Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 October 2022 doi:10.20944/preprints202210.0012.v1

A Method for Improving Interspecies Welfare Comparisons

Leigh P. Gaffney [1†], J. Michelle Lavery [2†], Martina Schiestl [3†], Anna Trevarthen [4†], Jason Schukraft [5], Rachael Miller [6,7], Alexandra K. Schnell [6], Bob Fischer [8,9]

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

[1] Fisheries Ecology and Marine Conservation Lab, Department of Biology, University of Victoria, Victoria, BC, Canada
[2] Campbell Centre for the Study of Animal Welfare, Department of Integrative Biology, University of Guelph, Guelph, ON, Canada
[3] Faculty for Veterinary Medicine, University of Veterinary Science, Brno, South Moravia, Czech Republic
[4] Independent researcher, Gloucestershire, UK
[5] Open Philanthropy, USA
[6] Department of Psychology, University of Cambridge, Cambridge, UK
[7] School of Life Sciences, Anglia Ruskin University, Cambridge, UK
[8] Rethink Priorities, USA
[9] Department of Philosophy, Texas State University, San Marcos, TX, USA

***Correspondence:** Leigh Gaffney; e-mail: lgaffney@uvic.ca

Abstract

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 invites 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 empirically-based 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. We begin our proposal with a theory of welfare. We operationalize that theory of welfare by identifying metrics that are defensible proxies for measuring welfare, including cognitive, affective, behavioral, and neuro-biological measures. We assign differential weights 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. Then we review the evidence and score its quality to ascertain whether a particular taxa may possess the proxies in question to construct a taxa-level welfare range profile. Finally, we use a Monte Carlo simulation to generate an overall estimate of comparative welfare range relative to our hypothetical index species - humans. Interspecies welfare comparisons will help facilitate empirically informed decision-making to streamline the allocation of resources and to ultimately better prioritize and improve animal welfare.

Keywords: animal welfare, welfare range, comparative cognition, interspecies comparisons

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 involved in intensive farm production (FAO, 2021; Franks et al., 2021). This trend has led to an increase in intensive farm production practices that significantly impact the welfare 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; Browning, 2020; Budolfson & Spears, 2019; Cohen, 2009; Kagan, 2019; McMahan, 1996; 2002; Singer, 2011; Wong, 2016). For example, to assess whether some species experience higher negative impacts from such practices than others.

There are wide ranging examples of how intensive farming can impact welfare. For instance, in intensive pork production systems, most piglets have their tails docked 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 (Sutherland et al., 2008; 2011). 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).

Beak trimming (i.e., the partial removal of the upper portion of a hen's beak) is a standard procedure performed on young hens in intensive egg production facilities. 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 & 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 salmon 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 behaviors such as swimming efficiency (Roques et al., 2010; Buckland-Nicks et al., 2012; Schroeder & Sneddon, 2017; Uglem et al., 2020; Thomson 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 & Blankenship, 2001).

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 welfare funders and affective altruism advocates, because they are largely responsible for making complex decisions about how limited resources are allocated. These stakeholders need to allocate funding in a way that maximizes welfare improvements. Other stakeholders must also be considered. For instance, funding may come in part from the general public (e.g., through charity or campaign donations, or taxes). Some members of the public wish to make informed decisions around their food and purchasing choices, particularly when deciding which causes to support. They may, for instance, choose to become vegetarian or vegan, or stop eating meat though continue to eat fish, or other variations. 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 - both for stakeholders and the general public - are invariably ad-hoc or subjective, and thus unlikely to maximize impact. Interspecies welfare comparisons can provide a pathway to develop the knowledge base to make informed decisions about which areas and which taxa to prioritize for funding.

Making interspecies welfare comparisons can have other implications, particularly in relation to identifying acceptable standards of animal welfare. Animal welfare concerns have primarily been directed at a narrow selection of species, such as pigs and chickens (e.g., Franks et al., 2021; Gaffney & Lavery, 2022). Many species used in intensive farming 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 numbers of these latter species farmed tend to amount to considerably more overall than the more ‘traditional’ ones. But such attitudes are based on arbitrary distinctions and humans tend to relate to species that are evolutionarily closer to us and often more familiar, like mammals, than those that are more distant and different, like insects. Shifting attention to include diverse species can help establish acceptable welfare standards across species and aid in determining which species, if any, should be prioritized.

Interspecies welfare comparisons can also improve best practice 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), ostensibly “less-sophisticated” animals, like zebrafish, are often used as a substitute for “more-sophisticated” animals, like mice (Hamilton et al., 2018). These decisions are based on the assumption that members of one species would be harmed less by the research than members of another. Inevitably, without interspecies welfare comparisons, such judgements lack rigor and informed decision-making.

Our objective is to outline an empirically-based methodology that aims to improve our ability to make interspecies welfare assessments. We propose investigating welfare range, which refers to how good or bad animals can fare at a time in order to maximize welfare across species.

1.2 A case for how best to make empirically-informed interspecies welfare comparisons

Following List (2003), we can distinguish between two types of interspecies welfare comparisons. The first type is the more basic: it concerns the valences of experiences - i.e., whether they are positive, negative, or neutral. Imagine, for instance, an unhealthy sow who cannot reach her piglets and a healthy chicken who's 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 better off than the sow.

The second type of interspecies welfare comparisons are level comparisons, that is, differences within a given valence. Here, things become more complicated. Imagine a pig who has just had his tail docked and a hen who hasn't eaten for, say, eight hours. Both animals are in circumstances that we would expect to cause 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, for instance, think about how we, as humans, may feel in a comparable situation, reflecting on our own experiences. However, without knowing whether or not, and if so, to what extent, other animals may 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.

An animal's welfare is an objective assessment of a subjective state (Sandøe & Jensen, 2011). However, subjective states are not directly measurable and, of course, we cannot ask animals directly how they feel; we are thus left with comparing proxies of welfare, rather than the state itself. As a result, we face difficult questions about how to validate particular proxies, a problem that 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]). Moreover, it is unclear how to 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 side-steps these problems, at least for now, by investigating animals' welfare ranges in the hope of creating a tool that could inform interspecies welfare comparisons. An animal's welfare refers to how good or bad an individual is faring (Broom, 1986); so, an animal's welfare range refers to the difference between how good or bad 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 "large" and "small" welfare range is a simplification, glossing over 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. Marian Dawkins (2021) has such a theory; she maintains 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. Hedonism is a monistic theory: it states that animal welfare is determined by the quality of their subjective experiences (Robbins et al., 2018), all and only positive experiences are good for animals, whilst all and only negative experiences are bad for them.

It would be inherently difficult to explore the possibility of differences in the welfare range for each of the many theories of welfare, so for the sake of simplicity, let's assume hedonism. Hedonism 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 significant variation in their cognitive, affective, and sensory capabilities. It seems plausible, then, that there would be significant differences in their experiential lives. Indeed, Birch et al. (2020) argue that there could be five dimensions of variation: Perceptual Richness, Evaluative Richness, Integration at a Time, Integration across Time and Self-Consciousness. They suggest that such variation may make it useful to create "consciousness profiles," where for different species different levels of consciousness are found.

If there are such differences, then it seems plausible that there could be 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 often variable in humans (Hu & Iannetti, 2019). However, there is a difference between variations in terms of 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 sometimes explicitly raised in the literature (e.g., Yeates, 2012). Due to remarkable differences in perceptual and evaluative richness across species, i.e., the number and variety of representations available to an organism, as well as the number and variety of evaluative options open to an organism, it is difficult to believe that there is not some variation in the maximum response capacity.

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

When we assess animals' welfare, we assess it relative to a species-typical neutral point. Given that neutral point, we can assess both valence and strength of valence. 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 units to assess welfare, like the duration of protective behavior, cortisol levels, time to return to normal feeding, 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, or
2. 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.

That is, we typically validate measures of welfare by making either individual-level or species-level comparisons; we assess the impacts of particular stimuli in terms of how good or bad they appear to make animals feel by comparing their response relative to another individual or species. These relative rankings are essential, as we cannot ask animals directly how good or bad they are faring. This implies, however, that when we turn to interspecies welfare comparisons, 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.

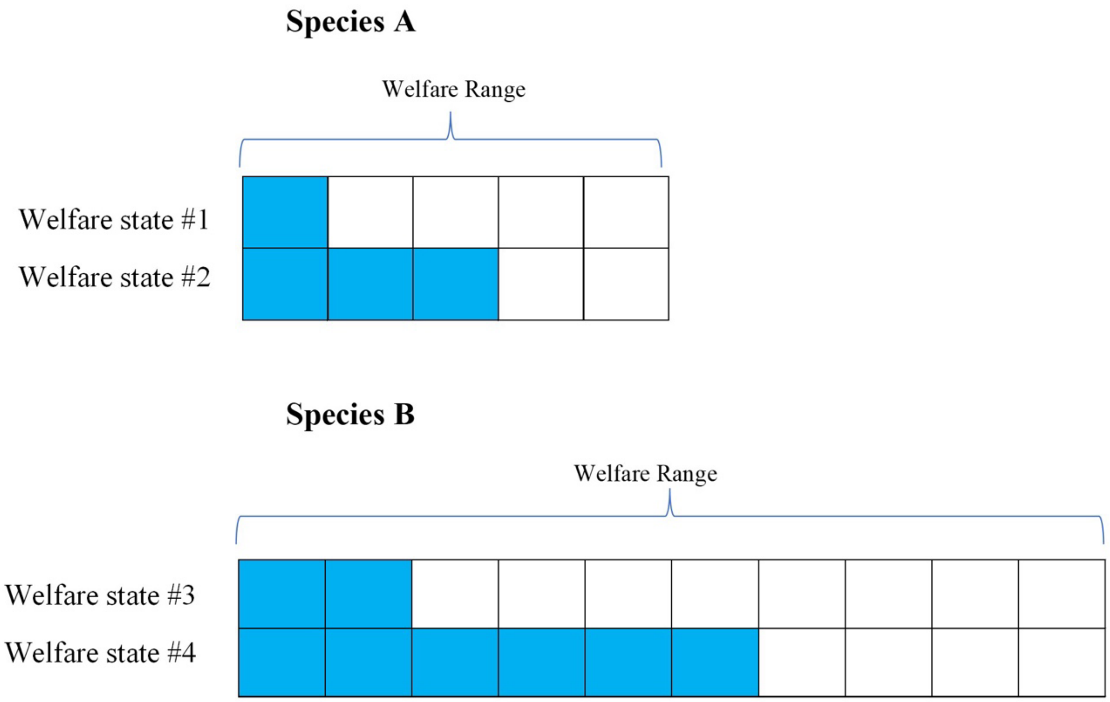

Figure 1. Theoretical figure to explain Welfare Range vs. Species-Relativized Welfare Impacts. Species A has a smaller welfare range than Species B, as represented by Species A having fewer total “welfare units” than Species B (i.e., the total number of cells per row). An ordinary welfare assessment method would compare the two welfare states (as indicated by the coloured cells) within each species, concluding that State #1 is worse than #2 for Species A and that State #3 is worse than #4 for Species B. Notably, though, such methods deliver proportional results: Welfare State #1 will seem about as bad for Species A as #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.

With this conceptual issue behind us, we turn to the main task. The usefulness of the notion 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. Notice that at the outset there is a tremendous amount of empirical uncertainty about the extent to which different animals display different welfare-relevant traits.

2. Proposed Methodology

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

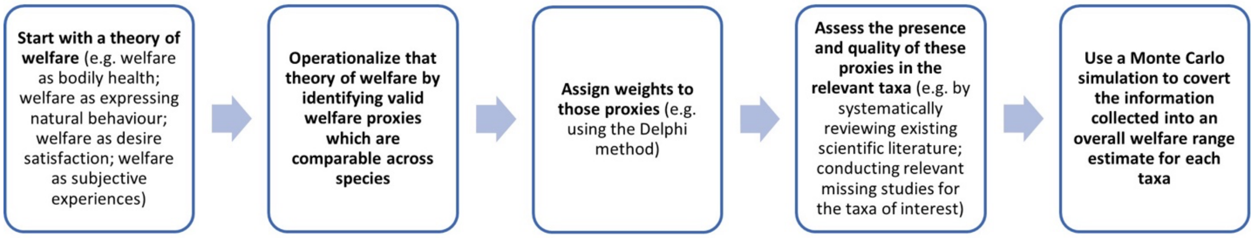

Figure 2. A summary of the proposed methodology for determining a welfare range estimate for each taxa of interest to enable interspecies welfare comparisons.

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. Here, for instance, are four theories of welfare 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 desire satisfaction: animals have positive welfare insofar as they get what they want (Dawkins, 2021).
4. Welfare as subjective experiences: animals have positive welfare insofar as they are experiencing sufficiently many positive affective states relative to negative affective states (Robbins et al., 2018).

We are not suggesting that these theories of welfare are equally plausible or that these four options represent the only possibilities available. Most obviously, these theories could be combined into pluralistic theories of some kind. The classic triadic theory discussed by Fraser (2008), for instance, proposes that welfare is jointly determined by bodily health, natural behavior, and subjective experiences. However, this is not the place to argue for a particular theory of welfare. Our goal here is simply to set out the methodology, not to defend choices within it. If, for instance, we conclude that welfare is determined by bodily health, we then turn to the task of operationalizing bodily health in ways that lend it to empirical investigation.

The *second and third* tasks involve turning the determinants of welfare enumerated during the first stage into measurable proxies. 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 & 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 & 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 *fourth* task involves assessing the evidence for these proxies in the relevant taxa. As a starting point, this task involves systematically reviewing and organizing the existing scientific literature. However, to apply this methodology in full, we would additionally 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 & 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 1). 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 both 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 - such as humans - that possesses the maximum observed value for any proxy that might matter for welfare. 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.

Table 1. Example 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 (Düpjan et al., 2013)	Likely yes (Crump et al., 2016)	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

The *fifth* task involves turning the data into overall welfare range estimates using a Monte Carlo simulation. Although other 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 run a survey of experts, using a formal, pre-registered, structured way of aggregating the results of the survey 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.

In brief, each iteration of the simulation will, given a theory of welfare and a set of welfare determinants:

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 research on particular proxies for particular species. At this juncture, we have a choice. One option is not to intervene, simply ignoring unknown values. As a result, when it comes time to do the weighted average calculation, the weight of the other proxies would be increased proportionally. 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 haven't 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.

First, 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 is 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 correlates roughly 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. The good news is that 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., 2022). 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 & Andolphs (2014) developed a framework for studying emotions across species. Such research provides reason for optimism about the potential of comparative cognition research.

That said, 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 & 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). 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 (Heyes, 1994; 1995; Anderson & 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 (Düpjan 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 & Clayton, [2019]; silkworms: Zhu et al., [2021]). As a result, including it would stack the deck heavily 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 more social 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. 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, impossible, given the literature constraints and other challenges discussed in this section. As such, our method provides only 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., Mogil, 2019; Dobromylskyj et al., 2000; 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., 2022). 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 & Lavery [2022]; conversations about "wicked problems", Bolton & 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 empirically-based interspecies welfare comparisons, while highlighting priorities for future research and promoting interdisciplinary collaborations to achieve this.

5. Acknowledgements

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. We would also like to extend our gratitude to Open Philanthropy and Rethink Priorities for facilitating and funding this work.

6. References

Adriaense JEC, Martin JS, Schiestl M, Lamm C, Bugnyar T. Negative emotional contagion and cognitive bias in common ravens (*Corvus corax*). *PNAS*. (2019) 116:11547–11552. doi: 10.1073/pnas.1817066116

Allen J, Perry GC. Feather pecking and cannibalism in a caged layer flock. *British Poultry Science*. (1975) 16:441–451. doi: 10.1080/00071667508416212

Alem S, Perry CJ, Zhu X, Loukola OJ, Ingraham T, Søvik E, Chittka L. Correction: Associative Mechanisms Allow for Social Learning and Cultural Transmission of String Pulling in an Insect. *PLoS Biology*. (2016); 14(12):e1002589.

Anderson DJ, Adolphs R. A framework for studying emotions across species. *Cell*. (2014) 157:187–200. doi: 10.1016/j.cell.2014.03.003

Anderson JR, Gallup GG. Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates*. (2015); 56(4):317-26

Bacqué-Cazenave J, Cattaert D, Delbecque JP, Fossat P. Social harassment induces anxiety-like behaviour in crayfish. Scientific reports. (2017); 7(1):1-7.

Bajer PG, Lim H, Travaline MJ, Miller BD, Sorensen PW. Cognitive aspects of food searching behavior in free-ranging wild Common Carp. *Environmental Biology of Fishes*. (2010); 88(3):295-300.

Bateson M, Desire S, Gartside SE, Wright GA. Agitated honeybees exhibit pessimistic cognitive biases. *Current biology*. (2011); 21(12):1070-3.

Béné C, Barange M, Subasinghe R, Pinstrup-Andersen P, Merino G, Hemre GI., et al. Feeding 9 billion by 2050–putting fish back on the menu. *Food Security*. (2015) 7:261–74. doi: 10.1007/s12571-015-0427-z

Birch J, Schnell AK, Clayton NS. Dimensions of animal consciousness. *Trends in Cognitive Sciences.* (2020) 24:789–801. doi: 10.1016/j.tics.2020.07.007

Birch J, Burn C, Schnell AK, Browning H, Crump A. Review of the evidence of sentience in cephalopod molluscs and decapod crustaceans. (2021) LSE Consulting. LSE Enterprise Ltd. The London School of Economics and Political Science. Available at: https://www.lse.ac.uk/News/News-Assets/PDFs/2021/Sentience-in-Cephalopod-Molluscs-and-Decapod-Crustaceans-Final-Report-November-2021.pdf

Botreau R, Bonde M, Butterworth A, Perny P, Bracke M, Capdeville J, et al. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: A review of existing methods. *Animal*. (2007) 1:1179–1187. doi: 10.1017/S1751731107000535

Bolton SE, Von Keyserlingk MA. The dispensable surplus dairy calf: Is this issue a “wicked problem” and where do we go from here?. *Frontiers in Veterinary Science*. (2021) 8: 660934. doi: 10.3389/fvets.2021.660934

Bracke MBM. Providing cross-species comparisons of animal welfare with a scientific basis. *NJAS-Wageningen Journal of Life Sciences*. (2006) *54*:61–75. doi: 10.1016/S1573-5214(06)80004-7

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

Broom DM. Animal welfare complementing or conflicting with other sustainability issues. *Applied Animal Behaviour Science*. (2019) 219:104829. doi: 10.1016/j.applanim.2019.06.010

Brown C. Convergent evolution of sentience?. *Animal Sentience*. (2020) 5:25

Browning H. If I could talk to the animals: Measuring subjective animal welfare [Doctoral dissertation, The Australian National University]. *ProQuest Dissertations Publishing*. (2020)

Bruckner DW. Animal welfare science, varieties of value and philosophical methodology. *Animal Welfare*. (2020) 29:387–397. doi: 10.7120/09627286.29.4.387

Buckland-Nicks JA, Gillis M, Reimchen TE. Neural network detected in a presumed vestigial trait: ultrastructure of the salmonid adipose fin. *Proceedings of the Royal Society B: Biological Sciences*. (2021) 279:553–563. doi: 10.1098/rspb.2011.1009

Budolfson M, Spears D. Quantifying Animal Well-Being and Overcoming the Challenge of Interspecies Comparisons. in B. Fischer (ed) *The Routledge Handbook of Animal Ethics* (pp. 92-101). (2019) Routledge

Bulloch MJ, Boysen ST, Furlong EE. Visual attention and its relation to knowledge states in chimpanzees, Pan troglodytes. *Animal Behaviour*. (2008) 76:1147–1155. doi: 10.1016/j.anbehav.2008.01.033

Burden N, Chapman K, Sewell F, Robinson V. Pioneering better science through the 3Rs: an introduction to the national centre for the replacement, refinement, and reduction of animals in research (NC3Rs). *Journal of the American Association for Laboratory Animal Science*. (2015) 54:198–208

Cauchoix M, Chow PKY, Van Horik JO, Atance CM, Barbeau EJ, Barragan-Jason G, et al. The repeatability of cognitive performance: a meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*. (2018) 373:20170281.doi: 10.1098/rstb.2017.0281

Cohen AI. Contractarianism and interspecies welfare conflicts. *Social Philosophy and Policy*. (2009) 26:227–257. doi: 10.1017/S0265052509090104

Crump A, Arnott G, Bethell EJ. Affect-driven attention biases as animal welfare indicators: review and methods. *Animals*. (2018); 8(8):136.

Davies AC, Nicol CJ, Radford A.N. Effect of reward downshift on the behaviour and physiology of chickens. *Animal Behaviour*. (2015) 105:21–28. doi: 10.1016/j.anbehav.2015.04.005

Dawkins MS. *The Science of Animal Welfare: Understanding What Animals Want.* (2021). Oxford University Press. USA

de Waal FBM. Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*. (2008) 59:279–300. doi: 10.1146/annurev.psych.59.103006.093625

de Waal FB. Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biology*. (2019) ;17(2):e3000112

Dicke U, Roth G. Neuronal factors determining high intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*. (2016) 371:20150180. doi: 10.1098/rstb.2015.0180

Dobromylskyj P, Flecknell PA, Lascelles BD, Livingston A, Taylor P, Waterman-Pearson A. Chapter 4 – Pain Assessment. *Pain Management in Animals*. (2000) 53–79. doi: 10.1016/B978-0-7020-1767-4.50007-2

Dubin AE, Patapoutian A. Nociceptors: the sensors of the pain pathway. *The Journal of clinical investigation*. (2010) 120:3760–3772

Duncan IJH, Slee GS, Seawright E, Breward J. Behavioural consequences of partial beak amputation (beak trimming) in poultry. *British Poultry Science*. (1989) *30:* 479–488. doi: 10.1080/00071668908417172

Düpjan S, Ramp C, Kanitz E, Tuchscherer A, Puppe B. A design for studies on cognitive bias in the domestic pig. *Journal of Veterinary Behavior*. (2013); 8(6):485-9.

Düpjan S, Krause A, Moscovice LR, Nawroth C. Emotional contagion and its implications for animal welfare. *CABI Reviews*. (2020) 15:1–6. doi: 10.1079/PAVSNNR202015046

Elder M, Fischer B. Focus on fish: a call to effective altruists. *Essays in Philosophy*. (2017); 18(1):107-29

FAO. *Crops and livestock products*. Food and Agriculture Organization of the United Nations. (2021). Available online at: www.fao.org/faostat/en/#data/QCL (accessed on August 10, 2022)

Farrar BG, Boeckle M, Clayton NS. Replications in comparative cognition: What should we expect and how can we improve?. *Animal behavior and cognition.* (2020) 7(1):1.

Fenwick N, Griffin G, Gauthier C. The welfare of animals used in science: How the “Three Rs” ethic guides improvements. *The Canadian Veterinary Journal*. (2009) 50(5):523

Forstmeier W, Wagenmakers E-J, Parker TH. Detecting and avoiding likely false-positive findings—A practical guide. *Biological Reviews of the Cambridge Philosophical Society*. (2017) 92:1941–1968

Franks B, Ewell C, Jacquet J. Animal welfare risks of global aquaculture. *Science Advances*. (2021) 7:eabg0677. doi: 10.1126/sciadv.abg0677

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

Gaffney LP, Lavery JM.Research before policy: identifying gaps in salmonid welfare research that require further study to inform evidence-based aquaculture guidelines in Canada. *Frontiers in Veterinary Science*. (2022) 1533. doi: 10.3389/fvets.2021.7685

Gallup Jr GG, Anderson JR. Self-recognition in animals: Where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, and Practice*. (2020); 7(1):46

Gibbons M, Sarlak S, Chittka L. Descending control of nociception in insects? *Proceedings of the royal society B*. (2022) 289:20220599. doi: 10.1098/rspb.2022.0599

Gordon GG. Chimpanzees: self-recognition. *Science*. (1970);167(3914):86-7

Hamilton TJ, Myggland A, Duperreault E, May Z, Gallup J, Powell RA, et al. Episodic-like memory in zebrafish. *Animal Cognition*. (2016) 19:1071–1079

Hamilton N, Sabroe I, Renshaw SA. A method for transplantation of human HSCs into zebrafish, to replace humanised murine transplantation models. *F1000Research*. (2018) 7. doi: 10.12688/f1000research.14507.2

Hammer SA, Lee Blankenship H. Cost comparison of marks, tags, and mark-with-tag combinations used in salmonid research. *North American Journal of Aquaculture*. (2001) 63:171–178. doi: 10.1577/1548-8454(2001)063<0171:CCOMTA>2.0.CO;2

Healey ML, & Grossman M. Cognitive and affective perspective-taking: evidence for shared and dissociable anatomical substrates. *Frontiers in neurology*. (2018) 9, 491

Herrando C, Constantinides E. Emotional Contagion: A Brief Overview and Future Directions. *Frontiers in Psychology*. (2021) 12:712606. doi: 10.3389/fpsyg.2021.712606

Held S, Mendl M, Devereux C, Byrne RW. Social tactics of pigs in a competitive foraging task: the ‘informed forager’ paradigm. *Animal Behaviour*. (2000) 59:579–576. doi: 10.1006/anbe.1999.1322

Held S, Mendl M, Devereux C, Byrne RW. Foraging pigs alter their behaviour in response to exploitation. *Animal Behaviour*. (2002a) 64:157–165. doi: 10.1006/anbe.2002.3044

Henderson SN, Barton JT, Wolfenden AD, Higgins SE, Higgins JP, Kuenzel WJ, et al. Comparison of beak-trimming methods on early broiler breeder performance. *Poultry science*. (2009) 88:57-60. doi: 10.3382/ps.2008-00104

Herculano-Houzel S, Catania K, Manger PR, Kaas JH. 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, Behavior and Evolution*. (2015) 86:145–163. doi: 10.1159/000437413

Heyes CM. Reflections on self-recognition in primates. *Animal Behaviour*. (1994); 47(4):909-19

Heyes CM. Self-recognition in primates: further reflections create a hall of mirrors. *Animal Behaviour*. (1995); 50(6):1533-42

Hu L, Iannetti GD. Neural indicators of perceptual variability of pain across species. *Proceedings of the National Academy of Sciences*. (2019) 116:1782–1791. doi: 10.1073/pnas.1812499116

Ioannidis JP. Why most published research findings are false. *PLoS medicine*. (2005) 2: e124. doi: 10.1371/journal.pmed.0020124

Jiménez-Morales N, Mendoza-Ángeles K, Porras-Villalobos M, Ibarra-Coronado E, Roldán-Roldán G, Hernández-Falcón J. Who is the boss? Individual recognition memory and social hierarchy formation in crayfish. *Neurobiology of Learning and Memory*. (2018); 147:79-89.

Kagan S. How to count animals, more or less. Oxford University Press. (2019)

Keeling L, Tunón H, Olmos Antillón G, Berg C, Jones M, Stuardo L, et al. Animal welfare and the United Nations sustainable development goals. *Frontiers in veterinary science*. (2019) 6:336. doi: 10.3389/fvets.2019.0033

Kohda M, Hotta T, Takeyama T, Awata S, Tanaka H, Asai JY, Jordan AL. If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals?. *PLoS biology*. (2019); 17(2):e3000021.

Kohda M, Sogawa S, Jordan AL, Kubo N, Awata S, Satoh S, Kobayashi T, Fujita A, Bshary R. Further evidence for the capacity of mirror self-recognition in cleaner fish and the significance of ecologically relevant marks. *PLoS biology.* (2022); 20(2):e3001529.

Kroese DP, Brereton T, Taimre T, Botev ZI. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*. (2014) 6:386–392. doi: 10.1002/wics.1314

Kuenzel WJ. Neurobiological basis of sensory perception: welfare implications of beak trimming. *Poultry Science*. (2007) 86:1273–1282. doi: 10.1093/ps/86.6.1273

Linstone HA, Turoff M. (Eds.) *The delphi method* (pp. 3-12). (1975) Reading, MA: Addison-Wesley

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

Lonsdale MB, Vondell RM, Ringrose RC. Debeaking at one day of age and the feeding of pellets to broiler chickens. *Poultry Science*. (1957) 36:565–571. doi: 10.3382/ps.0360565

Lundmark F, Berg C, Schmid O, Behdadi D, Röcklinsberg H. Intentions and values in animal welfare legislation and standards. *Journal of Agricultural and Environmental Ethics*. (2014) 27:991–1017

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, et al. The evolution of self-control. *Proceedings of the National Academy of Sciences*. (2014) 111:E2140–E2148. doi: 10.1073/pnas.1323533111

Mason GJ, Lavery JM. What Is It Like to Be a Bass? Red Herrings, Fish Pain and the Study of Animal Sentience. *Frontiers in Veterinary Science*. (2022) 9. doi: 10.3389/fvets.2022.788289

McMahan J. Cognitive disability, misfortune, and justice. *Philosophy & Public Affairs*. (1996) 25:3–35. doi: 10.1111/j.1088-4963.1996.tb00074.x

McMahan J. *The ethics of killing: Problems at the margins of life*. Oxford University Press, USA. (2002)

Mendl M, Burman OHP, Paul ES. An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the royal society B.* (2010) 277:2895–2904. doi: 10.1098/rspb.2010.0303

Mendl M, Neville V, Paul ES. Bridging the Gap: Human Emotions and Animal Emotions. *Affective Science*. (2022) 1–10

Miller R, Lambert M, Frohnwieser A, Brecht K, Bugnyar T, Crampton I, Garcia-Pelegrin E, Gould K, Greggor A, Izawa E, Kelly D, Li Z, Luo Y, Luong L, Massen J, Nieder A, Reber S, Schiestl M, Sepehri P, Stevens J, Taylor AH, Wang L, Wolff LM, Zhang Y, Clayton NS. Socio-ecological correlates of neophobia in corvids. *Current Biology.* (2022), https://doi.org/10.1016/j.cub.2021.10.045

Mlinarić A, Horvat M, Šupak Smolčić V. Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia medica*. (2017) 27:447–452. doi: 10.11613/BM.2017.030201

Mogil JS. The translatability of pain across species. *Philosophical Transactions of the Royal Society B*. (2019) 374:20190286. doi: 10.1098/rstb.2019.0286

Nawroth C, Ebersbach M, von Borell E. Are juvenile domestic pigs (Sus scrofa domesticus) sensitive to the attentive states of humans? – The impact of impulsivity on choice behavior. *Behavioral Processes*. (2013a) 96:53–58. doi: 10.1016/j.beproc.2013.03.002

Nawroth C, Ebersbach M, von Borell E. Juvenile domestic pigs (*Sus scrofa domesticus*) use human-given cues in an object choice task. *Animal Cognition*. (2013b) 17:701–713 Newberry RC. Exploratory behaviour of young domestic fowl. *Applied animal behaviour science*. (1999); 63(4):311-21

Nicol CJ, Pope SJ. Social learning in small flocks of laying hens. *Animal Behaviour.* (1994); 47(6):1289-96

Oostindjer M, Bolhuis JE, Mendl M, Held S, van den Brand H, Kemp B. Learning how to eat like a pig: effectiveness of mechanisms for vertical social learning in piglets. *Animal behaviour*. (2011); 82(3):503-11.

Plotnik JM, De Waal FB, Reiss D. Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*. (2006);103(45):17053-7

Prior H, Schwarz A, Güntürkün O. Mirror-induced behavior in the magpie (*Pica pica*): evidence of self-recognition. *PLoS biology*. (2008); 6(8):e202

Rethink Priorities. *Invertebrate Sentience Table*. (2020). Available online at: https://rethinkpriorities.org/invertebrate-sentience-table (accessed on September 7, 2022)

Raji JI, Potter CJ. The number of neurons in Drosophila and mosquito brains. *PLoS One*. (2021) 16: e0250381. doi: 10.1371/journal.pone.0250381

Reiss D, Marino L. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*. (2001); 98(10):5937-42

Robbins J, Franks B, von Keyserlingk MA. 'More than a feeling': An empirical investigation of hedonistic accounts of animal welfare. *PloS one*. (2018) 13:e0193864. doi: 10.1371/journal.pone.0193864

Roques JA, Abbink W, Geurds F, van de Vis H, Flik G. Tailfin clipping, a painful procedure: studies on Nile tilapia and common carp. *Physiology & Behavior*. (2010) 101:533–540. doi: 10.1016/j.physbeh.2010.08.001

Sandøe P, Jensen KK. The idea of animal welfare–developments and tensions. In *Veterinary & Animal Ethics: Proceedings of the First International Conference on Veterinary and Animal Ethics, September 2011* (pp. 19-31). Oxford, UK: Blackwell Publishing Ltd. (2012)

Schnell AK, Clayton NS. Cephalopod cognition. *Current Biology.* (2019) ;29(15):R726-32

Schroeder PG, Sneddon LU. Exploring the efficacy of immersion analgesics in zebrafish using an integrative approach. *Applied Animal Behaviour Science*. (2017) 187:93–102. doi: 10.1016/j.applanim.2016.12.003

Singer P. *Practical Ethics*. Third Edition, New York: Cambridge University Press. (2011).

Smith PL, Little DR. Small is beautiful: In defense of the small-N design. *Psychonomic bulletin & review*. (2018) (6):2083-101

Smith CL, Taylor A, Evans CS. Tactical multimodal signaling in birds: facultative variation in signal modality reveals sensitivity to social costs. *Animal Behaviour*. (2011) 82:521–527. doi: 10.1016/j.anbehav.2011.06.002

Stasiak KL, Maul D, French E, Hellyer PW, Vandewoude S. Species-Specific Assessment of Pain in Laboratory Animals. *Journal of the American Association for Laboratory Animal Science*. (2003) 42:13–20.

Sutherland MA, Bryer PJ, Krebs N, McGlone JJ. Tail docking in pigs: acute physiological and behavioural responses. *Animal*. (2008) 2:292–297. doi: 10.1017/S1751731107001450

Sutherland MA, Davis BL, McGlone JJ. The effect of local or general anesthesia on the physiology and behavior of tail docked pigs. *Animal*. (2011) 5:1237–1246. doi: 10.1017/S175173111100019X

Thomson JS, Deakin AG, Cossins AR, Spencer JW, Young IS, Sneddon LU. Acute and chronic stress prevents responses to pain in zebrafish: evidence for stress-induced analgesia. *Journal of Experimental Biology*. (2020) 223:jeb224527. doi: 10.1242/jeb.224527

Towner S. Concept of mind in non-human primates. *Bioscience Horizons*. (2010) 3:96–104. doi: 10.1093/biohorizons/hzq011

Uglem I, Kristiansen TS, Mejdell CM, Basic D, Mortensen S. Evaluation of large-scale marking methods in farmed salmonids for tracing purposes: Impact on fish welfare. *Reviews in Aquaculture*. (2020) 12:600–625. doi: 10.1111/raq.12342

Vettese T, Franks B, Jacquet J. The great fish pain debate. *Issues in science and technology*. (2020) 36:49–53

Von Bartheld CS, Bahney J, Herculano-Houzel S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *Journal of Comparative Neurology*. (2016) 524:3865–3895. doi: 10.1002/cne.24040

Weiss MN, Franks DW, Croft DP, Whitehead H. Measuring the complexity of social associations using mixture models. *Behavioral Ecology and Sociobiology*. (2019) 73:8

Wemelsfelder F. Animal boredom: is a scientific study of the subjective experiences of animals possible?. *In Advances in Animal Welfare Science* (1985) (pp. 115-154). Springer, Dordrecht

Wong K. Counting Animals: On Effective Altruism and the Prospect of Interspecies Commensurability [Doctoral dissertation, BA thesis, Princeton University, Princeton, NJ]. *Academia*. (2016)

Xu L, Yang X, Wu L, Chen X, Chen L, Tsai FS. Consumers' willingness to pay for food with information on animal welfare, lean meat essence detection, and traceability. *International Journal of Environmental Research and Public Health*. (2019) 16:3616. doi: 10.3390/ijerph16193616

Yeates JW. Brain-pain: Do animals with higher cognitive capacities feel more pain? Insights for species selection in scientific experiments. In K. Hagen, A. Schnieke, & F. Thiele (Eds.), *Large animals as biomedical models: Ethical, societal, legal and* *biological aspects*, 24-46. (2019) Europäische Akademie

Zhu Z, Tan Y, Xiao S, Guan Z, Zhao W, Dai Z, et al. Solitary living brings a decreased weight and an increased agility to the domestic silkworm, Bombyx mori. *Insects*. (2021) 12:809. doi: 10.3390/insects12090809