Section 1: Why Compare Welfare Across Species?

As outlined in the main summary, the core goal of the Moral Weight Project is to estimate how much welfare different animals can experience, so that we can make better decisions when prioritizing between species. This is not a hypothetical problem — it arises regularly in funding, policy, and advocacy. What this expansion adds is a clearer picture of *why* such comparisons are necessary, *how* they're often misunderstood, and *what* makes a structured approach preferable to intuition or assumption.

---

1. A Problem We Can’t Avoid

Whenever we decide how to allocate resources across species — whether to invest in broiler chicken reforms, shrimp welfare improvements, or interventions for farmed fish — we’re making tradeoffs. Often, these are made *implicitly* based on habits, heuristics, or what feels most salient or shocking.

But the absence of explicit comparison doesn’t mean comparison isn’t happening — it just means it’s happening unexamined. The Moral Weight Project aims to *replace vague intuitions with transparent, evidence-based estimates*. Even if imperfect, structured estimates are better than flying blind.

---

2. This Isn’t About Moral Status

One common misunderstanding is that estimating welfare ranges is the same as assigning *moral status*. It’s not. Moral status is about whose interests matter and how much — which varies by ethical theory. Welfare range, by contrast, is a descriptive estimate of how intense a being’s positive and negative experiences can be.

You can reject utilitarianism, or believe in rights-based or hierarchical views of moral value, and still find welfare range estimates useful. They tell you something about what’s *at stake* for each being — information that can inform many kinds of moral reasoning, even if it doesn't determine it.

---

3. Why Guesswork Isn’t Good Enough

Intuitively, many people think pigs matter more than chickens, who matter more than insects. But what drives those judgments? Often, it’s familiarity, perceived intelligence, physical

similarity to humans, or even visual appeal. These biases don't reliably track how good or bad an animal's life can be.

For example, most people would be surprised to learn that octopuses may have richer emotional lives than some vertebrates, or that bees show forms of learning and memory once thought unique to mammals. Without a more principled framework, these realities remain hidden from view — and from funding and policy decisions.

---

4. From Theory to Action

Welfare range estimates are more than theoretical tools — they're action-guiding. They can help:

- **Funders** choose between animal advocacy interventions (e.g., chicken vs shrimp campaigns)
- **Policymakers** assess tradeoffs in conservation (e.g., rodenticide harms vs predator control)
- **Researchers** make ethical decisions (e.g., using zebrafish instead of mice)

Even emerging fields like climate economics are beginning to explore how animal suffering — not just human impacts — could shape things like the *social cost of carbon*. For these purposes, some framework for interspecies comparison isn't optional. It's essential.

---

5. A Tentative but Necessary Start

The authors of the Moral Weight Project are clear: their numbers are provisional. They depend on assumptions like hedonism, valence symmetry, and comparability across species. But the alternative — acting as though all animals are equal, or that none of them matter, or that we can't even try — leads to more arbitrary, less accountable decisions.

Even if you disagree with some of the assumptions, the methodology remains valuable. The goal isn't to get it perfectly right, but to get it better than guessing — and to provide a foundation that can improve as the science advances.

Section 2: The Core Idea – Welfare Range

In the simple version of this section, we introduced welfare range as the difference between the best and worst possible experiences a being can have — a core idea in the Moral Weight Project. But that definition raises further questions: What kind of possibility are we talking about? How does this differ from similar concepts like moral status or welfare capacity? And how should we understand this idea in practice?

---

1. What Does "Can Experience" Really Mean?

The idea that welfare range refers to the difference between the best and worst states a being "can experience" might sound straightforward — but it depends heavily on what we mean by *can*.

The project considered several interpretations:

- **Logical possibility** is too broad — if something isn't contradictory, it counts. That would imply a mouse and a human could have identical ranges, which doesn't help us differentiate.

- **Physical or metaphysical possibility -** meaning what could happen under the laws of nature **-** is still vague and epistemically inaccessible.

- **Expected welfare range** (based on what we anticipate animals will experience) is useful for some decisions but hard to define rigorously.

The project settles on a more grounded interpretation:

> **Realistic biological possibility** — the difference between the best and worst conscious experiences that are *plausibly within reach*, given a being's cognitive, emotional, and physiological capacities.

This definition strikes a balance between theoretical coherence and empirical tractability. It asks: *Given what we know about the species' biology and psychology, what are the most intense experiences they are likely capable of having?*

---

2. What Welfare Range Is — And Isn't

The term "welfare range" is often confused with related ideas. Here's how they differ:

Term	What It Refers To
Welfare range	The difference between the best and worst experiences a being can plausibly have

Realized welfare	How well or badly a being is doing *right now*
Capacity for welfare	The total welfare a being can realize over its lifetime (which is proportional to its welfare range × lifespan)
Total Welfare	The cumulative welfare actually experienced by a being across time (this can be positive or negative)
Moral status	The normative importance of a being's interests (e.g. having rights or not)
Moral weight	A combination of welfare range, and ethical assumptions used in decision-making
Welfare profile	A descriptive list of cognitive, emotional, and behavioural traits (which may inform welfare range, but are not equivalent to it)

The differences between these terms are crucial. For example, a shrimp might have a low welfare range, but if you farm billions of them under poor conditions, the aggregate suffering might still be immense. Conversely, moral status theories might deny that animals' welfare *matters morally*, even if they have high welfare ranges.

---

3. Why This Matters, Even If You Disagree Ethically

Welfare range is a *descriptive concept*, not a moral conclusion. Different ethical theories will use it differently:

- **Utilitarians** might weigh interventions using expected welfare range × number of individuals.
- **Rights theorists** might care about protecting beings with certain capacities — and welfare range could help identify which capacities matter.
- **Hierarchical theorists** might argue humans matter more — but even then, knowing animals' welfare ranges can help weigh *lesser* interests or tradeoffs.

This modularity is a strength of the framework: even if you don't agree with the project's assumptions (like hedonism or valence symmetry), you can often adjust the inputs to suit your views. The concept of welfare range is not all-or-nothing — it's a flexible building block for ethical reasoning.

---

4. The Path Forward

Understanding welfare range requires humility. We can't directly measure the intensity of conscious experiences, but we can build structured estimates using neuroscience, behaviour, evolution, and proxies. Welfare range isn't a perfect measure — but it's far more principled than relying on species stereotypes or unexamined biases.

In short, the welfare range doesn't tell us *what to value*. But it helps clarify *what's at stake* — and that's essential if we want to make compassionate, informed choices.

Section 3: How the Estimates Were Made

The simple explainer outlined the broad approach used by the Moral Weight Project: identify traits that matter for welfare, score animals on those traits, and run simulations to generate estimates. This section goes deeper. It explains not just what the researchers did, but why they made the choices they did — and how those choices balance rigour with tractability.

---

1. A Four-Step Method for Estimating Welfare Ranges

The team developed a structured, modular method for estimating welfare ranges across species. It unfolds in four stages:

1.1 Specify a Theory of Welfare

They began by adopting *hedonism* — the idea that welfare consists in the balance of positive and negative conscious experiences. This makes valenced states (like pleasure and pain) the core currency of comparison.

But the method is flexible: if you prefer a different theory of welfare (like desire satisfaction or objective list views), you can re-run the process with that in mind. The same applies to moral uncertainty — you can assign credences to multiple views and combine the resulting estimates.

1.2 Identify Welfare-Relevant Proxies

Because we can't directly measure the intensity of valenced experiences across species, the project relies on proxies: observable traits that may inform us about a species' capacity to realise consciously felt positive or negative states.

But these proxies aren't assumed to be directly linked to the welfare range. Instead, they are selected based on their relevance to the functions that valenced states are believed to serve. Across competing theories of valence, these functions include:

- **Representing fitness-relevant information** (e.g. "this is bad for me"),
- **Comparing options through a common motivational currency** (e.g. weighing reward vs. risk),
- **Guiding learning and behaviour change** based on past outcomes.

Traits that support these functions may offer indirect evidence about how sophisticated or flexible an animal's valenced system is — and thus, about how broad its intensity range of experiences might plausibly be.

To capture this, the project draws from research in cognition, affective neuroscience, learning, and behaviour, grouping traits into seven functional categories:

- **Representational** (e.g., memory, mental time travel)
- **Agential** (e.g., goal-directed behaviour, inhibitory control)
- **Learning-related** (e.g., reversal learning, general intelligence)
- **Evaluative richness** (e.g., boredom, curiosity, frustration)
- **Pain-related** (e.g., hyperalgesia (extreme sensitivity to pain), response to analgesics)
- **Social** (e.g., empathy, mourning, bonding behaviour)
- **Neurophysiological** (e.g., neuron counts, brain complexity)

The model doesn't assume any one trait is decisive on its own. But each proxy is chosen because, under leading theories of how valenced states function, it provides theoretically grounded evidence about the animal's capacity to represent, regulate, or learn from affective experiences. Taken together, these traits offer a principled, multidimensional approximation of the systems that support a broad welfare range. The assumed connection is cautious — typically a weak positive correlation — but the structure allows these estimates to improve as evidence and theory advance.

1.3 Review the Literature and Score Each Species

For each proxy, the research team systematically reviewed the available literature across 11 species, including humans, pigs, chickens, carp, salmon, octopuses, crabs, bees, and black

soldier flies. Each trait was assigned a probability interval reflecting the team's confidence that the species possesses the trait in question. This allowed the model to incorporate uncertainty in a principled, quantitative way — especially useful when data was sparse or inconclusive.

The six scoring levels were as follows:

Score Label	Probability Interval	Interpretation
No	[0.00, 0.00]	Clearly absent or unknown (used as a default when data was insufficient)
Likely No	(0.00, 0.25)	Probably not present
Lean No	[0.25, 0.5)	Possibly absent
Lean Yes	[0.5, 0.75)	Possibly present
Likely Yes	[0.75, 1.0)	Probably present
Yes	[1.00, 1.00]	Clearly present

For example, if the evidence for a proxy like "reversal learning" in shrimp was suggestive but not conclusive, it might be scored as "lean yes" — meaning the model would treat it as present with a probability between 0.5 and 0.75. By contrast, if there was no relevant research on "episodic memory" in black soldier flies, it would be scored as "no," contributing a zero to that species' welfare range in that simulation. This scoring scheme allowed the simulations to reflect both what we know and what we don't. Scores were normalized to humans as the index species, who were assigned all proxies by default with certainty.

The use of binary scoring, while coarse, was necessary given inconsistent data quality and measurement methods across species. Future iterations may include more granular scales or weighted traits.

1.4 Aggregate Scores Using Monte Carlo Simulations

Once the proxy scores were gathered for each species, the team needed a way to account for all the uncertainty in the data. We don't know for sure whether a bee has episodic memory. We're not certain how important neuron count is compared to social bonding. And we can't agree on one single model of how all these traits should be combined.

To handle that, the researchers used something called a **Monte Carlo simulation** — a technique that helps you explore the full range of possible outcomes when you're dealing with a lot of unknowns.

Here's how it works:

Imagine you have a big bag of coins, and each coin represents a small decision:

- Does this species have Trait A?
- How much weight should Trait A carry?
- Should we use Model X or Model Y to combine the traits?

Instead of flipping each coin once and calling it done, the simulation flips all the coins thousands of times — trying out thousands of combinations of assumptions, data points, and weights. Each run gives one possible estimate of a species' welfare range. When you repeat that process across thousands of runs, you get a distribution: a picture of the range of plausible values, not just a single best guess.

Because humans are assumed to have every trait with full certainty, their average welfare range is set to 100%. Most other species have missing or uncertain data, so their average scores end up lower. But the model doesn't artificially cap anyone else. In a small fraction of simulation runs, species like octopuses — which score well on several key proxies — can even end up with welfare range estimates higher than humans. This doesn't mean octopuses are considered more morally important overall, but it reflects an important feature of the model: it allows for scientific uncertainty and biological variation, rather than simply reinforcing human-centric assumptions.

A welfare range estimate higher than humans is possible because, under Rethink Priorities' framework, welfare range is partly influenced by the rate of subjective experience — proxied by *flicker fusion frequency.* If an animal has a faster rate of perceptual processing than a human, the model allows that it might experience more "moments" of consciousness per unit time, which could widen its welfare range.

So instead of trying to guess one "correct" number, the project gives us a range of informed possibilities — and a way to reason clearly about where different animals might fall within it.

To account for uncertainty about how best to estimate welfare ranges, the researchers used nine different models, each based on a different set of assumptions about which traits are most relevant. These models drew from the Welfare Range Table and differed in which proxies they included and how they transformed the data. The models were:

- **Qualitative model** – counts all proxies equally
- **Qualitative-minus-social model** – same as above, but excludes proxies related to social behaviour
- **High-confidence (simple scoring)** – includes only proxies rated highly important, without weighting
- **Higher-confidence proxies (cubic model)** – includes only high-confidence proxies and cubes the welfare score
- **Cubic model** – includes all proxies and cubes the welfare score (amplifying differences between species)
- **Pleasure-and-pain-centric model** – emphasises hedonic traits like affective states and pain responses
- **Higher/lower pleasures model** – compares cognitive traits to hedonic traits (a nod to Mill's view of higher and lower pleasures)
- **Undiluted experience model** – inverts the above, focusing more on hedonic than cognitive traits
- **Neuron count model** – estimates welfare range purely based on the number of neurons in a species' brain relative to humans

Because there's no consensus on which model is "correct," the researchers combined all nine using a **mixture model**, assigning equal weight (1/9) to each. This approach spreads moral weight estimates across a range of plausible interpretations of the data and reduces overcommitment to any single view of what matters most for welfare.

These model weights are not fixed in principle — they reflect the researchers' best attempt to balance competing assumptions. Future users can explore how the estimates change if they give more or less weight to particular models.

---

2. Tradeoffs, Assumptions, and Caution

This method involves clear tradeoffs. It prioritizes:

- **Inclusivity** over premature pruning of traits
- **Transparency** over ad hoc judgments

- **Conservatism** over speculative optimism

For example, unknown traits were scored as zero, which likely underestimates the welfare ranges of less-studied species. But this was a deliberate choice to avoid overreaching in the absence of solid data. To partially counteract biases due to extremes, Rethink Priorities reported median welfare ranges (conditional on sentience) rather than means, as median values are less affected by extreme outlier runs.

Likewise, proxies were treated as independent, even though many are likely correlated. Future versions may incorporate proxy correlations or alternative aggregation strategies (e.g. power-law functions that amplify differences once certain thresholds are crossed).

---

3. A Transparent, Updateable Framework

The true strength of the methodology isn't the specific numbers it produces — it's the **framework itself**. The process is:

- **Transparent**: every assumption and weight is open to inspection.
- **Modular**: you can swap in new data, proxies, or moral views.
- **Updateable**: better science leads to better estimates, without starting from scratch.

It's not perfect. But compared to relying on raw intuition, species stereotypes, or cherry-picked metrics (like neuron counts alone), it's a major advance in bringing empirical structure to questions that have long been avoided.

Section 4: Key Findings

In the simple explainer, we introduced the welfare range estimates for various animals — values that represent how intense their best and worst possible experiences might be, compared to humans. This section digs deeper: What exactly do those numbers mean? How were they generated? And how confident should we be in them?

---

1. What the Estimates Represent

The project estimates each species' welfare range as a percentage of the human range, which is set to 100% by definition.

So, when we say pigs have a welfare range of ~50%, we're saying:

> The difference between the best and worst experiences a pig can plausibly have is about half as large as that same range for a human.

These are synchronic estimates — they describe momentary intensity, not lifetime wellbeing. To estimate capacity for welfare, you'd multiply this welfare range by lifespan. But the focus here is just: how intense can it get, right now, for this being?

Importantly, these numbers are not moral weights. They don't tell you how much a pig matters morally. They tell you what's at stake — how deep or rich the experience of life might be for them.

Species	Approx. Welfare Range (% of Human)
Humans	100% (reference species)
Pigs	~52%
Chickens	~33%
Octopuses	~21%

Carp ~9%

Bees ~7%

Salmon ~6%

Crayfish ~4%

Shrimp ~3%

Crabs ~3%

Black Soldier Flies ~1%

Silkworms ~0.2%

Some animals scored higher than many people might expect, while others were lower — but all likely still within one or two orders of magnitude of humans.

One striking finding is that these rankings were consistent across all models. Whether proxies were equally weighted, brain-focused, or simply summed, species' relative positions stayed largely the same.

---

3. Why There Are Multiple Models

Because there's no consensus on which traits best capture an animal's capacity for welfare, the researchers built nine different models to estimate welfare ranges. Each model used a different subset of traits — proxies — and combined them in different ways to reflect competing philosophical and scientific assumptions.

For example, some models focused more on cognitive abilities, while others emphasised pain and pleasure responses. One model excluded social traits; another only included traits judged to be highly relevant. An additional model estimated welfare range based purely on neuron counts.

Rather than choose a single model, the researchers created a **mixture model** that averaged across all of them, and each model contributed the same weight (1/9). This approach helps reflect deep uncertainty about how to interpret the available data.

These weights represent the researchers' best effort to balance competing views. While not adjustable by default, others are encouraged to rerun the models using different assumptions if they believe certain traits or modelling strategies are more important.

---

4. Accounting for Uncertainty

These are not crisp values — they're probability distributions generated by Monte Carlo simulations. For each species, the model runs thousands of iterations, sampling across uncertainty in trait presence, trait importance, and model choice.

Three key points on uncertainty:

- **Data gaps depress estimates.** When a proxy was marked "unknown," it was scored as 0 — a cautious choice that likely underestimates welfare ranges, especially for under-studied species like crabs or flies.

- **Sentience probability is treated separately.** Even if a species scores well on welfare-relevant traits, that only tells us what its welfare range would be *if* it's sentient. To adjust for uncertainty about sentience, the researchers assigned subjective probability distributions to each species — reflecting how confident they were that the species is conscious.

For example:

- **Pigs**: ~97%
- **Chickens**: ~90%
- **Octopuses**: ~78%
- **Bees**: ~42%
- **Fruit flies**: ~33%
- **Carp**: ~33%
- **Crayfish**: ~33%
- **Salmon**: ~33%
- **Crabs (**This same probability was used for **Shrimp**): ~31%
- **Black soldier flies**: ~22%
-

- These distributions weren't just guesses — they were based on current scientific evidence, reviewed across multiple experts. But they're still subjective, and users are encouraged to revise them if they disagree. The final welfare range estimate for each species is adjusted by multiplying the sentience-conditioned range by the probability of sentience.

  For example, if black soldier flies have a conditional welfare range of 1%, and a 35% chance of being sentient, their expected welfare range becomes just 0.35%.

- **Wider confidence intervals exist for invertebrates.** Species like crabs, flies, and bees tend to have more trait unknowns and lower confidence in sentience, which means their estimates vary more across simulations. Their low averages reflect this uncertainty — not certainty that they don't matter.

This structure lets readers be cautious without being dismissive — and update the estimates in light of new data or different beliefs.

---

5. How to Interpret the Findings

The authors emphasize that these are tentative, best-guess estimates — not final truths. But even with uncertainty, the direction of the findings is informative.

For example:

- It's unlikely that pigs can only feel 1/1000th as much as humans — even if you're conservative, the lower bound seems much higher than that.

- Likewise, even if black soldier flies have much smaller ranges and are less likely to be sentient, the scale of their use (billions per year) means their suffering could still matter significantly — if they are sentient.

So while you shouldn't treat these numbers as exact or definitive, they offer a far more principled and transparent alternative to relying on gut feeling, aesthetics, or intuition.

Section 5: Common Objections and Clarifications

The Moral Weight Project makes a bold attempt to estimate how much animals can suffer or flourish, relative to humans. Naturally, that raises questions. Below, we explore common concerns in more depth, drawing on the project's underlying methodology and philosophical foundations.

---

1. "Isn't this all just guesswork?"

It's true that we can't directly access another animal's conscious experience — but that's not unique to animals. We can't directly access other humans' experiences either, yet we still make judgments based on shared behaviour, neurobiology, and context.

What the Moral Weight Project offers is not certainty, but structured uncertainty. Rather than hand-waving or relying on intuition, it uses:

- Explicit theoretical assumptions,

- Literature-based scoring of nearly 100 traits,

- Monte Carlo simulations to incorporate uncertainty,

- And multiple models to reflect different views on what matters.

Every step is documented. You can inspect it, critique it, or adjust it. The alternative — guessing based on vibes, appearance, or cultural familiarity — is less transparent and more prone to bias.

---

2. “Aren’t humans just obviously more important?”

It’s a common intuition — most people feel that human lives matter more than animal lives. But it’s worth asking where that feeling comes from. Often, it reflects familiarity, emotional closeness, or cultural norms — not a careful comparison of what different beings can feel.

From the perspective of hedonism (the idea that welfare depends on the intensity of conscious experiences), many uniquely human traits — like our capacity for abstract thought — don’t matter unless they actually change how good or bad an experience feels. If contemplating philosophy doesn’t make your joy more intense than a pig’s contentment or a chicken’s fear, then it’s not relevant to welfare.

And in fact, the Moral Weight Project’s method is deliberately conservative:

- Humans are assumed to have *all* relevant traits with certainty — even if some animals might have capacities humans don’t.

- Unknown traits are scored as zero, which penalises under-studied species (especially invertebrates).

- Neurophysiological proxies that suggest large differences are included — even though they increase the gap between humans and others.

So while it may seem like the model is being generous to animals, it actually builds in multiple pro-human assumptions. If pigs or chickens still come out with non-trivial welfare ranges under those constraints, that suggests we shouldn’t be too quick to dismiss their moral importance.

---

3. “What about intelligence or language? Don’t those matter more?”

Cognitive traits like intelligence or language are often mistaken for morally relevant features. But from a welfare perspective, what matters is not how smart an animal is — but how deeply it can feel.

A dog doesn’t suffer less because it can’t solve equations. And most humans — infants, people with cognitive disabilities, and many adults — can’t perform abstract tasks either. Yet we still take their welfare seriously.

Some cognitive traits may correlate with welfare range (e.g. memory or future-planning could amplify suffering), but the project doesn't equate intelligence with moral value. It uses a wide range of traits — including emotional and sensory indicators — to build a much richer picture.

---

4. "What if we're wrong about animal consciousness?"

That's a real concern — and one the project takes seriously. We can't be certain which species are sentient. But rather than pretending the issue doesn't matter, the project incorporates that uncertainty directly. For each species, the welfare range is estimated conditional on sentience — and then adjusted by the researchers' best guess at the probability that the species is conscious.

For example, the model assumes a 35% chance that black soldier flies are sentient, based on current evidence. If you think that's too high, you can lower it — even drastically. The model is designed so you can plug in your own values and see how they affect the outcomes.

That said, it's hard to justify being extremely confident that certain animals aren't sentient at all — especially when the science is still developing. If there's even a modest chance they're conscious, and we can help them at scale, the case for acting is strong.

---

5. "I don't share the project's assumptions."

That's completely fine — the project is designed with that in mind. The estimates are conditional: they're based on specific assumptions like hedonism (that welfare is grounded in pleasure and suffering), unitarianism (that the same theory of welfare applies across species), and valence symmetry (that pleasure and pain matter equally if they are equally intense). If you disagree with any of those, you don't have to throw out the estimates — you can adjust them.

For example:

- If you think humans' welfare matters more per unit than animals', you can scale the animal estimates down.

- If you think welfare involves more than just pleasure and pain, you can treat the project's numbers as just one component — e.g., the *hedonic* portion of a broader theory.

The estimates are meant to be a tool, not a verdict. They give you a starting point that's explicit, transparent, and ready to be refined with your own views.

---

6. “So you’re saying a chicken is worth a third of a person?”

Not quite. The project doesn’t make any claim about how much a chicken’s life is worth relative to a human’s. It’s not estimating moral value — it’s estimating how intense a chicken’s best and worst experiences might be, compared to a human’s. Saying that chickens have a welfare range of around 33% means that the difference between their highs and lows might be about a third of the difference for humans — at any given moment.

That’s just one piece of the picture. If you want to compare individuals, you also need to factor in lifespan. For instance, if a human lives 80 years and a chicken lives 6 months, then the chicken’s total capacity for welfare would be a small fraction of a human’s — even if their moment-to-moment experiences are relatively intense.

And even if you believe that welfare range and capacity should factor into moral importance, that’s still just one input. People disagree widely about how to weigh welfare, rights, relationships, and many other values. The point of the estimates isn’t to dictate trade-offs — it’s to inform them more clearly.

---

7. “Should we stop helping humans altogether, then?”

No. This isn’t about abandoning humans — it’s about including animals in our moral reasoning. Many human-focused interventions remain cost-effective, urgent, and impactful.

But if your goal is to reduce suffering impartially, then it’s worth asking: *are there neglected areas where we could help more sentient beings for less cost?*

That’s what the project enables: comparisons. It helps you evaluate, not replace, your priorities. In practice, it supports a rebalancing of attention — not an exclusion of human causes.

That’s a fair concern. Some of the traits included in the model — like parental care, numerical cognition, or tool use — might seem only loosely connected to how much an animal can suffer or flourish. But these traits aren’t being treated as direct measures of welfare range. They’re used as proxies, drawn from theoretical accounts of how valenced experiences function.

The project works backwards from hedonism. If pain and pleasure evolved to help animals represent important information, guide decisions, or facilitate learning, then traits that reflect those capacities can give us clues about the potential range of their experiences. For

example, an animal with advanced memory or flexible learning might be able to experience more complex emotional states — even if we can't observe those states directly.

So there isn't a simple connection like "parental care means deeper suffering." Instead, the idea is that cognitive, social, and behavioural traits give us indirect insight into an animal's capacity to realise the functions associated with conscious affect. It's a long chain of reasoning — but one that's transparent, cautious, and grounded in existing science.

And because no single trait is decisive, the project includes a wide range to avoid overreliance on any one indicator. If some proxies turn out to be irrelevant, they can be dropped or down-weighted later. The method is designed to evolve — and to stay open to revision as better evidence or theories emerge.

---

9. "Don't these traits come in degrees?"

They do — and the researchers fully acknowledge that. Ideally, we'd score traits like memory, learning, or parental care on a sliding scale. But in practice, the scientific literature often doesn't provide that level of detail — especially across many species. For most proxies, the available studies only establish whether a trait is present, not how much of it an animal has.

To avoid injecting subjective judgments or arbitrary thresholds, the project used a present/absent/unknown system, attributing probabilities to the traits being present. This isn't ideal, but it's transparent and avoids smuggling in pro-human or pro-mammal bias. For example, trying to compare human and chicken parental care on a numerical scale might reflect our cultural norms more than any objective difference in experience intensity.

As more data becomes available, future versions of the model may include richer scoring systems. But for now, a cautious, binary approach helps preserve consistency and interpretability.

---

10. "I can't believe bees beat salmon!"

Some results may seem unintuitive — but that's not a reason to reject them out of hand.

Bees scored surprisingly well on certain proxies (e.g. learning, memory, social behaviour), while salmon had more "unknowns" due to data gaps. The project didn't fudge the numbers to fit expectations — it let the data speak, while noting that the estimates are conservative and provisional.

Surprising findings are a feature, not a flaw. They help identify where more research is needed, and they challenge assumptions we might never have questioned otherwise.

11. “Aren’t these conclusions a bit extreme?”

Some people worry that if we take these estimates seriously, they’ll lead to unsettling or radical conclusions — like prioritizing shrimp over humans, or focusing entirely on insect welfare. But it’s important to recognise where those conclusions actually come from.

The estimates themselves don’t say what we ought to do. They’re just conditional: *if* an animal may be sentient, *and* it may suffer to a certain degree, *then* we should take that seriously. Any strong conclusions come from combining those facts with particular moral theories — like utilitarianism, which adds impartiality, aggregation, and a focus on total welfare.

If that combination leads to uncomfortable conclusions, we should scrutinise the moral theory just as much as the empirical input. Ignoring facts because they’re inconvenient isn’t good reasoning — and it wouldn’t be acceptable in other domains (e.g. denying disease statistics because it would create moral pressure to intervene).

Ultimately, this project is just offering one part of the picture. If the implications seem demanding, the problem isn’t with measuring animals’ welfare ranges — it’s with deciding what we’re prepared to do about them.

12. “If you didn’t find many negative results, aren’t these estimates too high?”

It might seem that way — if few traits are explicitly ruled out, maybe animals got the benefit of the doubt too often. But that’s not actually how the model works. In the Moral Weight Project, both unknowns and negatives are scored the same way: as zero. So not finding evidence for a trait doesn’t help the animal — it leaves their score at the bottom either way.

That means the absence of negative results doesn’t inflate estimates. In fact, if anything, the model may underestimate animals’ capacities — because many traits are marked unknown when researchers simply haven’t looked yet. History suggests we’re often surprised by animal abilities once we start paying attention.

There’s also a deeper issue: should we penalise animals more harshly for lacking traits we haven’t studied? The researchers chose not to. If you wanted to include negative scores explicitly — say, subtracting points for confirmed absences — you’d risk implying some animals have negative welfare ranges, which doesn’t make conceptual sense.

So yes, negative results matter. But unless we also revise how we handle unknowns, incorporating more negatives might lower scores in a way that’s both theoretically murky and empirically lopsided.

13. “Why didn’t you include more moral theories or alternative aggregation models?”

Some readers may wonder why the project relies on the models that it did, when there are many other plausible ways to combine information — including theories that might produce much larger or smaller interspecies differences.

The authors acknowledge this limitation and fully agree that more models could and should be included. For example:

- **Power law aggregation** could be used to capture the idea that complex capacities build on one another non-linearly.
- A **Millian framework** could give greater weight to "higher" pleasures or cognitive sophistication.
- A **threshold model** might assume that only once a certain combination of traits is reached does high-intensity experience become possible.

Rather than picking a “correct” model, the project offers a flexible framework. The current estimates are based on a weighted mixture of 9 models as a starting point — not a final word. The authors explicitly invite others to add or replace models, test alternative assumptions, and apply their own normative frameworks to the outputs.

In short, if you think the theory behind the model is too narrow — you’re welcome to bring your own.

14. “Aren’t these scores based on adult animals, even when we’re farming the young?”

That’s true — and it’s an important limitation. For some species, especially insects like black soldier flies, animals are farmed and killed as larvae, not as adults. But most of the available scientific data — and therefore most of the trait scoring in the project — comes from adult stages.

It may be that larval animals may differ significantly from adults in their neurobiology, behaviour, and possibly their probability of sentience. For example, larval insects may lack some of the cognitive capacities that adults possess, which could affect both how likely they are to be conscious and how wide their welfare ranges might be.

Though the welfare range of silkworms - the larvae of silk moths - were estimated, the current model doesn’t attempt to score larvae and adults separately, simply because all of the relevant data doesn’t yet exist. But the authors flag this as a priority for future research. Ideally, welfare estimates should reflect the actual life stage at which animals are used — especially when that stage is shorter, simpler, or less likely to support sentient experience.

Section 6: Why This Matters for Advocacy

Welfare range estimates were developed to help animal advocates and funders make clearer, more principled decisions — not just about *whether* to help animals, but *which* animals to help, *how* to help them, and *how much* it matters. They offer a common language for comparing interventions that affect vastly different species, with vastly different capacities for experience.

Suppose you're choosing between a fish stunning campaign and a shrimp stocking reduction. Fish may have higher welfare ranges than shrimp, but shrimp are farmed in far greater numbers. Welfare range estimates give you a structured way to weigh that trade-off, rather than relying on gut instinct or familiarity.

Even if you're unsure about the sentience of certain species, the model accommodates that too: each estimate is multiplied by a probability of sentience, meaning low-confidence species like black soldier flies aren't ignored — they're just downweighted. This allows for moral precaution without overconfidence.

Welfare ranges also plug directly into cost-effectiveness calculations. If you know:

- How many animals are affected
- How long they're affected
- What their welfare range (conditional on sentience) might be
- How confident we are that they're sentient

…then you can start comparing different interventions using the same kinds of tools used for human-focused global health and development work.

How It's Being Used

The Moral Weight Project wasn't just designed as a philosophical exercise — it's already shaping how real-world decisions are made.

Animal Charity Evaluators (ACE) now incorporates MWP welfare ranges into its evaluation framework. In 2023, ACE revised how it compares the impact of helping different species by creating composite "welfare range scores" for each animal group. These scores combine MWP data (adjusted for ACE's own framework) with egalitarian baselines and staff estimates, and are now used to assess the *scale* of suffering addressed by each intervention. This, in turn, informs ACE's charity evaluations and cause prioritisation.

Open Philanthropy, one of the largest EA-aligned funders, also engaged early with the MWP, providing funding on the grounds that it could help "compare future opportunities within farm animal welfare [and] prioritize across causes." Program Officer Amanda

Hungerford has publicly stated that the research changed her views — she once didn't give insects much thought, but after reading the MWP work, she now sees them as morally important. This kind of shift could shape funding decisions going forward, especially for historically neglected taxa.

Rethink Priorities also reports early interest from **government agencies**. Their 2024 impact update mentions outreach from public bodies in the United States and the Netherlands about how welfare range estimates might inform public policy — for example, through animal-inclusive cost–benefit analysis. These applications are still emerging, but suggest early traction beyond the EA space.

The MWP has also been presented at institutions such as the **Stanford Humane & Sustainable Food Lab** and the **Shrimp Welfare Project**, indicating growing interest among researchers and advocates in aquatic animal welfare, alternative proteins, and high-impact campaign strategy.

Additionally, the project has earned the endorsement of Peter Singer, one of the most influential figures in animal ethics. Singer wrote that "never, in the fifty years I have been writing about ethics and animals, have I seen a project as philosophically and empirically daring as [Rethink Priorities'] attempt to develop a method for comparing welfare across species." That kind of praise not only legitimizes the work but signals to others that it deserves to be taken seriously.

A Starting Point, Not a Final Answer

The MWP isn't a set of final truths. Its welfare ranges are explicitly conditional, based on clearly stated assumptions. But this transparency is part of the value. You don't have to agree with the inputs — you can adjust them. If you think insects are less likely to be sentient, lower their weights. If you value only suffering, not pleasure, tweak the outcome range. If you hold different ethical theories, you can plug those in too.

What the project offers is a structure: a way to reason through hard moral trade-offs without relying solely on intuition. And for those who are already reasoning quantitatively — whether in grantmaking, policy, or strategy — the MWP provides a ready-made dataset and framework to plug in.

It's early days, but the estimates are already informing decisions across multiple domains: from evaluating charities, to shifting funder priorities, to developing new policy models. And as attention turns to neglected groups like fish and insects — whose moral weight was long assumed to be negligible — the MWP helps make those conversations more serious, more structured, and more evidence-based.

Section 7: Final Takeaways

The Moral Weight Project doesn't offer certainty. It doesn't claim to know exactly how much a crab can suffer, or whether a bee's joy is more muted than a chicken's pain. What it offers is something both more modest and more useful: a structured way to think about these questions, grounded in evidence and open to revision.

These numbers aren't final — but they're not arbitrary either

The estimates presented here are **conditional**: they depend on specific assumptions, limited evidence, and cautious scoring. But they're not pulled from thin air. They're based on a transparent method, using publicly available data, and tested under multiple models.

You can disagree with the moral framework. You can question the proxy list. You can adjust the model weights or input your own credences about sentience. That's the point: the estimates are a **starting point** for reasoning, not a final conclusion.

Uncertainty doesn't make the work irrelevant — it makes it necessary

If we knew exactly how much animals could suffer, decision-making would be easy. But we don't — and yet we still have to act. Billions of animals are being raised, fished, farmed, or experimented on right now. Policies are being made. Resources are being spent. Trade-offs are happening whether we acknowledge them or not.

So instead of pretending we know nothing — or pretending we know everything — the Moral Weight Project offers a **middle path**: cautious, structured reasoning that improves over time.

The estimates are surprisingly robust — even under conservative assumptions

The researchers emphasise their own uncertainty. They're clear that more work is needed on trait correlations, developmental stages, and theoretical models. But even with all those caveats, it's hard to believe that many animals have welfare ranges less than 1/100th of a human's — and 1/1000th seems wildly implausible.

That matters. It means that even if animals suffer less than we do, they likely suffer **a lot more** than most people assume — and that deserves moral attention.

Use this as a tool — not a doctrine

You don't have to believe the numbers are perfect. You don't have to accept every proxy or model. But if you think animal welfare matters, this framework helps you **reason about it more clearly** — and avoid being swayed by familiarity, appearance, or gut feeling.

The estimates can guide funding decisions, intervention design, research priorities, or cause selection. Or they can just serve as a reminder: animals are likely capable of deep experiences, and their welfare deserves careful thought.

The real takeaway: we need to take animals seriously

You don't need to believe a chicken is "one-third of a person." You just need to accept that chickens, pigs, fish, insects — all likely experience the world in ways that can go very badly or very well. And right now, for most of them, it's going very badly.

The precise numbers aren't the point. The point is that, given that they likely have non-negligible welfare ranges, their experiences matter — and that we can act on that, even while we're still learning.