their subjective experience. Others may be more circumspect, showing only small external

responses to large subjective changes. We have no way of testing for this possibility, and no *a priori* reason to rule it out.

Another example that might be familiar to many is that of the emotional reactions shown by young children. Toddlers are famous for having meltdowns over what we would consider to be very small problems, such as the shape of their sandwiches. When a toddler throws a tantrum, what we don't know is whether they are genuinely that deeply upset about the issue (i.e. as upset as we would have to be in order to show such a response), or whether they are just being particularly histrionic. Which of these interpretations we take will determine how we should react to the tantrums – should we show genuine concern for a deeply emotionally wounded child, or ignore an excessive display? Without further information to help us decide, either option is equally likely.

The overall problem is that under an observed difference in overall response, we don't know which of these factors – difference in level of welfare intensity, or in indicator response - is responsible for this, or indeed if both are varying simultaneously. Without such information, we cannot make comparisons.

6.3.2. Types of intersubjective comparisons

There are different types of intersubjective comparisons that can be made, and although the general problem as described above will affect all of them, the degree to which it will become a problem for a situation will differ depending on the type of comparison required. The types of comparisons link to the types of measurement scale discussed in Chapter Four.

The three types of intersubjective comparisons are valence comparisons, level comparisons and unit comparisons (List, 2003). Valence comparisons are the most basic, and simply require an assessment of welfare as positive or negative with reference to some intersubjectively relevant zero point - "i's utility in state x is greater than/equal to/less than a utility level of zero" (List, 2003, p. 5). The zero point represents a 'neutral' point of welfare, and if we can assign the same zero point to each individual, then we have a shared basis from which comparison can be made. Different individuals can be compared with regards to where they stand relative to this zero point. We can then make claims such as: both individuals are experiencing positive welfare, or one individual is positive and the other negative.

The problem of intersubjective comparisons does not hold for valence comparisons. For any individual animal, we are able to use indicators to make judgements about the valence of its welfare - whether it is experiencing positive or negative welfare. This is done in reference to a neutral 'zero' line, at which welfare is neither positive nor negative. The exact characterisation

of the zero line, and how to identify or measure it, are complex and interesting questions, but will not be explored here. Here I will take what I hope is a somewhat intuitive characterisation of the zero line as the subjective absence of both the characteristic positive and negative mental states (or an equal balance of both), that can be observed through the absence of indicators for either of these. Intersubjective comparisons of this type are made possible by the fact that this zero line is the same conceptual point for all individuals. What it means to be experiencing neither positive nor negative welfare is the same for all individuals – it does not make sense to think of one animal as having a more intense neutral experience than another. So, if we can identify when any individual is in a positive or negative welfare state, we can make comparisons of that to other animals – animal A is experiencing positive welfare while animal B is negative. This allows some basic level comparisons, in that we can assume that all animals experiencing positive welfare are at a higher level than those experiencing neutral or negative welfare.

Level comparisons are more general judgements that the welfare of one individual is higher or lower than that of another individual; of the form: "Person i's utility in state x is at least as great as person j's utility in state y" (List, 2003, p. 1); or more formally: $U_i(x) \geq U_j(y)$. This is a ranking of welfare states of different individuals from highest to lowest – similar to an ordinal scale of measurement. This type of comparison gives us no information about the valence of or degree of difference between the two levels.

The problem of intersubjective comparisons does not hold for valence comparisons. For any individual animal, we are able to use indicators to make judgements about the valence of its welfare - whether it is experiencing positive or negative welfare. This is done in reference to a neutral ‘zero’ line, at which welfare is neither positive nor negative. The exact characterisation of the zero line, and how to identify or measure it, are complex and interesting questions, but will not be explored here. Here I will take what I hope is a somewhat intuitive characterisation of the zero line as the subjective absence of both the characteristic positive and negative mental states (or an equal balance of both), that can be observed through the absence of indicators for either of these. Intersubjective comparisons of this type are made possible by the fact that this zero line is the same conceptual point for all individuals. What it means to be experiencing neither positive nor negative welfare is the same for all individuals – it does not make sense to think of one animal as having a more intense neutral experience than another. So, if we can identify when any individual is in a positive or negative welfare state, we can make comparisons of that to other animals – animal A is experiencing positive welfare while animal

B is negative. This allows some basic level comparisons, in that we can assume that all animals experiencing positive welfare are at a higher level than those experiencing neutral or negative welfare.

Unit comparisons are the most detailed, being judgements about the *amount* by which the welfare of one individual is higher or lower than that of another, as well as the relative difference of welfare change that will occur when individuals change conditions. These are of the form: "If we switch from state x to state y, the ratio of person i's utility gain/loss to person j's utility gain/loss is λ" (List, 2003, p. 1); or more formally: $U_i(x) - U_i(y) / U_j(x) - U_j(y) = \lambda$. Here we can compare the number of units of welfare experienced by different individuals in different situations, not just changes up or down, as in the quantitative (interval and ratio) scales. If it can be made, this type of comparison gives us more information about the magnitude of welfare differences, and is much more useful in many decision-making contexts. Comparisons of this type will require that measurements of welfare can be made to at least an approximate quantification of units; as mentioned in Section 6.1, this seems to be plausible through quantification of indicators.

The three applications of animal welfare comparisons described will generally require unit comparisons, though level comparisons might sometimes be sufficient. In making management decisions trading off current and future welfare, we will need to know the magnitude of the negative and positive experiences in order to compare. For use of animals in welfare science, we will at least need to know if some intervention has made a group of animals better or worse off than another, and often by how much. For management decisions about resource distribution, we want to know how much benefit we will be providing our different animals through provision of resources, in order to determine the optimal distribution.

Both level and unit comparisons, however, do seem to fall to the problem of intersubjective comparisons. In both cases, all we have is information about the conditions the animals are experiencing, and the indicators they are showing, but for the reasons laid out earlier, we cannot simply assume that animals are responding similarly to the conditions, nor that their indicators reflect similar levels of subjective experience. We may have slightly more justification in using level comparisons, at least in cases where the difference in condition or response is quite vast, but even these will rely on some similarity assumptions, which I will address shortly. The assumptions required to justify more detailed unit comparisons will be even more stringent.

6.4. Solving the comparison problem

Some writers have argued that intersubjective comparisons of welfare are meaningless and cannot be made (see examples in List, 2003). This is the strong claim that the subjective experiences of different individuals are of a different type, and are thus incommensurable – there is no common scale by which we could make comparisons (see Section 4.3). Intersubjective comparisons of welfare do not seem to be of this type. It makes sense to think that subjective experience is the same *kind* of thing between individuals, though its particular features and dimensions may vary. There is a first-person experience for any sentient animal, and it is in this experience that welfare is grounded. We do not think that the experience of individuals is of such a vastly different type that there is no common ground by which to understand it. At least, this should be the case between similar or closely related animals. We might have a different intuition when considering the minds of vastly different species. I will not argue for this claim here, but it seems intuitively plausible that we do not have an ineffably different state grounding experience for different individuals. The comparisons are not meaningless; however, they may not be possible.

Here I will take it as accepted that subjective experience is the same kind of thing between individuals; one may still want to push on this claim, but if experience is of a different type, then there will be no possibility of comparisons at all (see Chapter Seven for more on the similarities between different types of similarities). If subjective experience is of a single type, there will then be a common scale on which we can measure welfare. Instead of trying to compare weights and lengths, the process is more akin to trying to compare lengths measured in different units (i.e. centimetres and inches). However, when we are trying to compare measurements made in centimetres and inches, we have access to the required information for converting one to the other. The problem of intersubjective welfare comparison is that we might not be able to access the equivalent 'conversion formula'. We may have the measurements of welfare of one animal and those of another animal; both of which quantify welfare in relation to the scale for that animal. Due to the private nature of subjective experience, what we do not know, and possibly *can not* know, is how to convert units between the scales of each individual and compare them on a common scale, as we cannot differentiate between changes in measured indicators as a result of changes in welfare intensity or due to differences in responsiveness of these indicators.

6.4.1. Solutions for the human case

The problem of making interpersonal comparisons of welfare is not a new one, and this problem has been widely discussed in the literature on human wellbeing (see e.g. Elster & Roemer, 1991), though often only with a preference-satisfaction view of welfare in mind. However, this discussion has not been expanded to include welfare comparisons made of nonhuman animal species, or between different species. It seems that however bad the problem is in the human case, it is going to be even worse in the animal case. Firstly, we just don't have as much information about the minds of animals to work with. In the human case, we can use our knowledge of our own experience and the reported experience of others to make some justified assumptions about similarities and differences between individuals. With animals, as all our information about mental states is coming through indirect measurement of indicators, we cannot be anywhere near as certain. Additionally, we will often want to make comparisons between members of different species, and this will make the problem even worse, as the differences between individuals will be even larger.

There are three main classes of solution proposed in the human case – use of an 'imaginative empathy' introspective approach to imagine which of two welfare positions is likely to be greater than the other (Binmore, 2009; Harsanyi, 1955), a 'behaviourist' solution positing a connection between a measurable indicator and subjective experience (List, 2003), and lastly to simply move away altogether from the measurement of subjective welfare and to either measure something else we consider to be important in the questions of distribution under which these comparisons are usually required (such as resource availability), or use a different ethical or distributive principle in decision-making (Fleurbaey & Hammond, 2004). I will examine these here and argue that even if they are potentially useful in the human case (which, for most, is doubtful) they fail to meet the requirements for justifying intersubjective welfare comparisons in animals.

The first potential solution is to use 'imaginative empathy'. This involves an 'introspective' or 'imagining' approach to comparisons, in which an observer assesses the situations of the individuals under comparison and makes an introspective intuitive judgement about their comparative welfare; typically based on imagining themselves in both positions, with each individual's behaviour and desires. This can be of two types – a judgement of state (x is better off than y) or of preference (I would rather be x than y). This approach has some deep issues, particularly with reliability. Although we might gain information about the observer's judgement, why should we think that this tracks the fact of the matter about the comparative welfare of the individuals? In the human case, this method is given some (attempted)

justification through our understanding of what it is like to be a person under different conditions, with an assumption of similarity between individuals. This relies on our capability of truly imagining ourselves in the place of another, separate from our own desires and psychological biases, and judging between situations. This seems difficult for other humans, and probably impossible when it comes to other species. On what basis could I really make a meaningful judgement about whether a lungfish swimming in a tank is better off than a lion resting on the grass? Both are so far from my own experience, that my judgement is certain to reflect mostly my own preferences and biases. It also presupposes that there is some degree of similarity between the experiences of the individuals by which we could make the judgement, and this is partially what is at question. If I don't have information about the intensity of a particular animal's experience, how can I imagine myself in its position at all? In cases where there are disagreements between observers, there seems no further facts that can be appealed to in order to resolve the dispute, if this is all our comparisons are supposed to rest on.

The second option is to use a 'behaviourist' framework to "posit a fixed connection between certain empirically observable proxies and utility" (List, 2003, p. 1). That is, we use some external indicators of welfare as proxies for the subjective experience, and compare their levels or units instead. In the human case, this relies on using behavioural cues as proxies for internal states. For instance, we might use facial expressions as an indicator of happiness and compare the facial expressions of different individuals as a proxy for comparing their welfare. The problems with this approach should be immediately obvious, given what has already been said about the lack of justification for unreflectively using proxies in this way. This relies on an assumption that there is the same 'mapping' relation from subjective experience to indicator for all individuals; and the lack of justification of this assumption is precisely the problem we are dealing with. While this approach is the most promising, and is similar to the solution I will propose, it cannot be taken without further justification for believing in the reliability of the connection so posited.

The third option is to avoid the need to make welfare comparisons entirely, through either measuring and comparing something else or through using a different ethical or distributive principle for decision-making. A common response to the comparison problem is to use another measure to stand in for welfare. Not a direct measurement proxy, as in the previous suggestion, but some other proxy variable that could be considered valuable and could substitute for welfare. In this case, we are not really trying to make intersubjective comparisons of welfare at all, as there is no strong claim being made about the link between change in the proxy and change in welfare, but instead are comparing some other state or resource that we might

consider important in considerations of distribution (which is the context under which many comparisons are made). Examples might be availability of particular resources. In the human case this is often money; in animal cases it might be more basic resources, such as food or shelter. Other examples might be access to or realisation of some more basic 'objective goods' – for humans these are things like accomplishment, knowledge, friendship, autonomy; for animals they could perhaps include freedom, comfort, social relationships. Again, we are not here necessarily trying to compare welfare between individuals, but instead looking at comparisons of other goods that we might think are important, and can be more easily objectively measured and compared. As such then, this is not really a solution to the problem of intersubjective comparisons, but simply a workaround. It may be useful in some contexts, but not if there are situations in which we really want to measure and compare subjective welfare. As we don't know the relationship between these conditions and subjective experience (the very heart of the problem we are trying to solve), it may not help us much. I can measure whether Paddy and Sneezy have the same number of yabbies, but this is not telling me whether or not they have the same welfare, which was what we were originally looking to discover.

Finally, we have the option to bring in some other ethical or distributive principle to allow decision-making without comparisons. This differs from the previous point in that rather than trying to make intersubjective comparisons of something other than welfare, we instead try to find ways to get by without making comparisons at all. The type of comparisons required depends on the ethical or distributive principle we are operating under, and if we are able to find a principle that does not require comparisons to be made at all, we no longer have a reason to be concerned with the fact that we are unable to make such comparisons. Only if we are strong utilitarians will we assume that all our distributive decisions should involve absolute maximisation of overall subjective welfare. A utilitarian framework won't work in these cases, as it requires the use of unit comparisons which cannot be justified. Some potential alternatives are a maximin rule, sufficient threshold, Pareto and equal consideration of interests (Fleurbaey, 2016). Here, I will outline the first three, and detail why they would not work in the animal case. In Section 6.5 I will discuss the final option, which I think is the best contender for an alternative solution in cases where the one I will propose does not apply.

The maximin rule operates to ensure the worst off improves their situation. In this case this would mean that the resources should go to help whichever animal currently has the lowest level of welfare. This won't usually work in the cases we are interested in, as it still requires a level comparison to determine which animals are worst off to start with, something we won't have unless one or the other starts out below the zero line. Perhaps a modification of the

maximin rule, which applies to distribution of goods rather than distribution of welfare, could work here – the resources would go to the animal which has fewer overall. This would be a combination of the above 'proxy' solution with an alternative distributive principle. Although this would allow us to make decisions, there is a risk that it is not getting at what we might want in using our resources for welfare improvement.

Under a 'sufficient threshold' view (Nussbaum, 2000), we take each individual to have the right to access a sufficient threshold level of those components deemed necessary for a good life. This is similar to the proxies view as described above, in that we are looking at the components of welfare rather than the subjective state itself, but in this case, rather than quantifying the components as a proxy measurement for underlying welfare state, we are ensuring that each animal has access to some minimum threshold requirement. Like the valence requirement, this will only work in cases where we are making decisions between animals which do and do not already have their threshold level met. If both species are already over their threshold, then we have no way of choosing between them for further positive welfare, and if both are under their threshold then we have no way of prioritising which to raise first. We might think of using some additional principle like satisficing – rather than trying to make the theoretically best decision, we are instead trying to make an acceptable decision and so long as neither animal is below some minimum threshold then we feel like the decision was ok. But this doesn't help us much in deciding between alternatives.

For human cases, the most prominent contender for an alternative ethical principle is Pareto, which is commonly invoked in work on the interpersonal comparison problem in humans (Fleurbaey, 2016). Under the Pareto principle, we should only take actions that benefit all, or benefit some and leave others no worse off. This then requires only *intra*subjective level comparisons, to assess whether individuals are being made better off, but does not require any assessment between individuals. This is an intuitively appealing principle and seems like it would lead to good decision-making – every individual always ends up with higher welfare. However, given limitations on resources, it will be almost impossible to apply in most situations, and certainly not all. There are going to be cases where resource scarcity and competing interests make it impossible to improve the welfare of all individuals, and at times it may be the case that we have to accept a decrease in welfare of one individual or group to create a larger increase in welfare for another group. This principle may recommend that instead we would not act at all, so no improvements would be made, but no-one would be worse off; but this seems counterintuitive in cases where we could give a large benefit for only a small cost. Even a modification such as first implementing all improvements for animals with

negative welfare before assisting animals who already have positive welfare will not gain much, as there is still no way to prioritise between improvements within a valence class.

One proposed modification is that an action is considered acceptable if the 'gainers' could in theory compensate the 'losers' and still end up better off. Again, this does not require interpersonal comparisons, as it only need be the case that after the transfer, the winning group is still above their previous standing, according to their own ordering, and that the same is true for the losing group, also according to their own ordering. This solution works well in cases where there can be a quantifiable resource that could potentially be transferred to outweigh losses, but it seems there are likely to be cases where this won't apply (particularly for animals) – there may be no way of quantifying the value of the loss of one's health, for example; or dealing with any non-divisible resource (though economists might deny that there is anything in human life that can't be captured in this way). It also does not seem that we could use a similar process for determining which group would fare better in cases where there is no loss to outweigh, but where we are trying to decide on which group should receive a benefit (i.e. which would benefit more from it). We could compare how far up their own ordering each would go if given the resource and choose the larger benefit, but this requires some sort of interpersonal comparison if we are to assume that 'two steps up' one individual's ordering is to outweigh 'one step up' in another's. Unlike the previous case where all that is required is that the resource transfer would be sufficient to create a net gain in both cases, regardless of magnitude, the comparative magnitude of gain seems important in benefit-only cases. The solution seems particularly problematic in animal cases – even within a species, it is not clear what sort of hypothetical resource transfer we could imagine to balance the scales, and this is even more true across species. There is no obvious answer to how we could calculate whether there has been a net gain or loss in a decision involving different species. Perhaps we could use the value of the resources used initially – e.g. whether some of the money used to buy lion heating could instead buy lungfish logs – and compare whether in that case both groups would still benefit. However, this won't help in cases where the resources may be less quantifiable and still won't help us in the benefit-only cases described above.

It also doesn't allow for decisions between different potential allocations if all meet the Pareto standard – there may be some actions which leave many *much* better off, and some that leave them only slightly better off, but these differences cannot be assessed under this framework. We may, say, want to make a decision about which group to benefit (like whether we are to renovate and improve the lion exhibit or the lungfish exhibit) and want to know which group would have the higher welfare gain from this. Without a way of quantifying interpersonal

comparisons, it seems difficult to make such decisions in a non-arbitrary fashion. I outlined in Section 6.2 the decisions we are likely to need to make in animal welfare cases, and it seems that this solution is unlikely to be useful in many, if any, of them. As neither this, nor any of the other solutions discussed wholly suffice to justify making intersubjective comparisons of animal welfare, I will now describe a potential solution that does.

6.4.2. Making similarity assumptions

The problem of intersubjective welfare comparisons is that we are unable to make comparisons between the measures of welfare of different individuals, as the data underdetermines what the results might mean. There are two dimensions along which animals might vary – in their degree of welfare experience and in the relationship between their experience and the changes in indicators – and which we choose can affect how we interpret our results. Variation in observed reactions can be explained by variation along either of these dimensions, with no obvious and unproblematic way to choose between them, and thus no way to draw conclusions. Making comparisons of welfare based on the results of measurement such as described in the otter case, requires that we make some assumptions about the similarities between individuals. Here I will detail these assumptions and what may justify their use. It is important to keep in mind here, that comparisons are made for a reason, and we only need to be confident enough in our comparisons to serve the reason at hand; we usually do not require some stronger metaphysical certainty (though it might be nice!). Our level of confidence in the assumptions thus only needs to match what is required for the application.

There are two assumptions that we can make, based on the two dimensions of variation:

1. Similarity in degree
2. Similarity in response

By making either one of these assumptions, we are in essence holding fixed one of the dimensions and assuming that observed variation is explained through variation in the other. This then allows us to make the comparisons we require.

The first assumption is that the animals have the capacity to experience a similar *degree* of welfare. That is, that the animals are similar in respect to their level of welfare intensity - the amount of pleasure or suffering they can and do experience under different conditions. This assumes that the individuals have roughly equivalent minimum and maximum welfare intensities – the height of their highs and the depth of their lows - as well as similar degrees of change in between.

Making this assumption, we can go on to make comparisons using a zero-one method. The zero-one method is one under which we assign a score of 0 to the minimum level of welfare and 1 to the maximum level[28] for any individual. Here we assume that the maximum and minimum welfare levels are equivalent between individuals; this is what we are taking for granted under Assumption 1. This provides set points for conversion of individual results onto a common scale. For each individual, we can build up a welfare profile in which we can measure their level of response under a range of different circumstances to identify where they experience their maximum and minimum welfare levels (0 and 1), and the degree of indicator response they display at these extremes. This would involve exposing them to a range of conditions, both positive and negative, and recording their responses across a range of indicators, under these conditions. We could then identify their maximum and minimum response levels, and under what conditions these occur, which would then be used to create a scale for the individual, showing different conditions and indicator responses as proportions of their total, occurring along the 0-1 line. Regardless of the differences between the conditions and indicator responses for individuals, we can still express responses for each as a proportion of the maximum. We can then use our assumption about the common value of the 0-1 points to construct a common scale on which comparisons can be carried out. As will be discussed in Section 6.5, normalising the range in this way may also have a moral justification, in terms of fairness.

For example, think back to our otters, Paddy and Sneezy. We can begin by measuring their individual response profiles. We measure Paddy's heart rate and vocalisation under different conditions and find that her responses range from a minimum of 15 under her most unpleasant condition to a maximum of 350 in her favourite (presuming here for the sake of simplicity that all the measured indicators showed the same response profile; further on I will discuss the case in which this is not true). We then do the same for Sneezy and find a range of 2 - 180. If we were to just compare these as absolute responses, it would look like Paddy's highs were much higher than Sneezy's. However, if we are assuming that both their maximums and minimums represent an equivalent underlying welfare state, we can scale them accordingly. A score of 350 for Paddy represents the same level of experienced welfare as 180 for Sneezy. So while Paddy might show a response of 200 to yabbies, while Sneezy shows 100 – which on the surface makes it seem like Paddy likes them twice as much - when we scale to the 0-1 scale,

[28] When considering positive and negative welfare we might set these slightly differently – say 1 for best, 0 for neutral and -1 for worst, but the principle remains the same.

we find that both are around 0.6 of their maximum response, which would tell us they like them roughly equally (Figure 6.1). Paddy is in general more prone to a larger indicator response under all conditions, and so we see that Sneezy's lower absolute response is actually as high *for him* as Paddy's is for her, and can thus infer that he is actually enjoying the yabbies just as much. By making the assumption about similarity in degree of welfare (1), we can then use tests under different conditions to measure for differences in the indicator response profiles (2). We hold fixed the degree of welfare intensity and explain observed variation through differences in the response profiles.

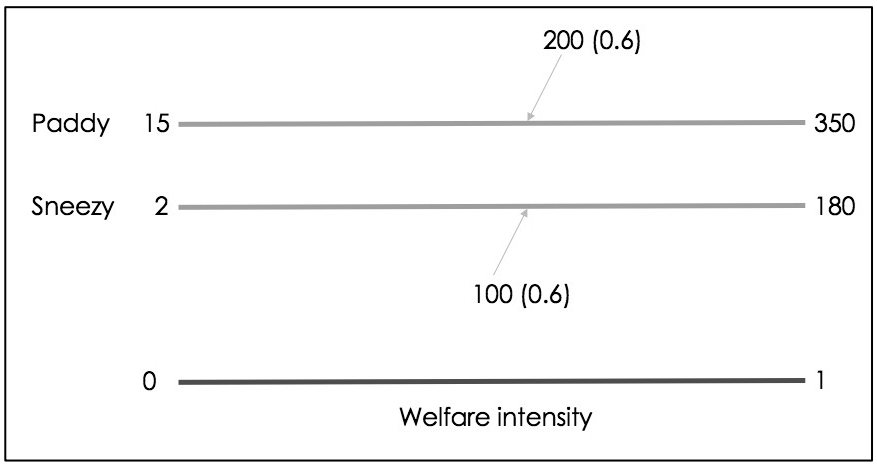


Figure 6.1: Comparison of welfare responses under Assumption 1

As described here, it may seem that use of this 0-1 rule requires a linear reaction function. That is, that we think that the level of welfare change between a 10% and 20% response level is the same as that between 70% and 80%. This, of course, may not be the case. It could be that animals are far more responsive at the lower or higher end of the scales (we might show a far more extreme response to intense pain than mild pain, for example), and that the response functions look something more like curves than lines. It is not the case that for the use of this assumption to work, that we require a linear response function. It is the case, however, that we require the shape of the individual welfare curves to all be roughly the same, so that when we map them onto the 0-1 scale, we are still able to make comparisons based on how far along the curve different responses are. This then raises the further issue of how we might determine individual response curves, to be sure we are justified in thinking they are similar between individuals. The details of this are beyond the scope of this work, but general testing mapping the degree of response along different conditions should give some idea of the shape of the response functions.

If we make the second assumption, we are assuming that the animals are similar with respect to the level of indicator response shown under the same state of experienced welfare, such as similar heart rate change for mild arousal. This is similar to the behaviourist solution for the human case, as described in Section 6.4.1. If we take this assumption, we can simply use behavioural and physiological data to determine the intensity of their subjective experience under different conditions (and map out the maximum and minimum overall levels). For example, take again the case of our otters. We map out their range of responses under different conditions, and find Paddy varies from 15 - 350 and Sneezy from 2 – 180. But this time, instead of scaling these responses to the same range of underlying welfare levels, we take them as directly representative of what is happening experientially, in the same way for each otter. Paddy's higher reaction levels suggests she is capable of experiencing more pleasure than Sneezy under a range of circumstances. Her 'highs' are higher, while his 'lows' are lower. Comparing their reactions again on receiving the yabbies, with Paddy showing a score of 200 to Sneezy's 100, we now take Paddy's more extreme reactions to mean that she is indeed experiencing more pleasure – twice as much - in receiving the yabbies (Figure 6.2). By making the assumption about similarity in indicator response (2), we are able to run tests to measure the differences in degree of welfare intensity experienced by individuals. We hold fixed the relationship between welfare experience and indicator response and explain observed variation through differences in the underlying experience of welfare.

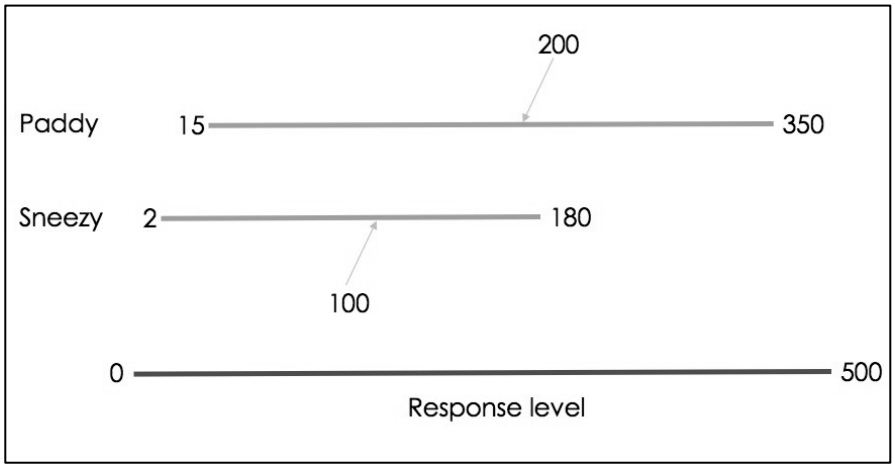


Figure 6.2: Comparison of welfare responses under Assumption 2

6.4.3. Justifying similarity assumptions

I have shown that by making either one of the similarity assumptions – similarity in degree of welfare experience or in profile of indicator response - we are then able to perform further tests and make inferences about variation along the other dimension, using this to make

meaningful intersubjective welfare comparisons. But this is only useful insofar as we are justified in making such assumptions. Here I will show how reasons of analogy and shared evolutionary history work to justify making these assumptions. The justifications provided will support either assumption with varying strength, depending on contextual details such as the particular indicators used, or the relatedness of the animals. In many cases, animals which are similar in the relevant regards (such as members of the same species) are likely to show similar response profiles, indicating that both assumptions will hold together. In cases such as our otters, where we observe difference in response, we can look either for which assumption holds the stronger justification, or use additional methods to decide between the assumptions. These will be detailed in the next section (6.4.4).

Reasoning by analogy holds that where animals are similar in terms of their underlying structures and mechanisms, they should also be similar in terms of the experiences and responses produced. Many animals have such similarities in their anatomy and physiology; the structures and mechanisms that give rise to both subjective experience and indicator responses. In terms of welfare intensity, similarities in brain structure and function would give us reason to think there is similarity in the subjective experience. The structure of the brain, and the way in which it functions, will determine the psychology of the individual, and these will vary depending on the inherited 'instructions' for development as well as the influence of the developmental environment. We then have good reason to think that individuals that are similar with regards to the relevant inheritance and developmental environments, will have similar types of psychology, with similar scope and boundaries. Insofar as subjective experience is a function of brain activity, and where there are neural correlates of experience, similarity in brain structure and function should then give us similarity in experience. This is much more likely to hold within species than across species.

Where neural structures directly mediate indicator responses, similarity in neural systems will also give us reason to think there will be similarity in these responses. Often, though, indicator response will also involve other physiological pathways and in these cases, we would also require similarities in the relevant response-producing mechanisms – such as the hormonal and neuronal outputs of the brain, and their impacts on bodily systems – to give us reason to think there is similarity in the responses produced.

The level to which we can trust the similarity assumptions on this basis will then depend on the level to which there is relevant underlying anatomical and physiological similarity. For example, the neural mechanisms for generating affect appear to be similar across most mammals (Kringelbach & Berridge, 2009). The level of similarity will in turn depend on the

degree of variation and plasticity within developmental processes, and further study on the precise mechanisms involved will help determine where the assumptions might hold. Effects of developmental environment, such as hormonal changes during foetal and infant development, and conditioning of particular behavioural responses throughout life, are also likely to play a strong role in determining both scope of experience and level of responsiveness, and so this type of similarity must also be considered. It may turn out that the assumptions only hold among related individuals that also share relevantly similar upbringing. Current and future research into the neural correlates of experience, and the development and function of both the mechanisms creating subjective experience and those linking experience to behavioural and physiological indicators, will shed further light onto which cases this justification will hold. In cases where these structures and processes are homologous, arising from the same evolutionary events, we would have even greater reason to think our assumptions justified on this basis.

The next justification is that of evolutionary history. Animals which have shared evolutionary history, as well as sharing the structures and function of their brains and bodies, also have shared selection pressures. If we take subjective experience, and the behavioural and physiological responses it produces, as being the products of selective processes (e.g. Ginsburg & Jablonka, 2019; Godfrey-Smith, 2017), then it makes sense that shared selection pressures will have led to similar experience and responses. Animals with shared evolutionary history (most particularly those of the same species) will have brains adapted to the same biological challenges, and it makes sense to infer that they will share similar psychology, with the same scope for welfare experience. The minds and the indicators of different individuals that evolved under the same conditions, are likely to be similar in scope and function. Physiological and behavioural responses to subjective welfare changes are going to depend in large part on evolutionary history. For behavioural responses this will include what was beneficial to communicate with others – for example, prey species are notoriously non-vocal when in pain as they do not wish to alert predators to their weakened status. For physiological responses this is likely to include those responses appropriate to ready the body to meet whatever particular challenges it is about to face, such as increased heart rate for fight or flight. These evolutionary pressures will be shared among members of the same species, and thus it is likely that the degree of response will be similar also. For example, when considering our two otters, we know they share a common ancestor and the same historical selection pressures that led to the development of their species-specific brain structure and function, as well as those which led to their propensity to vocalise under different circumstances. Their shared evolutionary history

gives us good reason to think they might be similar in their subjective experience under different conditions, and/or their set of indicator responses.

Finally, we have considerations of parsimony. This is using an inference to the best explanation; one that can describe our results within our best frameworks of understanding. This would include not positing differences without evidence of their existence – although it may be *possible* that different individuals vary widely in experience and response, if there is no positive evidence of this fact, then it does not seem likely that this is the case. As per Harsanyi (1955) - "If two objects or human beings show similar behaviour in all their relevant aspects open to observation, the assumption of some unobservable hidden difference between them must be regarded as a completely gratuitous hypothesis and one contrary to sound scientific method" (Harsanyi, 1955, p. 279). Overall, it seems far more likely that individuals with these similarities in anatomy, physiology and evolutionary/developmental history will have broadly similar minds than vastly different ones. Without evidence to the contrary, it seems more parsimonious to assume similarity in these cases. If the minds are sufficiently similar, then we are justified in making our assumptions and the comparisons which rely on them. As previously mentioned, we do not need to have an identical level of response; a broad similarity in the scope of subjective response is sufficient to justify comparison of measures in cases where this holds, as it will give us a close enough result for our needs. For most practical purposes, certainty is not required; just a reasonable assumption that we are getting close to the fact of the matter about comparative welfare experience.

There are several lines of evidence that can be taken to justify the similarity assumptions I've outlined. If the animals are sufficiently similar, then we are justified in making our assumptions and the comparisons which rely on them. As previously mentioned, we do not need to have identical level of response; a broad similarity in the scope of subjective response is sufficient to justify comparison of measures in cases where this holds, as it will give us a close enough result for our needs. For most practical purposes, certainty is not required; just a reasonable assumption that we are getting close to the fact of the matter about comparative welfare experience. In the next section, I will address some possible concerns with the similarity assumptions, before moving in the final part of the chapter to look back at the type of welfare comparisons we are likely to want to make and how the solution - use of similarity assumptions - applies to them.

6.4.4. Potential problems with similarity assumptions

There are two potential problems with making the similarity assumptions as described. The first is that these assumptions might be too fragile, only holding in a small range of cases. The second is that we need a method of determining which of the two assumptions we should make in any given case, if we know we cannot make them both together.

In the first instance, the upshot to the justifications provided for the similarity assumptions is that they will only hold for animals which share the relevant similarities of physiology or evolutionary history – typically those of the same species, or perhaps closely related species. This might also require those with similar developmental histories, which may mean for example sub-groups within species of age, sex or rearing type (wild vs captive). There are known effects of individual personality and temperament - as well as genetics and early experiences - on emotional responses to stimuli, and thus welfare (Boissy et al., 2007). This may make the assumptions too fragile, as some very small differences in anatomy, physiology or development could give rise to some quite large changes at the affective or response level. If this is true, then in order for our similarity assumptions to be justified, we would need to have quite a fine-grained level of underlying similarity in order to justify using either of our assumptions, and this would further narrow the range of cases for which we could use them.

What is key here, is that we need to understand what the relevant similarities are. We need to know how particular anatomical structures and biological processes give rise to both the subjective experience of welfare, and the indicators that we use to measure it. This requires further research into these structures and mechanisms. Only by understanding these processes, and how robust they are across different structures and different environments, will we understand which groups of animals possess the relevant similarities to make welfare comparisons – whether these are species groups or possibly further divided by classes such as sex, age group or type of developmental environment. This of course is quite a demanding research requirement, requiring a detailed causal model of the relations between the inputs, welfare changes and output indicators for a huge variety of taxa and the subgroups within. However, beginning with representatives of different higher-level taxa (e.g. felids, amphibians) and looking for the differences and similarities between them may help delineate where relevant similarities lie and how much further detailed work is required. Understanding the extent of similarity in structure, function and selection pressures across different groups will help us see how far we might extend this solution. For example, if we found that the mechanism linking welfare experience to changes in heart rate was one which arose fairly early in

evolution, shared across all vertebrates, then we could use this indicator to make comparisons between animals within this entire group.

One thing we might think is that the second assumption – that of similar response - is more fragile than the first, of similar scope of welfare. That is, we are more likely to find differences in response as we move towards less similar animals, than we are to find differences in scope of welfare experience. This is likely to be true if the mechanisms which cause response are less robust than those which cause experience; if they are affected to a greater extent by contextual factors and processes, such as developmental environment. There are many examples of these differences – urination and defecation in a new environment is a scent-marking behaviour in mice but a sign of fear in rats, and bulls show decreased corticosteroid response after tethering, while pigs show increased response (G. Mason & Mendl, 1993). It is plausible that in general, indicator mechanisms are more variable than those causing experience, as there is likely to be more homology in brain structures and processes than in the diverse range of indicators (see e.g. work by Kringelbach & Berridge, 2009 on the deep brain structures responsible for generating pleasure). If this is true, then in cases where we are unsure as to which assumption we should make, as discussed below, this will give us some reason to favour the first.

In the cases where the relevant similarities hold, whatever they might end up being, we are able to make the required assumptions and so can perform intersubjective comparisons of welfare. For now, I'll take it as safe only in the cases of the same or closely related species. This, however, leaves out a large class of comparisons we might like to make – particularly those between species. We do not have good reason to think that these similarities hold between quite dissimilar species; for example, it is not likely that the scope of welfare intensity for a lion is anything like that for a lungfish. Additionally, the types of indicators tend to be quite species-specific (especially behavioural indicators), and we should be quite circumspect in inferring similarity between species. In the final section, I'll look at what this might mean for the types of welfare comparisons we will want to make, as described in the beginning of the chapter.

The other important question we are left with is which of the two assumptions we should make in any given case, given that the same justifications hold for both. To begin with, we can test whether we can make both assumptions together, as in many cases this will be possible. We test this by mapping the overall response profile for our animals: finding their maximum and minimum response levels and the variation across different conditions. A full mapping of response profile would not necessarily be required for each individual animal – preliminary work mapping out response profiles for a representative group (such as species, age class, sex

etc.) could be used and if enough similarity seen, assumed to hold for the whole group. In cases where we see animals with a similar profile of indicator responses over different welfare conditions, we can make both assumptions together – that the animals have both a similar scope of welfare experience and a similar degree of response. This is a much more parsimonious explanation than the alternative - that both these factors are varying in opposite directions to give rise to the seeming similarity. The second scenario is far less likely, given that it requires correlated change in both systems, and there is no reason to think that even if both systems did vary to this degree between individuals, they would both do so in tandem. Without a plausible explanation as to why there would be such a hidden difference, we should think that the same responses under the same conditions reflect similarity in the underlying subjective experience, and that our two assumptions hold.

The best explanation of observed similarities between the behavioural and other responses of individuals is similarity in underlying mental states. It is far less likely (though not impossible) that these are the result of vastly different mechanisms working on different underlying states, so we should accept the more plausible hypothesis that there is a relevant similarity in underlying subjective states that can ground use of intersubjective comparisons. However, in cases as described above, where the animals were shown to have differing response profiles (with the example of Paddy ranging from 15-350 and Sneezy from 2-180), it must be the case that they differ along at least one of the given dimensions of welfare experience or level of response. And, as we saw, which assumption we make will lead us to a different result – making the first assumption tells us that the otters are enjoying their treat equally, while the second tells us that Paddy is experiencing almost twice as much pleasure from it as Sneezy. It matters a lot in these cases which assumption we choose, so how do we decide between them?

The answer to this is going to depend primarily on context. It will rely on the particular indicators we are using, and the proposed mechanism for linking these indicators to welfare. For indicators such as behaviour, which have a more flexible developmental pathway, we would be more inclined to assume that observed differences are a result of different response levels, where scope of welfare experience is held fixed (Assumption 1). For more deeply physiologically controlled indicators such as heart rate, we would be more likely to assume that different responses reflect different levels of experience, where response profiles are likely to be similar (Assumption 2). It will also depend on the mechanisms for creating subjective experience and how these operate within an individual, and across different types of development. The more robust these mechanisms, the more likely we are to think Assumption

1 holds. Once we are able to better identify the cognitive mechanisms responsible for giving rise to subjective experience, we can also look for how these are distributed across different species. Certain distributions will make it more likely that these were present in the common ancestors (as a more parsimonious explanation than multiple independent evolutionary events) and therefore we could use these to justify making comparisons between individuals within the given clade. There is interesting future work to be done here, linking the conditions for evolution of subjective experience and particular indicator responses, to what they can tell us about welfare of individuals and comparisons within and between species. The more we know about the conditions under which subjective experience arose, the mechanisms which create it, and how welfare experience links to changes in the measurable indicators, the better we will become at determining when and how the assumptions will hold.

Another way of deciding between the assumptions is to look for convergence between different types of indicators. By testing the response profile for an individual across a range of indicators, we can get a better idea as to whether observed variance is likely to be a result of variance in underlying welfare state or in indicator response. We would start by measuring welfare for an individual across different conditions and across time, using a range of indicators, which work through different mechanisms. This would give us a response profile for each of the indicator types for that individual. We could then compare these between individuals to look for where the inter-individual variation lies. If the different indicators give us similar results (e.g. one individual shows higher overall response across a range of indicator types), then this gives us reason to think that this is reflective of differences in underlying welfare intensity (Assumption 2). The alternative, that the individuals have differing responsiveness that operates equally across the range of indicators, despite their independent mechanisms, is far less likely. If instead we see different results across indicator types, this gives us reason to think that it is the indicator response profiles that are varying, and it is more likely that Assumption 1 holds and welfare intensity is similar.

One potentially promising line of evidence could be that from neuroimaging. If we are able to validate particular types of neural activity as representative of different intensities or types of subjective experience, then this could be taken as a reliable indicator of welfare state. This would require us to validate the images across a range of individuals to determine that there is a strong link between observed response and subjective experience, but if this is found then we could be confident that this data tells us something about variation in welfare state. We can plausibly assume that measurements of brain activity are more direct measures of the processes that cause subjective experience (granting that there are still currently assumptions in play in

producing and interpreting images), rather than of their downstream effects, so we would not expect to see differences in response profile for the same underlying experience. This means that unlike other indicators, these measures would not be subject to the same uncertainty regarding the source of variation – different measured levels of brain activity would be directly indicative of different underlying intensity of experience. There has been some promising work on this in humans, where both subjective report of intensity of experience, and behavioural responses, correlate with intensity of brain activity (Coghill, McHaffie, & Yen, 2003). Further work in this area could validate this link and perhaps give us a reliable direct measure of welfare experience that could be used to make comparisons.

6.5. Applying the solution to different comparison types

As described in Section 6.2, there are three types of comparisons we might want to make – individual management decisions, use of experimental animals and resource distribution decisions. Here I will examine how the solution applies in these different cases, given the scope and limitations of the similarity assumptions, and how we might go about making comparisons in cases that aren't covered by this method.

Management decisions for individuals require consideration of trade-offs between harms and pleasures – we might inflict current harms or prevent particular pleasures in order to decrease future harms or increase future pleasures; or we might just need to decide between sets of current conditions with different potential positives and negatives. This requires the use of intrasubjective welfare comparisons. These do not seem to be a different type of comparison in their own right, but rather a special case of intersubjective comparison, in which we consider the two 'subjects' to be two possible outcomes for the one individual, or different time-slices of the same individual. These seem like the least problematic type of comparison to make, as the similarity assumptions are most likely to hold in cases where we are talking about the same individual. Even if we think that individuals may change dramatically over time, we can use the inter-individual comparisons described in the next paragraph to make the necessary inferences about particular age groups, treating them as we would separate individuals.

In animal welfare science, measurements made on one group of animals will need to be compared to those made on different groups of experimental animals, as well as extrapolated to other members of the species. Both of these require that the similarity assumptions discussed in the previous section hold. In these cases, it seems highly likely that they do. Experiments of this type are almost always performed within a species, which gives us the justifications from analogy and shared evolutionary history. Further, they are often segregated for subgroups such

as age/sex classes to see if there are differences, which adds further strength to the similarity assumptions within these classes. Given our justifications for applying the similarity assumptions in these cases, we have good reason to be confident in the comparison of experimental groups in welfare studies, and the extrapolation of these results to other members of the species.

The final application of intersubjective welfare comparisons is not so simple. This is institutional (or individual) decision-making about the distribution of resources between animals in order to achieve the best outcome. In cases where a single type of animal is held, this shouldn't be an issue, as we can use the similarity assumptions to make comparisons and apply a basic utilitarian calculus. For example, on a dairy farm, we can use the similarity assumptions to make relevant comparisons to assess whether we would have greater overall welfare benefit through the provision of soft bedding for calving or higher quality food to pregnant cows, based on comparisons of the welfare increase under each condition and the number of animals likely to be affected. But many such decisions will involve multiple species, such as governmental investment in agriculture, interventions on wildlife, and in zoo management and husbandry. Think again of our zoo manager trying to decide between underfloor heating for a lion or aquarium furniture for a group of lungfish. Here we have two such disparate species that it is unlikely that the similarity assumptions will apply. It is perfectly plausible that lungfish and lions have completely different scopes for intensity of welfare; so that the heights and depths of lion experience may just be of a different scale to that of lungfish. It is also extremely improbable that there will be overlap in the types of indicators used to measure welfare in each species, let alone that they will be subject to the same processes linking subjective welfare to indicator outcomes. All we have is information about the benefit to the individuals – we know the lion will benefit from its heating and the lungfish from their furniture, but we cannot compare either their levels of welfare, or the degree of benefit each might receive.

There does not seem to be any objective standard to which we can appeal in order to convert units of lion welfare into units of lungfish welfare, and so we cannot make meaningful comparisons. But we still want to have some means of comparing the welfare gain to the lion from its underfloor heating to the gain of the lungfish of having new logs to explore and shelter in. As it stands, the only comparison we can make will be that of valence. As the zero-line of neutral welfare experience is the same for all individuals, we can at least rank lion and lungfish welfare according to whether it occurs above or below this zero line. For example, it could be that that the lungfish are currently experiencing negative welfare through the stress of being

exposed without exhibit furniture, and the lion positive welfare, being happy enough already but simply made happier with extra heating. Here we could make a simple valence comparison and say that the lion has better welfare, without needing to know anything about magnitude. However, this will not work in cases where both are above or below the zero line as we then cannot compare how far above or below they sit and so the solution is only of limited use.

One possible solution is to use some other capability as a proxy for welfare capacity – something like cognitive complexity. The idea here being that cognitive complexity might underlie the capacity to experience certain ranges of subjective welfare, so we could take a proportional increase in welfare for a more cognitively complex species to be greater than that for a less complex species. This sort of option would be particularly appealing to those who might think that cognitive complexity is a strong influence on the type and intensity of experiential states; in this chapter – indeed most of the thesis – I have been working with the assumption that (at least in the animal case) experiential states are somewhat simple and separate from cognition except insofar as it concerns which stimuli are likely to produce which experiences. This is not an issue there is space to explore here. Evidence to the contrary would not change much of what I have had to say throughout, except in making the case more complicated. In this instance, a strong link between cognition and affect would give us more reason to try and pursue this line of capacity-based proxies.

Cognitive complexity as a proxy is a commonly used division to separate out human welfare from that of other animals (McMahan, 2002). It is also becoming increasingly common in calculations of animal suffering under different production regimes, e.g. for use in Effective Altruism initiatives[29]. Different proxies have been considered here, including brain size, number of neurons, and connective complexity. There is perhaps some intuitive pull to the idea that size or complexity of the brain may relate to the potential breadth and depth of subjective experience. We might want to think that perhaps the pain of a cat just cannot reach the same intensity as that of a human.

The problem with this solution is that it relies on an assumption about the relationship between cognitive complexity and welfare, for which there does not seem to be a method of validation without running into the problems already described in this chapter. We do not have direct access to information about the welfare capacity of different organisms, thus we cannot simply check whether cognitive complexity correlates with welfare potential. We would instead need to make assumptions about these links, and these assumptions would then need to

[29] e.g. http://sandhoefner.github.io/animal_suffering_calculator

be made explicit, and justification provided for them. As it stands, we don't currently have strong reason to think that cognitive complexity correlates with welfare capacity (Browning, 2019b). Future work in understanding the mechanisms for production of sentient experience may provide answers leading to valid proxies, but this is not yet the case. What would be required would be an overall theory of cognitive function and affect, that could be tested independently of these problems.

Instead, in these situations in which we do not think the similarity assumptions hold, we may take the third solution described earlier for the human case and look at alternative methods of decision-making that do not involve direct comparisons between welfare. As described in Section 6.4.1, use of an alternative proxy is unlikely to help in the animal case. Neither are the alternative ethical frameworks of maximin, Pareto, or sufficient threshold. The most promising alternative is probably an equal consideration of interests view, where the interests of each individual are given the same weighting, regardless of absolute strength (Fleurbaey, 2016). This would fit in with a Kantian-style ethics in which each individual holds the same moral status (Sebo, 2018)[30]. This would ensure that a lungfish gets its best possible welfare and a lion gets its, despite potential differences in intensity between them. This is not because we have strong reason to think that the maximum welfare of a lungfish and that of a lion are of a similar level, but because we assign the same *value* (a moral value, rather than a measurable empirical welfare value) to the welfare of each. That is, we could say that allowing a lungfish to achieve its maximum welfare level is of equal moral importance as allowing a lion to achieve its, even if it turns out that the lion actually experiences three times the welfare intensity at its maximum than the lungfish do. In the absence of any decisive information about the relative welfare of the different species, this seems like perhaps the most sensible principle to apply.

Once we apply such a principle, we can then use something like the zero-one rule and assign the 0-1 scores to the maximum and minimum welfare levels of each animal, based on the relative *importance* of these states rather than their comparative value. Using these, we can then make our decisions through assessing different actions based on how far up their own scale each species might move – for instance, we might prefer the lungfish furniture if we have a 20% increase for each of the 10 animals, where the lion only has a 30% increase for the single individual. Instead of comparing improvements on a single objective scale of absolute 'welfare units', we would instead compare how much difference they make *relative to the individuals*

[30] There is also the possibility of exploring different degrees of moral status, though this is likely to rely on the types of considerations discussed throughout the chapter – if a lungfish and a lion have a different moral status, it is most likely to be a result of their different welfare capacity (which we've established we can't determine).

under consideration and rate them this way. This situation, while not empirically ideal, seems to capture much of what is important when making such decisions, such as giving equal weight to the interests of different individuals.

In the absence of the possibility of making determinate comparisons, an equal consideration view is perhaps the best we can do. Whatever principles we use for ethical decision-making in these cases of uncertainty about comparative welfare, they are likely to be specific to context and background values. Despite how one decides to make decisions in these cases, what is most important to highlight is that we shouldn't attempt to make direct comparisons of welfare, as this is highly unlikely to lead to reliable results of the type we want.

6.6. Conclusion

When measuring animal welfare, there is a problem in making comparisons between individuals, as doing so relies on background assumptions about similarity that require justification. In the case of animals of the same species, we have reasons of parsimony arising from analogy and shared evolutionary history that can justify the use of these assumptions. In the case of comparisons across species, we cannot justify such assumptions and instead need to use different ethical or distributive principles to make the decisions in which we may otherwise have wanted to use comparisons.

7. Chapter Seven – Integrating Subjective Experience into Welfare

7.1. Introduction – the problem

Animal welfare is often described as a single state, ranging on a continuum from good to poor. As I argued in Chapter Two, welfare should be understood as consisting in the subjective experience of animals - the way that they experience their own lives as going well or badly. Within the rest of the thesis, as I've discussed the measurement of subjective welfare, I have taken it as a single integrated entity that can be measured and compared. However, this is not obviously the case. The subjective mental states, or affects, that an animal can experience, both positive and negative, are hugely varied. Negative experiences include pain, fear, hunger, boredom, breathlessness, thirst, discomfort and nausea. Positive experiences include satiety, comfort, curiosity and companionship. These are sets of extremely heterogeneous states. From our own experience, we know that they are subjectively different, creating sensations of different types and with differing levels of impact. They are brought on by different internal and environmental causes, lead to different physiological and behavioural outputs, and are the results of differing brain pathways. This creates issues for how we understand the commensurability of these different affects, as discussed in Section 4.3 – how we can compare them, or integrate them into a single measure.

This problem has been noted in the literature on animal welfare science, particularly in attempts to create frameworks for measuring welfare that rely on sets of different conditions: "there is also a problem of comparing qualitatively different experiences associated with different treatments … until we have a common metric for unpleasantness, such comparisons are speculative at best" (Beausoleil & Mellor, 2015, p. 42). If we want to solve this problem, we must find some sort of underlying 'common currency' onto which we can convert different affects and with which we can determine the relative weightings in their contribution to overall welfare. The alternative is to consider welfare not as a real integrated state, but instead as construct: something like health, made up of multiple different components that, though individually real, do not together form any naturally existing state. In this case there would be no empirical facts to which we could appeal in making overall welfare assessments or calculations of trade-offs between affects and we would instead have to rely on pragmatic or other considerations in constructing weightings. Though these will still enable us to perform

many of the tasks we wish to in welfare measurement, this could weaken our confidence in the applicability of the resultant scores (see Section 7.4).

There is no doubt, of course, that we can individually measure different affects. We are able to measure both the causal and effect indicators (as described in Chapter Five) for a number of different affects. These can then be quantified either in terms of the absolute value of the particular indicators measured, or in terms of relative response on a scale of minimum to maximum. For example, if we are measuring hunger, we might do so by looking at how hard an animal would work for access to food. Comparing the measure at any one time to the possible minimum and maximum scores (representing complete satiety and extreme starvation respectively), we could then come up with a score representing the current state of hunger along that scale. This can be done using various indicators for a variety of affects such as pain, fear, curiosity etc. At the moment, there are many more well-validated indicators for negative states than positive ones, but current work is changing this and the range of measurable states is increasing.

However, having these simple measurements for each different affect is not useful in this context if there is no way of 'converting' the units between the different states to make comparisons or to combine them into a single measure. What we need is some way of quantifying the affects onto the conversion scale, and it is here the difficulty lies. This requires a common currency with which we can make comparisons. In this chapter, I will start by describing the problem in more detail, before moving on to provide reasons why we should be confident that there is such a common currency, and outlining how in practice we might go about assigning the relevant weightings to different affects through using different types of welfare assessments or through construction based on pragmatic concerns.

There are two primary reasons it is important that we have a common currency for comparing and combining the different affects that make up welfare. These are: to make overall welfare assessments of animals, and in making decisions regarding trade-offs in animal management. Assessing the overall welfare of animals is the target of many animal welfare measures and assessment frameworks. We want to know whether an animal is faring well or badly, not just with regards to some specific affect or another, but in a more general sense. We want our animals to be experiencing total positive, rather than negative, welfare. Making overall quality of life assessments allows us to make comparisons between how well animals are doing under different types of management (e.g. Littin, Fisher, Beausoleil, & Sharp, 2014 on comparing the humaneness of different pest-control techniques). It also allows us to make end-of-life decisions, based on the current and expected future welfare status of an animal. To

make these overall assessments we need some way of taking measures of the separate affects an animal is experiencing and combining them into a single total measure. In the best-case scenario, we want to do so by converting each experience into some sort of common currency, and weight each accordingly in combination; the alternative is the use of constructed weightings as will be discussed in Section 7.4.

In addition, when managing animals, we frequently need to make trade-offs. We might need to inflict some sort of pain or discomfort now, in the form of a medical procedure, to prevent some future suffering, or to allow for future positive experiences and wonder whether the positives can compensate the negatives. For example, we might want to decide whether it is worthwhile to catch up a male animal (causing stress) and castrate it (causing pain) in order to allow it to live in the future with a group of females (providing positive social interactions). Or we might need to decide on the best use of resources in order to maximise welfare. This may take the form of comparing different interventions to increase positive states: e.g. whether it is better for us to use resources to provide heating for an animal (increasing thermal comfort) or to provide a more varied diet (increasing consumption pleasure). It may also take the form of comparing interventions aimed at decreasing negative states (e.g. frustration) to those aimed at increasing positive states (e.g. providing favoured foods). To make these decisions, we require information about the relative strengths of different experiences, and how strongly they affect overall welfare.

Although there are two different applications described here, they reduce to the same problem. The first requires us to take a number of different affects and integrate them, somehow summing their differing contributions into a single measure representing overall welfare. The second requires us to take two or more different experiences and compare them to determine which may have the greater impact on welfare. But although the questions may arise in different contexts - as discussed above - the solution will be the same. That is, both will require us to find a common currency for welfare, and determine the relative weightings of different experiences, in order to either compare or combine them.

Where there could be a distinct problem is in making comparisons within and between valences (that is, positive and negative states). The first is that of combining the different experiences within a single valence (either positive or negative) into a total unidirectional score. For instance, can I bring together my experiences of comfort, pleasure and satiety into a single measure of positive experience, such as happiness? This would require both the ability to quantify each of these different affects, and to weight them relative to one another in total calculation. The second issue is that of comparing the different valences – the sum total of

positive experiences to that of negative experiences. That is, is there a particular amount of positive experience which could be considered to be equivalent to or compensatory for an amount of negative experience? It seems plausible to think that I would not like to experience, say, 10 units of suffering, even if it were to gain me 10 units of pleasure. One potential response to this is to just argue that one unit in either direction is the same, but that it is a lot easier to get a negative unit than a positive one – that negative experiences affect our welfare to a greater degree than positive ones do. Another potential answer is to say that negative experiences should have a greater weighting – that it would take, say, two units of positive to outweigh one unit of negative: "it takes a lot of pleasure or happiness to outrank a fairly small amount of pain or misery" (Griffin, 1986, p. 84). When we are trying to compare negative and positive welfare, we might just need to give a greater weighting to negative welfare as "a fairly small amount of misery will turn out to make life worse to a greater degree than a fairly large amount of happiness makes it better" (Griffin, 1986, p. 84).

In some sense, this is the same problem, in that it requires quantification of and weighting of different types of experience onto a single metric. However, in another sense it is a much deeper problem, as the positive and negative states are far more heterogeneous, and whatever underlying state they may combine to form, may not be of the same kind. We could in fact have two different common currencies in play here, and it may be much more difficult to find a way to weight them against one another. Indeed, Mellor's work on measuring animal welfare (e.g. Mellor, 2016) separates the two issues of comparing within and between valence, attempting the first through creating single measures for overall negative and positive welfare, but leaving the second by refusing to attempt to combine the two into a single unified score. However, he also discusses the concept of overall 'quality of life', which relies on the relative balance of positive and negative experiences, implying there is some meaningful way of comparing these, even referring to the "integrated subjective outcome" (2016, p. 15) of experiences, that represents the total welfare status. This suggests he is more concerned with the epistemic than the metaphysical problem I will describe below. Again, I will argue in Section 7.3 that these questions (comparisons with and between valence) are also reducible to the same issue – finding the common currency with which to weight the different experiences, and that this will hold both for with-valence and across-valence comparisons.

Solving the problem of comparing or combining welfare measures involves answering two different questions. The first is a metaphysical question about the existence of some underlying welfare state that allows us to create a common currency for measurement and comparison. The second is an epistemic question about if and how we could in practice convert our

measurements of different affects into this common currency to make comparisons or determine overall welfare state. These questions are separable, and the answer to one does not rely on the answer to the other. We may establish the existence of a common currency but be unable to gather the information required to convert and compare different affects using it. Or we may not find reason to think there is a common currency, but still feel that nonetheless we can have reliable and useful methods for comparing and combining different affects that are sufficient for our purposes (see Section 7.4). In the following sections I will describe each of these questions in more detail, arguing that we have good reasons to think there is such a common currency, and going on to describe some possible ways we might approach the question of measurement through combined use of two different types of welfare assessment, finishing with a brief outline of how we might proceed if we do not accept the presence of a common currency.

7.2. The metaphysical question – existence of a common currency

The first question is the metaphysical question relating to the existence of the common currency. That is, whether or not there is actually some single underlying state onto which we can map all the different affects, that would provide a basis for believing we can compare or combine them. If there is no such state, then they are incommensurable, and we would not be able to compare them (see discussion in Chapter Four). If we want to talk about welfare as a single entity as opposed to a heterogeneous collection of affects, there needs to be some underlying state that composes welfare and onto which we can map these various affects. This will form the common currency we would then use for measurement and comparison. This does not necessarily require a deep metaphysical commitment to a particular type of entity, as we can still do a lot by instead taking this underlying state of welfare as a construct and therein using constructed weightings for comparison (see Section 7.4). However, if we do have a meaningful underlying state that can form our common currency, this adds substantial strength to the work we do in measurement of welfare.

There are several reasons to think that such a common currency might exist – common usage/intuition, introspection, trade-offs and decision-making, and functional and structural similarities between affects. Note here that these reasons for thinking that there is an underlying state by which we can measure and compare different affects, work equally both within and between valences. Our common usage and intuition seem to allow us to make assessments combining positive and negative experiences, and trade-offs occur frequently between positive

and negative experiences. This means that, if there is a common currency, it is likely to be one that allows us to make both types of comparisons.

7.2.1. Common usage, intuition and introspection

The first line of evidence comes from common usage. The concept of welfare, as an overall integrated experience, is commonly used by both animal welfare experts and within the general population. We frequently refer to concepts like 'welfare status' or 'quality of life' when talking about ourselves or other animals. We can easily make sense of the idea of an individual having a state of welfare that is an integrated experience of a variety of affects; and often welfare is taken to mean something like the overall happiness of an animal. This concept is also used regularly within the animal welfare literature, for example: "animals' negative and positive mental experience, the overall balance of which underlies their welfare status or quality of life" (Littlewood & Mellor, 2016, p. 2) and the description of welfare as "the integrated balance of all sensory inputs to the animal's brain that are cognitively processed and experienced as emotions or feelings" (Beausoleil & Mellor, 2011, p. 456). This is the most common understanding of welfare found within this literature, with quotes such as these demonstrating the easy acceptance of this concept. This is of course no strong argument, but the fact that we can and do comfortably use the concept of welfare in this sense of integrated experience, should give us some increased confidence that there is such a state.

Indeed, intuitively it seems that there is some such overall welfare state, made up of different experiences, that we can and do understand and discuss. We can look at an animal and its condition and talk about whether it has overall positive or negative welfare, and where it might roughly sit on that continuum. For example, we might see a sheep standing in a field and observe its nearby social companions, lack of access to shade in the hot sun, plenty of grazing grass, and conclude it probably has acceptable, but not great, welfare.

Common usage and intuition may not be a great guide to the way the world actually is, as folk concepts do not always map onto real entities in the world. In particular, the above considerations would not rule out the possibility of welfare as a composite or construct, as opposed to a real integrated state. However, this intuition is reinforced through our experience of introspecting on our own state. I consider myself right now and the combination of states I am experiencing – mild hunger, physical comfort in my office chair, slight head pain from a lingering cold, anticipation of my upcoming lunch, some intellectual discomfort from trying to write this chapter, among other states. These are all different experiences - causing distinct sensations, and with differing impact on my overall wellbeing - not just additions on some

single continuum. But still there does seem to be some sense in which I am incorporating all these affects into some greater overall experience I could call my welfare, at least at a given moment in time. On reflection I judge my overall state as positive – I have several positive affects, only a few (mild) negative ones, and I feel like on balance, the positive are outweighing the negative. In an extremely informal way, I have combined the various types and degrees of the different affects into a single judgement on my welfare. However, it is more than just a judgement, it is an experience – overall, I *feel* good. There is some experience I am having that is more than the individual affects themselves, but instead is some sort of total state they are contributing to – what I might want to call my welfare. My subjective experience is one of integration of these varied affects. The same could be said for my diachronic assessment of my lifetime wellbeing so far – I can assess and integrate the different experiences I have had, both positive and negative, to estimate what I think my current welfare is in this regard.

This is, of course, also no proof of anything. My own introspective intuition that I have some sort of total welfare experience doesn't say much about how the world actually is. Introspection is often not the most reliable guide to the contents of our own minds, and so introspection cannot determine whether it is really the case that my experience is integrated in the way I feel it to be. It tells me even less about whether this may be the case for others when I can't access the contents of their minds, particularly nonhuman animals, whose minds are likely to be quite different from my own. We may be additionally concerned that similar considerations to those described above could also be applied to our usage of the term 'health', even while accepting that this is a construct. However, this provides at least some limited evidence that such an integrated state of welfare is possible and that there is some common currency mediating the different affective experiences. Along with the other considerations outlined below, it gives us initial reason to think that it might exist and that we might therefore be able to measure it, reason that is strengthened by these further lines of evidence.

7.2.2. Trade-offs and decision-making

We also think we can make trade-offs between different experiences, and do so frequently. For instance, we don't feed a lot of high-calorie food to captive animals, because we judge that their current sensory pleasure in receiving a preferred food is outweighed by the future negative experiences caused by ill health, such as nausea, discomfort and limited mobility. We will perform stressful catch-ups and painful veterinary procedures in order to decrease some sorts of future ill-effects on health or behaviour. In making these decisions, we are doing some sort of rough comparative calculation as to the relative strengths of these experiences and their

impact on overall welfare. Although these trade-offs are often very imprecise and context-specific, the fact that we are able to make them at all gives us reason to think there must be some sort of common currency we are comparing.

Animals will also make trade-offs like this for themselves, such as whether it is worth facing a fear of a novel object to obtain food or experience some amount of pain or discomfort to gain access to and explore a new area. They must constantly make decisions for action between competing motivations – for instance, a lizard which is currently hungry, thirsty and cold must decide whether to first eat, drink or find shelter. These are usually not conscious, rational deliberations in the same sense that we ourselves may make such decisions, but neither are they strictly instinctive behavioural patterns; they are flexible and responsive to different inputs and animal personalities. This sort of decision-making requires a common currency on which to weigh the differing motivations. Without it, an animal could not function in an environment of constant competing motivations. "The animal must therefore have a 'common (value) currency' for consistently evaluating types of world and body states in spite of their inevitable variations and for preferring one type of state over another type according to its value, which is context dependent" (Ginsburg & Jablonka, 2019, p. 374). The ability of animals to make decisions and trade-offs like these is strong reason to think there is some central metric they are using and by which we can compare different affects (as will be discussed in Section 7.2.3, probably something like 'pleasure').

In the next section, I will discuss the common evolutionary role for positive and negative affect. Importantly, it is likely that the integration of positive and negative affective states allows for animals to choose actions when faced with competing motivations, by assessing perceptual information and assigning hedonic value to different options (Gygax, 2017). This cannot be done directly through comparisons of different motivations (e.g. eating and drinking) without some further common currency, which is evaluated by some "structure, which is connected to both motivational systems" (Spruijt et al., 2001, p. 150). This additional structure compares the potential reward of different actions and will motivate the behaviour with the highest reward. This process is, of course, imperfect. There will not always be a single best action which obviously produces the highest reward, and at times this will create confusion or internal conflict. However, for the most part, animals are able to navigate complex situations and make appropriate choices. The ability of animals to make sophisticated decisions for action under the influence of various competing affective states, and decide on acceptable trade-offs, is strong reason to think there is a common currency by which they are assessing the strength of competing motivations.

7.2.3. Similarities between affects and the nature of the common currency

Even if we have good reason to think there is some common currency that makes the different affects commensurable, we might not think that we will be able to determine what exactly this is. In Section 7.3, I will argue that we don't always need to know the nature of the currency to allow us to perform the calculations we need. However, if we were able to determine the currency, this would give us a definitive positive answer to our metaphysical question, as well as provide guidance for the epistemic question, to give us methods for measurement; a much stronger position than the constructed weightings I will discuss in Section 7.4. Looking for this currency, we should look at what the different affects possibly have in common. If there is some way to measure all positive affects on a single scale, this must be because they have some common feature, either in terms of their intrinsic features, or a relational property such as a cause or effect onto some other state. The very fact that we are able to group 'positive' and 'negative' experiences together means they must have at least some common property that allows us to do so.

This line of support is strengthened through thinking about the evolutionary role for positive and negative affect. It is a common view that subjective experience evolved to enhance survival and reproduction action in animals (e.g. Cabanac, 1992; Dawkins, 1998; Fraser & Duncan, 1998; Mellor, 2012; Ng, 1995; Panksepp, 2005; Spruijt et al., 2001). Examples of this thinking are common in the animal welfare literature – for example: "pain and other negative affects evolved to guard us from danger, and equally important are the positive affects that evolved to attract us to things that will probably improve our lives" (Phillips, 2008, p. 291); "emotions refer to processes, which are likely to have evolved from basic mechanisms that gave the animals the ability to avoid harm/punishment or to seek valuable resources/reward" (Boissy et al., 2007, p. 377). As I will describe, under this view negative and positive affects serve to enhance learning and motivate behaviour; and as discussed in the previous section, their integration can then help decide between competing actions.

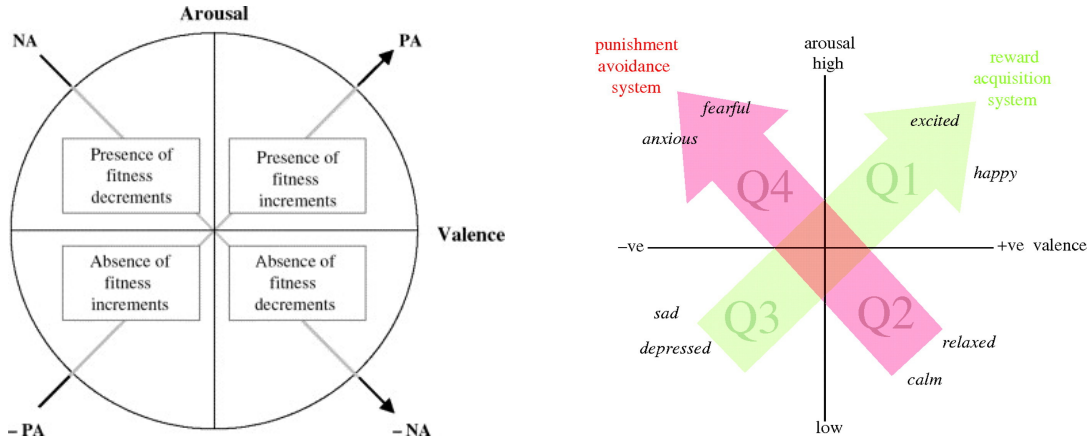
In this model of the evolution of affect, negative experiences arise from those conditions which adversely affect survival and reproduction (such as injury, thirst and disease) – so-called 'needs' situations (Fraser & Duncan, 1998) - and provide motivation to take appropriate action to reduce or remove these stimuli. These affects stimulate actions to correct whatever homeostatic imbalance they represent – breathlessness increases respiratory effort to increase oxygen supply, hunger stimulates eating to maintain nutrient balance, and pain stimulates withdrawal to minimise injury. These feelings increase in intensity - and motivation increases in urgency - as the situation becomes more critical to survival, and subside when the needs are

met. Higher intensity negative affects create a greater urgency to engage in the required behaviour. Positive experiences occur in those conditions that enhance survival and reproduction (such as mating or exploration) – 'opportunity situations' (Fraser & Duncan, 1998) - and encourage seeking out and participating in them. These experiences serve as proximate motivational mechanisms to promote behaviours serving more ultimate evolutionary ends. Positive and negative affects thus share a common motivational role, in that they all serve as motivators for action.

Recent work by Ginsburg and Jablonka (2019) argues for a similar model of the role of affect in motivation and goal-directed behaviour: "Felt values (pleasure and displeasure, although not all values need to be felt) guide the organism's behaviour and its ever-changing internal states and actions so that a homeostatic, fitness-promoting state is achieved" (Ginsburg & Jablonka, 2019, p. 101). Here, affect serves the evolutionary role of allowing organisms to ascribe value to various stimuli and actions, which influences learning and motivation through the process of reinforcement learning. Positive and negative experiences act as 'attractor' and 'repulsor' states which reward homeostasis promotion and punish its reduction (Ginsburg & Jablonka, 2019, p. 280). They describe a learning system containing a common 'reinforcement' unit that "assigns values to percepts and actions according to the deviation of the system from a state of homeostasis" (2019, p. 364). This reinforcement unit functions to categorise and integrate a variety of different stimuli in order to promote motivation and goal-directed behaviour. The presence of a central reinforcement unit such as this would provide a common role for different positive and negative affects in learning and motivation.

Similarities between the function and structure of different affects also gives us reason to think that they may not be as different as they appear and may indeed be reducible to a common currency. As described above, there seems to be a common evolutionary origin of positive and negative affect, and a common role in learning and motivation. Several authors have produced models in which different affects can be represented on a common schema. Both Burgdorf & Panksepp (2006) and Mendl et al. (2010) model emotional states in a two-dimensional space of arousal and valence (Figure 7.1). In these models, the four quadrants represent different effects on fitness (1a) and the types of emotional states that accompany these reward/punishment systems (1b). For example, a high-arousal, high-valence state would be related to reward-seeking and pleasure systems such as happiness and excitement, resulting from fitness rewards such as mating, or feeding. Low-arousal, low-valence states would relate to the absence of positive opportunities and be expressed as depression, or boredom. These models give us a way of understanding and comparing all emotional experiences along these

dimensions. This unified representational framework could underlie the common currency for measurement.



a. **b.**

Figure 7.1: Modelling positive and negative affect. a. Relation of fitness to affective states (from Burgdorf & Panksepp, 2006, p. 175). b. Mapping core affect space (from Mendl et al., 2010, p. 2896)

Different affects may thus share the same roles, and also underlying neurophysiology. Cabanac (1971, 1979, 1992) argues for pleasure as the 'common currency', with the maxim pleasant = useful (i.e. the perceived stimulus is useful to the organism), with displeasure relating to danger. He describes this as "internal signals modifying the conscious sensations aroused from peripheral receptors" (1971, p. 1107), such that the perceptual input is given a positive or negative 'gloss' by other processes, contingent on the current and past state of the animal. It is this 'gloss' that serves as the commonality between valenced experiences and provides motivation to action. "The tradeoffs between various motivations would thus be accomplished by simple maximization of pleasure . . . the displeasure of frustrating one motivation being accepted for the sake of a larger pleasure obtained in satisfying another one" (Cabanac, 1992, pp. 173–174). Spruijt et al. (2001) similarly argue that the common currency is pleasure, or satisfaction: "pleasure represented by, e.g. opioids and dopamine — or related receptor activated signals — is the currency of the brain" (Spruijt et al., 2001, p. 154). They also bring together the concepts of reward and aversion in terms of calculation between actual and expected state after an action – these are then just relative concepts acting through the same mechanism. More recently, the same line has been taken up by Ginsburg and Jablonka (2019): "pleasures and displeasures can be seen as such overall and general currencies of value, which can evaluate any percept or action" (2019, p. 374). A common currency like pleasure is flexible and allows complex evaluation and integration of different affects. With further evidence that

pleasure, or something similar, was in fact the common currency combining different affects into a single welfare state, we would have an answer to our metaphysical question.

More recent work provides the mechanisms by which this may occur, through similarities in underlying neurophysiology. Spruijt et al. (2001) claim that the 'central bank' that converts this currency of pleasure are the brain structures (such as the mesolimbic systems) which represent all these inputs. Berridge & Kringelbach (2013) claim that though different pleasures (such as food, or sex), feel subjectively quite different, they actually rely on similar underlying neural systems. They argue that "this overlapping pattern opens the possibility that the same hedonic generating circuit, embedded in larger mesocorticolimbic systems, could give a pleasurable gloss to all such rewards even when the final experience of each seems otherwise unique." (2013, p. 296) – the 'pleasurable gloss' being the commonality allowing commensurability. This is the same function described by Cabanac, with the addition of a potential mechanism. Although the perceptual pathways for each stimulus will be different, there appears to potentially be a single 'liking' system in the brain, mediated by opioid activity, that operates over a range of stimuli to create the sense of pleasure that accompanies them. These physiological similarities provide strong evidence for the existence of a common currency – most likely pleasure, or something similar - underlying all affective states. Further research into subjective experience, and its role and mechanisms, could help strengthen this case further. This still leaves open the possibility it would not do so – if it were instead to give us reason to doubt the underlying commonalities between affects, we would need to fall back on the constructed weightings I will discuss in Section 7.4.

7.3. The epistemic question – determining weightings

There are thus good reasons, even if not definitive, to think that there is some common currency, such as pleasure, onto which we are able to map different affects and through which they combine to create an overall experience of welfare. The second question to address is an epistemic question – whether or not, even if we accept the presence of a common currency, in practice it will be possible for us to know the weightings for the different components of welfare. It might turn out to be true that there is some meaningful single welfare scale, but that we are not actually able to measure the different components with enough precision to compare or combine them as we need to; and, perhaps even more likely, that we will not be able to determine the different weights we should allocate to the components to score their impact on overall welfare status. This is the position taken, for example, by David Mellor and colleagues (Mellor, pers. comm., Mellor & Beausoleil, 2015), who argue that we do not have the capacity

to know the relativities of these different affects either within or across valences, but that we can nonetheless do useful work in animal welfare and management. Here I will take a more optimistic approach and outline some methods by which we may be able to get some way towards determining what we want to know; in particular, through using a combination of different types of welfare assessment. In Section 7.4 I will also look at what we might be able to do in practice, even if we rejected the metaphysical claim about the existence of a common currency.

As I will describe in more detail in the following sections, there are two primary methods of assessing animal welfare – through whole-animal measures (or 'whole animal profiling' (Beausoleil & Mellor, 2011)) and through multi-criteria assessments (or 'systematic analytical evaluation' (Beausoleil & Mellor, 2011)). The former takes single measures to give an overall assessment of the welfare state of an animal, integrating multiple lines of information about an animal's behaviour and demeanour to give an overall rating for that animal's welfare status. The presence of such measures of overall welfare add further strength to the case for considering welfare to be an integrated state. However, it is limited in that it is unable to identify the specific conditions that are impacting welfare. By contrast, the latter uses measures of a number of indicators – usually causal indicators of the conditions an animal lives in – to bring together a complete picture of the potential welfare impacts of different situations. This can then be used to identify where welfare compromise may occur, to provide specific recommendations on which areas should be changed to improve welfare. The method is limited in that we need to know the weightings of the various welfare components in order to arrive at an accurate assessment. As I will discuss in Section 7.3.3, use of both of these assessments types together gives us the most useful way to find how to compare and combine different affects. Specific welfare impacts and their indicators can be measured through multi-criteria assessments, while their impact on overall welfare can be deduced using whole-animal measures.

7.3.1. Whole-animal measures

In many cases, we may not need to know the weightings of different welfare components at all. Whole-animal measures consist of a single indicator, used to represent the total quality of life for the animal. These include Qualitative Behavioural Assessment, cognitive bias tests and neuroimaging. These measures do not require us to know the relative weightings of the different affects, as we are taking a 'downstream' animal-based measure that gives us information arising from the integration that has already taken place within the mind of the

animal. For cases where we only want to know the overall welfare status of an animal - such as when comparing different housing systems - this will be sufficient. For cases where we want to make decisions regarding trade-offs etc., we will need further information about relative weightings, and this will be addressed in Sections 7.3.2 and 7.3.3.

7.3.1.1. Qualitative behavioural assessment

One commonly used whole-animal measure is Qualitative Behavioural Assessment (QBA) (Wemelsfelder, Hunter, Mendl, & Lawrence, 2001). Here, trained observers make assessments about the overall welfare status of an animal based on what they see of its behaviour, body language, vocalisations etc. QBA is carried out by having a set of observers assess the overall body language of an animal against either a set of fixed terms or through generating their own descriptive terms. They score the animal for each of these characteristics (e.g. nervous, alert, curious, excited) by marking a point on a line between the 'minimum' and 'maximum' for these traits (e.g. Figure 7.2). These scores can be applied either for individual animals or groups of animals, depending on the aim. These scores are then converted to numerical scores (0-100) by the researchers and analysed for general patterns.

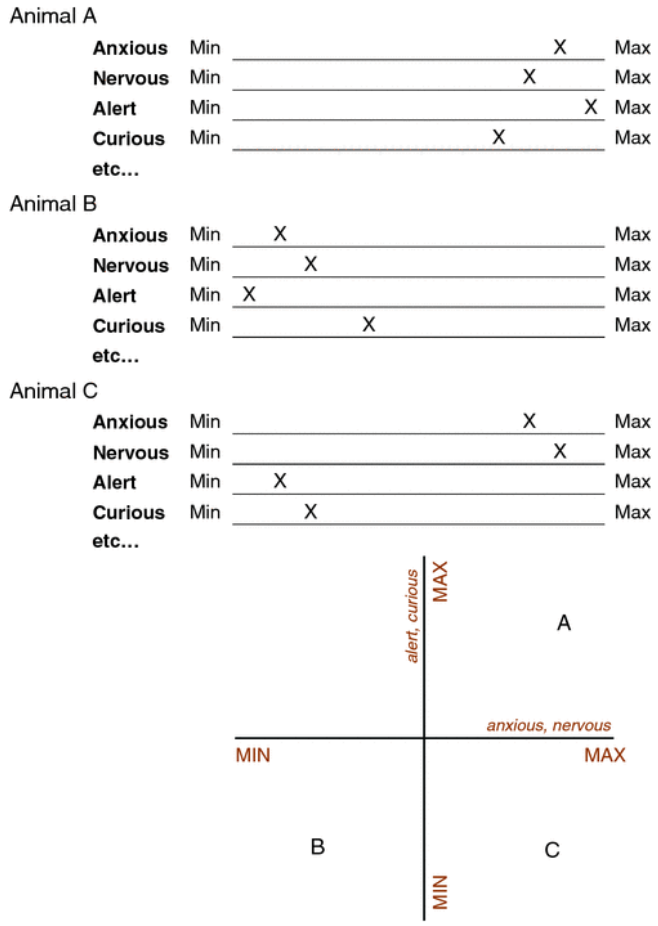


Figure 7.2: Application of QBA (from Fleming et al., 2016, p. 1570)

QBA uses the assessment of an observer to interpret the overall behaviour or demeanour of an animal – "in effect, the observers become the research instrument" (Beausoleil & Mellor, 2011, p. 457). It is an "integrative welfare assessment tool" (Wemelsfelder et al., 2001, p. 209), in which the observer is unconsciously integrating many pieces of information from the behaviour and body language of the animal. This approach is also integrative in that it is assessing the 'output' from the animal after it has already internally integrated its various experiences. The aim is to assess not so much the behaviours themselves, but the 'style' of the behaviour, as representative of the animal's overall condition (Wemelsfelder, 1997). Rather than looking at a single aspect of an animal, it is a whole-body approach that reflects how an animal is interacting with its environment. Such a measure should capture both the physical and psychological state of the animal.

This method has been shown to have high inter-observer reliability (Wemelsfelder et al., 2001) with similar scores produced across observers. Observers require only a little training and do not need to be experienced with the species in order to provide reliable assessments (Fleming et al., 2016). Importantly, it has been validated against other scientific measures of animal welfare, correlating with other relevant physiological and behavioural indicators (Wemelsfelder, 2007), illustrating that it is not merely subjective judgement but observation of real features of the animal. It is a quick and versatile measure that can be used in situations where it might be difficult to collect more detailed data (Fleming et al., 2016). It appears to be sensitive to subtle differences between animals, or over time (Fleming et al., 2016).

The primary drawback of QBA is that it may not be applicable over very many species. So far it has primarily been used for large mammals (and recently, chickens (Muri, Stubsjøen, Vasdal, Moe, & Granquist, 2019)). Given its reliance on human estimates of behaviour and body language, it may not be of much use for species very unlike ourselves or those we are not so familiar with, such as fish and insects (though Wemelsfelder (2007) thinks this is possible, just a matter of acquiring familiarity with and skill in assessing these more phylogenetically distant species).

7.3.1.2. Cognitive bias

Another emerging area of research in whole-animal measurement is cognitive bias testing. Cognitive bias is one promising measure which seems to account for overall animal welfare; cumulative welfare state, or a sum or averaging of positive and negative experiences up to that time (Mendl et al., 2010). In cognitive bias tests, animals are trained to expect a reward under one stimulus and a punishment under another. For example, if a light activates in the right-

hand corner of the room, they will receive food, and if it activates in the left-hand corner of the room they will receive an aversive stimulus (e.g. a loud noise or strong puff of air to the face). They are then presented with an ambiguous signal, such as a light somewhere in the middle of the room, to see how they react – whether they behave as though they are about to receive the reward (optimistic), or the punishment (pessimistic). By using signals with different levels of ambiguity (i.e. progressively closer to one side or other), we can map the differing levels of optimism and pessimism and infer degree of welfare. Individuals who have experienced primarily positive states, i.e. those who are likely to have lived in an environment providing fitness-enhancing rewards, will be likely to view ambiguous signals optimistically, as potential rewards. Conversely, individuals who have experienced primarily negative states (low reward opportunity environments) will be more likely to view ambiguous signals pessimistically, as potential threats (Mendl et al., 2010). State of welfare "may thus act as a heuristic device influencing cognitive processes and facilitating appropriate decision-making behaviour" (Mendl et al., 2010, p. 2900). This leads to a cognitive judgement bias varying with overall welfare experience. This has been established in human subjects, and is now the subject of a lot of work on animals, including mammals (Mendl et al., 2009), birds (Deakin, Browne, Hodge, Paul, & Mendl, 2016), fish (Laubu, Louâpre, & Dechaume-Moncharmont, 2019) and even honeybees (Bateson, Desire, Gartside, & Wright, 2011) across a range of situations. Validation of these measures must also take into account effects of personality on baseline levels of optimism and reactivity (similar to what was discussed in Chapter Six), but this can be accommodated with careful testing and appropriate sample sizes. This work is relatively new, but results so far give positive indication that this is an effective measure of overall welfare state.

7.3.1.3. Neuroimaging

One other potentially valuable future measure could be neuroimaging. Work identifying the regions of the brain responsible for generating positive and negative affect (Berridge & Kringelbach, 2013) could be used to identify the valence and intensity of affective reactions to stimuli, and possibly in assessing overall levels of pleasure or displeasure at a specific time. This work is still in the very early stages - mostly in humans – however, as discussed in Chapter Six, there have been some promising results where both subjective report of intensity of experience and behavioural responses correlate with intensity of brain activity (Coghill et al., 2003). Recently, Poirier et al. (2019) have argued that measures of biomarkers for the hippocampus (a part of the brain responsible for learning, memory and emotional regulation)

may give us information about the cumulative affective experience, or overall welfare status, of animals. Hippocampal volume, amount of local grey matter and neurogenesis (development of new neurons) have all found to positively correlate with other measures of welfare, such as subjective self-report in humans, and mood in animals (Poirier et al., 2019). This method has the advantage of measuring overall quality of life, and will not be sensitive to small changes induced through the process of restraint and measurement. If this type of measure can be properly developed and validated, it may be the strongest measure we could have, as it is accessing intensity of mental states more directly than any other measure.

7.3.2. Multi-criteria assessments

As demonstrated, whole-animal measures are highly effective for making overall welfare assessments, without the need to determine the weightings of different affects. The primary limitation of these measures is that they cannot tell us anything about which aspects of welfare are good or poor - and which conditions in the environment are responsible for creating good or poor welfare - and therefore cannot provide guidance as to what changes to make. Often, when assessing an animal, we do not want to know just how well or badly it is faring overall, so much as we want to be able to identify the primary positive and negative impacts to welfare, and how we can improve them. Whole-animal measures cannot tell us much about how to proceed in these situations, or in cases of trade-offs between affects, or choices of resource distribution. In these cases, we will prefer to use a more comprehensive multi-criteria assessment. These assessments take a variety of measures of the different conditions an animal is experiencing, and bring these together to form a total overall picture of welfare. Unlike the whole-animal measures described in the previous section, these assessments will require an answer to the epistemic question, needing at least a rough idea of the weightings of the different components in order to create an accurate picture of overall welfare. This problem will be addressed in Section 7.3.3.

Botreau et al. (2007) describe the multiple purposes for which such multi-criteria measures might be required – to advise farmers on how to improve welfare, to check compliance with legislative requirements, to implement welfare certification schemes and to compare systems to refine legislation. They argue that all these goals require combining different measures to form an overall assessment – either a relative or an absolute one. Frameworks of causal indictors[31] like this are often used because they are more easily measured on a large (whole-

[31] See discussion on types of indicators in Chapter Five.

farm) scale, and are highly repeatable (Bartussek, 1999). In general terms, for any multi-criteria assessment framework, the method will be to survey the various potential factors which can impact welfare in either a positive or negative direction (causal indicators). These factors will need to be quantified (to at least some degree, even if just ordinally such as 'low', 'medium', 'high') and then weighted in terms of their likely relative impact on overall welfare. This sort of method has the benefits of allowing us to try and make judgements in cases where more detailed information might not be available (or where it might be too time-consuming or costly to obtain), but is also very vulnerable to failures in implementation. In particular, it will not create an accurate score if some components have been left out or overlooked, or if the subjective assessments of those implementing it are not accurate. As described in Chapter Five, in any system using a set of causal indicators to assess a target state, the measurement will only be accurate if the set contains all and only those factors which actually impact the target. If we leave out a factor which is actually impacting welfare, or if we mistakenly include a factor which is actually neutral, our assessments may be quite wide of the mark. In this section, I will describe some of the multi-criteria assessment frameworks commonly used in animal welfare assessments, and their benefits and limitations, before moving on in Section 7.3.3 to outline how we might assign weightings to the different components. There are four primary multi-criteria frameworks used to assess animal welfare – the Five Domains, Animal Needs Index, Welfare Quality®, and SOWEL and related models. Here I will briefly describe the first three, and why I think they will fail to suffice for the purposes of determining weightings of different conditions or affects, before turning in more detail to the SOWEL-type models, which I think are most promising in this context.

7.3.2.1. The non-contenders

First, I will briefly discuss three multi-criteria frameworks that are commonly used in animal welfare assessment – Five Domains, Animal Needs Index and Welfare Quality® - and why they will not work for the purposes required here. The first framework is the Five Domains model (Mellor & Beausoleil, 2015), which is particularly commonly used in assessments for zoos and companion animals. In this model, the animal is scored for welfare impact over five different domains – four 'physical' domains which reflect the life conditions that affect welfare (nutrition, housing, health and behaviour), and the fifth 'mental state' domain which is what welfare is taken to consist in. This fifth domain counts both positive and negative mental states relating to provisions and lacks in the other four domains (e.g. lack of food causing hunger, or

lack of foraging opportunities causing frustration). All conditions are therefore represented with their relevant mental states, or affects, in mind. The underlying conceptual tools here are promising, but the model will not succeed in helping to determine weightings, as the impacts are only scored qualitatively, on an A-E scale, which only allows for ordinal rankings of states. No attempt is made to give a qualitative score, or to attempt to quantify the weightings of different components of the framework. This was deliberately built into the model to prevent over-precisification where the data does not support it: "numerical grading was explicitly rejected to avoid facile, non-reflective averaging of 'scores' as a substitute for considered judgment and to avoid implying, unrealistically, that much greater precision is achievable than is possible with such qualitative assessments" (Mellor, 2017, p. 10). Although the four scores are combined into a single score for the fifth (mental state) domain, this is done informally and based primarily on the knowledge and intuition of the assessor/s. Rather than functioning as a measurement tool, the model is instead aimed as a 'focussing' device, to gain a greater understanding of the welfare of an animal, and the conditions impacting it.

The Animal Needs Index (Bartussek, 1999) is a collection of proxy measures used to score and compare the conditions provided at a farm level, to ensure farms meet minimum provision standards for certification or labelling schemes, and to identify which areas should be improved. It scores indicators for five husbandry conditions that are considered to affect welfare – mobility, social contact, flooring, climate and stockman's care. Although these are taken as representing the most important welfare needs of the animals, they are not intended as a comprehensive set of measures. Additionally, measures at the farm level necessarily aggregate the welfare of individuals, and thus cannot tell us about individual welfare, which is typically the target of ethical concern (Houe et al., 2011). Although this ANI framework is linked to welfare assessment, as it is not a direct measure of welfare, it will not help in this case.

Another example of a multi-criteria assessment is the Welfare Quality® framework (Botreau, Bonde, et al., 2007), commonly used for farm animals. The model takes 12 welfare criteria, across four principles – feeding, housing, health and behaviour. Each of the criteria is then assigned several indicators, which are mostly effect rather than causal indicators, such as body condition score or incidence of aggressive behaviour. Multiple measures are taken for each criterion, which are aggregated into the criteria, followed by another aggregation into the welfare principles and finally bringing it all into an overall assessment of the quality of the facility, with regard to the welfare of its animals (Botreau et al., 2009). This is a more complete

framework than the Animal Needs Index, and produces cardinal scores, unlike the Five Domains model. However, the 12 criteria were taken from discussions with consumer-citizens as well as scientists and so do not necessarily represent welfare from the animal point of view. The aggregation weightings are also quite opaque, and seem to be based on expert opinion rather than measured effect on the animals (de Graaf et al., 2018). For this reason, this is still unlikely to be a useful model for determining actual weightings of welfare impact; and instead we should prefer the similar but more rigorously designed SOWEL-type models, as will be discussed below.

7.3.2.2. SOWEL-type models

The most comprehensive frameworks for measuring welfare and determining weightings of the different components are the SOWEL-type models. The SOWEL model was developed by Bracke et al. (2002 a; 2002 b) for assessing the welfare of breeding sows. The same framework has been used to create similar models for assessing welfare of chickens (the FOWEL model - de Mol et al., 2006) and cows (the COWEL model - Ursinus & Schepers, 2009).

These models are built by first selecting a number of needs (11 for the SOWEL model) which represent the different behavioural systems that are taken to contribute to subjective welfare – such as feed intake, thermoregulation and exploration. For each need, various attributes are identified (37 total in the SOWEL model), which are measurable conditions that contribute to that need (e.g. 'exposure to cold' for thermoregulation, or 'novel events per week' for exploration). Each attribute is given a discrete score between 0-1 depending on the quality of the measure (e.g. for number of eating places, the categories might be: sufficient, limited, and restricted, receiving scores of 1, ½ and 0). These cardinal scores allow for integration of all attributes into a single welfare score.

Finally, the attributes are given weightings. Unlike the opaque expert-assigned weightings used in Welfare Quality®, these are based on information available in the scientific literature. This evidence can be basic comments by researchers on animal behaviour, such as how strongly they think the animals want the resource, or can be more data-driven, such as the results of studies on animal preferences and willingness to work, or behavioural activity budgets. These weightings are not just assigned to the whole attribute, but different values within an attribute so that, for example, extreme pain may have a much larger 'pull' on total welfare score than mild pain. This allows for variation in the two types of weightings – weightings between conditions and weightings between different levels of a condition

(Norwood & Lusk, 2011). Weightings between conditions represent the relative strength of, say, food availability versus thermal discomfort on overall welfare. Weightings between levels represent the relative impact of mild versus strong discomfort on overall welfare. Once the weightings are set, a total score is then calculated as the weighted sum of the attribute scores.

This model is not perfect. One problem is in the scoring of the attributes. The attributes are scored ordinally (e.g. worst, mid, best) but the scores are then treated cardinally (0, ½, 1). This means that the relationship between the different levels of score is treated linearly (e.g. that the difference between sufficient eating places and limited is the same as between limited and restricted), with no justification for this actually being the case. Given that the categories are arbitrarily chosen, this may not often be true. Another issue is in the assignment of weightings – what counted as relevant data for these could vary from weighted preferences, to qualitative comments by scientists in their papers. There is no single objective and repeatable method used throughout to assign weightings, and so these will vary in accuracy and validity. However, these are issues with inputs, rather than with the framework of the model itself. The strength of these models is the transparency, and the ability to update and alter these scores and weightings as new data emerges. Although the current inputs into the model are not ideal, the underlying framework of the model is such that this could be easily improved and strengthened.

For this reason, the SOWEL-type models are currently the best available tool for accurately measuring welfare and determining weightings of different components. Their primary benefit, a feature which was deliberately built in, is in allowing for changes to be made as new information becomes available: “the decision support system is designed to be adaptable, that is, new insights can be incorporated when these become available” (Bracke, Metz, et al., 2002, p. 1835). The data in the model is directly linked to a table of the referenced data (e.g. comments in scientific papers) to allow for transparency, as well as making it updatable. If future studies show that something should be added or subtracted from the range of identified needs or attributes, or that the weightings should be altered, this can be easily accommodated. As this type of new information - and subsequent model changes - are highly likely to arise, this framework is strong in its ability to adapt accordingly. The framework can also be taken and applied to any species, with modifications needed only to the input data regarding species-specific needs and indicators. Use of this basic framework, with an eye to improving the quality of the inputs, will get us ever closer to an accurate welfare measure. In the next section, I will discuss some ways in which we could improve the scoring of weightings for these models.

7.3.3. Determining weightings

As demonstrated, there are a number of different frameworks that can be used to describe and assess welfare according to the variety of contributing conditions, the most comprehensive and flexible of which are the SOWEL-type models. As per the discussion in Chapter Five, we can take these frameworks to be valid measurements of welfare given the appropriate testing and justification for background assumptions. The primary problem with frameworks such as these is getting an answer to our epistemic question and determining the weighting of different components (Botreau, Bracke, et al., 2007). In these assessments we will want to know the weightings of different components of welfare, particularly when we need to make trade-offs between different areas of welfare impact and want to know which are acceptable for their impact on total welfare. In order to decide whether we are justified in inflicting a certain amount of pain or discomfort for some future benefit, or in deciding whether to invest resources in one benefit or another, we need to have at least a rough idea of the relative magnitudes of the costs and benefits under consideration.

Weightings are most often set by the opinions or intuitions of various practitioners or 'experts': "inevitably, the design of a practical assessment tool for animal welfare at farm level must primarily be the result of negotiation" (Bartussek, 1999, p. 186). The different backgrounds and expertise of these assessors will change the weightings assigned (Otten et al., 2017). For example, in another framework - the Animal Needs Index - the authors considered free movement as an essential prerequisite for other types of welfare experience, and so weighted this more highly (Bartussek, 1999). Mellor & Stafford (2008) describe a procedure used by McMillan for determining an animal's overall quality of life (welfare status). This is to: list all the feelings an animal is experiencing, weight these (by biological/survival value or urgency), alter the weightings for the particularities of the individual animal and finally construct a scale to assign weights to the adjusted feelings. They rightly criticise this approach as, with current methods, there is no objective way to set the values at each step – it may allow for ordinal, but not cardinal, scoring. This method does not take into account the effect on the animal, and to what degree each of these conditions actually do impact welfare. Instead, these subjective judgements will contain some value judgements about what counts in animal welfare – ethical as well as scientific judgements (Veissier et al., 2011).

Determination of weightings needs instead to be objective: in some sense, they need to be empirically constrained, to figure out what impact different experiences have on welfare, from the point of view of the animal rather than the researcher. There are multiple ways we can arrive at objective weightings for welfare components. We can do this either through testing

on the various whole-animal measures, or through preference tests. For the first, we would start by measuring the overall welfare of an animal at one point. We would then make an intervention we were interested in testing the effect of, say by changing food quality or amount of available shelter. Finally, we would measure overall welfare again, to observe the difference in the scores. This difference will help us determine the impact of this condition on overall welfare. Repeating this for many conditions would start to give us their relative weightings for the animal measured, and repeating over different animals can help find general species or group trends in overall weightings. We could also vary multiple conditions together to look for interaction effects between the impacts of different conditions. This information could then be used for making decisions about management trade-offs, or for integrating multiple weighted measures into a single welfare score. Even if rough, these weightings should still be sufficient for most applications.

In the second case, preference testing is another animal-based measure that can be used for determining the contribution of different affects to overall welfare. Using preference tests, animals can tell us which resources they prefer, and how hard they will work to get them (Dawkins, 1983). This information can then be used to scale the different components against one another, based on how the animal itself views such trade-offs. For example, we could look at whether an animal prefers x amount of resource A over y amount of resource B, or how hard an animal will work for A vs B. Either of these will give us a good idea of how much relative value the animal places on these resources, and the affects they create. Avoidance tests could give similar information for negative experiences. Preference tests also give the animal the ability to trade off positive and negative experiences, such as experiencing something noisy while accessing food, and thus allow us to determine relative weightings across valence as well.

One caveat here is that preference testing is not a perfect window into the affective state of an animal, as discussed in Chapter Two. Although very often, what an animal wants - and what they will work towards - will be those things that they find pleasurable and rewarding, this is not always the case, as the 'liking' (reward) and 'wanting' (motivation) systems have different neural substrates, and can operate independently of one another (Berridge, 1996). There can also be strong effects of history and individual temperament, and we cannot be sure that animals will maintain the same weightings over time as they age, or environmental conditions change. Validation through other measures of affect could help with this.

By recording the changes in whole-animal measures under specific changes in conditions, or by measuring something like relative strength of preference, we can gather information as

to the relative weightings of different affects on the welfare experience of the animal. This can then allow us to build up larger multi-criterion frameworks using these ratings to build an overall picture of animal welfare and its components. Using both of these types of assessment together allows us to get a picture of the overall welfare state of an animal, while still having sufficient detail about living conditions to allow us to determine where change is required. It also allows us to validate the measures against one another to make sure we have not missed anything on either side. This is possible because there is an overall integrated state of welfare, with a common currency onto which we can map the different affects that compose it; otherwise the relationship between whole-animal measures and multi-criteria frameworks would not hold. Even without knowing the exact nature of the common currency, this process allows us to determine the weightings of the different affects and thus make necessary welfare assessments and decisions.

7.4. A pragmatic proposal

I have flagged throughout this chapter that we might remain unconvinced about the metaphysical claim, regarding the existence of a common currency by which we can combine or weight different affects. However, even in this case, we would still want to make management decisions regarding welfare and resource allocation. Here, as mentioned, we might think instead of welfare as being a multi-component construct more like health. Here the different components may individually exist, but do not form an integrated state. Instead, they are brought together because of our categorisations, and interests. In this case, it would not be surprising that there is no metaphysical entity forming the common currency, or a single correct answer about the interactions between affects, and their weightings. Instead, we may have multiple accounts, each of which lays out a different way of combining affects into a single score, based on different background assumptions, normative commitments, and individual preferences. In this case, the problem of determining weightings is no longer an empirical one – there is no single privileged set of weightings that we are trying to discover. Instead, we are looking for a solution which best fits our purposes.

When considering why we want a method for integrating welfare components, we have the two reasons discussed in Section 7.1 – to make assessments of overall animal welfare, and to make considerations of trade-offs between different components of animal welfare. Most often, what we want is a way of determining how we should use our resources to maximise increases in welfare. To this end, we may consider using something like robustness reasoning (see Chapter Five). In these cases, we could compare the decisions recommended by multiple

different aggregation procedures and prefer those that are recommended by many or most different processes. Alternately, if simply considering whether a single proposed intervention is an appropriate use of allocated resources, we could examine whether it results in an increase in welfare across most or all of our different aggregation procedures. In cases where it does, we would proceed, and where it does not, we may instead look for an alternative. Something like the SOWEL model previously discussed could be extremely useful in these applications, as it allows for variation of input weightings, and could be used to test results across a range of acceptable parameters. In this way, even without the strong metaphysical commitment to a common currency and an integrated state of welfare as a real entity, we would still be able to make relevant decisions regarding animal husbandry and management.

This of course still leaves us with the issue of deciding on which set of aggregation procedures we consider reasonable. There are an infinite number of ways of weighting the different components of welfare and their contribution to the overall score, and so robustness reasoning of the type described above will not be of much use if we were to take all of these under consideration. Instead, we need to decide on a reasonable subset of those procedures. I will not here advocate for any definitive way of deciding on what would count as reasonable for these purposes, but I will raise a few relevant considerations.

Firstly, we would want to discount any procedure that places too much, or too little, weight on each component. If we are considering any particular condition or affect as a component of welfare, it is because we think it makes a significant contribution to welfare, and so we don't want excessively low weightings. Similarly, it does not seem right that any single affect will be primarily determinate for welfare, and so aggregation procedures with excessively high weightings should also be discounted.

This ties in to the second consideration, which is that our weighting procedures should be intuitively plausible. We have some sense from our own experience of which types of affects have greater and lesser impacts on our own feelings of wellbeing. Similarly, from our observations of other people and animals we have an idea of what influences them the most. An aggregation procedure that weighted transient boredom much higher than chronic pain, for example, would not be convincing. Intuitive plausibility must obviously be taken carefully. As discussed in Section 7.3.3, weightings set by expert opinion run the risk of being overly anthropomorphised, and missing those things which matter to the animals under consideration. However, when supplemented with appropriate knowledge about the animals, this can help us narrow down the set of appropriate procedures.

The primary drawback of treating welfare as a construct and using these procedures is that, compared to the case in which we believe we are measuring welfare using a real common currency, we would have weakened confidence in the applicability of our results. Our aggregation procedures would be based on human intuitions about plausibility, and the process of arriving at these intuitions is opaque (think of the Welfare Quality® framework). Where we are basing our sets of reasonable aggregation procedures on these considerations, we might fail to capture those things that matter more to the animals.

The comparison here is again with health. Health is a construct of this type, and we are frequently able to make sense of trade-offs between different components – say, taking a medication to improve symptoms of a disease, that at the same time will impact kidney function. However, there are two reasons why this case is not exactly analogous to animal welfare. The first is that we are making decisions for humans, and so our intuitions about how to weight acceptable trade-offs are much more likely to hit the mark than for animal species, particularly those quite distantly related and dissimilar to us. The second is that even in the human health cases, we might think that we are actually often appealing to some other common currency; weighting components of health in relation to our preferences for different types of disease or incapacity, or relating to our overall lifespan or quality of life. If this is true, then we still need common-currency thinking. This is not to say that the results of the procedures described above will not get us some way to achieving the ends we desire in welfare measurement and decision-making, but that they might be much more limited than we would ideally prefer.

7.5. Conclusion

Scientific measurement of subjective animal welfare is fraught with a number of problems, one of which is how we can make sense of integrating or comparing different positive and negative mental states. Despite the heterogeneity of these different affects, I have argued that we have good reason to think that these can be integrated into an overall welfare experience. Further, we can measure this overall experience using whole-animal measures, and use changes in these measures - along with preference testing - to see how animals make their own trade-offs to determine relative weightings of different experiences. This will thus allow us to make effective management decisions. Even if we remain unconvinced about the presence of a real common currency, we can still use robustness reasoning on different aggregation procedures to identify the best decisions under these circumstances.

8. Chapter Eight – Conclusion and Future Directions

Most people agree that animal welfare is important. It is bad for animals to suffer, and good for them to have happy lives, and where possible we should act to prevent the former and enable the latter. As Norwood & Lusk (2011) point out "almost everyone cares about the treatment of farm animals, to some degree" (p 5, Ch. 1). All else being equal, people would prefer animals receive better, rather than worse treatment. This means that this is an issue of moral concern. But we have limited resources available to us, and so are required to make decisions about where to prioritise our actions or interventions. All our actions come with an opportunity cost - that of some other action we could otherwise have taken that may have provided other benefits. Typically, we want to choose actions in such a way to have the greatest possible impact. If our goal is to reduce suffering, then we want to reduce suffering as much as we can. When considering animal welfare, making decisions like these then requires us to have an accurate understanding of the welfare of animals living under different conditions. Working to identify and implement our best interventions for improving animal lives requires accurate measures of animal welfare, in order to assess the current best and worst animal systems, as well as to compare different possible interventions for their effectiveness. Otherwise, we risk misapplying our resources to interventions that may fail to prevent, or may even increase, suffering. Although animal ethics has been well-explored by philosophers, animal welfare science and the measurement of animal welfare has largely been overlooked within the philosophy literature. The primary aim of this thesis has been to redress this oversight, and to make some steps towards opening up philosophical investigation into the measurement of animal welfare.

In Chapter One, I introduced the study of animal welfare, looking at the three primary questions in the field: regarding the moral status of animals, which animals are moral subjects and how we measure animal welfare. I described the field of animal welfare science, which aims to objectively measure the welfare of animals using behavioural and physiological indicators to track changes in welfare under different conditions.

Part One of this thesis aimed to establish the subjective conception of animal welfare, and its applications. In Chapter Two, I argued that we should understand the welfare of animals as consisting in their subjective experience over their lifetime. I argued that this welfare concept was able to fulfil both the normative and scientific roles required, as it is normatively

significant, fundamental and measurable, and defended it against some possible objections. I examined the other competing candidate welfare concepts – physical welfare, teleological welfare and preferences – and argued that they merely provide instrumental conditions for the realisation of welfare, important only with regard to their effects on subjective experience.

In Chapter Three, I introduced two cases in applied animal ethics. First, there was the issue of management euthanasia in zoos, where otherwise healthy animals are culled for reasons of space or resources. Second was de-extinction projects, in which cloning or back-breeding technologies are used to attempt to bring extinct species back into existence. Here I examined how the subjective welfare concept influences our decision-making in these areas, particularly regarding how welfare interacts with other competing values when assessing the acceptability of these practices.

In Part Two, I moved on to investigation of the measurement of animal welfare. Chapter Four examined whether subjective welfare is a measurable entity. Using measurement theory, I looked at the different types of measurement scales and argued that welfare would be best captured by either an ordinal or a ratio scale, depending on the necessary application.

In Chapter Five, I turned to the problem of validating indicators of welfare. As we cannot measure subjective experience directly, we must rely on the use of indirect indicators, but these have to be validated to ensure they are really measuring our intended target state. I described a four-step robustness procedure to validate these indicators in the absence of direct correlational data about changes in the target state.

In Chapter Six, I looked at the issue of making intersubjective comparisons of welfare. I pointed out that making comparisons of this type is difficult, as observed variation in the data can be explained either by variation in welfare experience, or by variation in individual responsiveness, with no way to tell between the two. I argued that in order to make such comparisons, we must make similarity assumptions regarding either scope of experience, or response level, justified by anatomical/physiological analogy and shared evolutionary history, and which assumption we prefer will depend on the context and details of the particular case. I finished by outlining some alternative decision procedures we might use in situations where these assumptions are not justified, such as in comparisons between species that are not closely related.

Finally, in Chapter Seven, I examined the problem of heterogeneity of mental states and what this might mean for commensurability and integration into single welfare measure. I argued that we have good reason to think that there is some ‘common currency’ by which we could do so, supported by our intuition and introspection, use of trade-offs and decision-

making, and the similarities in evolutionary role, function and structure. I described a process by which we might determine the relative weightings of the impact of different experiences on overall welfare, using a combination of whole-animal welfare measures and multi-criteria welfare assessments.

This is all introductory work, and there is much more to be done. At present the philosophy of animal welfare science is basically a non-existent field; though it is my hope that the work contained here will help stimulate interest in what will be a fruitful area for research. As this work is just beginning, there are still many core issues to explore, and connections to work in other areas. As indicated throughout the thesis, one of the primary fields I believe will have important ties to animal welfare research is animal sentience research. Animal welfare and sentience are tightly linked, as are the research programs, and further collaboration could help them inform one another.

Animal sentience research is a body of research that looks into animal sentience from a variety of perspectives, such as how it evolved, the mechanisms by which it develops and operates, and through which parts of the animal kingdom (or beyond) it extends. This is a multi-disciplinary field, drawing on - among others - cognitive science, neuroscience, animal behaviour, ecology and evolution, and philosophy. The aim is to understand how, when and why sentience occurs, and what forms it takes. Animal welfare science, by contrast, is examining the experience of animals from the perspective of what makes their lives go well or poorly. Both are emerging fields. A recent meta-analysis of published research in animal sentience found an almost 10-fold increase over the 20 years from 1990 – 2011 (Proctor et al., 2013). A similar recent meta-analysis of animal welfare publications over the past 20 years found an increase of around 10-15% per year, with over half having been released in the last four years (Walker, Díez-León, & Mason, 2014).

These links between the study of sentience and welfare, and the overlap in subject matter, give reason to think that both disciplines have something to offer one another, both conceptually and methodologically. Proctor claims that "the science of animal sentience underpins the entire animal welfare movement" (Proctor, 2012, p. 628). I see this connection occurring in three ways – the empirical question of which animals are sentient and thus targets of welfare concern; a methodological link, where the methods developed within each discipline can be used to assist the other, and the conceptual questions of the evolution and functioning of sentience which underpin the assumptions used in welfare science.

The most basic, and perhaps most important, way these two areas work well together is in identifying which animals should be the targets of welfare concern. As discussed, welfare is

taken here to consist in the subjective experiences of animals, their positive and negative experiences. It therefore follows that only those animals capable of such experience – those animals which are sentient – will be the appropriate targets of concern for welfare. "Acknowledgements of sentience would seem a prerequisite for concern about animals' well-being" (Walker et al., 2014, p. 80). Animal sentience research thus gives animal welfare science its targets – once there is sufficient evidence that an animal is sentient, and thus has welfare, welfare science will then work to investigate under what conditions the lives of these animals will be improved or worsened. We may also see the reverse effect – tracking those animals for which animal welfare science is successful or unsuccessful in measuring welfare can tell us something about whether or not they are sentient. Where welfare measures appear successful on an animal previously not considered sentient, or unsuccessful on an animal previously considered to be sentient, this would give us reason to revisit our assessment of sentience and perhaps revise methods accordingly. The current debates on the boundaries of sentience and the status of fish and invertebrates, as discussed in Chapter One, show how important use of strong established indicators is for both disciplines.

This sharing of methods, which can go both ways, is the second type of interaction between the two disciplines. We are still a long way from identifying conclusive behavioural or neurological markers of sentience. Current work in this area thus rests on untested assumptions, some of which can be tested or justified through developments in animal welfare science. Looking to welfare science for their well-validated indicators will also assist. Animal welfare science has over the years developed sound methodologies for measuring welfare. Particularly, they have identified and validated a range of both casual and effect indicators. The large amount of work done in welfare science to identify and validate these indicators for welfare gives sentience research a good pool of measures to draw from. Further, looking into which indicators are used, and why they work, could help develop understanding of sentience and how it functions. Taking from welfare science the well-validated behavioural and physiological indicators, these form a good starting point for investigation into the mechanisms by which affective states can produce the effects, which helps then to answer some of the questions raised below.

The final set of questions that animal sentience can answer, and which can assist welfare science, are some more conceptual questions about the evolution of and working of sentience. The answers to these questions will give us information both about the expected taxonomic distribution of sentience, and about the adaptive role sentience may have played. We can also learn about the structures and processes that give rise to different types of sentient experiences,

and the ingoing and outgoing pathways - both from stimuli to affect, and back from affect to physiological and behavioural responses. There are many areas of welfare science which rely on an understanding of the evolution and development of animal sentience, and the mechanisms by which it works, as have come up throughout this thesis.

As described in Chapter Five, part of the process of validating welfare indicators involves embedding within the best available theory. That is, if we understand the mechanisms working between welfare experience and the measured indicators, we have more reason to think that our measurements are mapping onto the right state of the world. So, for example, if we take the vocalisations of goats, we will have more confidence that this is mapping onto welfare experience if we can understand that goats are social animals that communicate their distress to conspecifics. If we take blood cortisol measurements, we will be more confident with their reliability if we understand the hormonal cascade that creates changes in cortisol and under what conditions it is triggered. We will also have reason to think we have made the right choice of conditions from which to test indicators. For example, understanding the evolutionary history of a stoat will help us to think that provision of water is a relevant positive stimulus, while for a tamarin presence of an aerial predator is a negative one. Animal sentience research helps provide understanding of these mechanisms, both in their operation and their evolution, and thus can help welfare science with the right choice of indicators.

In Chapter Six I outlined the similarity assumptions required for making comparisons of welfare between different animals. The justifications for these assumptions were based in appeal to evolutionary history and analogous anatomy and physiology. Here we can see again the role for sentience research. Understanding where sentience in general - or the particular types of affect or indicators in use - evolved will help us to know whether there were similar enough forces acting on the different individuals to create the same responses. Additionally, understanding the mechanisms by which sentience operates, how different affects and responses are created, will allow us to see whether these pathways are relevantly similar between individuals.

Lastly, in Chapter Seven I discussed making comparisons between positive and negative experiences, and the need to find a common currency. The ability to find such a common scale relies on understanding the neurological underpinnings of the different experiences and the evolutionary conditions through which they evolved; as well as those for sentience as a whole. Similar brain structures or chemicals involved gives us reason to think there are similar experiences. Similarity of evolutionary conditions, such as all negative affects being used for aversive learning, gives us reason to think there is a common currency. Similarities in either of

these areas supports the idea that we can be successful in comparing different types of experience in terms of their contribution to overall welfare.

These are some of the ways in which animal sentience research and animal welfare science could aid one another. This is not a revolutionary idea. Animal sentience research and welfare science have been overlapping and working together in various ways for decades, often with the same researchers working on both. All I have tried to do here is indicate some of the specific ways in which they have, or could, help one another. Understanding the application of sentience research in animal welfare science may help guide research programs into sentience; while following work in sentience will help welfare scientists develop new measurement indicators and identify new species for study. Drawing on the work already done in welfare science on developing indicators and methods of measuring welfare can help sentience research for work trying to identify sentient species, as well as understanding the causes and mechanisms of sentience. This is merely a call for ongoing collaboration in this area. Both disciplines can benefit from the other, and working together may help more quickly solve some of the problems both are investigating.

If we are able to accurately measure animal welfare, we will be much better placed to make decisions about when we should take action on behalf of animals, and of what type. We can compare the value of different interventions before deciding what to do. Philosophical analysis of the methods of animal welfare science is crucial for ensuring that the science is producing relevant results for use. The work in this thesis doesn't tell us anything specific about what we should do, but gives us better tools for figuring it out. Some of the types of actions that can be assessed include advocating for changes in production methods to improve welfare and encouraging consumer shifts between or away from animal products to reduce numbers or to change types of animals used. In particular, a recommendation arising from this work is that funding would be well-used in improvement of research; to further develop the techniques described for empirically measuring welfare, particularly in areas like cognitive bias testing and neuroimaging.

Accurate measurement of animal welfare is a crucial part of the process of making decisions for action to improve animal lives. This will require active engagement with the current science of animal welfare, as I have started to do in this thesis. Philosophical work in this area will examine how this science is practised: the concepts used, and the underlying assumptions in the methodologies. For the most part, this thesis has focussed on what is theoretically possible in the measurement of animal welfare; there are separate questions on what we should expect to be possible under our current pragmatic and epistemic constraints. There still remains much

more scientific and philosophical research to be undertaken in order to clarify and strengthen our understanding and measurement of welfare, but it is my hope that this thesis has played a role in opening up a new frontier for fruitful interdisciplinary investigation and developing the important role philosophy plays in such an integration.

9. References

Alexander, L., & Moore, M. (2016). Deontological ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Retrieved from https://plato.stanford.edu/archives/win2016/entries/ethics-deontological/

Asa, C. (2016). Weighing the options for limiting surplus animals. *Zoo Biology*, *35*(3), 183–186.

Baciadonna, L., & McElligott, A. G. (2015). The use of judgement bias to assess welfare in farm livestock. *Animal Welfare*, *24*(1), 81–91.

Barnes, H. (2014, February 27). How many healthy animals do zoos put down? *BBC News*. Retrieved from http://www.bbc.com/news/magazine-26356099

Bartussek, H. (1999). A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science*, *61*(2–3), 179–192.

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology*, *21*(12), 1070–1073.

Beauchamp, T. L. (2011a). Introduction. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 3–31). Oxford: Oxford University Press.

Beauchamp, T. L. (2011b). Rights theory and animal rights. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 198–227). Oxford: Oxford University Press.

Beauchamp, T. L., & Frey, R. G. (Eds.). (2011). *The Oxford Handbook of Animal Ethics*. Oxford: Oxford University Press.

Beausoleil, N. J., & Mellor, D. J. (2011). Complementary roles for Systematic Analytical Evaluation and qualitative Whole Animal Profiling in welfare assessment for Three Rs applications. *Proceedings of the 8th World Congress on Alternatives and Animal Use in the Life Sciences, Montreal, Canada*, 21–25.

Beausoleil, N. J., & Mellor, D. J. (2015). Advantages and limitations of the Five Domains model for assessing welfare impacts associated with vertebrate pest control. *New Zealand Veterinary Journal*, *63*(1), 37–43.

Beausoleil, N. J., & Mellor, D. J. (2017). Validating indicators of sheep welfare. In J. P. C. Greyling (Ed.), *Achieving Sustainable Production of Sheep* (pp. 327–348). Burleigh Dodds Science Publishing.

Beck, B. (1995). Reintroduction, zoos, conservation, and animal welfare. In B. G. Norton, M. Hutchins, E. Stevens, & T. L. Maple (Eds.), *Ethics on the Ark* (pp. 155–163). Washington and London: Smithsonian Institution Press.

Bennett, J. R., Maloney, R. F., Steeves, T. E., Brazill-Boast, J., Possingham, H. P., & Seddon, P. J. (2017). Spending limited resources on de-extinction could lead to net biodiversity loss. *Nature Ecology & Evolution*, *1*(4), 0053.

Bentham, J. (1879). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne & Son.

Berridge, K. C. (1996). Food reward: Brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, *20*(1), 1–25.

Berridge, K. C. (2009). Wanting and liking: Observations from the neuroscience and psychology laboratory. *Inquiry*, *52*(4), 378–398.

Berridge, K. C., & Kringelbach, M. L. (2013). Neuroscience of affect: Brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, *23*(3), 294–303.

Berridge, K. C., & Robinson, T. E. (1995). The mind of an addicted brain: Neural sensitization of wanting versus liking. *Current Directions in Psychological Science*, *4*(3), 71–76.

Binmore, K. (2009). Interpersonal comparison of utility. In H. Kincaid & D. Ross (Eds.), *The Oxford Handbook of Philosophy of Economics* (pp. 540–559). Retrieved from http://discovery.ucl.ac.uk/14435/1/14435.pdf

Boccia, M. L., Laudenslager, M. L., & Reite, M. L. (1995). Individual differences in Macaques' responses to stressors based on social and physiological factors: Implications for primate welfare and research outcomes. *Laboratory Animals*, *29*(3), 250–257. https://doi.org/10.1258/002367795781088315

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I. & Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*, *92*(3), 375–397.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M. B. M., Capdeville, J., & Veissier, I. (2007). Aggregation of measures to produce an overall assessment of

animal welfare. Part 1: A review of existing methods. *Animal*, *1*(8), 1179–1187. https://doi.org/10.1017/S1751731107000535

Botreau, R., Bracke, M. B. M., Perny, P., Butterworth, A., Capdeville, J., Van Reenen, C. G., & Veissier, I. (2007). Aggregation of measures to produce an overall assessment of animal welfare. Part 2: Analysis of constraints. *Animal*, *1*(8), 1188–1197.

Botreau, R., Veissier, I., & Perny, P. (2009). Overall assessment of animal welfare: Strategy adopted in Welfare Quality®. *Animal Welfare*, *18*(4), 363–370.

Bracke, M. B. M., Metz, J. H. M., Spruijt, B. M., & Schouten, W. G. P. (2002). Decision support system for overall welfare assessment in pregnant sows B: Validation by expert opinion. *Journal of Animal Science*, *80*(7), 1835–1845.

Bracke, M. B. M., Spruijt, B. M., Metz, J. H. M., & Schouten, W. G. P. (2002). Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. *Journal of Animal Science*, *80*(7), 1819–1834.

Brambell, F. W. R., & D S. (1965). *Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems*. Retrieved from http://edepot.wur.nl/134379

Briefer, E. F., Tettamanti, F., & McElligott, A. G. (2015). Emotions in goats: Mapping physiological, behavioural and vocal profiles. *Animal Behaviour*, *99*, 131–143.

Bringmann, L. F., & Eronen, M. I. (2016). Heating up the measurement debate: What psychologists can learn from the history of physics. *Theory & Psychology*, *26*(1), 27–43.

Broom, D. M. (1986). Indicators of poor welfare. *British Veterinary Journal*, *142*(6), 524–526.

Broom, D. M. (2016). Considering animals' feelings. *Animal Sentience*, *1*(5), 1.

Browning, H. (2018a). No room at the zoo: Management euthanasia and animal welfare. *Journal of Agricultural and Environmental Ethics*, *31*(4), 483–498.

Browning, H. (2018b). Won't somebody please think of the mammoths? De-extinction and animal welfare. *Journal of Agricultural and Environmental Ethics*, *31*(6), 785–803.

Browning, H. (2019a). The natural behavior debate: Two conceptions of animal welfare. *Journal of Applied Animal Welfare Science*, 1–13.

Browning, H. (2019b). What should we do about sheep? The role of intelligence in welfare considerations. *Animal Sentience*, *4*(25), 23.

Browning, H., & Maple, T. L. (2019). Developing a metric of usable space for zoo exhibits. *Frontiers in Psychology*, *10*, 791.

Burgdorf, J., & Panksepp, J. (2006). The neurobiology of positive emotions. *Neuroscience & Biobehavioral Reviews*, *30*(2), 173–187.

Cabanac, M. (1971). Physiological role of pleasure. *Science*, *173*(4002), 1103–1107.

Cabanac, M. (1979). Sensory pleasure. *The Quarterly Review of Biology*, *54*(1), 1–29.

Cabanac, M. (1992). Pleasure: The common currency. *Journal of Theoretical Biology*, *155*(2), 173–200.

Calcott, B. (2011). Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings. *Biology & Philosophy*, *26*(2), 281–293.

Camacho, A. E. (2015). Going the way of the dodo: De-extinction, dualisms, and reframing conservation. *Washington University Law Review*, *92*(4), 849–906.

Camerer, C. F. (2006). Wanting, liking, and learning: Neuroscience and paternalism. *The University of Chicago Law Review*, *73*(1), 87–110.

Campagna, C., Guevara, D., & Le Boeuf, B. (2017). De-scenting extinction: The promise of de-extinction may hasten continuing extinctions. *Hastings Center Report*, *47*, S48–S53.

Chang, R. (2015). Value incomparability and incommensurability. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 205–224). New York: Oxford University Press.

Clegg, I. (2018). Cognitive bias in zoo animals: An optimistic outlook for welfare assessment. *Animals*, *8*(7), 104.

Cochrane, A. (2012). *Animal Rights Without Liberation: Applied Ethics and Human Obligations*. New York: Columbia University Press.

Coghill, R. C., McHaffie, J. G., & Yen, Y.-F. (2003). Neural correlates of interindividual differences in the subjective experience of pain. *Proceedings of the National Academy of Sciences*, *100*(14), 8538–8542.

Cohen, S. (2014). The ethics of de-extinction. *NanoEthics*, *8*(2), 165–178.

Coko, K., & Schickore, J. (2012). Robustness, solidity, and multiple determinations. In L. Soler, E. Trizio, T. Nickles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 681–683). Dordrecht: Springer.

Crisp, R. (2006). Hedonism reconsidered. *Philosophy and Phenomenological Research*, *73*(3), 619–645.

Crisp, R. (2017). Well-being. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). Retrieved from https://plato.stanford.edu/archives/fall2017/entries/well-being/

Davis, M. A. (2000). "Restoration"—A misnomer? *Science*, *287*(5456), 1203–1203.

Dawkins, M. S. (1980). *Animal Suffering: The Science of Animal Welfare*. London: Chapman and Hall.

Dawkins, M. S. (1983). Battery hens name their price: Consumer demand theory and the measurement of ethological 'needs.' *Animal Behaviour*, *31*(4), 1195–1205.

Dawkins, M. S. (1988). Behavioural deprivation: A central problem in animal welfare. *Applied Animal Behaviour Science*, *20*(3), 209–225.

Dawkins, M. S. (1989). Time budgets in red junglefowl as a baseline for the assessment of welfare in domestic fowl. *Applied Animal Behaviour Science*, *24*(1), 77–80.

Dawkins, M. S. (1990). From an animal's point of view: Motivation, fitness, and animal welfare. *Behavioural and Brain Sciences*, *13*(1), 1–9.

Dawkins, M. S. (1998). Evolution and animal welfare. *The Quarterly Review of Biology*, *73*(3), 305–328.

Dawkins, M. S. (2003). Behaviour as a tool in the assessment of animal welfare. *Zoology*, *106*(4), 383–387.

Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal Welfare*, *13*(1), S3–S8.

Dawkins, M. S. (2006). A user's guide to animal welfare science. *Trends in Ecology & Evolution*, *21*(2), 77–82.

Dawkins, M. S. (2008). The science of animal suffering. *Ethology*, *114*(10), 937–945.

de Graaf, S., Ampe, B., Buijs, S., Andreasen, S.N, de Boyer des Roches, A., van Eerdenburg, F.J.C.M., Haskell, M.J., Kirchner, M.K., Mounier, L., Radeski, M., Winckler, C., Bijtterbier, J., Lauwers, L., Verbeke, W. & Tuyttens, F.A.M. (2018). Sensitivity of the integrated Welfare Quality® scores to changing values of individual dairy cattle welfare measures. *Animal Welfare*, *27*(2), 157–166.

de Mol, R. M., Schouten, W. G. P., Evers, E., Drost, H., Houwers, H. W. J., & Smits, A. C. (2006). A computer model for welfare assessment of poultry production systems for laying hens. *NJAS - Wageningen Journal of Life Sciences*, *54*(2), 157–168.

Deakin, A., Browne, W. J., Hodge, J. J. L., Paul, E. S., & Mendl, M. (2016). A screen-peck task for investigating cognitive bias in laying hens. *PLOS ONE*, *11*(7), e0158222.

DeGrazia, D. (1996). *Taking Animals Seriously: Mental Life and Moral Status*. New York: Cambridge University Press.

Diehm, C. (2017). De-extinction and deep questions about species conservation. *Ethics, Policy & Environment*, *20*(1), 25–28.

Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A Political Theory of Animal Rights*. OUP Oxford.

Donovan, J. (2003). Animal rights and feminist theory. In S. J. Armstrong & R. G. Botzler (Eds.), *The Animal Ethics Reader* (pp. 45–49). London: Routledge.

Duncan, I. J. (2009). Distress or suffering: What should be measured to determine animal well-being? *Sentience Collection*, *13*, 1–4.

Duncan, I. J., & Fraser, D. (1997). Understanding animal welfare. In M. C. Appleby & B. O. Hughes (Eds.), *Animal Welfare* (pp. 19–32). New York: CAB International.

Duncan, I. J., & Petherick, J. C. (1991). The implications of cognitive processes for animal welfare. *Journal of Animal Science*, *69*(12), 5017–5022.

Duncan, I. J., Slee, G. S., Kettlewell, P., Berry, P., & Carlisle, A. J. (1986). Comparison of the stressfulness of harvesting broiler chickens by machine and by hand. *British Poultry Science*, *27*(1), 109–114.

Elster, J., & Roemer, J. E. (1991). *Interpersonal Comparisons of Well-Being*. Cambridge: Cambridge University Press.

Eronen, M. I. (2015). Robustness and reality. *Synthese*, *192*(12), 3961–3977.

Fiester, A. (2005). Ethical issues in animal cloning. *Perspectives in Biology and Medicine*, *48*(3), 328–343.

Fleming, P. A., Clarke, T., Wickham, S. L., Stockman, C. A., Barnes, A. L., Collins, T., & Miller, D. W. (2016). The contribution of qualitative behavioural assessment to appraisal of livestock welfare. *Animal Production Science*, *56*(10), 1569–1578.

Fleurbaey, M. (2016). Economics and economic justice. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Retrieved from https://plato.stanford.edu/archives/win2016/entries/economic-justice/

Fleurbaey, M., & Hammond, P. J. (2004). Interpersonally comparable utility. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of Utility Theory: Volume 2 Extensions* (pp. 1179–1285). Boston: Springer US.

Fraser, D. (1999). Animal ethics and animal welfare science: Bridging the two cultures. *Applied Animal Behaviour Science*, *65*(3), 171–189.

Fraser, D. (2009). Animal behaviour, animal welfare and the scientific study of affect. *Applied Animal Behaviour Science*, *118*(3–4), 108–117.

Fraser, D., & Duncan, I. J. (1998). 'Pleasures','pains' and animal welfare: Toward a natural history of affect. *Animal Welfare*, *7*(4), 383–396.

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.

Frey, R. G. (2011). Utilitarianism and animals. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 172–197). Oxford: Oxford University Press.

Friese, C., & Marris, C. (2014). Making de-extinction mundane? *PLoS Biology*, *12*(3), e1001825.

Gamborg, C. (2014). What's so special about reconstructing a mammoth? Ethics of breeding and biotechnology in re-creating extinct species. In M. Oksanen & H. Siipi (Eds.), *The Ethics of Animal Re-creation and Modification* (pp. 60–76). London: Palgrave Macmillan UK.

Gamborg, C., Gremmen, B., Christiansen, S. B., & Sandøe, P. (2010). De-Domestication: Ethics at the Intersection of Landscape Restoration and Animal Welfare. *Environmental Values*, *19*(1), 57–78.

Ginsburg, S., & Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge: MIT Press.

Glatston, A. R. (1998). The control of zoo populations with special reference to primates. *Animal Welfare*, *7*(3), 269–281.

Godfrey-Smith, P. (2017). Animal evolution and the origins of experience. In D. L. Smith (Ed.), *How Biology Shapes Philosophy* (pp. 51–71). Cambridge: Cambridge University Press.

Gøtzsche, P. C., Liberati, A., Torri, V., & Rossetti, L. (1996). Beware of surrogate outcome measures. *International Journal of Technology Assessment in Health Care*, *12*(2), 238–246.

Grandin, T., & Johnson, C. (2009). *Making Animals Happy*. London: Bloomsbury.

Gray, J. H. (2017). *Zoo Ethics: The Challenges of Compassionate Conservation*. Cornell University Press.

Greely, H. T. (2017). Is de-extinction special? *Hastings Center Report*, *47*, S30–S36.

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon Press.

Gruen, L. (2011). *Ethics and Animals: An Introduction*. Cambridge: Cambridge University Press.

Gygax, L. (2017). Wanting, liking and welfare: The role of affective states in proximate control of behaviour in vertebrates. *Ethology*, *123*(10), 689–704.

Harrington, L. A., Moehrenschlager, A., Gelling, M., Atkinson, R. P. D., Hughes, J., & Macdonald, D. W. (2013). Conflicting and complementary ethics of animal welfare considerations in reintroductions: Welfare in reintroductions. *Conservation Biology*, *27*(3), 486–500.

Harrison, R. (1964). *Animal Machines: The New Factory Farming Industry*. London: Vincent Stuart.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321.

Hosch, W. L. (Ed.). (2011). *The Britannica Guide to Numbers and Measurement*. New York: Brittanica Educational Publishing.

Houe, H., Sandøe, P., & Thomsen, P. T. (2011). Welfare assessments based on lifetime health and production data in Danish dairy cows. *Journal of Applied Animal Welfare Science*, *14*(3), 255–264.

Hursthouse, R. (2011). Virtue ethics and the treatment of animals. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 119–143). Oxford: Oxford University Press.

Hutchins, M., Willis, K., & Wiese, R. J. (1995). Strategic collection planning: Theory and practice. *Zoo Biology*, *14*(1), 5–25.

IUCN. (1998). *Guidelines for Re-Introductions*. Retrieved from https://portals.iucn.org/library/efiles/documents/2013-009.pdf

Izzo, G. N., Bashaw, M. J., & Campbell, J. B. (2011). Enrichment and individual differences affect welfare indicators in squirrel monkeys (Saimiri sciureus). *Journal of Comparative Psychology*, *125*(3), 347–352.

Jensen, K. K. (2017). How should death be taken into account in welfare assessments? *Journal of Agricultural and Environmental Ethics*, *30*(5), 615–623.

Jones, K. E. (2014). From dinosaurs to dodos: Who could and should we de-extinct? *Frontiers of Biogeography*, *6*(1), 20–24.

Jones, R. C. (2013). Science, sentience, and animal welfare. *Biology & Philosophy*, *28*(1), 1–30.

Jørgensen, D. (2013). Reintroduction and de-extinction. *BioScience*, *63*(9), 719–720.

Kagan, R., Carter, S., & Allard, S. (2015). A universal animal welfare framework for zoos. *Journal of Applied Animal Welfare Science*, *18*, S1–S10.

Kagan, R., & Veasey, J. S. (2010). Challenges of zoo animal welfare. In D. G. Kleiman, K. V. Thompson, & C. K. Baer (Eds.), *Wild Mammals in Captivity: Principles and Techniques for Zoo Management* (2nd Edition, pp. 11-21). Chicago: University of Chicago Press.

Kasperbauer, T. J. (2017). Should we bring back the passenger pigeon? The ethics of de-extinction. *Ethics, Policy & Environment*, *20*(1), 1–14.

Key, B. (2016). Why fish do not feel pain. *Animal Sentience*, *1*(3), 1–34.

Khader, S. J. (2011). *Adaptive Preferences and Women's Empowerment*. New York: Oxford University Press.

Kirkden, R. D., & Pajor, E. A. (2006). Using preference, motivation and aversion tests to ask scientific questions about animals' feelings. *Applied Animal Behaviour Science*, *100*(1–2), 29–47.

Kringelbach, M. L., & Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences*, *13*(11), 479–487.

Korsgaard, C. M. (2011). Interacting with animals: A Kantian approach. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 91–118). Oxford: Oxford University Press.

Lacy, R. (1991). Zoos and the surplus problem: An alternative solution. *Zoo Biology*, *10*(4), 293–297.

Lacy, R. (1995). Culling surplus animals for population management. In B. G. Norton, M. Hutchins, E. Stevens, & T. L. Maple (Eds.), *Ethics on the Ark* (pp. 187–194). Washington and London: Smithsonian Institution Press.

Laubu, C., Louâpre, P., & Dechaume-Moncharmont, F.-X. (2019). Pair-bonding influences affective state in a monogamous fish species. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20190760.

Lawrence, A. B. (2008). Applied animal behaviour science: Past, present and future prospects. *Applied Animal Behaviour Science*, *115*(1–2), 1–24.

Lean, C. (2019). *Ecological Kinds and the Units of Conservation* (Australian National University). Retrieved from https://openresearch-repository.anu.edu.au/bitstream/1885/165694/1/THESIS%202019.pdf

Ledger, R. A., & Mellor, D. J. (2018). Forensic use of the Five Domains model for assessing suffering in cases of animal cruelty. *Animals*, *8*(7), 101.

Lindburg, D. G. (1991). Zoos and the “surplus” problem. *Zoo Biology*, *10*(1), 1–2.

Lindburg, D. G. (1999). Zoos and the rights of animals. *Zoo Biology*, *18*(5), 433–448.

Lindburg, D. G., & Lindburg, L. (1995). Success breeds a quandry: To cull or not to cull. In B. G. Norton, M. Hutchins, E. Stevens, & T. L. Maple (Eds.), *Ethics on the Ark* (pp. 195–208). Washington and London: Smithsonian Institution Press.

Lindenmayer, D. B., & Likens, G. E. (2011). Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems*, *14*(1), 47–59.

List, C. (2003). Are interpersonal comparisons of utility indeterminate? *Erkenntnis*, *58*(2), 229–260.

List, C. (2013). Social choice theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition). Retrieved from https://plato.stanford.edu/archives/win2013/entries/social-choice/

Littin, K. E., Fisher, P., Beausoleil, N. J., & Sharp, T. (2014). Welfare aspects of vertebrate pest control and culling: Ranking control techniques for humaneness. *Revue Scientifique et Technique (International Office of Epizootics)*, *33*(1), 281–289.

Littlewood, K. E., & Mellor, D. J. (2016). Changes in the welfare of an injured working farm dog assessed using the Five Domains model. *Animals*, *6*(9), 58.

Low, P., Panksepp, J., Reiss, D., Edelman, D., van Swinderen, B., & Koch, C. (2012, July). *The Cambridge Declaration on Consciousness*. Presented at the Francis Crick Memorial Conference, Cambridge, England.

Lukas, K. E., & Ross, S. R. (2005). Zoo visitor knowledge and attitudes toward gorillas and chimpanzees. *The Journal of Environmental Education*, *36*(4), 33–48.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114*(3), 533–541.

Manteca, X., & Deag, J. M. (1994). Individual variation in response to stressors in farm animals: Implications for experimenters. *Animal Welfare*, *3*(3), 213–218.

Maple, T. L. (2014, April 11). Copenhagen Zoo’s giraffe killing was wrong and disturbing. Retrieved February 16, 2018, from SFGate website: https://www.sfgate.com/opinion/article/Copenhagen-Zoo-s-giraffe-killing-was-wrong-and-5396183.php

Maple, T. L., & Perdue, B. (2013). *Zoo Animal Welfare*. Berlin: Springer.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning* (1st ed.). New York: Routledge.

Mason, C. (2017). The unnaturalness objection to de-extinction: A critical evaluation. *Animal Studies Journal*, *6*(1), 40–60.

Mason, G., & Littin, K. E. (2003). The humaneness of rodent pest control. *Animal Welfare*, *12*, 1–37.

Mason, G., & Mendl, M. (1993). Why is there no simple way of measuring animal welfare? *Animal Welfare*, *2*(4), 301–319.

McCauley, D. J., Hardesty-Moore, M., Halpern, B. S., & Young, H. S. (2017). A mammoth undertaking: Harnessing insight from functional ecology to shape de-extinction priority setting. *Functional Ecology*, *31*(5), 1003–1011.

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.

Mellor, D. J. (2012). Animal emotions, behaviour and the promotion of positive welfare states. *New Zealand Veterinary Journal*, *60*(1), 1–8.

Mellor, D. J. (2016). Updating animal welfare thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living." *Animals*, *6*(3), 21.

Mellor, D. J. (2017). Operational details of the Five Domains model and its key applications to the assessment and management of animal welfare. *Animals*, *7*(8), 60.

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.

Mellor, D. J., & Stafford, K. J. (2008). Quality of Life: A valuable concept or an unnecessary embellishment when considering animal welfare? *The Welfare of Animals – It's Everyone's Business*.

Mendl, M., Burman, O. H. P., Parker, R. M. A., & Paul, E. S. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, *118*(3–4), 161–181.

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.

Muri, K., Stubsjøen, S. M., Vasdal, G., Moe, R. O., & Granquist, E. G. (2019). Associations between qualitative behaviour assessments and measures of leg health, fear and mortality in Norwegian broiler chicken flocks. *Applied Animal Behaviour Science*, *211*, 47–53.

Narveson, J. (1973). Moral problems of population. *The Monist*, *57*(1), 62–86.

Ng, Y.-K. (1995). Towards welfare biology: Evolutionary economics of animal consciousness and suffering. *Biology and Philosophy*, *10*(3), 255–285.

Nicholls, S. (2018, January 11). Swedish zoo kills nine healthy lion cubs over six years. *Euronews*. Retrieved from http://www.euronews.com/2018/01/11/swedish-zoo-kills-nine-healthy-lion-cubs-over-six-years

Norton, B. G. (1995). Caring for nature: A broader look at animal stewardship. In B. G. Norton, M. Hutchins, E. Stevens, & T. L. Maple (Eds.), *Ethics on the Ark* (pp. 102–121). Washington and London: Smithsonian Institution Press.

Norwood, F. B., & Lusk, J. L. (2011). *Compassion, by the Pound: The Economics of Farm Animal Welfare*. Oxford: Oxford University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Nussbaum, M. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Nussbaum, M. (2004). Beyond "compassion and humanity." In C. Sunstein & M. Nussbaum (Eds.), *Animal Rights: Current Debates and New Directions* (pp. 299–320). New York: Oxford University Press.

Nussbaum, M. (2011). The capabilities approach and animal entitlements. In T. L. Beauchamp & R. G. Frey (Eds.), *The Oxford Handbook of Animal Ethics* (pp. 228–254). Oxford: Oxford University Press.

Osborne, S. R. (1977). The free food (contrafreeloading) phenomenon: A review and analysis. *Animal Learning & Behavior*, *5*(3), 221–235.

Otten, N. D., Rousing, T., & Forkman, B. (2017). Influence of professional affiliation on expert's view on welfare measures. *Animals*, *7*(12), 85.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80.

Parker, I. (2017, January 9). Killing Animals at the Zoo. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2017/01/16/killing-animals-at-the-zoo

Pearson, E. L., Lowry, R., Dorrian, J., & Litchfield, C. A. (2014). Evaluating the conservation impact of an innovative zoo-based educational campaign: 'Don't Palm Us Off' for orang-utan conservation. *Zoo Biology*, *33*(3), 184–196.

Penfold, L. M., Powell, D., Traylor-Holzer, K., & Asa, C. S. (2014). "Use it or lose it": Characterization, implications, and mitigation of female infertility in captive wildlife: Infertility and Sustainability. *Zoo Biology*, *33*(1), 20–28.

Perdue, B. M., Clay, A. W., Gaalema, D. E., Maple, T. L., & Stoinski, T. S. (2012). Technology at the zoo: The influence of a touchscreen computer on orangutans and zoo visitors. *Zoo Biology*, *31*(1), 27–39.

Phillips, C. (2008). Animal welfare: A construct of positive and negative affect? *The Veterinary Journal*, *175*(3), 291–292. https://doi.org/10.1016/j.tvjl.2007.05.015

Phillips, C. (2009). *The Welfare of Animals: The Silent Majority*. Dordrecht: Springer.

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.

Powell, D. M., & Ardaiolo, M. (2016). Survey of U.S. zoo and aquarium animal care staff attitudes regarding humane euthanasia for population management. *Zoo Biology*, *35*(3), 187–200.

Powell, D.M. & Bullock, E.V. (2014). Evaluation of factors affecting emotional responses in zoo visitors and the impact of emotion on conservation mindedness. *Anthrozoös*, *27*(3), 389-405.

Proctor, H. (2012). Animal sentience: Where are we and where are we heading? *Animals*, *2*(4), 628–639.

Proctor, H., Carder, G., & Cornish, A. (2013). Searching for animal sentience: A systematic review of the scientific literature. *Animals*, *3*(3), 882–906.

Regan, T. (1983). *The Case for Animal Rights*. London: Routledge & Kegan Paul.

Regan, T., & Singer, P. (Eds.). (1976). *Animal Rights and Human Obligations*. New Jersey: Prentice-Hall.

Rincon, P. (2014, February 10). Why did Copenhagen Zoo kill its giraffe? *BBC News*. Retrieved from http://www.bbc.com/news/science-environment-26118748

Robert, A., Thévenin, C., Princé, K., Sarrazin, F., & Clavel, J. (2017). De-extinction and evolution. *Functional Ecology*, *31*(5), 1021–1031.

Rohwer, Y., & Marris, E. (2018). An analysis of potential ethical justifications for mammoth de-extinction and a call for empirical research. *Ethics, Policy & Environment*, *21*(1), 1–16.

Rollin, B. E. (2006). *Animal Rights and Human Morality* (3rd ed.). New York: Prometheus Books.

Sandler, R. (2014). The ethics of reviving long extinct species. *Conservation Biology*, *28*(2), 354–360.

Sandler, R. (2017). De-extinction: Costs, benefits and ethics. *Nature Ecology & Evolution*, *1*(4), 0105.

Sandøe, P., Corr, S., Lund, T., & Forkman, B. (2019). Aggregating animal welfare indicators: Can it be done in a transparent and ethically robust way? *Animal Welfare*, *28*(1), 67–76.

Sandøe, P., Crisp, R., & Holtug, N. (1997). Ethics. In M. C. Appleby & B. O. Hughes (Eds.), *Animal Welfare* (pp. 3–18). New York: CAB International.

Sandøe, P., Hocking, P. M., Förkman, B., Haldane, K., Kristensen, H. H., & Palmer, C. (2014). The blind hens' challenge: Does it undermine the view that only welfare matters in our dealings with animals? *Environmental Values*, *23*(6), 727–742.

Sandøe, P., & Jensen, K. K. (2011). The idea of animal welfare—Developments and tensions. In C. M. Wathes, S. A. May, S. P. McCulloch, M. C. Whiting, & S. Corr (Eds.), *ICVAE - First International Conference on Veterinary and Animal Ethics* (pp. 11–17). The Royal Veterinary College, University of London.

Sarkar, S. (2002). Defining "biodiversity"; assessing biodiversity. *The Monist*, *85*(1), 131–155.

Sarle, W. (1997). *Measurement Theory: Frequently Asked Questions*. Retrieved from ftp://ftp.sas.com/pub/neural/measurement.html

Schickore, J., & Coko, K. (2013). Using multiple means of determination. *International Studies in the Philosophy of Science*, *27*(3), 295–313.

Schupbach, J. N. (2016). Robustness analysis as explanatory reasoning. *The British Journal for the Philosophy of Science*, *69*(1), 275–300.

Sebo, J. (2018). The moral problem of other minds. *The Harvard Review of Philosophy*, *25*, 51–70.

Seddon, P. J., Moehrenschlager, A., & Ewen, J. (2014). Reintroducing resurrected species: Selecting DeExtinction candidates. *Trends in Ecology & Evolution*, *29*(3), 140–147.

Shapiro, B. (2015). *How to Clone a Mammoth: The Science of De-Extinction*. Princeton University Press.

Shapiro, B. (2017). Pathways to de-extinction: How close can we get to resurrection of an extinct species? *Functional Ecology*, *31*(5), 996–1002.

Singer, P. (1995). *Animal Liberation* (2nd ed.). London: Plimlico.

Sinnott-Armstrong, W. (2015). Consequentialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition). Retrieved from https://plato.stanford.edu/archives/win2015/entries/consequentialism/

Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour*, *97*, 201–212.

Soler, L. (2014). Against robustness? Strategies to support the reliability of scientific results. *International Studies in the Philosophy of Science*, *28*(2), 203–215.

Špinka, M. (2006). How important is natural behaviour in animal farming systems? *Applied Animal Behaviour Science*, *100*(1–2), 117–128.

Spruijt, B. M., van den Bos, R., & Pijlman, F. T. A. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science*, *72*(2), 145–171.

Steen, E. (1968). Some aspects of the nutrition of semi-domestic reindeer. *Symposium of the Zoological Society, London*, *21*, 117–128.

Stegenga, J. (2012). Rerum concordia discors: Robustness and discordant multimodal evidence. In L. Soler, E. Trizio, T. Nickles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 207–226). Dordrecht: Springer.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp. 1–49). New York: John Wiley & Sons.

Stevens, S. S. (1959). Measurement, psychophysics, and utility. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and Theories* (pp. 18–63). New York: John Wiley & Sons.

Szokalski, M. S., Litchfield, C. A., & Foster, W. K. (2012). Enrichment for captive tigers (Panthera tigris): Current knowledge and future directions. *Applied Animal Behaviour Science*, *139*(1), 1–9.

Turner, D. D. (2017). Biases in the selection of candidate species for de-extinction. *Ethics, Policy & Environment*, *20*(1), 21–24.

Ursinus, W., & Schepers, F. (2009). COWEL: a decision support system to assess welfare of husbandry systems for dairy cattle. *Animal Welfare*, *18*(4), 545–552.

Veasey, J. S., Waran, N. K., & Young, R. J. (1996). On comparing the behaviour of zoo housed animals with wild conspecifics as a welfare indicator. *Animal Welfare*, *5*, 13–24.

Veissier, I., Jensen, K. K., Botreau, R., & Sandøe, P. (2011). Highlighting ethical decisions underlying the scoring of animal welfare in the Welfare Quality® scheme. *Animal Welfare*, *20*(1), 89–101.

Walker, M., Díez-León, M., & Mason, G. (2014). Animal welfare science: Recent publication trends and future research priorities. *International Journal of Comparative Psychology*, *27*(1), 80–100.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177.

Webster, J. (1994). *Animal Welfare: A Cool Eye Towards Eden*. Oxford: Blackwell Science.

Webster, J. (2005). *Animal Welfare: Limping Towards Eden*. Oxford: Blackwell Publishing.

Weijers, D. (2014). Nozick's experience machine is dead, long live the experience machine! *Philosophical Psychology*, *27*(4), 513–535.

Wemelsfelder, F. (1997). The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science*, *53*(1–2), 75–88.

Wemelsfelder, F. (2007). How animals communicate quality of life: The qualitative assessment of behaviour. *Animal Welfare*, *16*(1), 25–31.

Wemelsfelder, F., Hunter, T. E. A., Mendl, M., & Lawrence, A. B. (2001). Assessing the 'whole animal': A free choice profiling approach. *Animal Behaviour*, *62*(2), 209–220.

Williams, N. (2003). Death of Dolly marks cloning milestone. *Current Biology*, *13*(6), R209–R210.

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge: Harvard University Press.

Wimsatt, W. C. (2012). Robustness, reliability, and overdetermination (1981). In L. Soler, E. Trizio, T. Nickles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 61–87). Dordrecht: Springer.

Yeates, J. W. (2010). Death is a welfare issue. *Journal of Agricultural and Environmental Ethics*, *23*(3), 229–241.

Yeates, J. W., & Main, D. C. J. (2008). Assessment of positive welfare: A review. *The Veterinary Journal*, *175*(3), 293–300.

Zimmer, C. (2013). Bringing them back to life. *National Geographic*, *223*(4), 28–41.