



OPEN ACCESS

EDITED BY

E. Tobias Krause,
Friedrich-Loeffler-Institute, Germany

REVIEWED BY

Antje Schubert,
Friedrich-Loeffler-Institute, Germany
Hayley Randle,
Charles Sturt University, Australia

*CORRESPONDENCE

Leigh P. Gaffney
[✉ lgaffney@uvic.ca](mailto:lgaffney@uvic.ca)

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Animal Welfare and Policy, a section of the journal Frontiers in Animal Science

RECEIVED 05 October 2022

ACCEPTED 29 December 2022

PUBLISHED 16 January 2023

CITATION

Gaffney LP, Lavery JM, Schiestl M, Trevarthen A, Schukraft J, Miller R, Schnell AK and Fischer B (2023) A theoretical approach to improving interspecies welfare comparisons. *Front. Anim. Sci.* 3:1062458. doi: 10.3389/fanim.2022.1062458

COPYRIGHT

© 2023 Gaffney, Lavery, Schiestl, Trevarthen, Schukraft, Miller, Schnell and Fischer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A theoretical approach to improving interspecies welfare comparisons

Leigh P. Gaffney^{1*†}, J. Michelle Lavery^{2†}, Martina Schiestl^{3†}, Anna Trevarthen^{4†}, Jason Schukraft⁵, Rachael Miller^{6,7}, Alexandra K. Schnell⁶ and Bob Fischer^{8,9}

¹Fisheries Ecology and Marine Conservation Lab, Department of Biology, University of Victoria, Victoria, BC, Canada, ²Campbell Centre for the Study of Animal Welfare, Department of Integrative Biology, University of Guelph, Guelph, ON, Canada, ³Faculty for Veterinary Medicine, University of Veterinary Science, Brno, Czechia, ⁴Independent Researcher, Gloucestershire, United Kingdom, ⁵Open Philanthropy, San Francisco, CA, United States, ⁶Department of Psychology, University of Cambridge, Cambridge, United Kingdom, ⁷School of Life Sciences, Anglia Ruskin University, Cambridge, United Kingdom, ⁸Rethink Priorities, San Francisco, CA, United States, ⁹Department of Philosophy, Texas State University, San Marcos, TX, United States

The number of animals bred, raised, and slaughtered each year is on the rise, resulting in increasing impacts to welfare. Farmed animals are also becoming more diverse, ranging from pigs to bees. The diversity and number of species farmed invite questions about how best to allocate currently limited resources towards safeguarding and improving welfare. This is of the utmost concern to animal welfare funders and effective altruism advocates, who are responsible for targeting the areas most likely to cause harm. For example, is tail docking worse for pigs than beak trimming is for chickens in terms of their pain, suffering, and general experience? Or are the welfare impacts equal? Answering these questions requires making an interspecies welfare comparison; a judgment about how good or bad different species fare relative to one another. Here, we outline and discuss an empirical methodology that aims to improve our ability to make interspecies welfare comparisons by investigating welfare range, which refers to how good or bad animals can fare. Beginning with a theory of welfare, we operationalize that theory by identifying metrics that are defensible proxies for measuring welfare, including cognitive, affective, behavioral, and neuro-biological measures. Differential weights are assigned to those proxies that reflect their evidential value for the determinants of welfare, such as the Delphi structured deliberation method with a panel of experts. The evidence should then be reviewed and its quality scored to ascertain whether particular taxa may possess the proxies in question to construct a taxon-level welfare range profile. Finally, using a Monte Carlo simulation, an overall estimate of comparative welfare range relative to a hypothetical index species can be generated. Interspecies welfare comparisons will help facilitate empirically informed decision-making to streamline the allocation of resources and ultimately better prioritize and improve animal welfare.

KEYWORDS

interspecies comparisons, comparative cognition, animal welfare range, empirical methodology, effective altruism

1 Introduction

1.1 A case for the need to make interspecies welfare comparisons

The number of animals bred, raised, and slaughtered each year for food and other purposes is on the rise (Béné et al., 2015). On an annual basis, over 70 billion terrestrial animals and nearly a trillion aquatic animals, across a wide variety of species, are raised or captured for food (FAO, 2021; Franks et al., 2021). This trend has led to an increase in intensive production practices that significantly impact the welfare (see Table 1 for key definitions) of the various species involved and may lead to increased pain, suffering, and other negative experiences (e.g., Lundmark et al., 2014; Broom, 2019; Keeling et al., 2019; Xu et al., 2019). One major challenge for animal welfare science is the difficulty of making meaningful comparisons between the welfare impacts of certain practices on different species (Bracke, 2006; Cohen, 2009; Wong, 2016; Budolfson & Spears, 2019; Browning, 2020). That is, it is difficult to assess whether some species are made worse off by such practices than others.

There are many examples of how intensive production can impact welfare. Globally, for instance, most intensive pork production systems dock piglets' tails in their first week of life (Sutherland et al., 2008). This involves using clippers that are heated so that they both cut the tail and cauterize the wound at the same time. The procedure is done without anesthesia and can cause acute pain that disrupts normal behavior in the short run (2011; Sutherland et al., 2008). In the long run, tail docking can result in the growth of neuromas (i.e., nerve tumors) that are permanently sensitive (Sutherland et al., 2008). Production system managers argue that tail docking is necessary to reduce injury from other piglets, who often bite at tails if they are left long (Sutherland et al., 2008). In most

intensive egg production facilities worldwide, beak trimming (i.e., the partial removal of the upper portion of a hen's beak) is a standard procedure performed on young hens (Bessei, 2018). It involves removing roughly a third of the upper beak, or sometimes both the upper and lower beak (Lonsdale et al., 1957), with a hot blade that both cuts and cauterizes (Henderson et al., 2009). Like tail docking, beak trimming can cause acute pain that disrupts normal behavior (Duncan et al., 1989) and also result in the growth of neuromas that are permanently sensitive (Kuenzel, 2007). Production system managers argue that beak trimming is necessary to reduce feed waste and avoid pecking-related injuries that can lead to cannibalism and increase chicken mortality (Allen and Perry, 1975).

Mass marking of salmon by fin clipping (i.e., the partial or full removal of a fish's fins) is a procedure commonly used in intensive aquaculture and hatcheries to distinguish farmed or hatchery-reared salmon from wild salmon (Uglem et al., 2020). Similarly, to tail docking and beak trimming, fin clipping may cause pain and injury in fish and alter swimming efficiency (Roques et al., 2010; Buckland-Nicks et al., 2021; Schroeder & Sneddon, 2017; Thomson et al., 2020; Uglem et al., 2020). Production system managers argue that fin clipping is the easiest method to identify fish because it is inexpensive, quick, and requires minimal equipment and training (Hammer and Lee Blankenship, 2001).

These practices raise questions that need to be addressed to inform future directions in welfare in intensive production. For example, are the welfare impacts of tail-docking pigs worse than beak trimming chickens? Are the welfare impacts of beak trimming chickens worse than fin clipping salmon? Or are the welfare impacts equal? What empirical evidence exists that could be used to make this assessment? Considering whether one practice has greater welfare impacts than the other is a primary concern for animal advocates (see Table 1 for key definitions) who have to make choices about how to allocate limited resources. Many of these advocates, including

TABLE 1 Key definitions.

Term	Definition
Affective states	Affective states refer to the experience of feeling the underlying emotional state (Hogg et al., 2010).
Cardinal utility	Cardinal measurement of utility refers to the measurement (or expression) of utility in terms of units like 2, 4, 6 and 8. Cardinality means that utility can be measured in numbers (Baumol, 1958).
Effective altruism	Effective altruism is a research field and conceptual approach emerging from a social community, using data-driven reasoning, aiming to find the best ways to help others, and put them into practice (Broad, 2018).
Empirical methodology	Empirical research is a type of research methodology that makes use of verifiable evidence in order to arrive at research outcomes, meaning it relies solely on evidence obtained through scientific data collection.
Hedonism	The word 'hedonism' comes from the ancient Greek word for 'pleasure'. If hedonism is true, then what matters is how an animal feels - its subjective experiences (Weijers, 2011).
Proxy/Indicator	Proxies or indicators, are defined as measures relevant to cognition, behavior, welfare, and physiology (e.g., measures of brain power, working memory capacity, self-awareness) (Fischer, 2022).
Stakeholder	A person with an interest or concern in something, especially a business.
Valenced	Valence, or hedonic tone, is the affective quality referring to the intrinsic attractiveness - goodness (positive valence) or averseness - badness (negative valence) of an event, object, or situation. The term also characterizes and categorizes specific emotions (Lindquist et al., 2016; Bruckner, 2020)
Welfare	The term "welfare" refers to the state of an individual in relation to its environment, and this can be measured (Broom, 1991; Broom, 1996; Bruckner, 2020).
Welfare range	An animal's welfare range is the difference between how well and poorly that animal can fare at a time (Fischer, 2022).

effective altruists (see [Table 1](#) for key definitions), want to allocate funding in a way that maximizes returns on welfare investments (i.e., produces the largest welfare improvement per dollar spent). Likewise, many members of the general public wish to make informed decisions around their food and purchasing choices. Individuals may, for instance, choose to become pescatarians, vegetarians, or vegans, or simply avoid one kind of animal product while eating others (e.g., those who abstain from eating veal or foie gras). These decisions are largely based around their perceptions and understanding of the impacts of farming on different animals. However, without relevant empirical data, such decisions, for stakeholders (see [Table 1](#) for key definitions) of all types, are invariably *ad-hoc* or subjective, and thus unlikely to achieve their intended aims. Interspecies welfare comparisons can provide a pathway to make informed decisions about which areas and which taxa to prioritize for various purposes.

Making interspecies welfare comparisons can have other implications, particularly in relation to identifying bias in discussions of animal welfare. Animal welfare concerns have primarily been directed at terrestrial vertebrates used in agriculture, laboratory research, and as companion animals (e.g., [Russell & Burch, 1959](#); [Lundmark et al., 2014](#); [Cardoso et al., 2017](#); [Franks et al., 2021](#); [Gaffney and Lavery, 2022](#)). However, many species used in intensive production systems, such as fish, shrimp, and silkworms, have received little attention and consequently, their welfare is often regarded with less concern (e.g., [Elder & Fischer, 2017](#)). Furthermore, the production numbers of these latter species tend to amount to considerably more overall in comparison to the more ‘traditional’ ones ([Franks et al., 2021](#)). Such attitudes may be based on arbitrary distinctions, with humans tending to care more about species that are evolutionarily closer and often more familiar, like mammals, than those that are more distant and different, like insects. Or, there may be legitimate reasons to be less concerned about the welfare of some species compared to others. Nevertheless, without tools to compare welfare across species, it is difficult to answer these questions.

Interspecies welfare comparisons can also improve welfare guidelines for scientific research. Such comparisons become particularly important when implementing the imperative to “reduce, refine, and replace” (the 3Rs; [Fenwick et al., 2009](#)). For example, when possible, researchers are required to replace animal models with non-animal models ([Burden et al., 2015](#)). However, in situations where replacement is not possible (given research objectives), some scientists defer to using animals, which are thought to be “cognitively less-sophisticated” animals. For example, zebrafish are often used as a substitute for ostensibly “cognitively more-sophisticated” animals, like mice ([Hamilton et al., 2016](#); [2018](#)). These decisions are based on the assumption that members of one species would be harmed less by the research than members of another ([Schaeck et al., 2013](#); [Message & Greenhough, 2019](#); [Sloman et al., 2019](#); [Almstedt et al., 2022](#)). Inevitably, without interspecies welfare comparisons, such subjective judgements could introduce unjustified bias towards certain species over others.

Our goal is to outline a theoretical approach to improving interspecies welfare comparisons using an empirical methodology (see [Table 1](#) for definition and details). We propose investigating welfare ranges (see [Table 1](#) for key definitions), which refer to the differences between how well or poorly various animals can fare at a

time. This theoretical construct allows us to compare the severity of harms and benefits across species.

1.2 Conceptual issues associated with interspecies welfare comparisons

We need to consider several conceptual issues before turning to our method for making interspecies welfare comparisons.

First, we should acknowledge that there are many theories of welfare. For example, here are four that have had some influence in agriculture, conservation biology, animal welfare science, and philosophy:

1. Welfare as bodily health: animals have positive welfare insofar as their bodies are functioning properly ([Dawkins, 2021](#)).
2. Welfare as engaging in or expressing natural behavior: animals have positive welfare insofar as they exhibit (or can exhibit) natural behavior ([Bruckner, 2020](#)).
3. Welfare as subjective experiences: animals have positive welfare insofar as they are experiencing sufficiently many positive affective states (see [Table 1](#) for definition and details) relative to negative affective states ([Robbins et al., 2018](#)).
4. Welfare as hedonism/desire satisfaction: animals have positive welfare insofar as they “get what they want” ([Dawkins, 2021](#)).

Theories of welfare differ over the determinants of welfare. Nevertheless, these theories are sometimes combined: the classic triadic theory discussed by [Fraser \(2008\)](#), for instance, proposes that welfare is jointly determined by bodily health, natural behavior, and subjective experiences. Similarly, the Five Freedoms ([Webster, 1994](#)) has had considerable influence as a framework for animal welfare assessment in policy-making spaces and incorporates elements of subjective experience, bodily health, and natural behavior into its conceptualization of welfare. Balancing the overall valence of lifetime subjective experiences and incorporating aspects of hedonism, the concept of a “life worth living” ([FAWC, 2009](#); [Yeates, 2011](#)) has been used to determine minimum standards for the treatment of farm animals in some policies and guidelines. Many of these theories have received criticism (e.g., [Korte et al., 2007](#); [McCulloch, 2013](#); [Duncan, 2016](#)), but are generally unified by some degree of concern about an animal’s subjective experiences.

Second, aside from aligning with a theory of welfare, we must also consider the different types of interspecies welfare comparisons. [List \(2003\)](#) distinguishes between two types of comparisons. The first type is the more basic: it concerns the valences of experiences (see [Table 1](#) for key definitions)—i.e., whether they are positive, negative, or neutral. Imagine, for instance, a sow who is physically restricted (e.g., in a farrowing crate) and cannot reach her piglets and a healthy chicken who is pecking at some corn in a safe environment. It seems likely that the sow’s experience is negatively valenced whereas the chicken’s is positively valenced. So, we can plausibly conclude that, at least with respect to their experiential states, the chicken is faring better than the sow.

The second type of interspecies welfare comparisons are level comparisons, that is, differences within a given valence, which

introduces additional complexity. Imagine a recently tail-docked pig and a hen which has not eaten for eight hours. Both animals are likely to be having negatively valenced experiences (acute pain and some degree of hunger, respectively). However, while it may seem plausible that the docked pig is worse off than the hungry chicken, it is difficult to provide a detailed justification for this judgment. We may inherently think about how we, as humans, may feel in a comparable situation, reflecting on our own experiences. However, without knowing the extent to which other animals experience pain or hunger comparably to us (or to one another), we cannot accurately make such a distinction. At present, there is no agreed-upon method for making such interspecies welfare level comparisons.

Finally, it is important to recognize that our assessment of a given animal's welfare is based on objective measures of the animal's subjective state (Sandøe & Jensen, 2011). However, subjective states are not directly measurable, and we cannot ask animals directly how they feel. Thus, we are left measuring "indicators" or "proxies" of welfare (see Table 1 for key definitions), rather than the momentary state itself. Validation of such proxies of welfare is therefore of particular importance and is especially pressing in cases where we have a limited understanding of animals' physiology and behavior (e.g., the pain debate in fishes and insects; see Vettese et al., 2020 and Gibbons et al., 2022). Further, it is unclear how to theoretically aggregate proxies into a measure of overall welfare, even within a species (e.g., see Botreau et al., 2007 for a review).

Our proposed solution avoids these problems for now, by investigating animals' welfare ranges with the aim of creating a tool that could inform interspecies welfare comparisons. An animal's welfare refers to how well or poorly an individual is faring (Broom, 1986); so, an animal's welfare range refers to the difference between how well or poorly an animal can fare at a time. The contrast here is between the actual state of an animal (welfare) and possible states of that animal (welfare range). Animals with relatively large welfare ranges can be harmed to greater degrees than animals with relatively small welfare ranges. Notice that welfare range profiles can be created for animals at the individual-level, but our methods have been designed to create welfare range profiles at the species-level.

As the definition of welfare ranges suggests, talk of "larger" and "smaller" welfare range is a simplification, overlooking potential dissociations between the various dimensions and multiple theories of animal welfare (see review by Bruckner, 2020). According to a pluralistic theory of welfare, there are multiple determinants of welfare. Dawkins (2021) has such a theory, which states that animal welfare is determined by two factors: namely, animals being healthy and getting what they want. By contrast, a monistic theory of welfare suggests there is a single determinant of welfare, such as hedonism (see Table 1 for key definitions) This theory states that animal welfare is determined by the quality of their subjective experiences (Robbins et al., 2018), where all and only positive experiences are good for animals, whilst all and only negative experiences are bad for them.

While it is possible to investigate differences in welfare ranges assuming any theory of welfare, it is impossible to do that in a single paper. So, for simplicity, we assume hedonism. This theory of welfare is compatible with the view that it matters whether animals are healthy and whether they can express species-typical behaviors (Robbins et al., 2018). Following hedonism, we will assume that

welfare at a time is determined by the qualities of experiential states, i.e., the strength of how good or bad an animal's overall experience is. So, if there could be variation among species in terms of the potential intensity of their experience, then there could be differences in their welfare ranges.

Animals differ with respect to their evolutionary history, neurophysiology, and neurobiology. This seems to have led to variation in their cognitive, affective, and sensory capabilities. It seems plausible, then, that there would be considerable differences in their experiential lives. Indeed, Birch et al. (2020) argue that there are five dimensions of variation: Perceptual Richness, Evaluative Richness, Integration at a Time, Integration across Time and Self-Consciousness. They argue that traditional one-dimensional scales of consciousness neglect these important dimensions of variation across taxa. Using a multi-dimensional approach by investigating taxa against each proposed dimension would create "consciousness profiles" that capture variation and highlight where a taxon is likely to fit in the space of possible forms of experience.

If different species encounter differences in their experiential lives, then it is plausible that there are characteristic differences in the determinants of the qualities of experiential states. Differences in intensity are perhaps the most familiar to us, such as pain perception, which is variable in humans (Hu & Iannetti, 2019). However, there is a difference between variations in the strength of the stimulus to produce a given response and variation in maximum response capacity. Given apparent differences among humans, who broadly share social, affective, intellectual, behavioral, and neurobiological characteristics, it is not hard to imagine more profound differences among nonhuman animals, a possibility that is explicitly raised in the literature (e.g., Yeates, 2012).

1.3 Why could differences in welfare ranges be relevant to interspecies welfare comparisons?

In brief, we can use standard welfare assessments, interpreted with welfare ranges, can be used to estimate the relative badness of harms or the goodness of benefits. This is because, from a philosophical perspective, when we assess animals' welfare, we it is assessed it relative to a species-typical neutral point. Given that neutral point, we assess both valence and strength of valence is assessed. For example, we can say that a particular state is positive or negative and that it is more positive or negative than some other state (e.g., Mendl et al., 2010). So, while we use measures with cardinal utility are used (see Table 1 for key definitions) to assess welfare, such as the duration of protective behavior, cortisol levels, time to return to normal feeding behavior, and changes in time spent resting vs. active, we aggregate them to produce an ordinal ranking of welfare states (Botreau et al., 2007). When it comes to intraspecies welfare comparisons, what matters is not, for instance, the duration of protective behavior per se, but one of two comparisons:

1. The duration of protective behavior that one individual displays in response to a given stimulus compared to the duration of protective behavior that the individual displays in

response to a different stimulus, i.e., an individual-level focus, for example, using individual-based measures of welfare (see [Blokhuis et al., 2010](#)), or

- The duration of protective behavior that one individual displays compared to the typical duration of protective behavior that individuals of that species display in response to a range of stimuli and / or stimuli of that kind, i.e. a species-level focus, for example, using group-level measures of welfare (see [Main et al., 2003](#)).

We typically, validate measures of welfare are validated by making either individual-level or species-level comparisons; we assess the impacts of particular stimuli in terms of how they affect animals by comparing their response relative to another individual or species. These relative rankings are essential, as we cannot ask animals directly how they are faring. This implies, however, that when we make interspecies welfare comparisons are made, we are starting out with species-relative data. As such, it is safe to assume that apparently equivalent harms reduce the welfare of members of each species by an approximately equivalent percentage of their respective welfare ranges. To see this, consider [Figure 1](#).

[Figure 1](#)'s conceptualization of "welfare units" and welfare ranges provides a tentative way to quantify the relative welfare impacts of

different harms and benefits. While obviously imprecise, it may still be the case that they are useful for many practical purposes. That being said, the usefulness of welfare ranges depends entirely on our ability to empirically assess and quantify it. If there is no way to do that, then we cannot use welfare ranges to tackle the problem of interspecies welfare comparisons.

2 Proposed methodology

Our aim in this section is to propose a basic methodology for assessing welfare ranges. This is summarized in [Figure 2](#).

The *first* task is to specify features that are intrinsic, rather than extrinsic, determinants of welfare, and so of welfare ranges. This part requires selecting a theory of welfare; (see section 1.2).

Importantly, we do not suggest that the theories of welfare outlined in section 1.2 are equally plausible or that the options we mentioned represent the only possibilities available. Our goal here is to set out the methodology, not to defend particular choices within it. If, for instance, we conclude that welfare is determined by bodily health, we would then turn to the task of operationalizing bodily health in ways that lend it to empirical investigation.

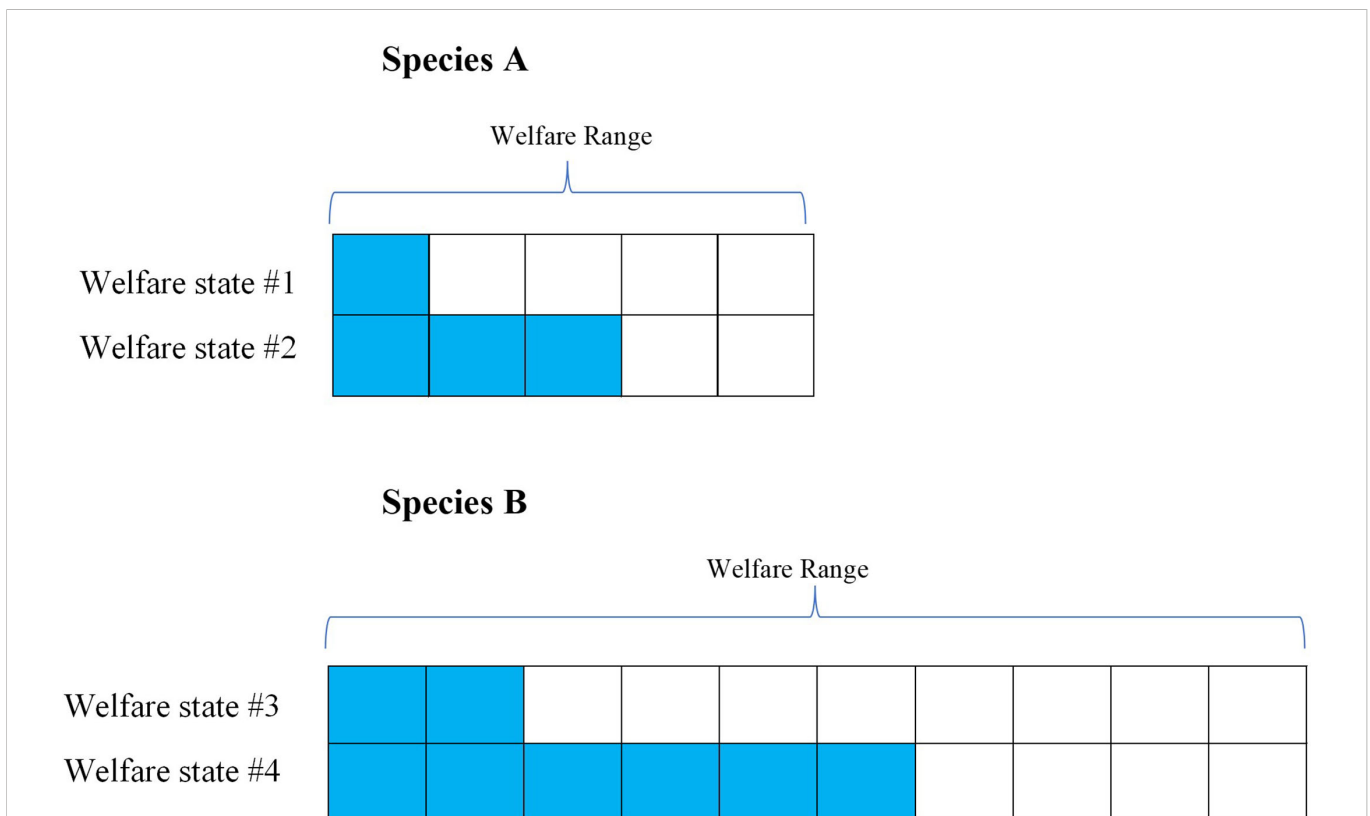
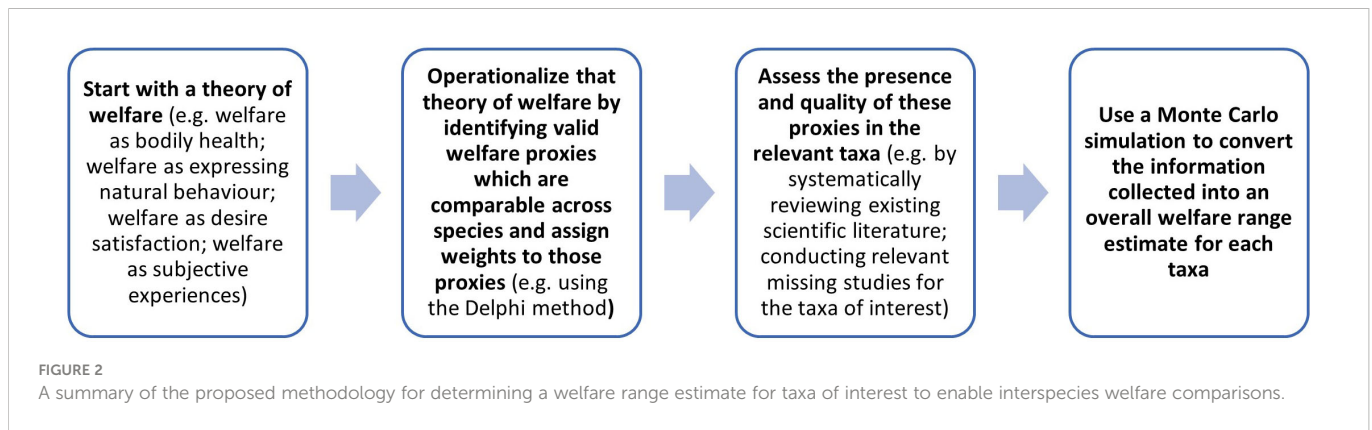


FIGURE 1
Theoretical figure to explain Welfare Range vs. Species-Relativized Welfare Impacts. Species (A) has a smaller welfare range than Species B, as represented by Species A having fewer total "welfare units" than Species B (i.e., the total number of cells per row). An ordinary welfare assessment method would compare the two welfare states (as indicated by the colored cells) within each species, concluding that State #1 is worse than State #2 for Species A and that State #3 is worse than State #4 for Species (B). Notably, though, such methods deliver proportional results: Welfare State #1 will seem about as bad for Species A as State #3 seems for Species B (20% of the welfare range), since those welfare states are just being compared to the best and worst state for each species. The outcome is that apparently equivalent welfare states are already scaled to welfare ranges, which means that if Species B has a greater welfare range than Species A, the members of species B are actually worse off in welfare states that appear equivalent. In other words, we can assume an apparently equivalent harm scale with welfare ranges, which makes welfare ranges a useful tool for interspecies welfare comparisons.



The *second* task involves turning the determinants of welfare enumerated during the first stage into measurable proxies. Notice that, at the outset, there is a tremendous amount of empirical uncertainty about the extent to which different animals display different welfare-relevant proxies. But that does not negate the value of describing a theoretical methodology built on such proxies, as it can assist in prioritizing research efforts such that our empirical certainty increases, and the estimates produced by the methodology are refined. These proxies should ideally be valid and amenable to operationalization, comparable across taxa, and chosen with an understanding of their ecological relevance to the taxa being compared. Further, there are considerable theoretical and practical challenges involved in comparing morally relevant features across phylogenetically distant animals. For example, the presence of nociceptors provides some evidence of the capacity for negative subjective experiences, but it is not definitive, since there can be nociception without any subjective experience at all in humans (Dubin and Patapoutian, 2010). Moreover, these proxies may relate to cognition, affect, behavior and neuro-biology. We therefore suggest that the best way forward is to weigh the chosen proxies in terms of the quality of the evidence they provide for the factors that are taken to be determinants of welfare. One way to select and provide these precise proxy weights is to use the Delphi method (Linstone and Turoff, 1975). In brief, the Delphi method is a form of structured deliberation. It begins with the selection of a panel of experts. Then the experts answer questionnaires in at least two revisions. After each revision, the experts send their answers to a facilitator who returns an anonymized summary of the experts' assessments to each member of the panel.

The *third* task involves assessing the evidence for these proxies in the relevant taxa. To begin, this task involves systematically reviewing the existing scientific literature. For more in depth knowledge about how this can be done, please refer to our pre-printed review about the relationship between cognition and welfare in 10 farmed animal taxa (Miller et al., 2022b pre-print). Notice that to apply our empirical methodology in full, we would likely need to conduct various relevant new studies that have not been completed for the taxa of interest. In primates, for instance, perspective-taking is associated with self-awareness, theory of mind, and empathy (Bulloch et al., 2008; de Waal, 2008; Towner, 2010). Specifically, perspective-taking involves reasoning about the mental states of others (e.g., their intentions, desires, and knowledge) and has been linked to possessing strong emotional capacities (Healey and Grossmann, 2018). Consequently,

perspective-taking may be considered a suitable proxy for some cognitive capacities that are either determinants of welfare or are themselves associated with determinants of welfare. There is ample evidence of perspective-taking in pigs: they can learn to follow other pigs who they recognize to have information about the location of food (Held et al., 2000), they can adjust their own behavior to prevent other pigs from exploiting their knowledge in this way (Held et al., 2002a), they can detect whether humans are paying attention to them *via* head cues (Nawroth et al., 2013a), and they can follow human hand signals to find food (Nawroth et al., 2013b). However, there is very little evidence as to whether chickens engage in perspective-taking (Smith et al., 2011), suggesting that additional research would be valuable.

Before we can draw any conclusions about the value of additional research, it is critical to identify the quantity and quality of the evidence that has already been published. For each publication found in the review, it would be important to record the estimate of the credibility of that paper and either its conclusion regarding the presence, absence or magnitude of the proxy, depending on whether the proxy is discrete or continuous. The strength of evidence could be rated along a scale. For example, a recent review of sentience in invertebrates used a scaled rating method ranging from 'lean no' to 'yes' (Rethink Priorities, 2020; Table 2). Another review on the evidence of sentience in cephalopod molluscs and decapod crustaceans used a scaled rating method that graded evidence in terms of how many of criteria for sentience were satisfied (8 criteria in total) (Birch et al., 2021). Specifically, evidence was graded as 'extremely strong' if 7–8 criteria were satisfied, 'strong' if 5–6 criteria were satisfied, 'substantial' if 3–4 criteria were satisfied, 'some' if only 2 criteria were satisfied, and 'unknown or unlikely' if 0–1 criteria were satisfied. Using scaled rating methods can generate welfare range profiles per taxa that simultaneously highlights the quality and quantity of evidence *and* identifies gaps in the current literature. We note that all estimates of scalar proxies should be normalized to a hypothetical index species that possesses the maximum observed value for any proxy that might matter for that particular welfare comparison. Since it is essential to compare all the values in the table to some reference value possessed by the index species, the absence of a proxy in the index species entails that the welfare range of other species goes to infinity, or some other arbitrarily large number.

The *fourth* task involves turning the data into overall welfare range estimates using a Monte Carlo simulation. Although other

TABLE 2 Examples of potential literature review output and rating scale for some example proxies and species, using the rating approach from **Rethink Priorities (2020)**.

Proxy	Species A (e.g., pig)	Species B (e.g., chicken)	Species C (e.g., salmon)	Species D (e.g., crayfish)	Species E (e.g., bee)
Judgment bias	Likely yes (Düpján et al., 2013)	Likely yes (Crump et al., 2018)	Unknown	Lean yes (Bacqué-Cazenave et al., 2017)	Likely yes (Bateson et al., 2011)
Social learning	Likely yes (Oostindjer et al., 2011)	Likely yes (Nicol & Pope, 1994)	Likely yes (Bajer et al., 2010)	Lean yes (Jiménez-Morales et al., 2018)	Likely yes (Alem et al., 2016)
Boredom-like behavior	Likely yes (Wemelsfelder, 1985)	Likely yes (Newberry, 1999)	Unknown	Unknown	Unknown

methods may also be possible, Monte Carlo methods are the preferred choice for modeling phenomena with significant uncertainty in inputs (Kroese et al., 2014). They reduce the need for using human judgment, which is often unreliable when dealing with complex questions. They also allow a complex probability density function to be presented as an output, rather than just a point estimate or a simple range, which is especially important for this project because it makes it easier to appreciate the degree of uncertainty in particular welfare range estimates. One way to proceed is to survey experts, using a formal, pre-registered, structured way of aggregating the survey results into a useful bottom-line estimate that preserves all information about the range of judgments that the experts make. This process reduces the need to make decisions about how to aggregate information that could influence or bias the results.

Each sample used as input for the Monte Carlo method is the judgment of one expert in the field, combined with the results of one paper that studies each proxy that the expert considers to be important. The result of this sample is plotted on a histogram and the process is repeated thousands of times. The resulting histogram represents the scale of possibilities for the welfare range typical for a given species, given different judgments and lines of evidence. This histogram can be used to produce averages, confidence intervals, and other ways of summarizing or reporting the data.

Given a specific theory of welfare and a set of welfare determinants, each repetition of the simulation will:

1. Randomly choose one expert in the Delphi panel. Then, assign a weight to each proxy based on that expert's estimates for the proxy weights.
2. Randomly choose one paper for each proxy, based on the credibility assigned to that paper. Pull a sample of the numerical value of that proxy from its adjusted distribution.
3. Calculate a weighted average of the capacity, using the values from Step 2 and the weights in Step 1.

The simulation should be run at least 10,000 times, producing a histogram of results. Again, this histogram will be the probability distribution of the species' welfare range as a fraction of the hypothetical index species.

There are bound to be gaps in the available proxy-relevant research for some species and we have a choice about how to manage this. One option is not to intervene, simply ignoring

unknown values. As a result, the weight of the other proxies (for which there is known information) would be increased proportionally when performing weighted average calculations. So, if a species has (average) values of 0.2, 0.3, 0.4, and unknown across four proxies, with equal weight on them all, the average would be 0.3. However, this has the effect of amplifying the significance of the other sources of variance.

A second option is to replace all unknown values with the corresponding values from the target comparison species. The hypothetical index species has the maximum observed value for each proxy across all actual species. So, entering values from the hypothetical index species would produce empirically implausible results, e.g., attributing cognitive capacities that we know a species lacks simply because its specific capacities have not been studied. For example, if pigs are compared to chickens, and there are lots of unknowns for chickens then we replace unknown values for chickens with the known values for pigs. This would have the effect of reducing the significance of the other sources of variance and would amount to a "curve" in favor of no variance. This would reflect the judgment that we should err on the side of welfare ranges being distributed more equally across the target taxa. Moreover, it may mean that we are unable to identify any differences in welfare ranges between some taxa, which will result in there being a narrower range of cases where we can draw on welfare range differences to make interspecies welfare comparisons. However, a narrower range of cases might still be a practically significant range of cases. Then finally, with our estimate in place, it is possible to make certain interspecies welfare comparisons.

3 Discussion

Our aim has been to propose a method for making interspecies welfare comparisons *via* estimates of comparative welfare ranges. We do not assume that this methodology will reveal differences (or similarities) in welfare ranges. Instead, we believe that *if* there are differences across taxa, ours is a promising method for discovering them. Furthermore, as our description suggests, this is a substantial research program that could only be completed over a significant period of time with extensive interdisciplinary collaboration. There are still some aspects of the method that deserve special attention, which we discuss below.

Depending on the theory of welfare used, the method could become more complex. If applying a pluralistic theory of welfare (such as Fraser, 2008's triadic theory) or using multiple theories of welfare at once, a separate Delphi method for each theory or component of the theory (e.g., bodily health, natural behavior, and subjective experiences; Fraser, 2008) would need to be conducted. The method can become more complicated because it might be necessary to use a different panel of experts appropriate to that theory or component. Empirical research would then need to be focused on the proxies, if any, that are shared across components or theories and are found by consensus to be important for each theory.

Depending on the proxies that are chosen and the taxa that are compared, a lack of relevant literature reporting evidence of those proxies may represent a significant limitation. Gaps in the literature may also make choosing proxies difficult. For instance, neuron counts (Herculano-Houzel et al., 2015; Raji & Potter, 2021) are relatively easy to compare across species and there are already data for many taxa of interest. However, it is not clear how neuron counts are linked to the welfare of an animal. To properly compare neurons, we need to know where they are located and how they are connected to each other. So, insofar as neuron counts are worth investigating and comparing, they must be handled carefully as proxies for other characteristics of interest (Von Bartheld et al., 2016). It may be, for example, that neuron count is associated with affective sophistication, intensity of valenced experiences, or general intelligence, though extensive research would be required before such conclusions could be drawn (Dicke & Roth, 2016). Our approach helps to identify where these gaps in the literature exist and highlights which proxies should be prioritized for future research.

Beyond a lack of literature, comparing phylogenetically distant taxa may pose additional challenges. For instance, if it turns out that sentience (assuming it is a feature relevant to the theory of welfare in use) is the product of convergent evolution, with multiple independent origins (Brown, 2020), then we might never find proxies that work across those taxonomic gaps. Even if it turns out that sentience is not the product of convergent evolution, we will end up relying heavily on the field of comparative cognition. Fortunately, there has been a recent surge of interest in comparing species across metrics that may bear on questions about welfare ranges (MacLean et al., 2014; Cauchoix et al., 2018; Miller et al., 2022a). There has been a concomitant surge in theoretical discussions about how to compare features across species, as seen in Weiss et al. (2019), which outlines a quantitative measure of social complexity that works across species. Similarly, Anderson and Andolphins (2014) developed a framework for studying emotions across species. Such research provides reason for optimism about the potential of comparative cognition research.

However, it should be noted that comparative cognition is a heterogeneous field with respect to the reliability and reproducibility of research findings. Some areas of comparative cognition research have been criticized for their low rates of reproducibility, largely owing to small sample sizes, inappropriate or noisy measurements, and implausible hypotheses (Forstmeier et al., 2017; Farrar et al.,

2020). By contrast, other areas of comparative cognition research appear to be less affected by low reproducibility rates due to the use of robust designs that can easily be replicated; for instance, the use of within-subject designs where subjects experience many trials multiple times (Smith and Little, 2018). The field of comparative cognition also bears hallmarks of the publication bias towards positive results. Specifically, the field is biased towards confirming more exceptional cognitive abilities in animals, since academic journals appear to favor papers with surprising results over papers which merely confirm the expected (Mlinarić et al., 2017).

Nevertheless, the unexpected is not always favored equally across species since there are differences in how abilities are perceived among different taxa. For example, a study recently demonstrated that a tiny fish, the cleaner wrasse (*Labroides dimidiatus*) passed the mirror mark test (Kohda et al., 2019), joining an 'elite' handful of other species including chimpanzees (Gallup, 1970), dolphins (Reiss & Marino, 2001), Asian elephants (Plotnik et al., 2006) and Eurasian magpies (Prior et al., 2008). Other animals such as pigs and parrots might be suitable candidates for passing the mirror mark test, as they are able to use a mirror as visual information to find hidden items (Pepperberg et al., 1995; Broom et al., 2009). The mirror mark test involves placing a mark on an animal in a location that can only be seen in a mirror reflection. Passing the mirror mark test involves performing self-directed behaviors in the mirror (i.e., exploring areas of the body that cannot be observed without the mirror), showing interest in the mark on the body and ultimately attempting to remove the mark. The test is considered a benchmark for investigating mirror self-recognition and self-awareness. The study on cleaner wrasse was strongly criticized and triggered debate about whether researchers included robust and appropriate controls to rule-out alternative explanations for the observed behaviors (Frans de Waal, 2019; Gallup & Anderson, 2020; but see Kohda et al., 2022). Moreover, skeptics were not convinced that self-scraping behavior in fish could be considered equivalent to mark-directed self-exploration with hands or trunks in humans, apes, and elephants. Notice that the interpretation of results from mirror mark tests in other animals are also subject to wide debate, particularly about the certainty with which behavioral responses during the test can be used as evidence of self-awareness (1995; Heyes, 1994; Anderson and Gallup, 2015). While it is important that all scientific findings are met with healthy skepticism, the response to the cleaner wrasse study hints that sophisticated cognitive capacities ascribed to intuitively perceived "lower-order" species can be met with stronger skepticism.

Our method could also be prone to bias if proxies are chosen without an understanding of their ecological relevance to the taxa of interest. Suppose we conclude, for instance, that the capacity for emotional contagion is a good proxy for the presence of certain subjective experiences that we take to be relevant to welfare (Düppjan et al., 2020). This proxy might be suitable for species that live in social groups or form affiliative relationships with conspecifics because sharing social experiences is thought to facilitate emotional contagion (Herrando & Constantinides, 2021). By contrast,

emotional contagion (Adriaense et al., 2019) might be practically useless for making interspecies welfare comparisons across relatively solitary species that do not form strong social bonds with other individuals (e.g., octopuses: Schnell and Clayton, 2019; silkworms: Zhu et al., 2021). As a result, including it would heavily bias against less social species, not because we have some positive reason to think that the relevant sorts of subjective experiences are absent, but because our method of assessment is skewed toward some species relative to others. However, this could be partially circumvented by building welfare range profiles at the class- or family-level rather than species-level. This becomes relevant when there is social variation within a taxonomic group of animals. For example, there are both solitary and eusocial species across the four main bee families. There are also both solitary (i.e., octopuses) and group living species (i.e., schooling squid) within the class Cephalopoda.

Other biases when choosing relevant proxies might arise because our human perspective may render the method prone to false negatives (e.g., Ioannidis, 2005). If this method does not uncover differences in welfare ranges between certain taxa, we caution against assuming that no differences exist. Regardless of the theory of welfare used, ultimately proxies will likely be chosen with some attention to what we perceive to be relevant determinants of welfare for humans. This anthropocentrism is present throughout animal welfare science. For example, many welfare indicators are validated using humans as a form of gold standard (e.g., Mendl et al., 2022). However, such decisions about which proxies to examine may introduce unconscious biases towards or against certain options and may indeed miss entire categories of proxies relevant for detecting differences in welfare ranges between taxa. A complete view of a given taxa's welfare range is, at present, difficult, given the literature constraints and other challenges discussed in this section. As such, our method provides an approximation that should be interpreted with care.

In any theory in which valenced experiences are determinants of welfare, it is plausible that differences in the possible intensity of those experiences will matter. Unfortunately, assessing potential differences in the intensity range of valenced experiences is a difficult task. Specifically, it is notoriously difficult to establish a scale and measure the intensity of an internal state, and harder still to do so across species. For example, it might be true that, in general, members of a species show shorter latencies to move toward more desirable rewards (Davies et al., 2015). However, there may be variation within species in terms of willingness to work for a reward that does not track the intensity of internal states. Across species, any number of factors may make it difficult to use differences in latency as a proxy, including ecological role (i.e., predator or prey) and physical anatomy (i.e., appendages that facilitate swimming, walking, crawling, or flying). This is true even for some closely-related species, but it becomes more pronounced as phylogenetic distance increases (e.g., Dobromylskiy et al., 2000; Mogil, 2019; Browning, 2020; and Stasiak et al., 2003). In

these cases, the use of careful controls in experimental design is critical, for instance, comparing a baseline latency with a test latency to construct a difference score per individual (Miller et al., 2022a). While there is little question about intensity of valenced experiences being a determinant of welfare, and intensity range being a factor that influences welfare range, it will be extremely difficult to make any progress on the problem of differences in intensity range. However, this is not necessarily a problem for the methodology. Experts can simply assign very low scores to any proxy for intensity, which means that while it will be included, its impact will be significantly attenuated. That is, even if there are large differences in the empirical assessments of that proxy across species, they will have only minor impacts on the overall welfare range estimate, with small or uncertain differences being almost irrelevant.

Finally, we foresee potential challenges in reaching consensus around which proxies are most relevant and how to weigh them. Using subjective, expert judgments in the Delphi method is an accepted, robust option as described in the previous section. However, in practice, such expert judgments may cause new tensions in already often politically-fraught conversations about animal welfare (e.g., the fish pain debate, Mason and Lavery, 2022; conversations about “wicked problems”, Bolton and Von Keyserlingk, 2021). To be clear, this is not a reason not to use this method; instead, it is a call to employ the results of the method with care for context, and with attention to how they may be received by diverse stakeholders.

4 Conclusion

From a theoretical perspective, the method we propose for assessing comparative welfare ranges is an attempt to answer fundamental questions about differences in the experiential lives of nonhuman animals. From a practical perspective, the method we propose is an attempt to improve daily judgments about how to allocate and prioritize resources to relieve animal suffering. We also acknowledge that there are risks and limitations to undertaking such a project. However, interspecies welfare comparisons are important and common: they are already being made on one basis or another, primarily without empirical evidence. Our methodological framework can facilitate comparisons which are based on a transparent and empirically informed process. Ultimately, interspecies welfare comparisons can help us direct our attention to issues that will be most important for improving estimates of comparative welfare ranges and allow us to conduct sensitivity analyses to determine where additional information has the highest value relative to that end. We hope that this methodology provides a starting point for developing empirical interspecies welfare comparisons, while highlighting priorities for future research and promoting interdisciplinary collaborations to achieve this.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

BF and JS contributed to the conception of the methodology. BF wrote the first draft of the manuscript. LG, JML, MS, and AT wrote sections of the manuscript and prepared it for submission. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to thank Richard Bruns, Marcus Davis, Adam Shriver, and Michael St. Jules for their discussion of the ideas presented in this paper and our reviewers for their constructive

References

- Adriaense, J. E. C., Martin, J. S., Schiestl, M., Lamm, C., and Bugnyar, T. (2019). Negative emotional contagion and cognitive bias in common ravens (*Corvus corax*). *PNAS* 116, 11547–11552. doi: 10.1073/pnas.1817066116
- Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Søvik, E., et al. (2016). Correction: Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLoS Biol.* 14 (12), e1002589. doi: 10.1371/journal.pbio.1002589
- Allen, J., and Perry, G. C. (1975). Feather pecking and cannibalism in a caged layer flock. *Br. Poultry Sci.* 16, 441–451. doi: 10.1080/00071667508416212
- Almstedt, E., Rosén, E., Gloger, M., Stockgard, R., Hekmati, N., Koltowska, K., et al. (2022). Real-time evaluation of glioblastoma growth in patient-specific zebrafish xenografts. *Neuro-oncology* 24, 726–738. doi: 10.1093/neuonc/noab264
- Anderson, D. J., and Adolphs, R. (2014). A framework for studying emotions across species. *Cell* 157, 187–200. doi: 10.1016/j.cell.2014.03.003
- Anderson, J. R., and Gallup, G. G. (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates* 56, 317–326. doi: 10.1007/s10329-015-0488-9
- Bacqué-Cazenave, J., Cattaert, D., Delbecq, J. P., and Fossat, P. (2017). Social harassment induces anxiety-like behaviour in crayfish. *Sci. Rep.* 7, 1–7. doi: 10.1038/srep39935
- Bajer, P. G., Lim, H., Travaline, M. J., Miller, B. D., and Sorensen, P. W. (2010). Cognitive aspects of food searching behavior in free-ranging wild common carp. *Environ. Biol. Fish.* 88, 295–300. doi: 10.1007/s10641-010-9643-8
- Bateson, M., Desire, S., Gartside, S. E., and Wright, G. A. (2011). Agitated honeybees exhibit pessimistic cognitive biases. *Curr. Biol.* 21, 1070–1073. doi: 10.1016/j.cub.2011.05.017
- Baumol, W. J. (1958). The cardinal utility which is ordinal. *Econ. J.* 68, 665–672. doi: 10.2307/2227278
- Béné, C., Barange, M., Subasinghe, R., Pinstrup-Andersen, P., Merino, G., Hemre, G. I., et al. (2015). Feeding 9 billion by 2050—putting fish back on the menu. *Food Secur.* 7, 261–274. doi: 10.1007/s12571-015-0427-z
- Bessei, W. (2018). Impact of animal welfare on worldwide poultry production. *World's Poultry Sci. J.* 74, 211–224. doi: 10.1017/S0043933918000028
- Birch, J., Burn, C., Schnell, A. K., Browning, H., and Crump, A. (2021). Review of the evidence of sentience in cephalopod molluscs and decapod crustaceans. LSE consulting. LSE enterprise Ltd. *London School Econ. Politic. Sci.* Available at: <https://www.lse.ac.uk/News/News-Assets/PDFs/2021/Sentience-in-Cephalopod-Molluscs-and-Decapod-Crustaceans-Final-Report-November-2021.pdf>. (Accessed August 10, 2022).
- Birch, J., Schnell, A. K., and Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends Cogn. Sci.* 24, 789–801. doi: 10.1016/j.tics.2020.07.007
- Blokhuys, H. J., Veissier, I., Miele, M., and Jones, B. (2010). The welfare quality project and beyond: safeguarding farm animal well-being. *Acta Agricult. Scand. Sect. A.* 60, 129–140. doi: 10.1080/09064702.2010.523480
- Bolton, S. E., and Von Keyserlingk, M. A. (2021). The dispensable surplus dairy calf: Is this issue a “wicked problem” and where do we go from here? *Front. Vet. Sci.* 8. doi: 10.3389/fvets.2021.660934
- Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M., Capdeville, J., et al. (2007). Aggregation of measures to produce an overall assessment of animal welfare. part 1: A review of existing methods. *Animal* 1, 1179–1187. doi: 10.1017/S1751731107000535
- Bracke, M. B. M. (2006). Providing cross-species comparisons of animal welfare with a scientific basis. *NJAS-Wageningen J. Life Sci.* 54, 61–75. doi: 10.1016/S1573-5214(06)80004-7
- Broad, G. M. (2018). Effective animal advocacy: effective altruism, the social economy, and the animal protection movement. *Agric. Hum. Values* 35, 777–789. doi: 10.1007/s10460-018-9873-5
- Broom, D. M. (1986). Indicators of poor welfare. *Br. Vet. J.* 142, 524–526. doi: 10.1016/0007-1935(86)90109-0
- Broom, D. M. (1991). Animal welfare: concepts and measurement. *J. Anim. Sci.* 69, 4167–4175. doi: 10.2527/1991.69104167x
- Broom, D. M. (1996). Attempts to cope with the environment. *Acta Agric. Scand. Sec. A. Anim. Sci. Suppl.* 27, 22–28.
- Broom, D. M. (2019). Animal welfare complementing or conflicting with other sustainability issues. *Appl. Anim. Behav. Sci.* 219, 104829. doi: 10.1016/j.applanim.2019.06.010
- Broom, D. M., Sena, H., and Moynihan, K. L. (2009). Pigs learn what a mirror image represents and use it to obtain information. *Anim. Behav.* 78, 1037–1041. doi: 10.1016/j.anbehav.2009.07.027
- Brown, C. (2020). Convergent evolution of sentience? *Anim. Sentience* 5, 25. doi: 10.51291/2377-7478.1617
- Browning, H. (2020). If I could talk to the animals: Measuring subjective animal welfare [Doctoral dissertation, the Australian national university]. *ProQuest Dissert. Pub.* 1–212. doi: 10.25911/5f1572fb1b5be
- Bruckner, D. W. (2020). Animal welfare science, varieties of value and philosophical methodology. *Anim. Welf.* 29, 387–397. doi: 10.7120/09627286.29.4.387
- Buckland-Nicks, J. A., Gillis, M., and Reimchen, T. E. (2021). Neural network detected in a presumed vestigial trait: ultrastructure of the salmonid adipose fin. *Proc. R. Soc. B: Biol. Sci.* 279, 553–563. doi: 10.1098/rspb.2011.1009
- Budolfson, M., and Spears, D. (2019). “Quantifying animal well-being and overcoming the challenge of interspecies comparisons,” in *The routledge handbook of animal ethics*, vol. 92-101. Ed. B. Fischer (New York, United States: Routledge).
- Bullock, M. J., Boysen, S. T., and Furlong, E. E. (2008). Visual attention and its relation to knowledge states in chimpanzees, pan troglodytes. *Anim. Behaviour* 76, 1147–1155. doi: 10.1016/j.anbehav.2008.01.033
- Burden, N., Chapman, K., Sewell, F., and Robinson, V. (2015). Pioneering better science through the 3Rs: an introduction to the national centre for the replacement,

feedback. We would also like to extend our gratitude to Open Philanthropy and Rethink Priorities for facilitating and funding this work. A preprint of this article (Gaffney et al., 2022) can be found on preprints.org.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- refinement, and reduction of animals in research (NC3Rs). *J. Am. Assoc. Lab. Anim. Sci.* 54, 198–208.
- Cardoso, S. D., Faraco, C. B., de Sousa, L., and Pereira, G. D. G. (2017). History and evolution of the European legislation on welfare and protection of companion animals. *J. Vet. Behav.* 19, 64–68. doi: 10.1016/j.jvbeh.2017.01.006
- Cauchoux, M., Chow, P. K. Y., Van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., et al. (2018). The repeatability of cognitive performance: a meta-analysis. *Philos. Trans. R. Soc. B: Biol. Sci.* 373, 20170281. doi: 10.1098/rstb.2017.0281
- Cohen, A. I. (2009). Contractarianism and interspecies welfare conflicts. *Soc. Philos. Policy* 26, 227–257. doi: 10.1017/S0265052509090104
- Crump, A., Arnott, G., and Bethell, E. J. (2018). Affect-driven attention biases as animal welfare indicators: review and methods. *Animals* 8, 136. doi: 10.3390/ani8080136
- Davies, A. C., Nicol, C. J., and Radford, A. N. (2015). Effect of reward downshift on the behaviour and physiology of chickens. *Anim. Behaviour* 105, 21–28. doi: 10.1016/j.anbehav.2015.04.005
- Dawkins, M. S. (2021). *The science of animal welfare: Understanding what animals want* (Oxford, United Kingdom: Oxford University Press).
- de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59, 279–300. doi: 10.1146/annurev.psych.59.103006.093625
- de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biol.* 17, e3000112. doi: 10.1371/journal.pbio.3000112
- Dicke, U., and Roth, G. (2016). Neuronal factors determining high intelligence. *Philos. Trans. R. Soc. B: Biol. Sci.* 371, 20150180. doi: 10.1098/rstb.2015.0180
- Dobromylyskij, P., Flecknell, P. A., Lascelles, B. D., Livingston, A., Taylor, P., and Waterman-Pearson, A. (2000). “Chapter 4 – pain assessment,” in *Pain management in animals* (Philadelphia, Pennsylvania, United States: W.B. Saunders), 53–79. doi: 10.1016/B978-0-7020-1767-4.50007-2
- Dubin, A. E., and Patapoutian, A. (2010). Nociceptors: the sensors of the pain pathway. *J. Clin. Invest.* 120, 3760–3772. doi: 10.1172/JCI42843
- Duncan, I. J. (2016). Is sentience only a nonessential component of animal welfare? *Anim. Sentience* 1, 6. doi: 10.51291/2377-7478.1023
- Duncan, I. J. H., Slee, G. S., Seawright, E., and Breward, J. (1989). Behavioural consequences of partial beak amputation (beak trimming) in poultry. *Br. Poultry Sci.* 30, 479–488. doi: 10.1080/00071668908417172
- Düpján, S., Krause, A., Moscovice, L. R., and Nawroth, C. (2020). Emotional contagion and its implications for animal welfare. *CABI Rev.* 15, 1–6. doi: 10.1079/PAVSNR202015046
- Düpján, S., Ramp, C., Kanitz, E., Tuchscherer, A., and Puppe, B. (2013). A design for studies on cognitive bias in the domestic pig. *J. Vet. Behavior* 8, 485–489. doi: 10.1016/j.jvbeh.2013.05.007
- Elder, M., and Fischer, B. (2017). Focus on fish: a call to effective altruists. *Essays Philos.* 18, 107–129. doi: 10.7710/1526-0569.1567
- FAO (2021) *Crops and livestock products* (Food and Agriculture Organization of the United Nations). Available at: www.fao.org/faostat/en/#data/QCL (Accessed August 10, 2022).
- Farrar, B. G., Boeckle, M., and Clayton, N. S. (2020). Replications in comparative cognition: What should we expect and how can we improve? *Anim. Behav. Cogn.* 7, 1. doi: 10.26451/abc.07.01.02.2020
- FAWC (2009). Farm animal welfare in Great Britain: Past, present and future. *Farm Anim. Welf. Council*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/319292/Farm_Animal_Welfare_in_Great_Britain_-_Past_Present_and_Future.pdf (Accessed August 15, 2022).
- Fenwick, N., Griffin, G., and Gauthier, C. (2009). The welfare of animals used in science: How the “Three Rs” ethic guides improvements. *Can. Vet. J.* 50, 523.
- Fischer, B. (2022) *The welfare range table. effective altruism forum*. Available at: <https://forum.effectivealtruism.org/posts/tNsg607cHFLc395/the-welfare-range-table> (Accessed November 16th, 2022).
- Forstmeier, W., Wagenmakers, E.-J., and Parker, T. H. (2017). Detecting and avoiding likely false-positive findings—a practical guide. *Biol. Rev. Cambridge Philos. Soc.* 92, 1941–1968. doi: 10.1111/brv.12315
- Franks, B., Ewell, C., and Jacquet, J. (2021). Animal welfare risks of global aquaculture. *Sci. Adv.* 7, eabg0677. doi: 10.1126/sciadv.abg0677
- Fraser, D. (2008). Understanding animal welfare. *Acta Vet. Scandinavica* 50, 1–7. doi: 10.1186/1751-0147-50-S1-S1
- Gaffney, L. P., and Lavery, J. M. (2022). Research before policy: identifying gaps in salmonid welfare research that require further study to inform evidence-based aquaculture guidelines in Canada. *Front. Vet. Sci.*, 1533. doi: 10.3389/fvets.2021.7685
- Gaffney, L. P., Lavery, J. M., Schiestl, M., Trevarthen, A., Schukraft, J., Miller, R., et al. (2022). A method for improving interspecies welfare comparisons. *Preprints* 8, 2022100012. doi: 10.20944/preprints202210.0012.v1
- Gallup, G. G. Jr. (1970). Chimpanzees: Self-recognition. *Science* 167 (3914), 86–87. doi: 10.1126/science.167.3914.86
- Gallup, Jr. G. G., and Anderson, J. R. (2020). Self-recognition in animals: Where do we stand 50 years later? lessons from cleaner wrasse and other species. *Psychol. Consciousness: Theory Res. Pract.* 7, 46. doi: 10.1037/cns0000206
- Gibbons, M., Sarlak, S., and Chittka, L. (2022). Descending control of nociception in insects? *Proc. R. Soc. B* 289, 20220599. doi: 10.1098/rspb.2022.0599
- Gordon, G. G. (1970). Chimpanzees: self-recognition. *Science* 167, 86–87. doi: 10.1126/science.167.3914.86
- Hamilton, T. J., Myggland, A., Duperreault, E., May, Z., Gallup, J., Powell, R. A., et al. (2016). Episodic-like memory in zebrafish. *Anim. Cogn.* 19, 1071–1079. doi: 10.1007/s10071-016-1014-1
- Hamilton, N., Sabroe, I., and Renshaw, S. A. (2018). A method for transplantation of human HSCs into zebrafish, to replace humanised murine transplantation models. *F1000Research* 7, 1–19. doi: 10.12688/f1000research.14507.2
- Hammer, S. A., and Lee Blankenship, H. (2001). Cost comparison of marks, tags, and mark-with-tag combinations used in salmonid research. *North Am. J. Aquacult.* 63, 171–178. doi: 10.1577/1548-8454(2001)063<0171:CCOMTA>2.0.CO;2
- Healey, M. L., and Grossman, M. (2018). Cognitive and affective perspective-taking: evidence for shared and dissociable anatomical substrates. *Front. Neurol.* 9, 491. doi: 10.3389/fneur.2018.00491
- Held, S., Mendl, M., Devereux, C., and Byrne, R. W. (2000). Social tactics of pigs in a competitive foraging task: the ‘informed forager’ paradigm. *Anim. Behaviour* 59, 579–576. doi: 10.1006/anbe.1999.1322
- Held, S., Mendl, M., Devereux, C., and Byrne, R. W. (2002a). Foraging pigs alter their behaviour in response to exploitation. *Anim. Behaviour* 64, 157–165. doi: 10.1006/anbe.2002.3044
- Henderson, S. N., Barton, J. T., Wolfenden, A. D., Higgins, S. E., Higgins, J. P., Kuenzel, W. J., et al. (2009). Comparison of beak-trimming methods on early broiler breeder performance. *Poultry Sci.* 88, 57–60. doi: 10.3382/ps.2008-00104
- Herculano-Houzel, S., Catania, K., Manger, P. R., and Kaas, J. H. (2015). Mammalian brains are made of these: a dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass. *Brain Behav. Evol.* 86, 145–163. doi: 10.1159/000437413
- Herrando, C., and Constantinides, E. (2021). Emotional contagion: A brief overview and future directions. *Front. Psychol.* 12. doi: 10.3389/fpsyg.2021.712606
- Heyes, C. M. (1994). Reflections on self-recognition in primates. *Anim. Behav.* 47, 909–919. doi: 10.1006/anbe.1994.1123
- Heyes, C. M. (1995). Self-recognition in primates: further reflections create a hall of mirrors. *Anim. Behav.* 50, 1533–1542. doi: 10.1016/0003-3472(95)80009-3
- Hogg, M. A., Abrams, D., and Martin, G. N. (2010). “Social cognition and attitudes,” in *Psychology*. Eds. G. N. Martin, N. R. Carlson and W. Buskist (Harlow: Pearson Education Limited), 646–677.
- Hu, L., and Iannetti, G. D. (2019). Neural indicators of perceptual variability of pain across species. *Proc. Natl. Acad. Sci.* 116, 1782–1791. doi: 10.1073/pnas.1812499116
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2, e124. doi: 10.1371/journal.pmed.0020124
- Jiménez-Morales, N., Mendoza-Ángeles, K., Porrás-Villalobos, M., Ibarra-Coronado, E., Roldán-Roldán, G., and Hernández-Falcón, J. (2018). Who is the boss? individual recognition memory and social hierarchy formation in crayfish. *Neurobiol. Learn. Memory* 147, 79–89. doi: 10.1016/j.nlm.2017.11.017
- Keeling, L., Tunón, H., Olmos Antillón, G., Berg, C., Jones, M., Stuardo, L., et al. (2019). Animal welfare and the United Nations sustainable development goals. *Front. Vet. Sci.* 6. doi: 10.3389/fvets.2019.0033
- Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J. Y., et al. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLoS Biol.* 17, e3000021. doi: 10.1371/journal.pbio.3000021
- Kohda, M., Sogawa, S., Jordan, A. L., Kubo, N., Awata, S., Satoh, S., et al. (2022). Further evidence for the capacity of mirror self-recognition in cleaner fish and the significance of ecologically relevant marks. *PLoS Biol.* 20, e3001529. doi: 10.1371/journal.pbio.3001529
- Korte, S. M., Olivier, B., and Koolhaas, J. M. (2007). A new animal welfare concept based on allostasis. *Physiol. Behav.* 92, 422–428. doi: 10.1016/j.physbeh.2006.10.018
- Kroese, D. P., Brereton, T., Taimre, T., and Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *Wiley Interdiscip. Reviews: Comput. Stat* 6, 386–392. doi: 10.1002/wics.1314
- Kuenzel, W. J. (2007). Neurobiological basis of sensory perception: welfare implications of beak trimming. *Poultry Sci.* 86, 1273–1282. doi: 10.1093/ps/86.6.1273
- Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., and Barrett, L. F. (2016). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb. Cortex* 26, 1910–1922. doi: 10.1093/cercor/bhv001
- Linstone, H. A., and Turoff, M. (1975). “The delphi method,” in *Reading (MA: Addison-Wesley)*, 3–12.
- List, C. (2003). Are interpersonal comparisons of utility indeterminate? *Erkenntnis* 58, 229–260. doi: 10.1023/A:1022094826922
- Lonsdale, M. B., Vondell, R. M., and Ringrose, R. C. (1957). Debeaking at one day of age and the feeding of pellets to broiler chickens. *Poultry Sci.* 36, 565–571. doi: 10.3382/ps.0360565
- Lundmark, F., Berg, C., Schmid, O., Behdadi, D., and Röcklinsberg, H. (2014). Intentions and values in animal welfare legislation and standards. *J. Agric. Environ. Ethics* 27, 991–1017. doi: 10.1007/s10806-014-9512-0
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., et al. (2014). The evolution of self-control. *Proc. Natl. Acad. Sci.* 111, E2140–E2148. doi: 10.1073/pnas.1323533111

- Main, D. C. J., Kent, J. P., Wemelsfelder, F., Ofner, E., and Tuytens, F. A. M. (2003). Applications for methods of on-farm welfare assessment. *Anim. Welf.* 12, 523–528.
- Mason, G. J., and Lavery, J. M. (2022). What is it like to be a bass? red herrings, fish pain and the study of animal sentience. *Front. Vet. Sci.* 9. doi: 10.3389/fvets.2022.788289
- McCulloch, S. P. A. (2013). Critique of FAWC's five freedoms as a framework for the analysis of animal welfare. *J. Agri. Environ. Ethics* 26, 959–975. doi: 10.1007/s10806-012-9434-7
- Mendl, M., Burman, O. H. P., and Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proc. R. Soc. B.* 277, 2895–2904. doi: 10.1098/rspb.2010.0303
- Mendl, M., Neville, V., and Paul, E. S. (2022). Bridging the gap: Human emotions and animal emotions. *Affect. Sci.* 3 (4), 703–712. doi: 10.1007/s42761-022-00125-6
- Message, R., and Greenhough, B. (2019). “But it’s just a fish”: understanding the challenges of applying the 3Rs in laboratory aquariums in the UK. *Animals* 9, 1075. doi: 10.3390/ani9121075
- Miller, R., Lambert, M., Frohnwieser, A., Brecht, K., Bugnyar, T., Crampton, I., et al. (2022a). Socio-ecological correlates of neophobia in corvids. *Curr. Biol.* 32 (1), 74–85. doi: 10.1016/j.cub.2021.10.045
- Miller, R., Schiestl, M., Trevarthen, A., Gaffney, L., Lavery, J. M., Fischer, B., et al. (2022b). From pigs to silkworms: Cognition and welfare across 10 farmed taxa. *bioRxiv* 11, 516141. doi: 10.1101/2022.11.11.516141
- Mlinarić, A., Horvat, M., and Šupak Smolčić, V. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. *Biochem. med.* 27, 447–452. doi: 10.11613/BM.2017.030201
- Mogil, J. S. (2019). The translatability of pain across species. *Philos. Trans. R. Soc. B.* 374, 20190286. doi: 10.1098/rstb.2019.0286
- Nawroth, C., Ebersbach, M., and von Borell, E. (2013a). Are juvenile domestic pigs (*Sus scrofa domestica*) sensitive to the attentive states of humans? – the impact of impulsivity on choice behavior. *Behav. Processes* 96, 53–58. doi: 10.1016/j.beproc.2013.03.002
- Nawroth, C., Ebersbach, M., and von Borell, E. (2013b). Juvenile domestic pigs (*Sus scrofa domestica*) use human-given cues in an object choice task. *Anim. Cogn.* 17, 701–713. doi: 10.1007/s10071-013-0702-3
- Newberry, R. C. (1999). Exploratory behaviour of young domestic fowl. *Appl. Anim. Behav. sci.* 63, 311–321. doi: 10.1016/S0168-1591(99)00016-7
- Nicol, C. J., and Pope, S. J. (1994). Social learning in small flocks of laying hens. *Anim. Behaviour* 47, 1289–1296. doi: 10.1006/anbe.1994.1177
- Oostindjer, M., Bolhuis, J. E., Mendl, M., Held, S., van den Brand, H., and Kemp, B. (2011). Learning how to eat like a pig: effectiveness of mechanisms for vertical social learning in piglets. *Anim. Behav.* 82, 503–511. doi: 10.1016/j.anbehav.2011.05.031
- Pepperberg, I. M., Garcia, S. E., Jackson, E. C., and Marconi, S. (1995). Mirror use by African grey parrots (*Psittacus erithacus*). *J. Comp. Psychol.* 109, 182. doi: 10.1037/0735-7036.109.2.182
- Plotnik, J. M., De Waal, F. B., and Reiss, D. (2006). Self-recognition in an Asian elephant. *Proc. Natl. Acad. Sci.* 103, 17053–17057. doi: 10.1073/pnas.0608062103
- Prior, H., Schwarz, A., and Güntürkün, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): evidence of self-recognition. *PLoS Biol.* 6, e202. doi: 10.1371/journal.pbio.0060202
- Raji, J. I., and Potter, C. J. (2021). The number of neurons in drosophila and mosquito brains. *PLoS One* 16, e0250381. doi: 10.1371/journal.pone.0250381
- Reiss, D., and Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proc. Natl. Acad. Sci.* 98, 5937–5942. doi: 10.1073/pnas.101086398
- Rethink Priorities (2020) *Invertebrate sentience table*. Available at: <https://rethinkpriorities.org/invertebrate-sentience-table> (Accessed September 7, 2022).
- Robbins, J., Franks, B., and von Keyserlingk, M. A. (2018). ‘More than a feeling’: An empirical investigation of hedonistic accounts of animal welfare. *PLoS One* 13, e0193864. doi: 10.1371/journal.pone.0193864
- Roques, J. A., Abbink, W., Geurds, F., van de Vis, H., and Flik, G. (2010). Tailfin clipping, a painful procedure: studies on Nile tilapia and common carp. *Physiol. Behav.* 101, 533–540. doi: 10.1016/j.physbeh.2010.08.001
- Russell, W. M. S., and Burch, R. L. (1959). The principles of humane experimental technique. *Methuen*.
- Sandoe, P., and Jensen, K. K. (2011). “The idea of animal welfare—developments and tensions,” in *Veterinary & animal ethics: Proceedings of the first international conference on veterinary and animal ethics* (Oxford, UK: Blackwell Publishing Ltd), 19–31.
- Schaeck, M., Van den Broeck, W., Hermans, K., and Decostere, A. (2013). Fish as research tools: alternatives to *in vivo* experiments. *Altern. to Lab. Anim.* 41, 219–229. doi: 10.1177/026119291304100305
- Schnell, A. K., and Clayton, N. S. (2019). Cephalopod cognition. *Curr. Biol.* 29, R726–R732. doi: 10.1016/j.cub.2019.06.049
- Schroeder, P. G., and Sneddon, L. U. (2017). Exploring the efficacy of immersion analgesics in zebrafish using an integrative approach. *Appl. Anim. Behav. Sci.* 187, 93–102. doi: 10.1016/j.applanim.2016.12.003
- Slovan, K. A., Bouyoucos, I. A., Brooks, E. J., and Sneddon, L. U. (2019). Ethical considerations in fish research. *J. Fish Biol.* 94, 556–577. doi: 10.1111/jfb.13946
- Smith, P. L., and Little, D. R. (2018). Small is beautiful: In defense of the small-n design. *Psychon. Bull. Rev.* 6, 2083–2101. doi: 10.3758/s13423-018-1451-8
- Smith, C. L., Taylor, A., and Evans, C. S. (2011). Tactical multimodal signaling in birds: facultative variation in signal modality reveals sensitivity to social costs. *Anim. Behav.* 82, 521–527. doi: 10.1016/j.anbehav.2011.06.002
- Stasiak, K. L., Maul, D., French, E., Hellyer, P. W., and Vandewoude, S. (2003). Species-specific assessment of pain in laboratory animals. *J. Am. Assoc. Lab. Anim. Sci.* 42, 13–20.
- Sutherland, M. A., Bryer, P. J., Krebs, N., and McGlone, J. J. (2008). Tail docking in pigs: acute physiological and behavioural responses. *Animal* 2, 292–297. doi: 10.1017/S1751731107001450
- Sutherland, M. A., Davis, B. L., and McGlone, J. J. (2011). The effect of local or general anesthesia on the physiology and behavior of tail docked pigs. *Animal* 5, 1237–1246. doi: 10.1017/S175173111100019X
- Thomson, J. S., Deakin, A. G., Cossins, A. R., Spencer, J. W., Young, I. S., and Sneddon, L. U. (2020). Acute and chronic stress prevents responses to pain in zebrafish: evidence for stress-induced analgesia. *J. Exp. Biol.* 223, jeb224527. doi: 10.1242/jeb.224527
- Towner, S. (2010). Concept of mind in non-human primates. *Biosci. Horizons* 3, 96–104. doi: 10.1093/biohorizons/hzq011
- Uglen, I., Kristiansen, T. S., Mejdell, C. M., Basic, D., and Mortensen, S. (2020). Evaluation of large-scale marking methods in farmed salmonids for tracing purposes: Impact on fish welfare. *Rev. Aquacult.* 12, 600–625. doi: 10.1111/raq.12342
- Vettese, T., Franks, B., and Jacquet, J. (2020). The great fish pain debate. *Issues Sci. Technol.* 36, 49–53.
- Von Bartheld, C. S., Bahney, J., and Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* 524, 3865–3895. doi: 10.1002/cne.24040
- Webster, J. (1994). “Assessment of animal welfare: The five freedoms,” in *Animal welfare: A cool eye towards Eden* (Oxford, UK: Blackwell Science), 10–14.
- Weijers, D. (2011). *Hedonism. Internet encyclopedia of philosophy* (London, Ontario, Canada: The PhilPapers Foundation).
- Weiss, M. N., Franks, D. W., Croft, D. P., and Whitehead, H. (2019). Measuring the complexity of social associations using mixture models. *Behav. Ecol. Sociobiol.* 73, 8. doi: 10.1007/s00265-018-2603-6
- Wemelsfelder, F. (1985). “Animal boredom: is a scientific study of the subjective experiences of animals possible?,” in *Advances in animal welfare science* (Dordrecht: Springer), 115–154.
- Wong, K. (2016). Counting animals: On effective altruism and the prospect of interspecies commensurability [Doctoral dissertation, BA thesis, Princeton university, Princeton, NJ]. *Academia*. 1–67. Available at: <http://arks.princeton.edu/ark:/88435/dsp0144558g929> (Accessed August 20, 2022).
- Xu, L., Yang, X., Wu, L., Chen, X., Chen, L., and Tsai, F. S. (2019). Consumers’ willingness to pay for food with information on animal welfare, lean meat essence detection, and traceability. *Int. J. Environ. Res. Public Health* 16, 3616. doi: 10.3390/ijerph16193616
- Yeates, J. W. (2011). Is ‘a life worth living’ a concept worth having? *Anim. Welf.* 20, 397–406. doi: 10.1017/S0962728600002955
- Yeates, J. W. (2012). “Brain-pain: Do animals with higher cognitive capacities feel more pain? insights for species selection in scientific experiments,” in *Large Animals as biomedical models: Ethical, societal, legal and biological aspects*. Eds. K. Hagen, A. Schnieke and F. Thiele (Europäische Akademie), 24–46.
- Zhu, Z., Tan, Y., Xiao, S., Guan, Z., Zhao, W., Dai, Z., et al. (2021). Solitary living brings a decreased weight and an increased ability to the domestic silkworm, *bombyx mori*. *Insects* 12, 809. doi: 10.3390/insects12090809