Wild Animal Suffering
Mike Johnson
Braindump for Evan's Wild Animal Welfare Project Discussion FB group (may polish & crosspost on EA forum if positive response)

The following are my personal intuitions as to how one might go about quantifying wild-animal suffering. Epistemic status: exploratory; highly incomplete and at points very speculative.

I. WAS is a plausible problem

As David Pearce, Brian Tomasik, and others have argued, wild animal suffering is probably very large; given
(1) the sheer number of wild animals (and insects);
(2) the affective carrot-and-stick balance that nature seems to favor; and
(3) the harshly unmerciful 'state of nature' of the average animal life,
it seems broadly plausible (though not universally accepted- see Plant 2016; Yudkowsky 2014) that wild animals experience net suffering and that the majority of suffering on earth is wild-animal suffering.

Helping alleviate some of this— or at minimum, better *understanding* it— seems important.

II. There is broad object-level agreement on when humans are suffering

With humans, we can simply ask people, "are you suffering?" — and they will give an answer. Mostly, there's consensus that these answers are pretty accurate.

III. There is little meta-level agreement on what suffering is

Essentially, the metaphysics of sentience/suffering/moral patienthood are very murky, and there's substantial divergence on predictions about edge-cases.

A first pass at parametrization: on one hand, we have qualia realists/formalists (e.g., QRI, David Pearce, Max Tegmark); on the other, we have functionalists (e.g., Brian Tomasik, most of FRI, Eliezer Yudkowsky?). The realists believe qualia (such as suffering) are real and objective, in the same way an electron is real and objective; the functionalists believe defining suffering is definitional/relational, in the same way defining chairs is definitional/relational ('words such as suffering are nodes in the network of language and get their meaning from it; change the network, change the meaning of the node'). These camps can be further subdivided: QRI thinks the Symmetry Theory of Valence is plausible, whereas David is skeptical; Brian thinks there's plausibly a gradient of sentience & moral patienthood even down to the level of fundamental physics, whereas Eliezer thinks there's a hard threshold somewhere between non-human primates and humans.

IV. It's unclear how to adjudicate disagreements within and between theories of sentience

In science, theories are judged by (1) their predictive power, and (2) their elegance/parsimony. It's unclear if we can use the first principle (predictive power) in the context of ethics, since there are deep disagreements as to whether there are objective moral facts to predict (see: the view from formalism; the view from functionalism). The second principle (elegance/parsimony) is slightly more promising, but different metaphysics (and even different factions within a given metaphysics) seem to have substantially different intuitions as to what constitutes parsimony.

In light of these difficulties, philosophers often fall back on common sense to adjudicate disagreements about ethics & moral patienthood— but as Eric Schwitzgebel notes, "Common sense is incoherent in matters of metaphysics. There's no way to develop an ambitious, broad-ranging, self-consistent metaphysical system without doing serious violence to common sense somewhere. It's just impossible. Since common sense is an inconsistent system, you can't respect it all. Every metaphysician will have to violate it somewhere."

It would greatly help WAS research if we could clarify exactly what's happening when people holding different positions talk about sentience. I.e., a lot of confusion happens because people can't see each others' ontological commitments, definitions of evidence, and definitions of elegance/parsimony. This is true in general, but *especially* true when talking about consciousness/sentience.

V. There's good existing literature on WAS

Brian Tomasik has written extensively on WAS: see e.g., The Importance of Wild-Animal Suffering; Should We Intervene in Nature?; Intention-Based Moral Reactions Distort Intuitions about Wild Animals. His provisional conclusion is that WAS is a real problem and is sensitive to human actions (e.g., pesticides, habitat destruction), and thus could be a plausible intervention point for EAs.

Other notable landmarks include

- Luke Muehlhauser (2017). 2017 Report on Consciousness and Moral Patienthood for OpenPhil, which is broadly consistent with Brian's approach (Muehlhauser assumes physicalism+functionalism+illusionism+fuzziness);
- David Pearce's notes on more speculative interventions such as reprogramming predators;
- Thomas Metzinger (2017). Suffering, in which he identifies some desiderata for a theory of moral patienthood and what might count as suffering;
- Georgia Ray (2017). Which Invertebrate Species Feel Pain?
- Robert Jones (2014). Can They Suffer? Pain in Insects, Spiders and Crustaceans

(Thanks to Evan Gaensbauer and the Wild Animal Welfare Project Discussion facebook group for many of these links.)

VI. Moving forward

To date, most evaluations of whether and how much various animals can suffer have been based on their *behavioral and anatomical similarity to humans*. E.g., Georgia Ray notes that "Prawns will groom antenna that are crushed or exposed to noxious chemicals"; Robert Jones argues lobsters should be considered moral patients since "crustaceans possess nociceptors, ganglia (nerve cell clusters associated with sensing pain), and nociceptor-to-ganglia pathways." (Based on these sorts of heuristics, Luke offered the following probabilities of morally-relevant consciousness in various species.)

Now, it seems intuitive that behavioral and anatomical similarity could be a good proxy measure for sentience. But it's unclear *how* good this is as a proxy measure, and under *which conditions* this correlation will *break down*. And I'd suggest that behavioral & anatomical comparisons often tend to be haphazard, involving cherrypicked criteria based on subjective intuition of similarity, and are at risk of inadvertently 'essentializing' the wrong things. So— what we have is good and let's not throw it out, but there's a limit to how far we can push it; let's also look for more angles.

The following are my personal intuitions on a possible path for expanding and formalizing research into WAS:

1. Distinguish sentience vs suffering. First, I think it makes sense to explicitly distinguish sentience (degree of consciousness / moral patienthood) from valence (suffering).
2. Focus on valence/suffering instead of sentience. Most of the WAS literature is about sentience, trying to clarify which animals might count as moral patients. There's relatively less research on trying to quantify *how happy* an organism like a lobster, fish, or chipmunk is, under various conditions. The former task is important, but I think this latter task could be *more important*, *more neglected* and *easier* than the former.
3. Examine evolutionary ecology and extract general patterns. There seem to be general ecological patterns which might bias classes of organisms toward positive or negative valence. E.g., Brian has noted (somewhere!) that prey need to be relatively more hypervigilant about threats than predators need to be hypervigilant about hunting opportunities, which may bias prey animals' valence downward relative to predators. Similarly, Dawkins has noted that evolution has reasons to limit suffering insofar as it would impair future adaptive behavior, but this no longer applies when outside of the 'adaptive window' - e.g. being eaten or parasitized.
4. Keep up with affective neuroscience. I'm an enormous fan of Morten Kringelbach & Kent Berridge's work in affective neuroscience, and I think their work on the interplay between neuroanatomy & neuroendocrinology (e.g., their model of the "affective keyboard") could form the basis for nuanced proxies for suffering, & general models for how various internal dynamics might bias certain organisms toward positive or negative affect. I also think there are some EA scientists and bloggers working on adjacent topics that would be worth speaking with— e.g., a discussion about predictive coding & suffering with

Adam Safron or Scott Alexander could be very generative. Speaking as a philosopher, I'd suggest seeking out scientists instead of philosophers to talk about this with.

5. Explore novel affective metrics. QRI is working on quantifying valence from fMRI data using Selen Atasoy's "connectome-specific harmonic wave" paradigm. Basically, this consists of roughly mapping someone's connectome, then calculating the natural resonances ('brainwave eigenvalues') of it, using this to transform brain activity into a Fourier series, then evaluating its 'consonance/dissonance/noise signature' (CDNS). More dissonant brain states should involve more suffering. Importantly, if this 'first-principles' approach works for people, it should work for animals too. (Sidenote: I'd love to see somebody attempt to 'out-QRI QRI', to try to one-up what we're doing, to figure out a better way of measuring valence from first principles. Research competition is healthy!)

VII. Endnotes

I think this is a really cool cause area. There are much better literature reviews out there (I like the google doc that's forming-- my apologies to the people I didn't cite) and of course people like Brian know this field a heck of a lot better than I do (hopefully I didn't misquote you too badly!). And thanks Evan for getting things going.