RETHINK
PRIORITIES

What Neuron Counts Can and Can’t Tell Us about Moral Weight

Adam Shriver, adam@rethinkpriorities.org

Key Takeaways

- Academic research has explored the relationship between the number of neurons different organisms possess and the cognitive abilities of those organisms. Several authors influential in the EA community have endorsed using neuron counts as a rough proxy that can help determine moral priorities.

- Some advantages of using neuron counts as a proxy for moral weight are: (1) neuron counts are a clear improvement over ignoring the interest of animals during the aggregation of welfare (2) they are quantifiable, (3) they are in-principle measurable, (4) they correlate to some extent with cognitive abilities that are plausibly relevant for moral standing.

- Neuron counts and related metrics roughly track people’s intuitions about the intelligence of different species (though for any easily measurable metric there are also examples where intuitions do not align with the suggested rankings). Given the rough correlation between neuron counts and intuitions about the moral status of other species, it might be thought that neuron counts can provide useful information in cases where we do not have clear intuitions, might update our views in cases where certain species turn out to have larger or smaller numbers of neurons than expected, and might provide partial support for our intuitions.

- The raw number of neurons does not fully determine the brain’s processing power. Other features, such as the firing rate of neurons, the number and properties of connections between neurons, and the spatial density of neurons also influence how much information can be processed in a given amount of time. Additionally, there are other forms of information processing in brains that do not depend on the firing of action potentials in neurons. Many neurons are also specialized and do not contribute to general processing capacity but rather to specific functions that are not welfare relevant (for example: larger bodies tend to require greater numbers of neurons regardless of intelligence for basic sensory and motor functioning). As such, the term “neuron count” is an imprecise indicator of the brain’s overall processing capacity.

- There is no straightforward empirical evidence indicating that relative differences in neuron counts within species changes functional capacities. In general, it is often the case that the pattern of activation in particular brain regions matters more for determining behavior than the raw number of neurons involved. However, differences between members of the same species are generally small compared to differences between species, and it certainly is the case that every cognitive capacity in biological organisms requires at least some minimal number of neurons.

- There are surprisingly few explicit arguments spelling out precise reasoning that explains why more processing power would lead to greater moral weight. Many ways of trying to make this connection seem incompatible with how the brain is likely to work. The strongest candidates for endorsing the neuron count approach are: (1) the belief that neuron counts are correlated with intelligence and intelligence is correlated with moral weight, (2) the idea that neuron counts might increase the number of subsystems in the brain which themselves each carry moral weight, (3) the idea that additional neurons result in “more consciousness” or “more valenced consciousness” and (4) the idea that increasing numbers of neurons are required to reach thresholds of minimal information capacity required for morally relevant cognitive abilities

- In regards to intelligence, we might question **both** the extent to which more neurons are correlated with intelligence and whether more intelligence truly suggests greater moral weight. In regards to the former, there certainly is some correlation between neuron counts and intelligence in biological organisms (after all, minimal information processing capacity is required for anything brains do), but there have been many reasons put forward questioning the extent to which large numbers of neurons are required for sophisticated cognitive abilities (given that organisms with relatively small numbers of neurons, such as bees, often demonstrate surprisingly sophisticated abilities). In general, there is great uncertainty about the degree to which neuron counts are correlated with intelligence, though how strong we view this correlation will depend on how intelligence is defined and what is thought to count as a “strong” correlation. In regards to the connection between intelligence and moral standing, there also is much uncertainty about the extent to which intelligence matters for moral weight.

- In the case of conscious subsystems in the brain which themselves carry moral weight, the argument made generally is not that it is likely that there are conscious subsystems, but rather that even with a relatively small probability of being true this consideration is likely to swamp other considerations. However, it is not clear that this “low probability but high expected value” proposition is not balanced out by other “low probability but high expected value” propositions that push in the opposite direction. Many of the arguments in favor of conscious subsystems conflict with the widespread practice in the science of consciousness of assuming that self-report can differentiate conscious from non-conscious states.

- In regards to the claim that more neurons result in more valenced experience, some ways of understanding this claim seem to be incompatible with current knowledge of how the brain works, given that we can add neurons without changing central affective processes and functional outputs. The claim also does not seem to follow directly from most leading theories of consciousness, which are silent on the relationship between consciousness and neuron number.

- Certain degrees of information processing capacity are required for various cognitive functions. If those cognitive functions are relevant for moral weight, this provides a connection between neuron counts and moral weight. However, if we think the cognitive functions are what matters, it is not clear what advantage neuron counts provide above and beyond simply looking for the presence of the capacities of interest.

- All in all, using neuron counts individually as a proxy for moral weight would be an improvement over many current methods, but also is likely to be misleading in a way that supports anthropocentric biases. Neuron counts do seem to provide some useful information, but since other indicators also seem relevant for moral weight, neuron counts would best be used as one part of a combined metric that included other information for the determination of overall moral weight.

Introduction

Can the number of neurons an organism possesses, or some related measure, be used as a proxy for assigning moral weight to that organism? Over the past few decades, there has been a large scholarly literature examining whether various measures of “brain size” can be predictive of purported measures of intelligence across different species (Reader and Laland 2002, Burkart et al 2017). More recently, several authors have endorsed the idea of using neuron counts in a manner that could help us to assign weights in moral decisions that involve aggregating the welfare of members of different species. For example, Budolfson and Spears write that neuron counts might be used as a proxy, “for estimating well-being potentials across species, analogous to the use of consumption as the basis for estimating well-being across humans“ (Budolfson and Spears 2020).

The idea also is reasonably well-represented in the effective altruism and rationalist communities. In a draft of his book *What We Owe The Future*, William MacAskill writes, “As a very rough heuristic, we could weigh animals’ interests by the number of neurons they have” and suggests, as a point in favor of the view, that “It’s not clear to me whether weighting by neuron count overweights or underweights human experience.” This approach leads to the tentative conclusion that, “if we allow neuron count as a rough proxy, we get the conclusion that the total weighted interests of farmed animals are fairly small compared to that of humans, though their wellbeing is decisively negative” (forthcoming, p. 15).

And Scott Alexander of Slate Star Codex has also illustrated how weighting by neuron count might work:

> “Might cows be "more conscious" in a way that makes their suffering matter more than chickens? Hard to tell. But if we expect this to scale with neuron number, we find cows have 6x as many cortical neurons as chickens, and most people think of them as about 10x more morally valuable. If we massively round up and think of a cow as morally equivalent to 20 chickens, switching from an all-chicken diet to an all-beef diet saves 60 chicken-equivalents per year.” (2021)[1]

This above reasoning is meant to be applied to comparisons between any organisms, including mammals, birds, fish, and insects. And in principle, it might be used as a way of evaluating the moral weight of artificial intelligence, if neuron count is meant to be simply a representation of information-processing capacity. However, the comparison with AI is beyond the scope of this report.

A few points of clarification are worth noting before going further: none of the above authors suggest that there is a *direct* relationship between neuron counts and moral weight. Rather, they suggest that neuron counts might be useful proxies that can be used in decisions that involve aggregating welfare across large numbers of members of different species. Some of them suggest that neuron counts might be better thought of as temporary stand-ins until more precise metrics are available. And some suggest that neuron counts are only useful proxies in very particular situations.

Moreover, though the authors use different terminology about what exactly is being correlated with neuron counts (“capacity for welfare” vs “suffering” vs “the weighing of interests”), the upshot of each claim can be cashed out roughly as follows: assuming a broadly consequentialist ethical perspective where certain states of the world can contain either positive or negative value, the weighting refers to the relative contribution a given being’s states make towards creating positive or negative value. As such, the idea is that as neuron counts increase, this increase roughly correlates to an increase in the amount of value (or disvalue) that results from

[1] It’s worth noting that Alexander’s estimate for cows is not based on a direct study of them, but rather on measurements of neurons in “cow-sized ruminants.” Regardless of whether this is a legitimate way to estimate the number of neurons an average cow possesses, the reasoning is clear enough for purposes of illustration.

an organism's experiences. And greater value leads to greater priority in consequentialist decision-making, or other frameworks that hold the consequences are relevant for moral decisions.

This report examines various possible reasons for treating neuron counts as proxies for moral weight. The general conclusion is that arguments attempting to link neuron counts to moral weight, even loosely, are speculative and do not reliably track other proposed indicators of moral status. However, these arguments cannot be dismissed entirely. The end of the paper examines which types of decisions might be good candidates for using neuron counts and proposes that neuron counts could be included as part of an assessment of moral weight but should not be relied upon as the sole indicator.

Note: other parts of the Moral Weight Project are focused on estimating welfare ranges. This report is on moral weight more broadly, rather than welfare ranges specifically. Of course, one way that neuron counts could be significant is if they prove to be good proxies for welfare ranges, but even if not, they could prove to be good proxies for something else that matters morally. So, we take a more inclusive approach here, investigating a range of ways that neuron counts could matter.

How might neuron counts be used

For the purposes of aggregating and comparing welfare across species, neuron counts are proposed as multipliers for cross-species comparisons of welfare. In general, the idea goes, as the number of neurons an organism possesses increases, so too does some morally relevant property related to the organism's welfare. Generally, the morally relevant properties are assumed to increase linearly with an increase in neurons, though other scaling functions are possible.

So, for example, say you need to decide whether it is better to choose an option that will increase ten humans' welfare from six on a scale from 1 to 10 to eight on a scale from 1 to 10, or an option that will increase five thousand chickens' welfare from six on a scale of 1 to 10 to eight on a scale of 1 to 10. Do you treat each unit of increased welfare as equally valuable across species? Using a neuron count multiplier, we could say that the average number of neurons in a human (86,000,000,000) is 390 times greater than the average number of neurons in a chicken (220,000,000) so we would treat the welfare units of humans as 390 times more valuable. Or, alternatively, we could say that we could treat the welfare units of chickens as worth 0.002564 of a unit of human welfare.

So, for the above welfare comparisons, we get the following:

Humans: 10 humans X 2 units of welfare = 20 additional units of welfare

Chickens: 5000 chickens X 2 units of welfare X 0.002564 = 12.82 additional units of welfare

So, according to this logic, the option with the human welfare intervention would be preferable, despite the large difference in the number of individuals affected. This is one simplified example of how neuron counts could be used to make comparisons across species.

Who has suggested this and why

What is the proposed value of using neuron counts as a proxy for moral weight? Many decisions, particularly those in the effective altruism movement, depend upon our ability to accurately aggregate the moral value or disvalue that results from changing the welfare of large numbers of individuals across different species. To take one example, it has previously been suggested that it is better to eat beef than to eat chicken, for the following reason:

> The average cow is very big and makes 405,000 calories of beef; the average chicken is very small and makes 3000 calories worth of chicken. So each year (assuming equal calorie consumption of each), I kill about 0.3 cows and about 42 chickens, for a total of 42.3 animals killed….Suppose that I stop eating chicken and switch entirely to beef. Now I am killing about 0.6 cows and 0 chickens, for a total of 0.6 animals killed. By this step alone, I have decreased the number of animals I am killing from 42.3/year to 0.6/year, a 98% improvement. (Alexander 2021)

If, hypothetically, it turned out that cows have 10 times more neurons than chicken, this equation changes. In this particular case, it doesn't change enough to suggest that purely from an animal welfare perspective it would be better to eat chicken, but it might be enough to change the "all things considered" judgment if we factor in, say, the impacts of cattle production on climate change.

Another proposed use of neuron counts is not for making tradeoff decisions between different non-human species, but rather for figuring out how to weigh the interests of non-human animals along with those of humans. For example, Budolfson and Spears as well as Stawasz look at neuron counts as options that could allow economists to include estimates of the welfare effects on animals as part of the cost-benefit analysis of different interventions. It should be clear, then, that the precise weight we assign to the interests of various animals has important implications for cause prioritization.

Which measures best track the general idea behind neuron counts?

Thus far, I've been using the term "neuron count" to indicate the feature of the brain presumed to be relevant for comparisons across species. However, in the debate over general intelligence, there have been several different features of the brain proposed as potential indicators. For example, one might think that "brain size" is a rough approximation of the processing power of a particular organism. However, brain size turns out not to directly correlate with the number of neurons in an organism, because neurons themselves can be different sizes, and because brains can contain differing proportions of neurons and connective tissue. Moreover, different types of species have divergent evolutionary pressures that can influence neuron size. For example, avian species tend to have smaller neurons because they need to keep their weight down in order to fly. Aquatic mammals, however, have less pressure than land mammals from the constraints of gravity, and as such can have larger brains and larger neurons without as significant of an evolutionary cost. Brain size itself can also be measured in different ways (brain volume vs brain mass) that may have different implications.

Because brain size and neuron number tend to increase as body size increases for reasons that seem unrelated to any additional processing power in the brain (see below), some researchers have proposed that a measure called encephalization quotient (a measure of the ratio of brain volume to body volume compared against the average ratio across species) is a more relevant metric for measuring brain size that can be put to work in estimating more general cognitive processing ability, though this metric is typically only used on vertebrates. And given that neurons primarily transmit information to other neurons and that there can be a lot of variation in how many connections particular neurons have to other neurons, some researchers have suggested that it is better to use a measure of the connectivity between neurons or the number of synapses than to simply count the number of neurons.

Given these various different potential measures of brain capacity, which metric would be most relevant as a proxy for moral weight? I consider this question from two different perspectives. First, I’ll discuss some theoretical considerations relevant to thinking about the role of neurons. And second, I’ll examine relevant empirical evidence about the relationship between metrics of intelligence and neuron counts which is relevant for some claims about the relationship between these metrics and moral weight.

Theoretical Considerations About the Contribution of Neurons

This report assumes that moral weight depends on consciously experienced valenced states. It adopts a broadly consequentialist perspective, such that moral weight depends entirely upon the goodness or badness of particular states, and further assumes that goodness is connected to consciously experienced positively valenced states and that badness is connected to consciously experienced negatively valenced states. Given these assumptions, the idea that neuron counts will contribute to moral weight requires an assumption that neuron counts influence the intensity of conscious experiences of positive and negative states.

Two of the main competitors among philosophical accounts of the mind/body relationship are *type identity* accounts of consciousness, which claim that consciousness is identical to physical properties (typically neurophysiological properties), and *functionalism*, which claims that consciousness is equivalent to specific functional properties instantiated by the physical systems of the brain. Many contemporary scientific theories of consciousness fit under the banner of one of these two broad families of theories.

If consciousness is identical to the lower-level physical properties as identity theory suggests, this would imply that some (as yet unknown) properties of neural circuits, individual neurons, or perhaps even smaller units that are parts of neurons, contribute to conscious experiences. Given this idea, one might think that, if properties of neurons contribute to consciousness, and consciousness gives rise to moral weight, it follows that having *more* neurons results in more moral weight. However, it's important to note that identity theory is an account of consciousness more generally, rather than an account of *morally relevant* conscious states, and as such morally relevant conscious states might require more than simply the presence of consciousness. For example, there appear to be many conscious states that are neither affectively positive nor affectively negative and may therefore be value-neutral. Given this, we might suggest that it is not the total number of neurons, but the number of neurons in particular brain regions (perhaps the regions responsible for pleasure or pain, or those responsible for representing a "common currency" of different valenced states) that are relevant for moral weight. Without knowing exactly which regions are the relevant ones, one possible (though potentially dubious) simplifying move would be to assume that as the total numbers of neurons increase, the number of neurons in relevant brain regions is likely to increase proportionately.

Functionalism is the view that the identity of mental states is determined by its causal relations to sensory input, other mental states, and behavioral outputs (SEP). It is not quite as clear how increasing the numbers of neurons would result in more moral weight on a functionalist account. Various possible explanations will be considered below. But one possible explanation is that a certain degree of functional capacity is required to meet various thresholds, which themselves are relevant for moral weight (see further discussion below). Another possibility is that as the number of neurons increases, the system's representational capacities become more sophisticated, leading to richer or perhaps more expansive types of suffering and enjoyment. On such accounts, increasing the number of neurons might increase moral weight in virtue of increasing the overall processing capacity of a system. A further possibility is that having more neurons increases the number of subsystems that themselves have moral weight. This will not be discussed further in this document, but an additional report on the possibility can be found in our conscious subsystems report.

Considerations relevant to the type identity accounts and the threshold accounts will be discussed below. Given that the increased processing capacity account has the most in common with attempts to connect general intelligence to neuron counts, in the next section I review some relevant empirical evidence about intelligence.

Empirical evidence of the connections between intelligence and neuron metrics

How well do various measures of brain size track overall processing capacity? First, it must be noted that there's quite a bit of debate about whether and how something like general intelligence can be measured at all, and there are several strong arguments questioning whether anything like "domain-general" intelligence exists. But, assuming for the sake of argument that general intelligence can be reliably measured, here's how Susan

Herculano-Houzel (2012), a leading expert on the relationship between intelligence and brain metrics, summed up the situation:

> Given that neurons are the computational units of brains, it makes intuitive sense to expect larger brains to be made of larger numbers of neurons, in which case the larger the brain, the larger its computational abilities should be, such that the mammal with the largest brain is expected to be the most cognitively able…Oddly enough, however, the most cognitively able brain – according to ourselves – is not the largest. Mammalian brains vary in size over 100000 times, but the human brain does not rank first in size, and not even an honorable second: at about 1.5 kg, it is 2–3 times smaller than the elephant brain, and 4–6 times smaller than the brains of several cetaceans. Humans also do not rank first, or even close to first, in relative brain size (expressed as a percentage of body mass), in absolute size of the cerebral cortex, or in gyrification. At best, we rank first in the relative size of the cerebral cortex expressed as a percentage of brain mass – but not by far: the cerebral cortex represents 75.6% of the whole brain in humans, against 73.0% in the chimpanzee, 74.5% in the horse and 73.4% in the short-finned pilot whale.
>
> …a largely accepted alternative explanation for our cognitive superiority over other mammals has been our supposedly extraordinary brain size compared to our body size – that is, our large encephalization quotient…However, the notion that higher encephalization correlates with improved cognitive abilities has recently been disputed, in favor of absolute numbers of cortical neurons and connections, or simply absolute brain size. If encephalization were the main determinant of cognitive abilities, small-brained animals with very large encephalization quotients, such as capuchin monkeys, should be expected to be more cognitively able than large-brained but less encephalized animals, such as the gorilla. However, the former, smaller, are outranked by the latter in cognitive performance.

Dicke and Roth (2016) reach a similar conclusion:

> There is no clear correlation between absolute or relative brain size and intelligence. Assuming that absolute brain size is decisive for intelligence, then whales or elephants should be more intelligent than humans, and horses more intelligent than chimpanzees, which definitely is not the case. If it were relative brain size that counted for intelligence, then shrews should be the most intelligent mammals, which nobody believes.

Dicke and Roth suggest that information processing capacity (IPC) might be the most relevant metric for understanding intelligence, where IPC is a measure of the amount of information that can be processed during a period of time. If this is the case, as seems plausible, then there are a number of other features of the brain that influence the IPC. For example, neurons in vertebrates communicate primarily through the firing of action potentials, which send signals to the dendrites of other neurons. For this reason, the number of synapses is relevant to IPC, since additional synapses increase the capacity for communication between different neurons. Dicke and Roth also suggest the following:

> another factor that is important for cortical IPC is processing speed, which in turn critically depends on (i) interneuronal distance, (ii) axonal conduction velocity and (iii) synaptic transmission speed. (p. 6)

In the paper *Are Bigger Brains Better*, Lars Chittka additionally emphasizes the importance of local metabolism for “absolute computational power,” which can be read as IPC in this context:

> In terms of absolute computational power, relative brain size provides little information. The number of computations within a given time that can be supported by neural tissue is dependent upon absolute brain size, the number and size of neurons, the number of connections among them and the metabolic rate of the tissue. The upper rate of action potentials that can be sustained is determined by the specific metabolic rate, which will be higher in smaller brains. Smaller brains can therefore maintain a higher density of computations. Thus, relatively large brains from animals with small body mass are likely to have higher specific metabolic rates than similar sized brains from larger animals. Hence, although the absolute numbers of neurons and connections primarily determine the computational power, the energy available for neural processing, which is affected by the specific metabolic rate, is also important. (p. R1004)

Finally, it must be noted that though action potentials are the primary form of neuronal communication in vertebrates, some invertebrates rely more on neurons that signal without the use of “all or nothing” action potentials that either send a signal at a particular time or do not. This further complicates comparisons of IPC between these invertebrates and species that rely more on action potentials.

In short, if what we are interested in is the overall ability of the brain to process information in a given amount of time, there is no one single metric that seems to fit the bill. Rather, the overall processing power of the brain depends upon multiple factors which aren’t entirely understood. At the very least, this suggests that the relevant metric isn’t as easily measurable as has previously been suggested as a primary pragmatic advantage of neuron counts (see Section 4.3 below). More dramatically, this implies that much smaller brains with fewer neurons could potentially have *more* processing capacity than larger brains with more neurons, which certainly would be unwelcome news for the overall project of using neuron counts as a proxy for morally relevant traits.

Nevertheless, even if the number of neurons isn’t the *only* thing relevant for intelligence, it still is the case that, even with these qualifications, it is *in general* true that adding neurons would tend to increase IPC, assuming that each neuron carries some additional information processing capacity.

The review in this section has been primarily concerned with the relationship between neuron counts and intelligence, but it’s hard to see the situation looking better with moral weight. If anything, it seems *more* plausible that neuron count correlates with general intelligence than with capacity for welfare. Given that we can dissociate cognition and welfare, it’s possible and even likely that there will be some neurons that are relevant for general cognition but not for valenced affective states. As such, we can further restrict the discussion to the IPC of *relevant* cognitive systems. This, presumably, is the most plausible metric which might track morally relevant properties. But if we assume that the number of neurons in the relevant brain areas scale proportionally to the total number of neurons, or at least scaled proportionally after a minimum number of neurons required for basic cognitive functions is achieved, we might still believe that neuron counts can be used as a proxy. In what follows, I’ll assume that the best overall metric would be relevant information processing capacity (RIPC), but the term I will primarily be using, neuron counts, is meant to be a rough approximation of this measure.

Critically evaluating arguments for neuron counts as proxies

Very few people think that we can literally assign moral weight to an organism based only on the number of neurons they possess. Budolfson and Spears, for example, are very clear that this is meant only as a proxy. As such, this document will primarily be concerned with the idea that we can use neuron counts as proxies that roughly correlate with moral weight in the animal kingdom. Nevertheless, in order to endorse the idea that neuron counts are a good proxy, we must have at least some plausible explanation for why more neurons might lead to increased moral weight.

The literature is surprisingly thin on explicit arguments connecting neuron counts to sentience, welfare capacity, or moral weight. As such, this document needs to be a bit speculative about specific arguments connecting neuron counts and moral weight, including arguments that may sound implausible. The hope, however, is that by being as explicit as possible, the relative strengths and weaknesses of using neuron counts as a proxy can be evaluated by the reader, and that supporters of the use of neuron counts will be motivated to offer more explicit arguments for the connection to moral weight.

The argument from common sense intuitions

The argument from common sense intuitions goes as follows. We have an intuitive sense about which animals have greater moral weight. For many of the animals we are familiar with where we have reliable estimates of neuron counts, these counts nicely track our intuitions about which animals have more or less moral weight. Humans have more neurons than chimpanzees. Chimpanzees have more neurons than dogs. Dogs have more neurons than chickens. Chickens have more neurons than fruit flies. And so on. Given this close alignment, we may then use neuron counts in cases where we are uncertain about how much weight to give to particular animals or types of animals. So, for example, if we're unclear how much weight to give to a mink, we might measure the number of neurons in the average mink and use this to form our weighting multiplier.

There are other things to note about this approach. First, though there is some question about which precise metric to use as a proxy, the previous section demonstrated that whichever metric we choose will not align perfectly with common sense. Elephants have three times as many neurons as humans and have bigger brains. Shrews have greater brain-to-body ratios. Some new-world monkeys have a greater encephalization quotient than certain great apes. So, at the very least, it is clear that neuron counts do not infallibly track commonsense intuitions about moral status. We have evidence that any metric will diverge from common sense intuitions in certain cases. However, it might still be thought that neuron counts are close enough to tracking such intuitions that they can be useful in certain contexts, just as consumption might be a highly fallible but nevertheless still helpful measure of human welfare in economics. How closely neuron counts track our intuitions might depend on how "rough" of a rough measure we are looking for.

Finally, one might question whether we should place much stock in our common sense intuitions about nonhuman animals. Humans have a long history of assigning special value to their own characteristics and making bad assumptions about the uniqueness and importance of human traits. It seems very likely that our intuitions about moral weight are biased towards those things that we value as parts of our distinctively human lives. Beyond just our anthropocentric biases, we also have other biases that often lead us astray. Consider a hypothetical proposal that moral weight should be assigned based on cuteness. We might check our intuitions and find that chimpanzees are more cute than pigs, who are more cute than chickens, who are more cute than shrimp, etc. But clearly, this is a poor way of reasoning, and there's a fairly straightforward scientific story to tell about how these biases are simply a relic of the types of faces and features humans have been designed to respond to through evolution. It is therefore important not to place too much weight on the intellectual equivalents of "cuteness" in our evaluations of moral weight. The point is not that all intuitions about moral status are necessarily as unreliable as those that closely track cuteness, but rather that we should be cautious about assuming that agreement with pretheoretical intuitions is a strong point in favor of a view, particularly if those intuitions are likely to come from people who may be influenced by a lot of assumptions built into the current culture.

Empirical evidence that more neurons increase valence

One possible source of evidence linking increased neuron counts to moral weight would be neuroscientific evidence indicating that larger numbers of neurons result in stronger positive or negative valenced experiences. I consider two possible arguments along those lines below.

Brain Volume and Valence

Studies of humans provide the best opportunity of linking verbally self-reported positive and negative states to the number of neurons in particular brain regions. There are at least some studies suggesting a relationship where particular brain areas show increased volume in connection with greater intensity of either pleasure or pain.

For example, looking at happiness, consider various studies with titles like, "Meditation led to increased brain volume in areas associated with happiness," "Researchers found a link between subjective happiness scores and gray matter volume in the right precuneus," and "Decreased brain volume is associated with depression."

Or if we consider pain, Davis and Moayedi write:

> "Teutsch and colleagues demonstrated that 20 min of noxious stimulation over eight days in healthy subjects induced increased gray matter volume in the premotor cortex, MCC, S1, inferior parietal lobule, and the medial temporal gyrus (Teutsch et al. 2008). Additionally, a study in our lab (Erpelding et al. 2012) found that cortical thickness in S1 is correlated with individual subjects' heat and cold pain sensitivity, and cortical thickness of the MCC correlated with heat pain sensitivity."

However, selectively citing a few of these studies can be misleading. For there also are a large number of studies showing an inverse relationship between brain volume and the intensity of particular affective states. In particular, when it comes to chronic pain, by far the most commonly cited relationship between brain volume and chronic pain is a decrease in brain volume in regions commonly associated with the experience of pain. For example, see the Davis et al. (2008) article "Cortical thinning in IBS: implications for homeostatic, attention, and pain processing".

Thus, in general, there are many examples of brain volume in certain regions increasing with certain affective states, and brain volume decreasing with certain affective states, but no clear pattern governing a more general relationship between numbers of neurons and the intensity of those states. As such, what we know about the relationship between self-report of affective states in humans and brain volume doesn't currently support the idea that more neurons in general correlates with increased valence, either comparing within individuals or comparing across individuals. It may turn out that increased numbers of neurons in particular regions of the brain inevitably correlates with increased valence intensity, but this remains to be seen. At present, no simple pattern such as this has been identified.

Moreover, even if it were discovered that the number of neurons in a particular brain area did reliably correlate with subjective ratings of valence, there are alternative explanations available that do not posit a direct relationship between the total number of neurons and intensity. One such explanation is that it is not the total number of neurons, but rather the proportion of active neurons that matters for valence intensity. Brian Tomasik (2013), for example, has suggested that a simple organism with a large proportion (but a relatively small number) of neurons active during pain might suffer far more than a larger organism with a smaller proportion but a larger total number. Thus the fact that more neurons in a particular brain region is correlated with greater valence could be taken as evidence that increasing the proportion of neurons in the brain devoted to a particular task increases the experienced valence of that task. However, this suggestion might also be vulnerable to some of the points raised in the following sections where we consider adding additional neurons that are not playing any particular role in pain.

Another potential explanation for the available data, one that is consistent with functionalist accounts of mind, is the widespread idea that it's not how many neurons, but rather what the neurons are doing, that matters most. The fact that increased pain can correlate with increased brain volume in some areas but decreased brain volume in others suggests that what the neurons are doing is changing under different circumstances and that this is far more important than simply the total number of neurons involved. In fact, many studies of neuronal coding of intensity focus on the *firing patterns* of neuronal ensembles rather than simply the firing rates or numbers of neurons involved. This idea will be discussed further below.

Empirical evidence that increased activation increases valence

Functional brain imaging studies are replete with examples of increasing levels of activation[2] in particular brain areas correlated with valenced affective states. There are studies linking increased activation in various brain areas with pains, pleasures, anxiety, depression, fear, nausea…pretty much any valenced mental state one can think of. Could this be taken as evidence that increasing the number of neurons also increases valence?

Again, the evidence paints a more complicated picture. To use pain as an example again, researchers have known for decades that general brain regions tend to show increased activation in correlation with increased self-reports of pain, including those specifically linked to increased ratings of the unpleasantness of pain. Nevertheless, despite knowing that in general pains tended to activate the anterior cingulate cortex, the primary somatosensory cortex, and the insula cortex, to name a few regions, pain scientists nevertheless were not able to simply look at the fMRI results from an individual patient and say definitively, "this person has a pain at a level of 8 out of 10." There were too many individual differences in the relationship between pain and activation levels in individuals to make these types of judgments. For example, though it is true as a general rule that as pain increased activation in those brain regions increased, one person's 8 out of 10 might look very different in a brain scanner than another person's 8 out of 10.

This was thought to have changed when a breakthrough was made in a 2013 article in the New England Journal of Medicine by Wager et al. The researchers in that study used a machine learning algorithm to determine an activation pattern that could determine, with over 90% accuracy, whether someone was in acute physical pain by looking only at that person's pattern of brain activation. What is important for present purposes, however, is that this correct

[2] In the case of fMRI, "increased activation" refers to increased levels of oxygenated blood flow, which is believed to be correlated with neural activity (though perhaps more with local field potentials than with actual numbers of action potentials).

determination depended not only on finding voxels (the smallest 3D units in brain imaging output) where activation was positively correlated with physical pains but also on finding voxels where activation was negatively correlated with physical pain. In other words, in order to successfully predict pain, researchers needed to look for patterns that included both increased activation in some areas and decreased activation in other areas, rather than just a blanket pattern of “more neuronal activation.” So again, the picture seems to be that a simple story of “more neuronal activation leads to more pain” is far less plausible than a different account where pains typically represent activation in some brain areas and inhibition in other brain areas.

The “each neuron independently has moral weight” claim

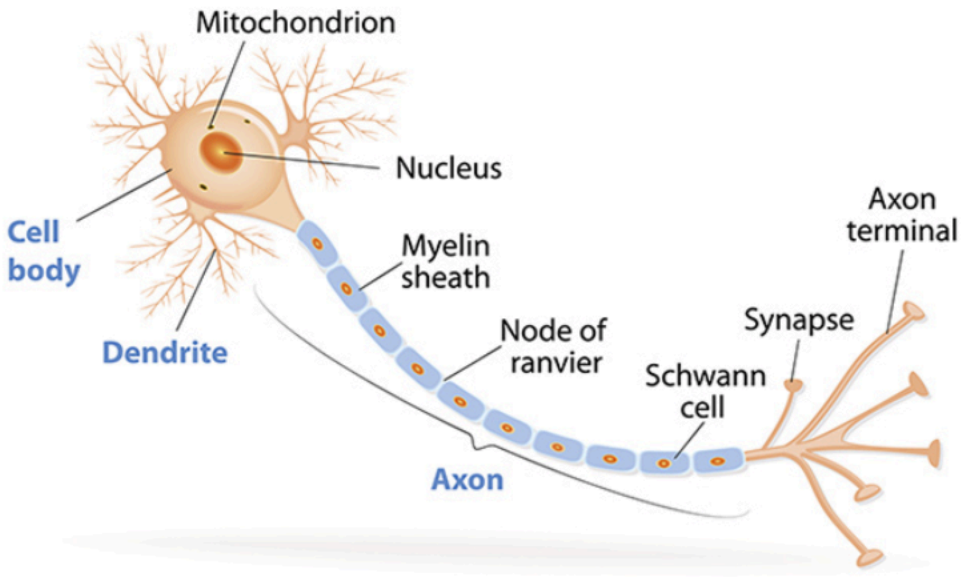
One potential argument for linking neuron counts and moral weight is that if moral weight depends upon the properties of individual neurons, then having more neurons results in more weight, based on a summation of the contributions of individual neurons. This idea should be kept separate from the idea that each neuron contributes to the magnitude of the overall weight of the system, but that the actual weight is determined by what the system does, which will be considered later.

Why should we think that individual neurons have more weight? Starting from a sentientist view where only consciously experienced valenced states have moral weight, we might think that individual neurons are conscious because we hold a view like panpsychism, where consciousness is ubiquitous, or a view where every “computation” or processing of information is conscious. On these views, even an individual neuron might have some degree of consciousness, and as such might be taken to provide at least a little moral weight.

These views do not seem to be very widely held. But even if we think that the evidence on balance weighs against such a view, we might nevertheless at least think that we should assign some credence to the possibility, which would alter how we prioritize different species. If we think there’s a 1/10 chance that each individual neuron could contribute to moral weight, that would influence how we should weigh the welfare of humans vs other species.

However, this particular way of claiming that individual neurons might matter seems flawed, even as it relates to assigning credence.

Why? Consider an individual neuron of a type that would be found in the central nervous system. The neuron receives input in its dendrites, and when a certain threshold is crossed, an action potential is triggered that propagates a signal down its axons and releases neurotransmitters, which thereby increases the odds of other downstream neurons firing.



Assume for the sake of argument that the neuron’s contribution to behavior is entirely dependent upon the information transmitted via action potentials (though this assumption will be relaxed below). Thus, there is a binary signal of “firing “ or “not firing” at any given moment and the overall contribution of the neuron can be summarized as the number of firings per some finite amount of time, based on what input it receives.

This neuron could be part of an overall pattern of activity that constitutes a pain experience, such that it typically serves as a relay for nociceptive information that is transmitted up the spinal cord and ultimately results in pain behavior. As such, we might be tempted to think that this neuron contributes to the potential negative valence that could be experienced by an organism, and thus to the overall moral weight we should assign to that organism.

However, the neuron, which receives a certain input and as a result of that input sends out its own signal in the form of action potentials, could also play a role in pleasure. That is, the very same neuron, considered individually, could have played a role in the experience of pleasure. We can think of this in terms of a hypothetical (that is, we imagine taking a neuron that previously was in the pain circuitry of the brain but then moved it to pleasure circuitry) or literally (as in some neurons do in fact seem likely to be active during both painful and pleasurable experiences).

Moreover, the neuron also could be involved in information processing that merely results in neutral experiences (neither positively nor negatively valenced).

If there are additional ways that an individual neuron contributes to the nervous system in a way that influences eventual behavior, we can substitute those into our hypothetical example as well. For example, say that the neurons communicate via some interaction with the glial cells that surround neurons. There is no particular reason to believe that neurons communicate with glial cells differently when involved in pleasure or pain, so there doesn't seem to be a reason for saying that we couldn't still substitute the hypothetical neuron in either type of circuit.

As such, it seems like a mistake to say that an increase in the number of neurons increases the capacity for suffering or the capacity for enjoyment via the contribution of individual neurons. An individual neuron, considered by itself, is not any more likely to have a negative weight than a positive or neutral weight. So adding an individual neuron by itself shouldn't alter the expected utility of any particular system.

In fact, this argument need not be limited to individual neurons. We might say that for any neural circuit that can be equally substituted in either pain, pleasure, or neutral processing, we have no reason to assign moral weight to that circuitry.

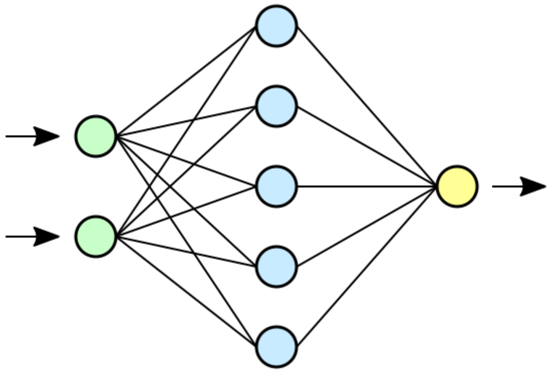
How many neurons need to be linked together before this substitution is possible? This is unclear. But it seems like the neurons need to at some point be doing something that uniquely connotes pain or pleasure before we would have any reason to assume that the neurons themselves might even plausibly carry moral weight.

The "each neuron contributes to the overall weight of the system" argument

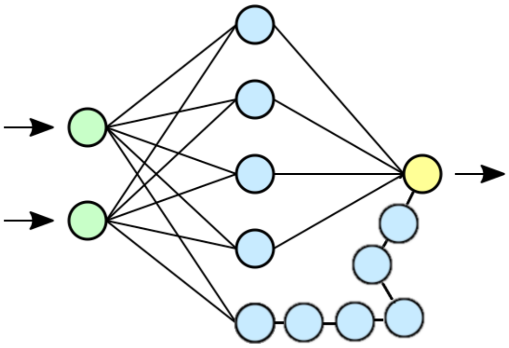
As a different possibility, consider the idea that each additional neuron contributes to a "valence capacity" of a welfare-capable system, but that the overall valence is determined by the state of the system. On this framing, individual neurons don't themselves generate positive or negative valence, but they do increase the overall capacity of the system which thereby increases moral weight.

In response to this idea, as a thought experiment, consider the following two systems.

One, a simple neural network, which transforms an input into two nodes into an output from a different node.



The second, a neural network where one of the direct connections between two nodes is replaced by a chain of five neurons. Each neuron firing in the chain leads to a 100% chance of the following neuron in the sequence firing. In other words, the relationship between inputs and outputs of the two overall systems is identical, and the addition of the five additional neurons doesn't in any way change how the system processes information, despite increasing the total number of neurons by 62.5%.



This example suggests that, on the picture considered above, the bottom system might have 62.5% more moral weight despite there being no possible observable difference in behavior of the overall system. This seems implausible and would suggest that the theory is essentially untestable.

Granted, the above example is not wholly realistic. The brain has ways of eliminating redundant neurons and a tendency to do so. However, considered as a thought experiment, it suggests that it is implausible that merely adding additional neurons that are not changing the input/output relationships adds to the moral weight of the overall system.

This example helps illustrate the attractiveness of the functionalist picture that what a particular neural circuit does (which can be precisely specified in terms of input/output mappings) is far more important for how any particular subsystem of the brain works (including valenced states) than the raw number of neurons. And while the above example is not realistic, if we think of neural circuits in the brain functionally as primarily transforming particular inputs into particular outputs, then there are innumerable numbers of potential neural circuits that could perform identical operations. So there do exist more realistic examples where systems with different numbers of neurons are performing identical operations, at least if we assume that neuronal transmission is the sole source of information transmission, though they would be too complex to represent here.

One response to this argument is to say that such cases are extremely unlikely, since both evolutionary fitness constraints more generally and "neural pruning" more proximately both imply that wholly redundant neurons are unlikely to exist in developmentally complete animals. As such, it might be argued that more neurons in the brain in fact tend to contribute to the processing capacity of the overall system, even if there are thought experiments that show that neurons needn't do so, which could lend at last some credence to the neuron count hypothesis.

Neuron counts correlate with intelligence, and intelligence correlates with moral weight

The argument for neuron count proxies that connects with the largest body of scientific literature is the following: neuron counts positively correlate with intelligence, and intelligence positively correlates with moral weight. Therefore, neuron counts positively correlate with moral weight. We can, of course, question both links in this chain.

As was discussed above, though there is not a large literature connecting neuron counts to sentience, welfare capacity, or valenced experience, there is a reasonably large scientific literature examining the connection of neuron counts to measures of intelligence in animals. The research is still ongoing and unsettled, but we can draw a few lessons from it.

First, it seems hard to deny that there is one sense in which the increased processing power enabled by additional neurons correlates with moral weight, at least insofar as welfare-relevant abilities all seem to require at least some minimum number of neurons. Pains, for example, would seem to minimally require at least some representation of the body in space, some ability to quantify intensity, and some connections to behavioral responses, all of which require neurons. As such, each welfare-relevant capacity requires at least some minimum threshold of neurons.

But aside from the ability to cross certain morally relevant thresholds, things become less clear. Setting aside, temporarily, the idea that increased intelligence increases experiential richness (which will be discussed in the next section), it's worth noting that intelligence is typically defined as something along the lines of the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria. But on this definition, there's no obvious linkage between intelligence and sentience. It seems conceptually possible to increase intelligence without increasing the intensity of experience, and similarly possible to increase the intensity of experience without increasing intelligence.

Furthermore, as Peter Singer among others has pointed out (Singer 2011), it certainly is not the case that in humans we tend to associate greater intelligence with greater moral weight. Most people would not think it's ok to dismiss the pains of children or the elderly or cognitively impaired in virtue of them scoring lower on intelligence tests.

And finally, it's worth noting that some people have proposed precisely the opposite intuition: that intelligence can blunt the intensity of certain emotion states, particularly suffering. According to this account, our intelligence can sometimes provide us with tools for blunting the impact of particularly intense experiences (Chavez and Barber 1974, Rybstein-Blinchik 1979), while other less cognitively sophisticated animals may lack these abilities. It also is possible that this relationship is not entirely linear. For example, at the low end of neuron counts/intelligence, a reduced total number of neurons could feasibly prevent the capacity for suffering, but in the middle range, there may be some kind of positive correlation/step function where increased neurons/intelligence generate an increased ability to suffer.

So, at any rate, more explanation seems to be required for an account arguing that neuron counts influence moral weight purely through intelligence.

Experiential and Evaluative Richness

It seems plausible that increasing processing ability will increase the potential richness of experiences in organisms. In fact, some of the ideas that have been used to dismiss the value of neuron counts as correlates of general intelligence, such as noting the fact that many additional neurons are needed simply to represent larger body sizes, may not apply equally in relation to richness. After all, if more neurons are needed to represent more points in space on the body or in the surrounding environment, this nevertheless suggests that the experiences are richer than they would be if fewer points were represented. So if the richness of experience bears directly on moral weight, this is another potential avenue for connecting neuron counts to moral weight.

How might an increased richness of experience that comes from more processing power influence moral weight? This section considers two possibilities: first, that it influences the phenomenal richness of experience, which adds value to the experience as expressed through the preferences of an agent. And second, through what might be called evaluative richness, where the evaluative capacity of the organism becomes more fine-grained.

To set up these two separate ideas, let’s assume for the sake of argument that there’s an informational or sensory component of experience that can be dissociated from an evaluative component. That is, in comparison to one person getting joy from eating fresh blueberries, we can imagine another person who experiences the same amount of joy from the experience but who has far more fine-grained ability to discern subtle variations in the taste and texture of the overall experience. The ability to discern the taste and texture we might call the sensory component of the experience, and the joy that accompanies this is the evaluative component.

Assuming that greater numbers of neurons result in greater representational capacity, could this add moral weight purely in virtue of influencing the sensory dimensions of experience rather than the evaluative dimensions? It seems as though the answer to this should be “no” on a hedonistic account of moral weight, given the fact that the affective and sensory components of pain have been shown to be experimentally dissociable.

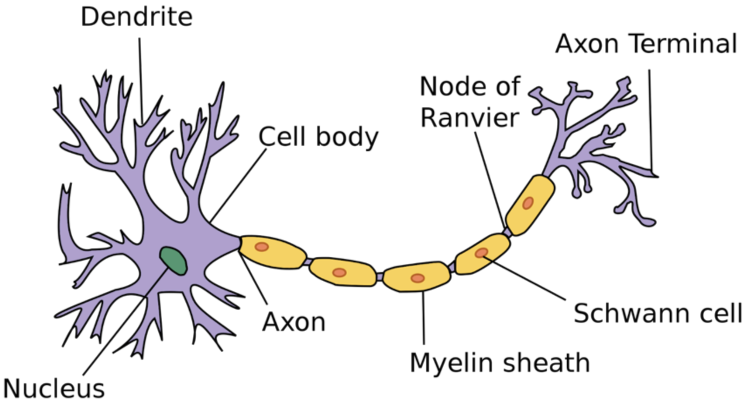
Let’s turn, then, to evaluative richness, which is defined as sensitivity to variations in experienced valence (Birch, Schnell, & Clayton 2020). Evaluative richness can be further subdivided into evaluative bandwidth and evaluative acuity: “Rich affect-based decision making takes many inputs into account at once (evaluative bandwidth) and is sensitive to small differences in those inputs (evaluative acuity)" (ibid). Here the connection to welfare-relevant states is clearer since changes in valence are directly relevant for hedonistic accounts.

Nevertheless, it's not immediately obvious that simply having a more discerning ability to evaluate will also entail having a wider range of experiences.

Henry Shevlin, in an unpublished draft, proposed two possible accounts of what happens when we add evaluative richness. On the "compression account," the ends of the spectrum stay as far apart as they were previously, but we add precision in how small the units can be existing between the two polls. However, on the "expansion account" adding more evaluative states pushes the ends of the spectrum outward. As such, on the expansion account, adding richness would seem to actually increase the range of evaluative states, and hence, assuming that moral weight tracks the capacity of welfare, then this would give us one account of how adding richness could increase moral weight. To be clear, there was no particular reason to assume that the expansion rather than the compression account is correct, but this does offer one story of how increased neuron counts in affective parts of the brain may increase moral capacity.

A General Argument Against Thinking Neuron Counts are Proxies for Moral Weight

This section presents a general argument against thinking neuron counts can be used as proxies for moral weight based on general principles of how the nervous system works.



Most neurons communicate with each other primarily through the firing of "all-or-nothing" action potentials, electrical signals that travel down axons before causing the release of neurotransmitters to downstream neurons. Action potentials are "all or nothing" because a certain input threshold is required to trigger them, but once triggered they follow a relatively defined pattern. In other words, if the threshold is crossed just barely or by a lot, the output of the action potential is the same.

Neurons can be thought of as performing computations. As various inputs from other neurons send signals to the dendrites or cell body of a neuron, these inputs influence how likely the neuron is to fire an action potential. After each action potential, there is a "refractory period" where the neuron can’t fire, but the pattern of inputs over time generally determines how many times the neuron fires over that period of time.

Since action potentials can be thought of as computations, and since the number of action potentials over time of a given neuron is its primary influence over other neurons, it can be tempting to think of the number of action potentials within a certain neural system as serving as a quantification of the total amount of “mental stuff’. Moreover, this implies that more neurons also leads to “more mental stuff” because more neurons would lead to more firing of action potentials. However, both of these ideas rest on a mistaken conception of how the brain uses information to determine behaviors that maximize an organism’s chances of survival.

Empirical Evidence

First, consider the empirical evidence. It can be tempting to look at a brain imaging study, such as Rainville et al. 1999, that showed that increased activation in the anterior cingulate and insula are correlated with increases in reported unpleasantness of pain, and conclude that “more neurons firing” in those regions leads to more pain unpleasantness, and therefore we can use the number of neurons in pain regions as a proxy for total welfare capacity. This, however, is a mistake. As noted above, the scientists studying brain imaging patterns of pain are very clear that these patterns can only be used to identify the intensity of pain or pain unpleasantness *within* individuals. It is not a measure that can be used to make comparisons across individuals. That is, you cannot look at one individual and say this person has more ACC and insula neurons active during pain and therefore they experience more pain than another individual who has fewer neurons active. Studies looking at individual differences in the experience of pain tend to focus on the robustness and likelihood of activation in particular regions, but they do not simply focus on the total amount of activation in particular regions to make comparisons across individuals. In other words, one individual might be “more sensitive” to pain because their affective pain regions are more likely to be activated for an identical stimulus, but they are not considered to be “in more pain” merely because they have more neurons active when both patients are showing strong activation, aside from at a very coarse-grained level of comparison. Some potential reasons for this will be discussed below.

But equally important, neuroimaging researchers are highly invested in finding an objective biomarker for pain, and in being able to predict individual differences in pain responses. They have performed hundreds of experiments looking at brain activation during pain. The current consensus is that there is not yet a reliable way of identifying pain across conditions merely by looking at brain scans, and no leading neuroscientists have said anything that suggests that the mere number of neurons in a particular region is predictive of “how much pain” a person might experience. I think this is compelling evidence that the claim that “more pain neurons = more pain” is not suggested by currently available evidence. It might, of course, turn out to be the case that there exists a subset of neurons in particular neural circuits that really do have this relationship with pain, and which haven’t yet been identified due to the limited spatial resolution of brain imaging, but it seems ill-advised to put any weight on this idea in the absence of any substantial evidence.

Theoretical considerations

As noted above, a given neuron N performs a computation by “deciding” whether to fire an action potential based on the set of inputs. Upstream neurons send signals to N by releasing neurotransmitters across a synaptic cleft onto the dendrite or cell body of N, which alter the electric and chemical gradients between the inside and outside of the cell. When a particular threshold is reached, an action potential occurs which propagates a signal down N’s axon to release neurotransmitter to downstream neurons.

The important feature for this report is that there is a wide variety of differences between any given “contact point” between an upstream neuron and N, such that different input neurons can have very different influences on how likely it is that N fires an action potential. One input neuron I_1 might lead to a 100% chance of N firing, while I_2 only increased the likelihood of firing by 25% and I_3 only increases the likelihood by 10%. In fact, there are inhibitory neurons that decrease the likelihood of downstream neurons firing.

Why is this important? Because there’s no fixed relationship between the number of neurons upstream of N or the rate of firing upstream of N and N’s likelihood of firing. A given system can be designed to be triggered only if a large number of neurons fire, or could be designed to fire only if a small number of neurons fire frequently, or designed to fire only if a set of neurons fires according to a certain timing pattern or any of many other possibilities.

All of this is also true of any output system. Take any behavior B generated by the experience of pain. B could be designed so that it only occurs if a certain number of neurons are firing at a certain rate, but it also could be designed to be triggered by only a very small number of neurons being active, or to be triggered if neurons from brain regions X, Y, and Z but not Q are active, etc. From an evolutionary perspective, there’s no particular reason for B to be sensitive to the total number of neurons firing. What’s important is that B is sensitive enough to different signals that the organism behaves appropriately in each circumstance.

Ultimately, this is true for any observable (in the scientific sense) output of the pain system. For any potential pain output system, there is a range of possible responses, ranging from a minimum to a maximum. It could be the case that “more pain neurons firing” always reliably pushes the output towards the maximum, but note that at some point the maximum would be reached and there would no longer be any observable relationship between more pain neurons being active and any given observable pain behavior. Moreover, there’s no particular reason for the system to be designed in a way that has the simple relationship of more pain neurons = bigger output. We know that brains are sensitive to all manners of additional inputs from expectations to mood to previous experiences. And, as noted above, neuroimaging experts who study pain look at overall firing patterns rather than raw numbers of neurons as the best potential methods for estimating differences in the experience of pain between individuals.

So the upshot is the following: to the extent that more pain neurons matter for observable effects that result from conscious pain, the best evidence suggests that they matter insofar as they influence the variety of signals that can be used to influence outputs. Having more pain neurons does increase the complexity and subtlety of translating potential inputs into different outputs because they allow for more complicated computations. And being more sensitive to different signals presumably allows more flexible behavior for an organism and thereby could contribute to fitness by providing a wider range of possible responses to pain (and in particular pain in the context of many other competing environmental demands). However, there does not seem to be any evidence for, or any clear theoretical reason in favor of, the hypothesis that there's a simple relationship between the number of neurons and the amount of pain, or any other mental stuff, generated by the brain. The translation from potential inputs to potential outputs could take place any number of ways, and while it makes sense for an organism to be able to rate the relative valence of different noxious stimuli in order to respond accordingly, there's no particular reason to think that the magnitude or number of neurons firing has any particular objective relationship to the overall amount of pain experienced since there's no evolutionary reason for the latter to expand indefinitely.

When might neuron counts be useful?

As stated previously, whether or not we should use neuron counts as a proxy for moral weight will ultimately depend on context. Below are some thoughts about this context.

Neuron Counts Are Better Than Nothing

First, it should be noted that in one of the primary domains in which neuron counts have been proposed as a weighting tool, there currently is no consideration given to the interests of animals. In the field of economics, when welfare is used as a measurement, outcomes are generally evaluated by aggregating only the welfare of human beings. The welfare of animals is not taken into consideration.

In such cases, if we think that the welfare of animals is a morally relevant consideration, then it would be preferable to use neuron counts as an alternative to the status quo, since it is unlikely that neuron counts would be so far off as a measure that they would be worse than assigning all animals a weight of 0. If a pig has some welfare, then assigning some weight is better than assigning none. If a mealworm has some welfare, then assigning some is better than assigning none. The only way neuron counts would lead to a less accurate system of weighting than ignoring animal interests would be if it overestimated the moral weight of animals by more than assigning zero underestimates their moral weight. Given that neuron counts strongly prioritize human interests (with some exceptions already noted), this seems fairly unlikely.

However, the fact that neuron counts are preferable to the status quo is only a limited point in their favor. The "better than nothing" consideration, for example, doesn't give us any reason to believe that neuron counts are better than many other potential methods for assigning nonhuman animals a nonzero moral weight, including simply relying on pre-theoretical intuitions, since those also are better than nothing. So whether it is good to use them will depend on what alternatives are available (and what alternatives could be created in the relevant timeframe).

Neuron Counts are Quantifiable

Some decisions require weighing outcomes involving large numbers of individuals of different species. For these decisions, we need some way of assigning a numerical value to particular individuals. Because there is a fact of the matter as to how many neurons any individual has, and a fact of the matter as to what the average number of neurons there are for members of particular biological taxa, neuron counts provide a method for assigning a quantified measure for use in comparisons involving members of different species. As such, they have a practical advantage over other methods for assigning moral weight that do not easily translate into numerical values, at least in certain contexts.

A similar rationale presumably exists for using income as a proxy for welfare in economics. No one thinks that income perfectly correlates with human welfare, but the assumption is that it correlates closely enough that using income to provide a precise quantity allows economists to make calculations involving large numbers of people in different circumstances.

Of course, the appearance of precision can also be a flaw in certain contexts. If neuron counts are only an extremely rough predictor of welfare, then treating them as precise indicators could lead to mistakes. As a general rule, it would be better if decisions were sensitive to the level of uncertainty around welfare when this is possible.

Neuron Counts are Measurable

Another important difference between neuron counts and many other purported weighting criteria is that not only is there a fact of the matter about how many neurons any particular individual has, we can in principle measure this number using current technology. In fact, in practice, we are likely to be able to come up with at least reasonably good estimates for the average number of neurons for members of particular species within a range that would be useful for comparisons.

However, though we could reasonably expect to use modern neuroscientific techniques to measure “the total number of neurons” in an organism, things get more complicated if we recall that “neuron count” is actually best viewed as an approximation of a quantity of relevant information-processing per unit of time. We can, in principle and likely in practice, approximately measure the number of neurons in the brains of particular species. However, measuring the number of connections between neurons and the strength of such connections is much harder. And including an understanding of the local metabolism which helps determine the number of action potentials that can be sent per unit of time adds a further layer of complexity. Since these other characteristics are relevant to how much information can be processed, measuring neuron counts may be importantly misleading about the actual capacity for information processing, while measuring all of the relevant criteria is not currently possible for whole brains.

Alternative Frameworks

In the last section, it was argued that using neuron counts as a proxy for moral weight is likely better than assigning zero moral weight to all nonhuman animals. But what other alternatives exist as competitors to neuron counts?

Moving beyond the anthropocentric assumption that only human welfare matters leads to various possibilities for taking the experiential states of other animals into consideration. Perhaps one of the most straightforward ways of adopting this position is to argue that, once an animal has some welfare and as such some moral weight, the animal's subjective welfare translates to moral considerations in the exact way that humans welfare does; that is, that the pain of a shrimp is equally morally important as a relevantly similar pain in humans. This is perhaps one reading of Peter Singer's famous comment, "in suffering, the animals are our equals" (Singer 1975).

In addition to the intuitiveness of the idea that all beings that matter morally matter equally, there are some developed arguments. For example, Visak (forthcoming), citing Cabanac (1996) and Rayo and Becker (2007), suggested that experiential states of well-being are representative of the degree to which current states reflect how fitness-promoting or fitness-threatening particular circumstances are, relative to the fitness of other options. Since it seems as though a beetle can have states that are equally fitness promoting or fitness threatening relative to other options as those of a human (after all, there is only one state at any given time that is maximally fitness promoting for each organism), then provided that there are no dramatic differences in how those states are represented in experience, it would follow that the beetle's experiences are equally important, from a moral point of view, as those of humans.

How does this perspective which has at least some arguments supporting it, compare to that of neuron counts? Neuron counts certainly offer a weighting system that more closely tracks the intuitions of the folk when it comes to assessing the moral status of animals, since people do not default to the view that all animals matter equally. Neuron counts are also more sensitive to the fact that there are at least some differences where it appears that having larger brains facilitates morally relevant behavior. Nevertheless, the advantages of neuron counts are not so persuasive that the "equal weight" view should be discounted entirely. If nothing else, it seems as though a proxy for moral weight that used neuron counts should also include some hedging that accommodates the possibility of the equal weight view being true.

The same point is true for relying directly on folk intuitions about animals' moral status. One could in principle simply poll the public about how they would weigh the interests of different animals and subsequently use these results directly for assigning weights. These intuitions do seem to very roughly track some differences in intelligence and capacities. However, they also are very likely to be biased towards features that humans find valuable in themselves as well as morally irrelevant features such as visual appearances that resemble human features. Again, it seems as though any assignment of moral weight should assign at least some probability to the chance that the "equal weight" view is correct...relying solely on human intuitions is likely to tip the scales in humans' favor. And these intuitions, unlike neuron counts, are not connected to an objectively measurable metric that may actually yield surprising results that conflict with our original beliefs

In addition to the above suggestions which have at least somewhat concrete proposals for assigning weight, there have been numerous suggestions of possible capacities that count as significant thresholds for moral consideration but which do not directly lend themselves to precise methods for weighting. For example, Varner (1998) has suggested that reversal learning may be an important capacity related to moral status. Colin Allen (2004) has suggested that trace conditioning might be an important marker of the capacity of conscious experience. Birch, Ginsburg and Jablonka (2020) have raised the possibility that unlimited associative learning is the key indicator of consciousness. And Gallup (1970) famously proposed that mirror self-recognition was a necessary condition for self-awareness. All of these views entail claims that are potentially relevant for assessing moral weight, but none directly translates into a weighting function.

Finally, in addition to using particular metrics as proxies of moral weight, it is of course possible to come up with a composite function that incorporates multiple possible metrics. For example, one might develop a method for assigning moral weight that takes into consideration neuron counts in addition to evidence of other capacities such as episodic memory, unlimited associative learning, etc. These methods would face difficult questions about how each component contributes to the final weight, but would have the clear advantage of being receptive to more potentially relevant information than individual metrics alone. This approach acknowledges that there might be some expected increase in value correlated with additional neurons, but also recognizes limitations in this assumption and includes additional methods that might capture instances where organisms' behavioral complexity diverges from their raw number of neurons.

Summary and Conclusions

Neuron counts appear to be an appealing proxy for moral weight estimates because of the relative simplicity of the metric and the likelihood of the metric tracking something of importance. However, as we more closely examine what the metric would need to be measuring in order to provide a good proxy for moral weight, the metric loses its simplicity. In particular, it is not the number of neurons that seems most likely to matter, but rather the amount of information that can be processed in a given amount of time in relevant brain areas, and this later metric depends on numerous other factors such as the firing rate, the number of connections between neurons, and the precise properties of those connections. While some techniques exist to estimate the number of neurons in particular organisms, no techniques are yet available that can capture this broader suite of relevant information for an entire brain.

To summarize, the attraction of the metric is its simplicity and reliability relative to particular aims. However, the more simple a metric we choose, the less reliable it is, and the more reliable a metric we choose, the less we are currently able to measure it. As such, the primary attraction of neuron counts is an illusion that vanishes once we attempt to reach out and grasp it.

Moreover, there have been a number of specific capacities proposed as potential indicators of moral status in nonhuman animals, such as the capacity for reversal learning, trace conditioning, or self-recognition. Neuron counts do not seem to reliably predict success on these tasks nor on proposed tests of intelligence. There are numerous studies showing some small-brained animals demonstrating the capacities such as trace conditioning, transitive inference, concept learning, and more. As such, the plausibility of neuron counts as a metric runs counter to many other proposals of moral status in other animals.

Given this, we suggest that the best role for neuron counts in an assessment of moral weight is as a weighted contributor, one among many, to an overall estimation of moral weight. Neuron counts likely provide some useful insights about how much information can be processed at a particular time, but it's unclear if they are a better predictor for moral status than simply relying on intuitions, and it seems especially unlikely that they would provide more useful information individually than a function that takes them into account along with other plausible markers of sentience and moral significance. Developing such a function has its own difficulties, but it is preferable to relying solely on one metric which deviates from other measures of sentience and intelligence.

Acknowledgments

This research is a project of Rethink Priorities. It was written by Adam Shriver. Thanks to Bob Fischer, Jason Schukraft, Marcus Davis, Meghan Barrett, Gavin Taylor, David Moss, Joseph Gottlieb, and Michael St. Jules for helpful feedback on earlier versions of this post. If you're interested in RP's work, you can learn more by visiting our research database. For regular updates, please consider subscribing to our newsletter.