Heather Browning[1]

The Measurability of Subjective Animal Welfare

Abstract: *One of the most challenging questions surrounding subjective animal welfare is whether these states are measurable: that is, is subjective welfare an appropriately quantifiable target for scientific enquiry and ethical and deliberative calculation? The availability of several different types of measurement scale raises important questions regarding whether subjective experience has the right properties to be meaningfully represented on the types of scale required for different applications. This methodological question has so far received scant attention in the animal welfare literature. In this paper, I address this omission by examining the types of measurement scale we can reasonably expect to apply to animal welfare measurements, and which we will actually need for our applications. I argue that our different applications will require variously ordinal, interval, and ratio scales, and that we have sufficient reason to believe that subjective welfare is a target with the appropriate characteristics to justify the practice of representing it using each of these types of scales.*

Correspondence:
Email: drheatherbrowning@gmail.com

1 London School of Economics and Political Science, Centre for Philosophy of Natural and Social Science, Houghton Street, London, WC2A 2AE, UK.

1. Introduction

This paper is aimed at answering the question: *how can we measure the subjectively experienced side of welfare?* There are two ways of approaching this question. The first is to look at the methods we might use to achieve this goal. There are large numbers of different indicators — physiological, behavioural, and cognitive — used by animal welfare scientists and we might want to examine how well these perform in measuring subjective welfare. However, this is not the approach I will be taking here. Instead, I will approach this question on a more conceptual level, to ask whether subjective welfare is measurable, in terms of possessing the right properties to be meaningfully represented on the types of measurement scale required for different purposes.

Subjective welfare refers to the integrated set of positively and negatively valenced feelings, or affects, that an animal experiences — feelings such as hunger, pain, and joy. Welfare is increased when an animal experiences positive affects, and decreased through the experience of negative affects. This is sometimes taken to be one among multiple components of welfare, the others most commonly seen within animal welfare science being biological functioning, natural living, and satisfaction of preferences (Dawkins, 2021; Mellor and Stafford, 2008; Veit and Browning, 2021a). There are benefits and drawbacks to each of these conceptions and it isn't necessary for this paper to take a stance on which one is the correct, or best, overall (though for discussion see Browning, 2020a,b). Some authors take a combined approach, highlighting the links between these (e.g. Broom, 1986; 1998; Broom and Johnson, 2019; Fraser, 2008; Fraser *et al.*, 1997). Here, they emphasize that subjective experiencing is important but also deeply entangled with natural biological functioning, in terms of the causes of feelings, their evolutionary function, their behavioural and physiological effects, and the indicators we use to measure them.

However, the specific target of this paper is subjective welfare, which can be discussed as conceptually distinct from other aspects of welfare even if in practice it is strongly connected with natural physical functioning. As well as being common in the animal welfare science literature (Duncan, 2002; Grandin and Johnson, 2009; Mellor and Beausoleil, 2015), the subjective conception of welfare is common throughout animal ethics, grounding Singer's possession of interests (Singer, 1995), forming part of Tom Regan's 'subject of a life' criterion for rights (Regan, 1983), and even forming part of the

necessary capacities for flourishing in approaches focused on natural living (Nussbaum, 2011). For the sake of simplicity, in this paper I will use the term 'welfare' to refer to subjective welfare, while acknowledging that in practice the distinction may not be as clear-cut. Whether or not one endorses a fully subjective conception of welfare, measurement of subjective states will be important for any conception that takes subjective experience to form at least some part of welfare. Almost all agree that subjective experience is a necessary component of welfare, and this makes its measurement an important part of animal welfare science (Veit and Browning, 2021b).

Subjective welfare is of particular interest because its measurement is indirect and less obviously justified than for physical functions. There is a long history of disagreement about the measurability of psychological properties and sensations, frequently centred around whether or not subjective states are accessible to science at all, i.e. whether they are detectable (Browning and Veit, 2020). Particularly in animal welfare science, subjective experience has been questioned as an appropriate target for measurement (Dawkins, 2017; 2021). Here, I will leave aside scepticism and take for granted that subjectively experienced mental states have some effects on behaviour and physiology that make them accessible to measurement. However, formal measurability goes beyond just our ability to detect subjective states but instead reflects our ability to represent them on different numerical scales.

Measurement is 'the assignment of numerals to objects or events according to rule — any rule' (Stevens, 1959, p. 19). It involves mapping the well-understood mathematical relationships between numbers onto the attributes of some state or object (in this case, subjective welfare) so that we can then use the mathematical rules to better understand the properties of whatever it is we are measuring. Numerals are assigned through the relation of isomorphism — similarities in empirical relations between objects and numerical relations between numerals. So long as we have reason to believe that the empirical properties of what we are measuring are mirrored by the mathematical properties of the assigned numerals, then we can gain understanding of the system of interest. An entity is thus measurable if it possesses the necessary features for mapping onto a numerical scale.

The measurability of welfare is important because judgments about welfare are used in many scientific, ethical, and management contexts (e.g. Browning, 2018a,b). There are various decision contexts that take welfare as part of their calculations, such as social, legal, and

moral deliberations, institutional and governmental policy, and even individual actions. In particular, utilitarian calculations looking to identify the actions that maximize value will require accurate measurements of welfare to provide appropriate information for guiding action. Whatever our decision procedure, if it uses facts about welfare, it will be important to have measurable inputs to ensure empirically-informed deliberation.

We have several reasons to be optimistic that welfare will be a measurable entity, i.e. one that can be meaningfully represented on one (or more) of the available measurement scales. We speak of the 'quantity' of welfare possessed by some individual, without any seeming oddness (Pettigrew, 2019) and we are able to make generalizations connecting particular goods to changes in welfare (Alexandrova, 2017). Indeed, given our practices and uses of the term, it would seem strange if it turned out that welfare was not measurable. The more interesting question, then, is what kind of measurement welfare lends itself to. Is welfare measurable in the right kind of way for the required applications? If we use measurement scales unreflectively, we risk using the wrong scales, and drawing unjustified conclusions from our measures; a point emphasized in recent work by Narens and Skryms (2020). For example, using numerical scores for ordinal scales creates a risk of unintentional attribution of the numerical properties to the objects they are representing, and mistakenly performing unjustified operations, such as averaging results. Thus, it is crucial to establish that a particular target is actually measurable using the desired scales. If the wrong measurement scales are used for welfare, this can lead to outcomes that don't actually have the intended effects.

Different measurement scales represent different ways of assigning numbers to a measured attribute. There are four primary scales of measurement — nominal, ordinal, interval, and ratio (Stevens, 1951). These are defined by the empirical operations used to create them, and the mathematical transformations its constituents can be subjected to without altering the character of the scale (the properties of the relationships between the items measured). As these scales increase in complexity, they can provide us with ever more sophisticated mathematical analyses, but are also more demanding regarding the necessary properties of the measured attribute. It is important to choose the right scale to match the features of the attribute or object being measured, such 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. Our choice of scale will affect which calculations can be meaningfully performed using measures of our target.

The three measurement scales of interest for animal welfare are ordinal, interval, and ratio. Ordinal scales are rankings of objects in order, regarding some property on which an attribute is ranked as higher than, lower than, or equal to another within a list. These are then assigned corresponding values, reflecting the ordering between the members. Interval scales are a type of cardinal measurement that provides further information beyond just the relative ordering or rankings of a set, but also the size of the gaps between the items — the difference between adjacent items or *how much more* some item is above another. The choice of units and zero point, however, are arbitrary. Finally, ratio scales have a fixed non-arbitrary zero point and can thus also provide information about the ratios between the items on the scale (e.g. 'x is twice as much as y').

Establishing that subjective welfare is measurable thus requires demonstrating that it possesses the necessary properties for measurement on at least one of these scale types. This process contains two parts: identifying the scale required for the goals of the measurement programme, and examining the features and characteristics of the target to see whether they line up with the required scale (Cartwright and Bradburn, 2011). We have particular knowledge or assumptions about the properties of our target and we need to check whether these align with the features of the scale. In this paper I will be performing both of these tasks for the different types of measurement scale as they relate to subjective welfare — identifying the goals of different types of measurement, and showing that we have sufficient reason to treat subjective welfare as possessing the required properties to justify use of scales of that type.

This is a distinct project from a discussion of the properties and applications of the specific methods or indicators that are used to measure welfare. We must first establish that the target variable itself has the right properties for representation on a particular type of measurement scale before we can look at the procedures for performing measurement acts in practice. Importantly, we must not conflate the properties of a measurement method with the properties of the measurement target. There can be multiple methods of measurement for a single target and the scale that applies to the method of measurement used, particularly if it is an indirect indicator, may not be similarly applicable to the target. For example, just because

measurement of blood cortisol levels can be done according to a ratio scale, this does not automatically mean that welfare itself is a ratio-scaled quantity.[2] The properties of the indicators cannot be read onto the target. It is important that the scales constructed using the observed data really reflect the underlying attribute and are not simply artefacts of the measurement procedure. Describing welfare as measurable via an ordinal or ratio scale is a claim about the properties of subjective experience itself, not about the outputs of the methods used to measure it. There is then the additional question about which methods or indicators are best for creating specific types of outputs, as will be briefly discussed in the conclusion.

In this paper, I will be looking at the types of scale we might use for animal welfare and how we can justify their use, given the properties of subjective experience. Importantly, I will be aiming to establish justification for the use of different scale *types* rather than some *particular* scale or another, akin to justifying the use of an interval scale measure for temperature without going on to construct Celsius or Fahrenheit scales. This paper thus represents a first step in a more rigorous analysis of the measurement of subjective animal welfare, in showing that we would be justified in constructing or using such a scale, without yet providing specific proposals on how this should be done.

Though there has been a lot of work on measurement of human welfare and utility, particularly within economics and psychology, this has not yet extended to the measurability of animal welfare. While there has been much fruitful overlap between the work in economics and animal welfare, this particular area remains unexplored. While the practice of animal welfare science involves the measurement of animal welfare via a range of methods and proxy indicators, there are currently few if any explicit scales of measurement applied to subjective welfare. Instead, measures are taken and typically used to make rough ordinal judgments about a small set of situations, such as 'situation A is better for the welfare of this animal than situation B'. To make progress in justifying these judgments, and to go on to perform more sophisticated quantitative assessments, we need to establish a justified scale of measurement for welfare. In this paper I will bring together different strands of discussion from the literature on

---

2   There is an additional issue regarding whether blood cortisol is really a measure of welfare at all (Dawkins, 1998), but even if it was, this problem still applies.

human well-being to apply them to animal welfare, assessing whether the same types of strengths, weaknesses, and limitations apply. In particular, I will show that many of the complexities and objections that commonly arise in the discussions of human subjective well-being do not apply in the animal case, and thus the justification for use of the more demanding scale types is simpler.

   As I have described, there are several different measurement scales, each with different properties and that allow for different types of transformation and statistical analysis. There isn't, however, necessarily a single unique representation for each target system. Which scale we use is, in large part, a response to our goals. In what follows, I will examine the adequacy of each of the three types of scale — ordinal, interval, and ratio — for animal welfare; describing for which applications in animal welfare science and ethics they are useful and outlining the relevant features of subjective welfare that allow its measurement using these scales. I will argue that we are justified in representing welfare on each type of scale, up to and including a ratio scale, but that which type of scale we use will be determined by the context-dependent requirements of each case.

2. Ordinal Scales

An ordinal measurement system is one in which ordering relations hold; each object can be positioned on the scale as greater than, less than, or equal to each other item. The assigned numerical values reflect these ordering relationships, e.g. those with higher levels of the attribute are assigned higher numbers, but the values themselves otherwise mean little — the differences between the values don't represent anything about the differences between the attributes. Ordinal measurement of welfare is perhaps the most basic view. It has been particularly popular within economics, largely because of the use of preferences as a measure, which inherently lend themselves to orderings. However, ordinal measurement does not have to imply a preference-based conception of welfare and can be equally well used to rank subjective states.

2.1. Applications of ordinal measurement

There are a number of applications for ordinal measurement within animal welfare. They are appropriate for one of the most common applications in animal welfare science — the comparison of different resources or housing systems to determine which are better for

welfare, such as different types of flooring for laying hens (Hughes and Black, 1973) or stalls vs. group housing for sows (Karlen *et al.*, 2007). With an ordinal scale we can rank different interventions according to their impact on welfare, identifying those that give the greatest welfare increase.

Another common use of welfare measurement is in welfare assessments. These typically occur within practical animal management contexts, to find out about the quality of the situation particular animals are in. Ordinal scales can be used for individual animal welfare assessments, where we compare an animal's welfare at different points in time. The same is true for larger-scale institutional assessments of farms or other types of animal management, to assess the housing and husbandry practices in terms of their impact on the welfare of the animals housed within. In most cases, an ordinal scale will be sufficient, as a farm can be ranked in relation to others, or according to a fixed comparison point or set of graded categories.

Sometimes we may also want to compare multiple animals to see which have higher or lower welfare. This can be done with an ordinal scale, though with limited applications. If we could determine which animal had the lower welfare, we could then make them a priority for interventions (if we were using a decision rule such as maximin, which prioritizes improving things for the worst off). However, other decision rules, such as typical utilitarian calculations, would require further information on how much better or worse off the animals are.

2.2. Ordinal properties for subjective welfare

After identifying the welfare measurement applications for ordinal scale measurement, we next need to determine that subjective welfare has the relevant ordinal properties to allow this type of measurement. The most important property to establish is the ordering property — whether different instances of welfare can be judged as greater than, less than, or equal to one another. Additional properties will depend on the strength of ordering we wish to use, but include completeness, reflexivity, transitivity, and anti-symmetry. Here I will show that subjective welfare possesses the necessary properties for representation on an ordinal scale.

First, subjective welfare possesses the ordering property. The quality of subjective experience can be better or worse along the welfare scale, and particular instances can be ranked against one another. We can assess, at the very least, whether some welfare situation is

better or worse than some other. We shouldn't have much trouble
deciding, for example, that a bear rolling around in a pool with its
companions is in a better state than a malnourished lion pacing on
concrete. This illustrates that, at least in principle, subjective experi-
ence can be ordinally ranked, though there may be some less obvious
cases.

The next important property is completeness (necessary unless we
want an incomplete or partial ordering). 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. Whether or not we take subjective experience to
possess this property depends on whether we think there are cases for
which there is genuinely no fact of the matter regarding their relative
ordering. Consider, for example, a more closely-matched case than the
one described above, comparing the welfare of a lion hunting prey
against a bear catching fish. It may not be immediately obvious how
to rank these states against one another. Whether or not this is a
challenge to completeness will depend on whether we take the diffi-
culty to be a result of incommensurability, vagueness, or measurement
imprecision.

A deep incommensurability between individuals (as will be
discussed further in Section 3.4) is implausible for the level com-
parisons that ordinal scales require — there are obvious cases in
which the welfare of one animal is higher than another, even if there
are some cases in which it is unclear, and thus this cannot explain the
difficulties with closely-matched cases. It could be a real property of
the values themselves, a vagueness that cannot be resolved by
measurement, a possibility that led Griffin to posit 'rough equality'
(similarly described by Chang, 1997, as items being 'on a par') where
'the roughness is not in our understanding but ineradicably in the
values themselves' (Griffin, 1986, p. 81). He takes rough equality as
inherent to measurements of well-being, but as he also holds a
preference-based view, this does not have to speak to the measura-
bility of subjective welfare.

It is instead also explicable as a practical difficulty rather than a
failure of ordering — a result of measurement imprecision rather than
a case of strict incommensurability. We can say that there is some fact
of the matter about whether A > B, we just lack sufficiently precise
measurements to tell. Although it may be difficult at times to judge
the exact ranking of some states of welfare, it does not necessarily
follow from this that such states are unrankable. In particular, people

tend to consider such cases unrankable when relying on intuitive estimates. These may indeed be quite imprecise, but this is no reason to think that once empirical data is collected there won't be some fact of the matter regarding which welfare state is actually higher. I may not be able to judge the difference in volume between my similarly-sized cup of water and mug of tea, but don't take this to mean that there is no empirical fact regarding their ordering. Where our measurement instruments cannot give us precise enough values to determine the ordering, we can treat the items as roughly equal, in the knowledge that future measurement may allow us to be more precise and alter the ordering.

This conclusion is challenged in the case of human subjective well-being. In particular, there are worries about completeness when it requires the integration of multiple affects into a single subjective experience, and the weightings of different components into a single measure. This is not such a problem for the animal case as neither of the sources of concern about commensurability strongly apply there. The first is as a methodological artefact of the typical measures of human subjective well-being — most commonly summary assessments arising from self-report on life satisfaction surveys. There is scepticism regarding whether these surveys are able to accurately represent overall well-being, when individuals may have difficulty in deciding on how to compare different experiences, or weight the different aspects of their experience (Benjamin *et al.*, 2020; Ferrer-i-Carbonell and Frijters, 2004; Haybron, 2011). Though it is questionable to what degree this tells us anything about the features of subjective well-being: people's failure to intuitively rank different sets of conditions speaks more to a failure of our ability to accurately introspect on our own actual or potential experiences than a deep problem with the ordinal nature of the items. And as these subjective reports are not used for animals, instead replaced by external behavioural or physiological measures (some of which will be discussed in Section 5), the particular methodological worries will not apply.

However, one may still be concerned that there is deeper incommensurability between different affects, that can't be meaningfully integrated into a single welfare experience. While this is a complex topic that requires more attention than this paper allows, I have argued elsewhere that the shared evolutionary function of different affects is what justifies their integration into a single welfare experience, and that we have measures that can track this overall state

(Browning, 2020a). It is thus not important whether an individual is able to judge for themselves how different affects integrate or trade off, but that the 'output' of such an operation within the animal is able to be measured to determine how it has been performed.

The additional properties required for an ordinal scale are reflexivity and transitivity. 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), which is not a problem for subjective welfare. Transitivity means that, for example, if A > B and B > C, then A > C. Transitivity for subjective welfare may fail in the cases of vague ordering described above, 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. However, as discussed, it is only if we take this to be a result of measurement imprecision rather than ontological vagueness that transitivity will then fail. In cases where we have sufficiently precise measurements to determine the differences between closely matched items and these properties will be met, this will give us a weak ordering. A strong ordering requires the additional property of anti-symmetry, meaning that no two items can occupy the same place in the ordering, but for welfare there seems no reason to presume that two cases could not be of the same welfare level.

Ordinal scales are thus consistent with welfare measurement and will meet our goals in many cases. However, these scales are limited in that when comparing interventions or welfare levels they cannot provide any further information regarding *how much* impact or the size of the difference in welfare. When operating in circumstances of limited resources (which reflects most real-world decision contexts), we cannot always simply aim for the best option, but must instead consider trade-offs. In situations where we need to make trade-offs or comparisons, we will need to measure magnitude of welfare effects and for this we will need a quantitative scale, such as an interval or ratio scale.

3. Interval Scales

Interval scales go beyond ordinal, providing additional information regarding the magnitude of the differences between items. The exact units selected are arbitrary, as is the zero point. Examples of interval

scales are the temperature scales of Celsius and Fahrenheit. Here we can see the arbitrary selection of scale — although both recognize the same relative difference between points on the scale (for example, two units between 20° and 22°), they differ in the size of these units relative to the measured quantity and in the arbitrary position of the zero point. Importantly, use of an interval scale gives the possible rate of substitution between two values — we can now talk meaningfully about trade-offs between items with equivalent units.

3.1. Applications of interval measurement

I have described how ordinal scales are useful for judgments about which conditions or systems are better or worse, or whether an animal's welfare is improving. However, I also indicated that these judgments will be limited as they cannot tell us *how much* better or worse. When we need to prioritize between animals or compare the relative benefits of different interventions, we require further cardinal information. For example, we might have limited resources and can either increase the size of an otter exhibit to give them more space to play, or add a pond to a tiger exhibit to allow them to swim. If our aim is maximizing the welfare outcome, then this sort of decision requires quantitative comparisons to determine which option has the greater welfare benefit; comparing the magnitudes of different effects. When assessing the relative benefit of some intervention over another, an interval or ratio scale will be necessary.

An interval scale will also be necessary for most utilitarian applications, that use the sum of individual utilities to calculate the total value of an action. Transformation of utilities using affine transformations (permissible for interval scales) preserves the relationships between the totals, producing the same ordering of results — no matter which interval scale we used, we would still prefer the same outcomes.[3] Animal welfare applications in ethics and policy commonly involve the use of utilitarian calculations, contexts for which we need at least an interval scale. For example, we might want to look at the global welfare impact of climate change, or conduct a cost–benefit analysis regarding where to invest a charitable donation for the greatest welfare benefit. If subjective welfare failed to demonstrate the properties

---

[3]   Though this only works for calculations involving the same number of individuals, as I will detail in Section 4.

required for interval scale measurement, these calculations would no longer be justified. However, as I will show in the next section, subjective welfare does possess the necessary properties and its measurement can thus be used in utilitarian calculations.

3.2. Interval properties for subjective welfare

Application of an interval scale is more difficult than an ordinal scale — we're no longer just trying to rank states as better or worse than others, but trying to determine *by how much* they are so. This requires quantification of the underlying states, transforming subjective experience into 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.' Justifying use of an interval scale means having reason to think that subjective welfare possesses this interval property — that it is an entity composed of equally-spaced units that can add or subtract in determining the whole (note that this doesn't commit us to taking the units as discrete rather than continuous — temperature is a continuous variable, while still measured with units).

It is tempting to rely on intuitive plausibility about the features of the target. Thinking about welfare, it *feels* like the sort of thing that moves up or down along a linear scale. We can seemingly make sense of statements like 'I am twice as happy today as I was yesterday', or 'It takes two pizzas to make me as happy as one ice cream'. Though this lends credence to the measurability of welfare on a cardinal scale, these intuitions can be faulty. As an example, this method of assuming the measurable properties of sensation has arguably led to problems in psychophysics, as the subsequent data have not supported these intuitions (Laming, 1997).

However, our intuitive understanding of the target can help ground our expectations of its properties. There is a sense in which we should be able to grasp the interval scale measurement of welfare. For example, it is possible to say we are better off in our current situation (A) than if we were in this state with a painful ant bite (B), but that this would still be better than if we were to be thrown into sulfuric acid (C), and further that the difference between A and B is less than between B and C (Ng, 1997). Klockseim (2010) makes a similar argument regarding comparing the reduction of pain in curing one toothache to that of curing three. Such a judgment cannot be made using an ordinal scale alone; thus if we think it is meaningful, we

accept that a cardinal (interval or ratio) scale measure can be used for welfare, in at least some cases.

To confirm this, we need to look at the actual properties of the measurement target, and the types of transformations it can undergo. One line of evidence in favour of cardinal measurement of welfare is our ability to make trade-offs. We can choose to suffer some amount of negative experience in order to gain something positive, or to forego some positive experience for more of something else. That we can conceive of rates of substitution, or trade-offs, in welfare gives reason to think there are meaningful intervals between items on the scale. This intuition may be resisted in the human case, if one takes our ability to make such trade-offs as imperfect, such that there is sometimes no fact of the matter regarding our rates of substitution between different experiences. However, it seems that this problem typically arises in the human case because human subjective well-being is complicated by the range of abstract concepts and objects that humans are able to value and form preferences for, even outside of their direct effects on (hedonic) subjective welfare — for example, achievement, or freedom. Indeed, some researchers explicitly describe subjective well-being as inclusive of cognitive states such as life satisfaction, as well as affective states (Haybron, 2011).

This variety increases the sense of incommensurability — it is very difficult to determine, for example, how much satisfaction of achievement one is willing to sacrifice in order to gain the gustatory pleasure of eating chocolate ice cream. While the relationship between such preferences, values, and subjective well-being is complex (e.g. whether they are partially constitutive or merely instrumental), it is clear that human judgments of their own well-being are influenced by these factors. There is currently no evidence that the same is true for any non-human animals, where the value of such states seemingly resides solely in their felt effects, and thus no reason we should treat them as such. Where it is the case for any particular animal that they can conceive of and value more abstract states, these animals may also fall into this more complex category.

Additionally, we should be cautious about inferring from our inability to introspect regarding our own rates of substitution to a strong claim about the existence of such rates. As with the discussion in Section 2.2 about the methodological sources of scepticism, it is not at all clear that humans are actually accurate judges of their own experiences, or of predicting the impact of change on their overall well-being. There could be some fact of the matter regarding the

weightings of different factors as contributors to one's subjective well-being, even when that individual is unable to introspect or report on how they would perform the task.

Indeed, many animals have been demonstrated to make trade-offs regarding their own welfare — for example hermit crabs (Appel and Elwood, 2009), rats (Cabanac and Johnson, 1983), and iguanas (Balasko and Cabanac, 1998). This aligns with most accounts of the evolution of valenced subjective experience that take it to function as a common currency when making decisions between competing motivations requiring trade-offs (Cabanac, 1992; Ginsburg and Jablonka, 2019; Gygax, 2017; Spruijt, van den Bos and Pijlman, 2001). By using external, or objective, behavioural and physiological measures of welfare, we are able to measure these trade-offs without requiring subjective judgments to track them.

We can then perform tests to establish empirically whether the properties of additivity and associativity hold for welfare measures. For example, we could measure how much welfare an animal gains when receiving each of two resources and whether receiving both together gives the same number of units as the sum of each individually (additivity); also whether this is subject to an ordering effect (associativity). Similarly, assessing the trade-off behaviour of an animal both in a normal condition and when experiencing a negative affect such as pain could show an additive effect of the pain experience on the trade-off calculations. This has been observed in humans (Lang, 1995) and calves (Adcock and Tucker, 2021); further work aimed directly at investigating these effects for the purpose of establishing additivity could strengthen this case.

3.3. Setting the unit

I have demonstrated that there is sufficient reason to believe that subjective welfare possesses the necessary properties to justify its representation on an interval scale. This leads to another important question regarding how we establish the unit of measurement. Though the units are arbitrarily chosen, we still need to fix a stable reference unit and go on to determine how many of these are present in particular cases. It thus doesn't matter particularly what unit we choose, so long as it can be applied uniformly. I will describe two methods: the first looking for a principled single unit measure, the other setting two fixed points that are stable across cases and using the divisions between these to create the units. As I will show, given the current

methods of measurement available in animal welfare science, the
second is likely to be the more feasible possibility for setting a
measurement unit for welfare.

The first method is setting a fixed unit for measurement that can
then be compared across cases. For example, the metre was originally
fixed by reference to the length of a specific platinum bar (and now
fixed by the speed of light). The problem for applying this method to
animal welfare is in finding an appropriate fixed unit for reference.
One possibility is the Just Noticeable Difference (JND), a measure
used in psychophysics for quantifying measurement of sensation, such
as loudness, brightness, temperature, pressure, or even pleasure. The
JND is the minimum increment between two sensations that are
distinguishable by a subject, typically determined through self-report.
One unit of welfare would then be one JND of pleasure (or pain).

However, there are a number of reasons why this unit is unlikely to
be useful for animal welfare science, at least without substantial
further development. First, psychophysical methods have only been
developed for perceptual states and not yet validated even for
measurement of human welfare; though some writers have proposed
their use for measurement of happiness through just perceivable
increments of pleasure/pain (Ng, 1996). Second, the method of self-
reporting — criticized in the human literature for lack of clarity
regarding the method by which subjects are judging sensation, and the
accuracy of the judgments made (Laming, 1997) — will run into even
more problems for animals, when 'reports' could only be taken
behaviourally and not validated through investigation of the details of
individuals' reported experiences. Finally, we have no reason to think
that JNDs represent equally sized units along the spectrum of intensity
(i.e. that a JND at the lower end is the same 'amount' of sensation as a
JND at the higher end) — this has been assumed rather than
demonstrated, and some results in psychophysics seem to suggest
otherwise (Laming, 1997; Narens and Skyrms, 2020).

Thus, unless we are able to develop and validate JNDs for the
valenced experiences of animals (which, for the reasons mentioned,
does not seem promising), we need to turn to another method. The
second method — which I take to be most promising in this context
— is to set two fixed points and use the divisions between these for
constructing units. Remembering that units on an interval (or ratio)
scale are arbitrary, it does not matter specifically what these fixed
points are, so long as they are stably identifiable to set units for com-
parison and measurement across cases. In particular, if we have a

fixed zero, minimum, and/or maximum, we can use these to set the units as divisions between them (e.g. setting minimum as one and maximum as 100). For example, the units of the Celsius scale for temperature were derived by divisions between the set points of the freezing point of water (0°C) and its boiling point (100°C).

This method relies on an assumption of scale linearity. If the indicator or measure used is not one with a linear relationship to changes in welfare, the divisions in the measure will not represent equal intervals with regard to welfare and we risk making inaccurate calculations and unjustified inferences (Wodak, 2019). Although there are other methods for constructing e.g. logarithmic interval scales, these will then have different applications and require different types of calculations. This can be addressed in part by checking different measures against one another to establish the relative shapes of the response curves. For practical purposes within animal welfare science, we could approximate such a carving through the proxy measurements we use, such as physiological and behavioural variables. Many of these can be quantified and compared — blood concentrations of hormones, frequency of behaviours, changes in heart rate or body temperature. If we can assume that these proxies stand in the appropriate quantitative relationship to the central target of subjective experience, then we can give at least a rough cardinality to measurements of welfare.

An alternative method, developed by von Neumann and Morgenstern, and still commonly used in economics, is construction of an interval scale of utility using preferences over lotteries. While this method is targeted at utility understood directly in terms of preference satisfaction, a version more appropriate for subjective welfare has recently been proposed by Narens and Skyrms (2020) to measure utility through preferences over hedonic episodes. They use the duration of episodes as analogous to the probability of outcomes — e.g. an episode with 1/3 of the duration at intensity A and 2/3 at intensity B would play the same functional role as a gamble with 1/3 probability of A and 2/3 probability of B. An individual can then rank their preferences over sets of outcomes with different intensities and durations, and these orderings can be used to set the unit for a scale of hedonic utility in the same way as the von Neumann-Morgenstern lotteries could be used for a scale of preference-based utility (with the additional benefit of providing a natural zero point, as will be discussed in the next section).

However, the problem with these methods is their reliance on preference behaviour as a marker of utility, as the underlying assumptions linking preferences and hedonic utility don't necessarily hold. In large part this is due to the distinction between 'wanting' and 'liking' — these arise from separate brain systems that, although closely linked, can operate independently (Berridge, 1996) and thus preferences and welfare will not always track one another. It's possible to want things that one won't enjoy, and to enjoy things one does not hold a preference for. When thinking about subjective welfare, we should be careful in ascribing too much weight to measures based on choice or preference as these will only imperfectly represent hedonic experience. There is also currently no work testing whether animals hold preferences over lotteries or hedonic episodes, or how we could identify them. While there is a lot of work on measuring animals' preferences and willingness to pay (via 'work' or similar) (Kirkden and Pajor, 2006), it is far less clear how we could test these more complex attitudes. There is already scepticism as to whether humans satisfy the rationality axioms required for von Neumann-Morgenstern-type preference orderings, such as completeness and transitivity of preferences (Angner, 2013), and this seems even less likely in animals. Alongside the conceptual divide between preference and subjective experience, this gives reason to believe that these methods won't reliably give the right results for measuring subjective welfare. Instead we should prefer the method described above, of using intervals between fixed points, for setting units of welfare.

3.4. The problem of interpersonal comparisons

The final problem in determining the units for measurement of welfare is that they may not be the same across individuals (particularly across species). That is, no matter how we set our units, the same level of response given by different individuals on a particular measurement indicator (whether choice behaviour or something else) may not represent the same underlying intensity of experience; in essence each individual could have a different represented unit. This is the problem of interpersonal comparisons, which has occupied much of the literature on measurement of human welfare. If each individual has a different unit, then we cannot make comparisons or trade-offs using multiple individuals, which includes most utilitarian calculations. This has led some to reject the use of utilitarian sums at all (e.g. Narens and Skyrms, 2020, who favour the 'product utilitarianism'), and for a long

time led many in economics to reject interval scale measurement in favour of ordinal (and utilitarian calculation in favour of Pareto-optimal decisions) (Fleurbaey and Hammond, 2004).

This is only a problem if we don't have a way of 'converting' the units between individuals. For instance, returning to the earlier example of a tiger and an otter, if we know that the welfare units of an otter are double that of a tiger, we are then able to perform the necessary conversions to combine or compare them, in the same way as we would be able to compare a temperature in Celsius to one in Fahrenheit, so long as we know the formula. What matters then is determining if or how we can perform such calculations — whether we can acquire enough information on the actual relationship between individual utilities.

Some solutions proposed for the human case are: use of imaginative empathy, extended preferences, or stipulation of similarity in psychology or behaviour (Binmore, 2009; Harsanyi, 1955), or setting these values via convention (as suggested by Narens and Skyrms, 2020). These solutions are unlikely to be useful in the case of interspecies comparisons. They primarily rely on postulating strong psychological or behavioural similarities between individuals, or on our ability to introspectively understand the mental states of others, which seem difficult enough even for other humans. However, this doesn't mean the problem is insoluble. I won't argue for a particular solution here, but note that there are possible options, including use of proxies (e.g. neurophysiological complexity), reliance on deeper similarities between individuals (e.g. anatomical structure or evolutionary history of the valence system), or even setting values via convention relying on our (educated) intuitive estimates regarding comparative welfare (a version of extended sympathy) or moral weighting of different individuals.

A lot will depend on whether we take the comparison problem to be metaphysical or epistemic. That is, if there is some natural fact of the matter about the relative utilities of different individuals, but where we may in practice find it difficult to determine exactly what these are. If the problem is an epistemic one, then there is — at least in principle — some solution to the comparison problem, even if it's challenging to discover. If the problem is a metaphysical one, there will be no such answer and we need to develop alternative ways of comparing or aggregating utilities (e.g. Narens and Skyrms, 2020). If we can make sense of at least the most coarse-grained comparisons between species (as I argued in Section 2), then we should take it to be

the case that there is some fact of the matter regarding relative welfare and thus that the problem is more likely to be epistemic rather than metaphysical. In these cases we may take intuitions to the contrary to simply be a remnant of dualist thinking about the mind, where a materialist should accept that interpersonal comparisons are possible, at least in principle (Ng, 1997).

There is still a lot of work to do here to establish this as fact, and to determine the appropriate conversion ratios (or to set them via convention or by using our best proxies). This is a complex set of issues that cannot be addressed in this paper (see discussion in Browning, 2020a), but it doesn't provide an insurmountable case against unit measurement of this kind. Depending on how strong one takes the interpersonal comparison problem to be, and how convinced of the case for or against, will then lead to the selection of different measurement scales and ethical/decision procedures to account for this.

4. Ratio Scales

Although interval scales can give us good quantitative information, their limiting feature is that they don't have a true 'zero' point and so we cannot say anything about the ratios of different values. A ratio scale, by contrast, is set around a non-arbitrary origin (zero) point and thus allows these transformations. On a ratio scale, the ratios between the assigned values (doubles, triples, etc.) will then stand in for corresponding ratios between quantities of the attribute being measured, which allows comparisons such as 'twice as much as'. However, though the origin is fixed, the units of measurement are still arbitrary and can be varied.

4.1. Applications of ratio measurement

Ratio scales allow us to perform the same calculations regarding aggregative welfare and trade-offs as described above for an interval scale. However, they also allow for additional calculations — those requiring multiplicative or ratio judgments. This includes judgments like 'this animal has welfare twice as good as last week' or 'this animal's welfare is 80% of its maximum', but also for the more formal assessments described below.

When we compare interventions across different numbers of individuals, we cannot use an interval scale, as our arbitrary choices of zero and unit will change the outcomes of our deliberations. Here, interval-scale transformations will fail to preserve the ordering

between states, while ratio scales will not. This application is a common one within utilitarian animal welfare ethics, where we make decisions about how best to distribute our resources, say choosing between improving conditions for caged chickens or farmed salmon. Within these calculations, it isn't just the average welfare level of individuals in these conditions that matters, but the total number of animals affected, and in these cases we will need a ratio scale measure.

Finally, we require ratio scales to make quality of life assessments, determining whether an animal has a 'life worth living'. This is very common for aged animals with failing health, when their carers want to decide at which point euthanasia is the most humane option; or for determining the permissibility of bringing an animal into existence where it may have a life of net-negative welfare. Here, we compare welfare to some fixed zero point below which it is negative and above which it is positive. Unlike an interval scale this zero point is not arbitrary, as the decision to move it will affect our decisions regarding which lives are and are not worth living. In utilitarian calculus, this dividing line between positive and negative utility, or 'pleasure' and 'pain', is often very important. Any judgments about welfare level relative to a non-arbitrary zero, even if these are only level rather than unit judgments, will require a ratio scale as this is the only scale type that contains such a zero.

4.2. Ratio properties for subjective welfare

I described in the previous section that welfare can be measured cardinally, and how we might fix the unit. For a ratio scale, we also need to establish that these units hold the necessary multiplicative properties — for instance, whether doubling a resource leads to a measured doubling in welfare. Even if these calculations are not always relevant to the applications described (e.g. quality of life judgments), the properties must be established to use a ratio scale at all. The possibility of ratio-scaled measurement of hedonic experience is a position that is common within some work in human psychology, such as the use of 'objective happiness' scales that integrate momentary well-being reports over time (Kahneman, 2000). In this section I will show how subjective welfare meets the most important criterion for ratio-scale measurement — the possession of a zero point.

The additional important difference between an interval and a ratio scale is that a ratio scale has a fixed, non-arbitrary zero point.

Establishing that welfare can be measured on a ratio scale requires establishing that there is a natural zero point, fixed across all cases, regardless of the choice of unit. It is intuitive to think that there is such a zero point: the neutral point between positive and negative welfare, the point at which experience is evaluatively neither positive nor negative. Welfare, particularly with utilitarian frameworks, is typically conceived of as a continuum from negative to positive with a neutral zero point that separates the two types of experiences. At least in principle, then, we can make sense of a natural zero point for welfare. One may worry that this is not a stable reference point, if it relies on humans' intuitive judgments. After all, the phenomenon of response shift, or scale norming (Fabian, 2019; McClimans, 2011), has shown that individuals can change their understanding of the meaning of the reference points for judgments about well-being. However, it is not obvious that we would observe the same shift for the zero point. Unlike other points on the scale, it is not relative; instead more stably conceived of as the tipping point between net-positive and net-negative experience. Importantly, this may not represent the point at which one has an equal amount of positive and negative experiences, but rather the point at which their impact on overall welfare is balanced out. However, even if this is accepted as a conceptual set point, the problem still remains of how to fix it in practice. Here I will assess a few potential ways of doing so.

One proposal is that the natural zero point is the 'null episode': the value of an episode of zero duration (Narens and Skyrms, 2020). This can then be used to identify the zero point for intensity, as the intensity such that an episode of some positive duration of it is equal in value to that of the null episode. This is identified through finding episodes for which we have no preference of its occurrence over experiencing nothing at all; though this then has the problems associated with taking preferences as a guide to welfare experience, as already discussed. Another proposal is to set the zero as the value of non-existence, such that all lives worth living are positive, and lives not worth living are negative (Adler, 2012); which could be seen as a similar framework but for evaluating entire lives rather than single episodes. A life that one would be indifferent toward living as compared to not existing would thus be a neutral or 'zero-scored' life, though this proposal also runs into the difficulties in using preferences.

An objection to taking a central zero point is whether positive and negative welfare actually form a common continuum. Critics point out

that they may be separate dimensions — it is possible to simulta-
neously feel pleasure and pain, so these may be different systems that
can simultaneously be active (Narens and Skyrms, 2020), though
others point out the similarities in the neurochemical basis of each
(Berridge and Kringelbach, 2013; Spruijt, van den Bos and Pijlman,
2001). The potential heterogeneity of positive and negative affect is
too complex a topic to cover here (see discussion in Browning,
2020a), however there are good reasons to think that they can be
integrated into a single welfare experience, such as the fact that indi-
viduals are able to use both their positive and negative affect in
calculating trade-offs for decision-making regarding their actions. One
can be willing to undergo some amount of a negative experience in
order to have a positive experience. If positive and negative experi-
ence can be used as a 'common currency' where negative experiences
can be traded off against positive, this gives us reason to believe there
is a continuum.

   One other potential solution is to take positive and negative welfare
as two different ratio scales (Narens and Skyrms, 2020). However,
this leads to the obvious problems in preventing comparisons between
or integrations of positive and negative experiences. We will be
unable to use both sets of experiences to create a total welfare score
for an animal, either at a time or over a lifetime. We also couldn't
compare changes over time (or between individuals) if these changes
shift an animal from positive to negative welfare. These comparisons
can still proceed through ordinal measurement, or through an interval
scale (e.g. an animal going from –2 to +1 has increased its welfare by
3 units) which don't require separate scales, but these cannot provide a
principled distinction between positive and negative.

   To overcome these problems, the second option would be to instead
set the zero point as an absolute zero and have all states of welfare as
positive increases from there. This could, for example, represent an
ethical zero, as a minimum standard of 'unbearable pain' we would
allow any animal to exist at.[4] This would then assess all interventions
on how much they raise welfare above this baseline — all improve-
ments from the worst state will be positive. However, this would only
be a useful score for applications that aim at assessing the change
created by interventions. Other applications that aim at simply
measuring welfare will fail, as an animal could conceivably be

---

4   Richard Bradley (personal communication).

measured below this zero line, which would not be permissible. We can also again run into the comparison problem, if we want to take this as being the same across all species.

It could instead represent an empirical zero, the lowest possible welfare score an animal could receive. However, it still fails as a characterization of the total welfare of an animal. Here, all animals with welfare above the minimum would be represented as having positive welfare, regardless of whether their experience is overall positive or negative. There is no way of differentiating between these states and thus it would not permit the quality of life judgments described above. It also does not allow us to assess which experiences will add to total welfare and which will detract from it — when all are assigned a positive score, they will all increase the total regardless. There is no distinction between pleasures and pains; it fails to capture our intuitive grasp of the ethical and empirical significance of the dividing line between these states. There is also a concern about the cross-applicability of this zero across individuals or species. While a neutral point or null episode of welfare is intuitively the same across individuals, the worst possible experience may depend on specific features of capacity for suffering or pleasure.

While there is an intuitively plausible zero point for welfare — a non-arbitrary distinction between positive and negative experiences — it is still not entirely clear the best way of setting this zero. For this reason, in many instances it may be best to use an interval scale. However, given that it is counter-intuitive to arbitrarily set the zero point wherever one wishes (thus changing what counts as positive and negative welfare), an interval scale also seems insufficient. Wherever the principled dividing line between positive and negative experience is important, we should thus prefer a ratio scale.

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 several common measurement scales that can be applied — ordinal, interval, and ratio. Importantly, the scale chosen must match the properties of the measured attribute and play the appropriate role for the use of the measurements. In this paper I have given reasons for believing that subjective welfare possesses properties such that we are justified in representing it on all types of

scale up to and including ratio scales. These reasons include intuitive plausibility of cases and some indirect empirical evidence, which justify treating subjective welfare in this way, even where we do not yet have more definitive evidence for a stronger metaphysical claim regarding the properties of subjective experience. Which scale we require will depend on the application, and which we use will also depend on the properties of the specific measurement method used.

Cartwright and Bradburn (2011) describe three stages in the process of measurement. They are: *characterization* of the measurement target, defining the appropriate system of *representation*, and fixing the measurement *procedures*. In this paper, I began by characterizing the measurement target — subjective welfare. Throughout, my focus was on laying the groundwork for the second stage — determining the appropriate system(s) of representation for welfare, in terms of establishing which of the available types of measurement scale are suitable for use with subjective welfare. The next part of this step would then be to develop an explicit measurement scale (or scales) of the type(s) discussed above. Ideally, then formalizing this using a representation theorem to demonstrate that the choice is correct, using axiomatic assumptions about the underlying empirical structure of the phenomenon of subjective welfare, which can then be checked empirically.

The third stage would then be establishing the procedures for applying the representation system to produce the measurement results — the empirical investigation of specific cases and assigning of a value to the item relative to the scale. This requires identifying methods of measurement that can give results that accurately reflect the state of the target (i.e. producing high scores when welfare is high, decreasing scores as welfare is reducing, etc.) as well as producing values that match the representation system required (i.e. ordinal, interval, ratio). While this paper has focused on the justification for the use of these scales for welfare measurement in theory, to justify their potential use in practice there is the additional question of which of the range of available measurement methods would be the best fit for the different applications, and which could map onto the different scale types. There is a large range of methods currently used for the measurement of animal welfare, the description and assessment of which is beyond the scope of this paper. However, it is important to keep in mind when selecting a measurement indicator that the properties of the indicator must also match the properties of the scale we are mapping welfare onto for the context or application. For

instance, as we have seen, treating ordinal measures as cardinal can lead to inaccurate results and predictions.

The subjective welfare concept is common in animal ethics and animal welfare science, and it is important that this is measurable. Here I have shown that we can justifiably perform welfare measurement on either ordinal, interval, or ratio scales. At minimum, we can rank welfare states on an ordinal scale, which will allow us to make many of the required judgments 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. We can then use this zero as one of the two set points for determining the welfare units through divisions within the measured range.

Which type of scale we choose will depend on which applications we require, as well as how pressing we take problems such as interpersonal comparison to be. However, in each specific instance, the type of scale in use should be specified and the relationship made explicit between the properties of the measured indicators and the measurement scale for welfare. In the end, all the scale types have various benefits and drawbacks, and their use will depend on whether we are able to accept the problems for the applications we need. This isn't surprising, given the general move in the philosophy of science towards a pluralist perspective regarding models and concepts (Veit, 2020; Veit and Browning, 2021a). Although I have shown that all the described scales are, at least in principle, justified in their application to welfare, we should always aim for the simplest one necessary for our purposes and be aware of the context — there is no 'one size fits all' solution.

Acknowledgments

References

Adcock, S.J.J. & Tucker, C.B. (2021) Injury alters motivational trade-offs in calves during the healing period, *Scientific Reports*, **11** (1), art. 6888. doi: 10.1038/s41598-021-86313-z

Adler, M.D. (2012) *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*, Oxford: Oxford University Press.

Alexandrova, A. (2017) *A Philosophy for the Science of Well-Being*, Oxford: Oxford University Press.

Angner, E. (2013) Is it possible to measure happiness? The argument from measurability, *European Journal for Philosophy of Science*, **3** (2), pp. 221–240. doi: 10.1007/s13194-013-0065-2

Appel, M. & Elwood, R.W. (2009) Motivational trade-offs and potential pain experience in hermit crabs, *Applied Animal Behaviour Science*, **119** (1), pp. 120–124. doi: 10.1016/j.applanim.2009.03.013

Balasko, M. & Cabanac, M. (1998) Behavior of juvenile lizards (Iguana iguana) in a conflict between temperature regulation and palatable food, *Brain, Behavior and Evolution*, **52** (6), pp. 257–262. doi: 10.1159/000006570

Benjamin, D., Cooper, K., Heffetz, O. & Kimball, M. (2020) Self-reported well-being indicators are a valuable complement to traditional economic indicators but are not yet ready to compete with them, *Behavioural Public Policy*, **4** (2), pp. 198–209. doi: 10.1017/bpp.2019.43

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

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

Binmore, K. (2009) Interpersonal comparison of utility, in Kincaid, H. & Ross, D. (eds.) *The Oxford Handbook of Philosophy of Economics*, pp. 540–559, Oxford: Oxford University Press.

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

Broom, D.M. (1998) Welfare, stress, and the evolution of feelings, in Møller, A.P., Milinski, M. & Slater, P.J.B. (eds.) *Advances in the Study of Behavior: Stress and Behavior*, vol. 27, pp. 371–404, San Diego, CA: Academic Press.

Broom, D.M. & Johnson, K.G. (2019) *Stress and Animal Welfare: Key Issues in the Biology of Humans and Other Animals*, Cham: Springer Nature.

Browning, H. (2018a) No room at the zoo: Management euthanasia and animal welfare, *Journal of Agricultural and Environmental Ethics*, **31** (4), pp. 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), pp. 785–803.

Browning, H. (2020a) *If I Could Talk to the Animals: Measuring Subjective Animal Welfare*, [Online], https://openresearch-repository.anu.edu.au/handle/1885/206204.

Browning, H. (2020b) The natural behavior debate: Two conceptions of animal welfare, *Journal of Applied Animal Welfare Science*, **23** (3), pp. 325–337.

Browning, H. & Veit, W. (2020) The measurement problem of consciousness, *Philosophical Topics*, **48** (1), pp. 85–108. doi: 10.13140/RG.2.2.23082.75207

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

Cabanac, M. & Johnson, K.G. (1983) Analysis of a conflict between palatability and cold exposure in rats, *Physiology & Behavior*, **31** (2), pp. 249–253. doi: 10.1016/0031-9384(83)90128-2

Cartwright, N. & Bradburn, N. (2011) A theory of measurement, *The Importance of Common Metrics for Advancing Social Science Theory and Research: A Workshop Summary*, pp. 53–56, Washington, DC: National Academies Press. doi: 10.17226/13034

Chang, R. (1997) Introduction, in Chang, R. (ed.) *Incommensurability, Incomparability, and Practical Reason*, pp. 1–34, Cambridge, MA: Harvard University Press.

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

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

Dawkins, M.S. (2021) *The Science of Animal Welfare: Understanding What Animals Want*, Oxford: Oxford University Press.

Duncan, I.J. (2002) Poultry welfare: Science or subjectivity?, *British Poultry Science*, **43** (5), pp. 643–652.

Fabian, M. (2019) Scale norming undermines the use of life satisfaction scale data for welfare analysis, Preprint, [Online], https://osf.io/preprints/socarxiv/cg8n9/

Ferrer-i-Carbonell, A. & Frijters, P. (2004) How important is methodology for the estimates of the determinants of happiness?, *The Economic Journal*, **114** (497), pp. 641–659. doi: 10.1111/j.1468-0297.2004.00235.x

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

Fraser, D. (2008) *Understanding Animal Welfare: The Science in its Cultural Context*, Oxford: Wiley-Blackwell.

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), pp. 187–205.

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

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

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

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

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

Haybron, D.M. (2011) Taking the satisfaction (and the life) out of life satisfaction, *Philosophical Explorations*, **14** (3), pp. 249–262. doi: 10.1080/13869795.2011.594959

Hughes, B.O. & Black, A.J. (1973) The preference of domestic hens for different types of battery cage floor, *British Poultry Science*, **14** (6), pp. 615–619. doi: 10.1080/00071667308416071

Kahneman, D. (2000) Experienced utility and objective happiness: A moment-based approach, in Kahneman, D. & Tversky, A. (eds.) *Choices, Values, and*

Frames, pp. 673–692, Cambridge: Cambridge University Press. doi: 10.1017/ CBO9780511803475.038

Karlen, G.A.M., Hemsworth, P.H., Gonyou, H.W., Fabrega, E., David Strom, A., & Smits, R.J. (2007) The welfare of gestating sows in conventional stalls and large groups on deep litter, *Applied Animal Behaviour Science*, **105** (1), pp. 87–101. doi: 10.1016/j.applanim.2006.05.014

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), pp. 29–47.

Klocksiem, J. (2010) The amenability of pleasure and pain to aggregation, *Ethical Theory and Moral Practice*, **13** (3), pp. 293–303. doi: 10.1007/s10677-009-9200-8

Laming, D. (1997) *The Measurement of Sensation*, Oxford: Oxford University Press.

Lang, P.J. (1995) The emotion probe: Studies of motivation and attention, *American Psychologist*, **50** (5), pp. 372–385.

McClimans, L. (2011) Interpretability, validity, and the minimum important difference, *Theoretical Medicine and Bioethics*, **32** (6), pp. 389–401. doi: 10.1007/s11017-011-9186-9

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*, *Proceedings of the Australian Animal Welfare Strategy International Conference*, Conrad Jupiters, Gold Coast, Queensland, Australia, 31 August–3 September 2008..

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), pp. 241–253.

Narens, L. & Skyrms, B. (2020) *The Pursuit of Happiness*, Oxford: Oxford University Press.

Ng, Y.-K. (1996) Happiness surveys: Some comparability issues and an exploratory survey based on just perceivable increments, *Social Indicators Research*, **38**, pp. 1–27.

Ng, Y.-K. (1997) A case for happiness, cardinalism, and interpersonal comparability, *The Economic Journal*, **107** (445), pp. 1848–1858. doi: 10.1111/j.1468-0297.1997.tb00087.x

Nussbaum, M. (2011) The capabilities approach and animal entitlements, in Beauchamp, T.L. & Frey, R.G. (eds.) *The Oxford Handbook of Animal Ethics*, pp. 228–254, Oxford: Oxford University Press.

Pettigrew, R. (2019) *Choosing for Changing Selves*, 1st ed., Oxford: Oxford University Press. doi: 10.1093/oso/9780198814962.001.0001

Regan, T. (1983) *The Case for Animal Rights*, London: Routledge & Kegan Paul.

Singer, P. (1995) *Animal Liberation*, 2nd ed., London: Plimlico.

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), pp. 145–171.

Stevens, S.S. (1951) Mathematics, measurement and psychophysics, in Stevens, S.S. (ed.) *Handbook of Experimental Psychology*, pp. 1–49, Hoboken, NJ: John Wiley & Sons.

Stevens, S.S. (1959) Measurement, psychophysics, and utility, in Churchman, C.W. & Ratoosh, P. (eds.) *Measurement: Definitions and Theories*, pp. 18–63, Hoboken, NJ: John Wiley & Sons.

Veit, W. (2020) Model pluralism, *Philosophy of the Social Sciences*, **50** (2), pp. 91–114. doi: 10.1177/0048393119894897

Veit, W. & Browning, H. (2021a) Perspectival pluralism for animal welfare, *European Journal for Philosophy of Science*, **11** (9), pp. 1–14.

Veit, W. & Browning, H. (2021b) Phenomenology applied to animal health and suffering, in Ferrarello, S. (ed.) *Phenomenology of Bioethics: Technoethics and Lived-Experience*, vol. 84, pp. 73–88, New York: Springer International Publishing. doi: 10.1007/978-3-030-65613-3_6

Wodak, D. (2019) What if well-being measurements are non-linear?, *Australasian Journal of Philosophy*, **97** (1), pp. 29–45. doi: 10.1080/00048402.2018.1454483