

# Everyday Utilitarian

[Home](#)[About](#)[Posts](#)

## Interview with Brian Tomasik

07 AUG 2014 • PETER HURFORD

Some people have committed a great deal of their lives to trying to best make the world a better place. I'm trying to sit down with some of these people and learn more about their thoughts and motivations.

Today, I sit down with Brian Tomasik. Brian worked at Microsoft as a computer programmer, earning a sizable salary that he then donated to important causes. Throughout this time, he also wrote on his blog, [Utilitarian Essays](#), and he is noted for being one of the first people to consider [earning to give](#) and [donating to vegan outreach](#), two concepts that are now commonly found in the “effective altruism” community.



**Peter Hurford:** Hi Brian. When we last talked, you were working on Bing at Microsoft. But I've heard that you've since left your job there to pursue other projects. What changed?

**Brian Tomasik:** Microsoft was a great place to work. My job was intellectually interesting, allowed a flexible schedule, paid well per hour, and brought me in contact with really smart people. I appreciated Microsoft's \$12K/year matching-donations limit (now \$15K) and the fact that I could use a treadmill desk in my office without people complaining. While earning money was great, I began to realize how many open altruism-related questions I faced in deciding where to donate. Because my approach to ethics is somewhat unique, I didn't know of anyone else I could pay to work on these questions for me. I knew that these topics would probably take many years to explore and were beyond what I could do in my free time. So, in early 2013, I began plans for starting my own organization to work on basic-research questions in altruism. My collaborators and I eventually decided to call it [Foundational Research Institute](#) (FRI). I left Microsoft in October 2013 to spend more time reading, writing, and thinking about issues like artificial intelligence, consciousness, cooperation, and future trajectories from the perspective of aiming primarily to reduce suffering.

-

**P:** That sounds like a lot to give up. With all your compensation at Microsoft, you could be funding a lot of people to do research instead of doing it yourself. Why do you think it might be higher value for you to work on these projects instead?

**B:** It's a tough decision, and I still think a lot about where my efforts would accomplish the most. Right now, FRI is my top priority for a few reasons.

Figuring out what you value, what actions best advance those values, and what charities best

contribute toward those actions is a lifelong process of exploration. It's easier if you have relatively mainstream values, because then you might take the recommendations of a charity evaluator or other friends who have thought hard about this question. However, that's less feasible in my case. I'm not aware of any existing organization with the foremost goal of reducing suffering that has systematically explored interventions to shape the far future in light of artificial intelligence and other crucial considerations. This is new ground that needs breaking.

Potentially I could have tried to hire others to study these questions for me, but in practice, I'm not aware of many other people I could have hired. (I'd love to find more!) Many of my collaborators on these questions feel that my contributions to this field have been non-replaceable. These people helped persuade me that my ideas were sufficiently valuable that they would be worth giving up my Microsoft income to explore further. I also personally feel that I contribute some perspectives that others have overlooked, and I need to exercise some degree of due diligence rather than just trusting the opinions of others who explore these questions.

Finally, it's worth remembering that the cost to hire a researcher is often more than her salary; it also includes the opportunity costs of what she would have been doing instead, which might include earning to give in her own right. The best researchers are likely to have higher opportunity costs. So it's not clear that my Microsoft income could have paid for *several* good FRI researchers.

The situation may change in the future. Perhaps in 5 years FRI will have additional researchers to explore ways to shape the future with an eye toward reducing the most suffering. It's also likely that I'll have a better overall picture of the world and diminishing marginal insights. At that point I can re-evaluate my options, and earning to give would be on the table. By that point the question of where to donate would be more clear, since I would then be able to donate to a self-sustaining FRI.

Alan Kay famously [said](#), "The best way to predict the future is to invent it." Analogously, if you have unusual values, one way to find an effective charity is to create it. I don't mean that FRI is likely to be more efficient by objective metrics like output or influence per dollar than other charities. The principle of mediocrity argues against that. Rather I mean that adjusted for value differences relative to other organizations, FRI's work is likely to be among the most useful per dollar for those whose focus is reducing suffering. And even if not, then FRI can point to other

charities that seem to be more effective than itself.

-

**P:** So it seems like you're intensely interested in the question of reducing suffering. Most of my friends aren't like this – instead, they're concerned with a specific kind of suffering, like getting rid of cancer, and even then, that concern doesn't inform *everything* they do. Why is suffering in general so interesting to you, as opposed to specific kinds of suffering, and why is that concern so dominating in your life?

**B:** I'll start with the second part – why is suffering so dominating? Subjectively, the reason is that when I see or imagine extreme suffering – such as being eaten alive or fried to death in a [brazen bull](#) – it seems overwhelmingly apparent that preventing such experiences is the most important thing in the world, and nothing else can compare.

This intuition seems clear enough to most of us when we imagine the suffering happening nearby. If someone was being tortured in a way that could be prevented in the room next door, few of us would hesitate to stop whatever we were doing and go help. But when distance and uncertainty stand in the way, this intuition fades, and people become preoccupied with goals like ensuring interesting, complex, and [awesome](#) futures. Of course, I get plenty distracted as well – doing so is only human, and it's necessary for our emotional wellbeing. But I know in my heart that these other pursuits are only instrumentally valuable, and nothing besides reducing extreme suffering really matters when people and animals are being tortured as we speak.

What I've described is analogous to Peter Singer's [drowning-child](#) thought experiment. But in addition to making the case for altruism, my analogy further argues for a kind of prioritarianism in which preventing extreme suffering is more urgent than, say, creating additional happy lives or beautiful art or other forms of value. Because torture-level suffering will likely always exist in some form somewhere within the reach of our influence to prevent it, prioritarianism is in practice equivalent to a focus on reducing suffering over and above other goals.

While I see this argument as relatively convincing on its own merits, I seem more persuaded by it than many people. I think one reason is that I experienced a great deal of physical pain in 2002, when I was [dealing with](#) a condition called esophagitis. The pain continued in mild form for a few years thereafter. I think this has rewired my brain to show me just how overriding suffering is compared with other goals. Matt Ball [expressed a similar sentiment](#) when describing his youth:

*Back then, I worried about abstractions and words and principles; I argued about exploitation, oppression, liberation, etc. I didn't take suffering seriously. Now, knowing what suffering really is, and knowing how much there is in the world, all my previous concerns seem - well, to put it kindly, ridiculous.*

It's easier to answer why I care about any form of extreme suffering rather than just particular forms. I can generalize from knowing that particular experiences I've seen or witnessed were agonizingly terrible to realizing that similar sorts of experiences from other causes would also be agonizingly terrible. I have not personally suffered nearly as much as most torture victims or dying wild animals, but I can take what I have experienced and multiply it in my imagination to attempt some degree of comprehension of the horror they have endured.

Based on my own experience and intuitions of others, I think extreme suffering is in some sense qualitatively different from mundane suffering. There's something so much worse about extreme suffering that even comparing torture with, say, stubbing your toe seems to trivialize torture. Only the really intense forms of suffering demand moral urgency.

-

**P:** Have you come to any important realizations from your research yet? What do you think the impact of your research is so far?

**B:** For concreteness, I'll define "my research" as things I've realized and written about in the last 1.5 years. There have been many course-changing insights. They've affected not just me but many of my Swiss colleagues and some other people I know through Facebook. The most tangible impacts may have been to shift where some people donate and where FRI applies future research resources, but the long-run changes to our frameworks for thinking about these questions are likely much bigger.

Some examples of insights:

- I now think the work of [the Machine Intelligence Research Institute \(MIRI\)](#) is more likely to reduce expected suffering than increase it, and I now think MIRI is among the best places to donate. In early 2013, I had no idea whether MIRI's work was more likely good or bad. I still have significant uncertainty, but I'd say there's maybe a ~70% chance I'll continue to support MIRI ten years from now.
- I think cooperation – both at the [international level](#) and the [interpersonal level](#) – is more

important than I had previously realized. AI arms races may be a significant driving force behind dangerous AI in coming decades. Interventions that improve cooperation, social norms, discourse, reflectiveness, and institutions for positive-sum improvements seem reasonably likely to be [robustly positive](#). I now support the principle of [coherent extrapolated volition](#) (CEV) for reasons of strategic compromise. Even though CEV will never happen in practice, we can work toward approximations, like democracy and philosophical sophistication.

- I've downshifted my probability for a [hard AI takeoff](#) in favor of [soft-takeoff](#) scenarios. This means that society as a whole is more likely to play a role in AI development than I would have thought, and military, economic, political, and social dynamics around AI are likely to be important.
- I've changed from seeing consciousness as having a narrow, relatively clear cutoff to being a broad, fuzzy continuum. I now realize that sentience should be considered a [graded property](#) of systems, and [every system is at least slightly conscious](#). Subjectively, I feel this has been one of the biggest Copernican shifts of my life, and I think I have a much better understanding of the issues at play than a year ago, even if other philosophers disagree with my conclusions. In practical terms, this realization has made it more apparent that the instrumental computations of even a rogue, uncontrolled AI would be somewhat morally problematic. I also now care a little bit about [plants and bacteria](#), although not nearly as much as I care about [insects](#).

These are just some highlights, and in reality, many of the pieces I've written since early 2013 represent updates in my views. That said, I have noticed some decline in insights per unit time recently, which may suggest increased convergence in my thinking. On the other hand, I thought my views were pretty well converged in 2010-2012, but then I had a huge number of discoveries in 2013. I expect that many priority shifts due to known and unknown unknowns remain on the horizon.

-

**P:** How did you get started at writing on [utilitarian-essays.com](#)?

**B:** In 11th grade, I had an English teacher who focused his class around philosophy. Most high schools never teach philosophy, so I was fortunate to be exposed to the subject. My teacher assigned his students a 500-word journal entry every week, typically on a philosophical topic.

This got me into the habit of writing philosophical essays, and I continued doing so on my own after the class ended.

One of the philosophers that we learned about was Epicurus, and his idea of subjective experience as the ultimate good resonated with me. I noticed how, selfishly, I only cared about my own feelings, and hence altruistically I would care about the feelings of everyone who had feelings. I only really thought about humans at this point, because I mistakenly believed that animals probably didn't have feelings due to lacking language and true self-reflection. (This position is not dissimilar from [Eliezer Yudkowsky's](#), though obviously I didn't have his degree of sophistication at the time.)

I also noticed how people were often irrational in apportioning resources (e.g., fighting terrorism getting vastly more funding than addressing world hunger, even though the latter kills orders of magnitude more than the former). I suggested using cost-benefit analysis to evaluate policies, taking into account opportunity costs and counterfactual reasoning.

In spring 2005, my public-policy teacher used the word "utilitarian" in class, and I didn't know what it meant. Once I got home, I looked it up, and I discovered a philosophy called "utilitarianism", including a wonderful site of Utilitarian Philosophers by Pablo Stafforini. I began eating up the Peter Singer articles on that site. One topic that I found particularly important was Singer's defense of animal sentience, which revolutionized my view of the world's priorities. I began writing more about utilitarianism in my private essay collection.

Caring about animals led me to discover Vegan Outreach, from which I contacted Matt Ball, who passed me along to another utilitarian friend of his. That friend introduced me to the ideas of Yew-Kwang Ng, Ray Kurzweil, Nick Bostrom, David Pearce, and many others.

In June 2006, I read David Pearce's [Hedonistic Imperative](#), after which I wrote to David with comments. I included the collection of utilitarian-related essays that I had been accumulating since roughly 2004. David wrote back: "My first (and utilitarian!) response was: how can I encourage you to get a website?" David soon gave me the domain utilitarian-essays.com and still helps maintain my site to this day.

-

**P:** It seems like you might have more influence if you wrote for a wider audience, such as at [LessWrong](#). Why haven't you chosen to do that?

**B:** Writing on LessWrong or related forums is a reasonable suggestion. I do share my essays on Facebook when they're published. The reasons I haven't posted on LessWrong more are:

- I think my essays look more authoritative on a stand-alone site than in a forum setting
- Sometimes what I write about is not of sufficient interest elsewhere, and I don't want to bore people
- It might look bad if I didn't reply to comments, but I'm also not sure that reading comments and engaging in discussions is the best use of time. It may or may not be depending on the issue. In general I try to read more famous authors or at least peer-reviewed publications, which I find can be more insightful per paragraph than comments on Internet forums, though this depends a lot on the topic and the forum. For some relatively uncharted questions, forum comments are like gold.

-

**P:** A lot of your writing these days seems to focus on Wikipedia, and you have racked up an impressive list of contributions there. Could you briefly outline why you think editing Wikipedia is such an important outlet for your research?

**B:** I wrote [an essay](#) listing many reasons to favor contributing to Wikipedia on altruism-relevant subjects, including not reinventing the wheel by writing what has already been written, large readership, credibility, peer review, and being forced to maintain a balanced perspective on a topic. When I want to learn about an issue, Wikipedia is typically the first place I go, and I suspect the same is true for many people reading this interview. Wikipedia is an immense public good that we may often take for granted.

I'm trying to take a big-picture view of many issues before getting deeply engrossed in technical minutiae, so that I can pick areas of focus that I find most important. This contrasts with the standard approach in academia where you often have to zero in on a microscopic piece of one subject area in order to make novel contributions. Because I'm mostly not making novel contributions but rather am absorbing knowledge that already exists, Wikipedia is a great place for me to be. It allows me to accomplish something with my reading beyond just privately improving my own understanding of the world.

On a few occasions I've had the experience of proposing a research topic and then hearing people reply, "Oh, that has been discussed in papers A, B, C, D, and E." I think to myself, "Thanks

– that’s great. But I need to get the information from papers A-E into my head. That takes time. In addition, there are probably papers F-Z that would also be relevant that I haven’t found yet.” Most of the effort in cause-prioritization research is not in inventing novel findings but rather in accumulating nuggets of insight from a vast sea of information that already exists. Wikipedia is a natural way to attack this problem because Wikipedia is the world’s definitive summary of a topic. If you read papers A-E, you should consider adding a sentence or two describing each of them to relevant Wikipedia articles, so that others who want an overview can read your summaries, get a big-picture understanding, and have links to the papers if needed. And because Wikipedia is coherently organized, readers will be able to absorb the insights in a sensible and content-addressable fashion. In many ways, Wikipedia is the ultimate platform for altruism research.

Some altruists prefer to summarize existing knowledge on their own websites in order to increase traffic to their organizations. In general, effective altruists may underestimate the value of Wikipedia because it mainly benefits people outside their movement. I think these ideas rely on a [false assumption](#) that effective-altruist organizations are vastly more important than what everyone else in the world is doing. It’s easy to see your own work with rose-tinted glasses and assume that the efforts of others aren’t as useful. This is a bias on our part. Many people throughout the world are doing immensely valuable projects, even if they don’t realize it or label those projects as altruism. Wikipedia contributions help many people whom we will never meet and whose future contributions we’ll never find out about; that doesn’t mean we didn’t facilitate their work.

Of course, the world contains many harmful and risk-increasing projects as well. For this reason, I recommend “differential Wikipedia progress” in which we contribute mainly on topics where additional knowledge seems likely to be used beneficially. That said, I suspect ~95% of Wikipedia articles are either net good for the world (e.g., the [wild animal suffering](#) article) or basically neutral (celebrity biographies, TV episode guides, etc.). A few of the contributions I’ve made to Wikipedia have been just for fun or my own personal edification rather than altruistic impact.

-

**P:** Contributing to Wikipedia feels like a good idea for someone doing a lot of research on their own, but do you think there are ways a more typical person could still contribute? Would that be worthwhile?

**B:** It's actually not hard to make small contributions to Wikipedia. If you read an academic paper or news article on a topic you want to share with others, you can do a search like {author-surname keyword site:wikipedia.org} to see if what you want to cite is already mentioned somewhere. If not, you can find a relevant article and add a sentence or two with your citation. The whole process could take just a few minutes. If it feels daunting at first, that may be just because you're not used to it.

I think Wikipedia is one of the easiest ways to make a meaningful contribution in one's free time, and the work can be neatly packed into small, discrete chunks. But of course there are other things you can do as well. I have [a page](#) describing some ideas for spreading concern for wild-animal suffering, but many of those suggestions apply to other causes. And you can probably invent many ideas on your own.

I find that it's often better to just try something on your own and see how it goes than to over-analyze whether it's the best thing to do or create bureaucracy in order for it to get done. Keep learning, experimenting, and changing course based on how things turn out. I'm thinking of my personal experience when saying this, but we can see similar patterns in companies, where big companies stereotypically may be slow to innovate because they over-analyze and over-bureaucratize. Wikipedia's model seems better: Be bold, just do what you planned to do, and let other people get involved in an organic, asynchronous fashion. That said, I can imagine legitimate debate about these issues. For instance, if you don't have bureaucracy, it may be harder to divide labor with comparative advantage in mind. There are structural and efficiency reasons why small, fast companies become big and slow.

-

**P:** I also saw that you've recently been working on a [small video series](#). What do you think you've gotten out of this?

**B:** At the moment I have two videos on YouTube. I created them because I woke up one day and decided it would be fun to make a video and see how it would come out. This relates directly to what I said in my previous answer: I like to just try stuff in a low-cost fashion and learn from the experience, without worrying about its apparent effectiveness. These particular videos were decently successful (together they got more than 500 views), but even if they weren't, it was a fine use of a few hours to create them.

This approach of trying stuff to learn what it's like and see how it goes is also a reason why I

decided to submit a paper to an academic journal, build [AdWords campaigns for some organizations](#), and create the [FRI podcast](#). Doing many things can improve my understanding of what different parts of altruism work are like. This is also something I appreciate about being on the board of [Animal Charity Evaluators](#): I can get a first-hand glimpse at the internals of what a charity does in a way that I wouldn't have if I were just an external donor.

-

**P:** Cool. With all this writing and time to yourself to do whatever you think is high priority, you must have a less typical schedule. What would you say your typical day is like?

**B:** I don't work on a plan or schedule. Instead I do whatever I'm in the mood for at a given moment. This makes work effortless and probably improves productivity. Sometimes I get in the flow of writing an essay or Wikipedia article and try to ignore everything else until I'm done with it. On other days, I don't have anything specific to finish and instead try to "pick up the crumbs", i.e., do those little tasks that built up while I got distracted with a bigger project.

Taking an average across all days might give something like the following breakdown, though I don't keep track explicitly:

- 10 hours: sleep
- 1.5 hours: physical maintenance and manual chores (during some of which I can listen to podcasts)
- 2.5 hours: reading and responding to emails and Facebook discussions
- 1.5 hours: exercise while watching YouTube videos or a movie
- 2 hours: random small tasks, including little jobs for other people and monitoring Google Grants, Wikipedia updates, website traffic, my to-do list, etc.
- 2.5 hours: tangible altruism-related work like fixing a website, Animal Charity Evaluators board meetings, or discussing strategy
- 4 hours: reading and writing (with more emphasis on the reading part).

I think it's important to read a lot (or listen to audio content) in order to produce useful writing. [Elijah Young has said](#): "If you can't think of anything to write, you're probably not reading

enough.” Of course, some of my writings are based on personal experience, long-term observations, and/or theory. But especially on empirical topics, good writing generally requires good background knowledge.

-

**P:** Some people feel like “saving the world” all the time is too much and they need some balance in their life, so they might give only 10% of their income and 10% of their time, instead of all of it. It’s comforting that you make sure to spend a large amount of time on self-maintenance and sleep. What are your thoughts on this balance?

**B:** I think the question assumes a premise that is (happily) misguided. There’s not a hard distinction between “save the world” work and everything else we do. Our altruism stands on the shoulders of the rest of our lives. We need to [learn](#) how the world works at a deep level to direct our altruism in helpful directions. This learning takes many decades and is happening all the time through our experiences. Likewise, earning money and developing social connections are other resources that we build as we go through life. Certainly some uses of time are more valuable per hour than others, but even relaxation is important when it keeps our moods positive and ensures that we enjoy our altruistic activities.

I think altruism should generally be fun, and if it’s not, you might consider how else you can contribute to reducing suffering that is fun. There are so many ways to make a difference that the process doesn’t need to be a burden. There’s no law stating that good accomplished is proportional to amount of sacrifice on your part (even if perhaps social praise increases with degree of sacrifice because of the need to generate incentives for unpleasant tasks like military service or providing emergency aid). And if altruism is fun, then the problem of how much of it to do compared against other leisure activities becomes less potent.

One way to make altruism more exciting is to become friends with other altruists. It’s natural to be more interested in what your friends are interested in. For me, the boundary between socializing and saving the world often breaks down; they are often the same thing.

Finally, as noted previously, I “waste” some time on so-called non-altruistic activities as well. In the past I found this troubling, but I’ve grown to realize that it’s part of being a finite human who needs many forms of rejuvenation. If you feel burned out, you’re probably pushing yourself too hard. Let yourself relax a bit without feeling guilty.

On the flip side, if you think about altruism very little, try increasing that somewhat. 10% of your effort seems a reasonable goal. Then you can evaluate where you are and see if you're inspired to do more. But remember that "altruism time" is not clearly defined. Ask yourself whether one of your ideal hobbies or careers is altruistically valuable even if it's not what most effective altruists spend their time doing.

**P:** Thanks, Brian. I really appreciate that closing thought. I'm glad to have followed the work you're doing and wish you continued success! I'll eagerly be following what you do next.

TAGGED: INTERVIEWS

18 Comments **Everyday Utilitarian**

 Login ▾

Sort by Best ▾

Share  Favorite 



Join the discussion...



**Pablo Stafforini** · 4 months ago

Based on my own experience and intuitions of others, I think extreme suffering is in some sense qualitatively different from mundane suffering. There's something so much worse about extreme suffering that even comparing torture with, say, stubbing your toe seems to trivialize torture. Only the really intense forms of suffering demand moral urgency.

What's your response to the "continuity argument" that I and others have advanced against this conclusion?

Also, how do your views about the unique moral importance of extreme suffering cohere with your belief that the sort of panpsychism you defend has such moral significance? Plants, bacteria, video game characters, etc. presumably endure only very mild forms of suffering, so your position would imply that there is no moral urgency to help them. Thus, I'm not sure I understand why your realization that every system is at least slightly conscious "has been one of the biggest Copernican shifts" of your life.

Finally, have you considered writing a Wikipedia article on "differential Wikipedia progress"? :-)

1 ^ | v · Reply · Share ›



**Brian Tomasik**  Pablo Stafforini · 4 months ago

Hi Pablo :)

You've already seen discussion of continuity here: <http://felicifia.org/viewtopic...> I

may be willing to throw out the continuity axiom out of the set of four. After all, physics is discrete, and it's plausible that one extra neuron somewhere could lead to a qualitatively different valuation. Indeed, it may be that a single neuron can lead to qualitatively different brain behavior in other contexts, such as on the boundary between whether a stimulus becomes globally broadcast ( <https://en.wikipedia.org/wiki/...> ). In any case, I would emphasize that for practical purposes this doesn't really matter. Whether you accept continuity or not, as long as you disvalue suffering a lot, similar conclusions follow.

Your second question has been asked before, as a result of which I wrote this: <http://www.utilitarian-essays....>

I could potentially write a Wikipedia essay on differential progress, but obviously an article on that topic would lack notability. I feel as though my current essay on differential intellectual progress ( <http://foundational-research.o...> ) combined with <http://www.utilitarian-essays....> are sufficient for now, but I welcome someone else writing more. :)

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago

Your second question has been asked before, as a result of which I wrote this: <http://www.utilitarian-essays....>

Interesting. But what is the rationale for assigning independent weight to suffering relative to the organism's typical levels, in addition to absolute levels of suffering? I thought the reason you (together with Dave and others) thought that suffering above a certain threshold was so much more morally important had to do with how these experiences felt like. Yet what you say in this section of your essay implies (if I understand you correctly) that two phenomenally indistinguishable instances of suffering could be valued differently depending on the complexity of the subject that experiences them. Thus, a very mild form of pain would matter when experienced by a very simple form of consciousness, but not when experienced by a more complex one (because that painful episode would be much more intense than the typical experiences of the former, but not of the latter).

1 ^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

It's a subtle point, but our difference stems (like so many things :) ) from our differing views on consciousness.

I added two paragraphs to clarify: <http://www.utilitarian-essays....>

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago



I see. It's still not clear to me how on your view the value of an instance of suffering depends both on absolute and relative levels of suffering. It seems that you should only care about how important the experience is relative to its subject, regardless of how simple or complex it is. If the video game character cares just as much about avoiding torture as does the human who plays video games (relative to other things that they both care about, such as alleviating a headache), why give so much more weight to the concerns of the latter?

1 ^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

"It seems that you should only care about how important the experience is relative to its subject, regardless of how simple or complex it is."

Well, that is a tempting view, but then the preferences of superstrings would dominate our calculations. Also, keep in mind that there is no such thing as objective interpersonal comparisons of utility. Still, we do have some rough intuitions. For instance, it seems obvious that an animal living in a factory farm is vastly more painful than someone enjoying chicken nuggets is pleasurable.

The reasons for my favoring bigger brains more are those discussed in <http://www.utilitarian-essays....>, though I admit I am uncomfortable with this enterprise and am not as enthusiastic as many LessWrongers who take it as obvious that valuation should scale with brain size/complexity.

1 ^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago

After all, physics is discrete, and it's plausible that one extra neuron somewhere could lead to a qualitatively different valuation.

Mmh, I'm not sure I understand this reply. How exactly is the discreteness of physics related to the continuity argument? The latter says that it's implausible that two different but arbitrarily similarly intense painful experiences differ so much in value that a single instance of one could be more disvaluable than arbitrarily many instances of the other. This is not an argument intended to establish the conclusion that it's physically impossible for a brain to make these seemingly implausible value judgments.

(Feel free to ignore this comment if you are not interested in further discussing this topic.)

^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

Sorry for the confusion. :) I added this footnote to elaborate on what I was thinking: <http://www.utilitarian-essays...>

In other words, the function that maps physical states to "suffering" values may be discontinuous given the finiteness of physics (or, more relevantly, neurons). Actually, given the finiteness of physics, this function is necessarily discontinuous between any two physical configurations, but I guess I mean it could have a really steep change at some point.

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago

Thanks for the clarification. I agree that such "bifurcations" are possible, but I still don't see how they constitute an answer to the argument. You can imagine yourself as currently being presented with the pair of painful experiences that would trigger the discontinuous moral judgment. You'd still have to give an answer to the person that confronts you with the continuity argument.

Maybe your position is that this judgment is a brute fact, that is not susceptible to further justification. The problem with that view is that we make all sorts of judgments about specific cases that are inconsistent with other specific judgments we make. Even if you think that all judgments are ultimately arbitrary, to remain consistent you need to make sure that these judgments do not conflict with some of the judgments that you'd make in other contexts.

^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

Sorry for the confusion. :) (I should make that sentence my email signature or something.)

I meant that the bifurcation can be in the dynamics of how a brain responds to a given objective stimulus. Say between a 10 mm pin and an 11 mm pin, the difference in number/intensity of pain signals triggers qualitatively different neural dynamics. Those qualitatively different dynamics might be such that they feel astronomically worse. So I'm talking about (the possibility of) a bifurcation in the experience of suffering.

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago



I'm talking about (the possibility of) a bifurcation in the experience of suffering.

Ah! But the continuity argument involves making a series of comparisons involving pains *that feel very similar in intensity*. Maybe your brain is wired in such a way that you are constitutionally incapable of experiencing some of the required pain intensities. But surely this is something contingent on the architecture of your particular brain, or the brain of particular species. All that would show is that you wouldn't in practice be able to experience all the pains postulated by the argument. How does that differ from, say, the inability of a person in a wheelchair to be able to rescue the drowning child in Singer's famous argument?

1 ^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

"Maybe your brain is wired in such a way that you are constitutionally incapable of experiencing some of the required pain intensities."  
Yes, that's the possibility I'm suggesting, though I don't hang my (possibly) threshold-utilitarian hat on it being true.

"But surely this is something contingent on the architecture of your particular brain, or the brain of particular species."

Well, if you're talking about agent-relative utilities, then for each organism that has a bifurcation like this, there will be a corresponding threshold.

Could there be organisms that have a different mechanism of suffering that lacks the bifurcation? Plausibly, especially if one's view of suffering is sufficiently cosmopolitan. In such a case maybe they would always be either above or below threshold? This is a very good point, my dear sir, and I should think about it further. Fortunately, I already have a hat rack.

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago

Okay, last email for now.

Well, if you're talking about agent-relative utilities, then for each organism that has a bifurcation like this, there will be a corresponding threshold.

I don't see why this is so. The continuity argument requires you to consider a series of increasingly painful experiences, and to ask yourself whether it's plausible that some moral discontinuity exists

along that phenomenal continuum. If you are constitutionally incapable of having some of those experiences yourself, this may interfere with your ability to represent those experiences in your imagination, which in turn might interfere with your ability to evaluate them morally. But it has no implications for the substantive moral discussion over whether a sudden jump in value can occur along a continuum of phenomenal experiences.

An analogy. Suppose you had a moral theory on which light is bad the brighter it feels. And suppose you thought that just one flash of a very bright light was worse than any number of very dim light flashes. Then suppose someone presented you with an analogous continuity argument. Would it make sense for you to reply that, because the way your visual system is constituted, at certain light intensities an increase in just one lumen (or whatever) is associated to an "astronomical" increase in subjective brightness? This seems only indirectly relevant to the issue at hand. At best it would show that your imaginative capacities are inadequate for you to vividly picture all the steps in the argument. But it's not really central: the argument is not about how lumens relate to brightness, but about how brightness relates to value.

^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

"If you are constitutionally incapable of having some of those experiences yourself, this would interfere with your ability to represent those experiences in your imagination, which in turn may interfere with your ability to evaluate them morally. But it has no implications for the substantive moral discussion."

Relative to my brain construction, such intermediate phenomenal experiences cannot exist (assuming the bifurcation hypothesis). It's not just that I can't imagine them, but there is no such thing as my brain experiencing an intermediate value. One could construct different minds lacking bifurcations that experience intermediate values, but then the threshold question comes up for them, not for me.

"Would it make sense for you to reply that, because the way your visual system is constituted, at certain light intensities an increase in just one lumen (or whatever) is associated to an 'astronomical' increase in subjective brightness?"

Yes. :) Not sure why you disagree.

If the objection is that the objective brightness is continuous, this doesn't carry over to the case of suffering, because the whole point is to observe the subjective intensity, not the objective intensity.

to observe the subjective intensity, not the objective intensity.

^ | v • Reply • Share ›



**Pablo Stafforini** → Brian Tomasik • 4 months ago

I though I was going to sleep soon, but given that I'm still awake, I might as well write another reply. :-)

If the objection is that the objective brightness is continuous, this doesn't carry over to the case of suffering, because the whole point is to observe the subjective intensity, not the objective intensity.

What do you mean? Let's go back to the case of suffering, since my analogy wasn't that illuminating (no pun intended).

Assume an absurdly crude model on which one neural firing in some brain area is associated to a painful experience of intensity 1 (in arbitrary units), two neural firings are associated to a painful experience of intensity 2, and so on. In normal brains, this relationship holds all the way to the most painful episode humans can experience. But in abnormal brains, the relationship breaks down at a specific point. When one thousand neurons in that person's brain fire, the intensity of the associated experience is not 1,000, but, say, 1,000,000,000. And when more than a thousand neurons fire, the intensity of the associated experience is still higher than 1,000,000,000.

So what? The continuity argument is not concerned with the relationship between matter and consciousness, or more specifically, between number of neural firings and intensity of subjective experience. Rather, the continuity argument is concerned with the relationship between subjective experience and moral valuation. If your brain is such that you can't experience a painful episode of 1000 units of intensity, then you are just not very well positioned to engage in the imaginative exercise needed to assess the argument. But as far as I can see, no substantive moral conclusions follow from this fact.

1 ^ | v • Reply • Share ›



**Brian Tomasik** → Pablo Stafforini • 4 months ago

Ok, I see your point now. :)

What I was trying to explain was that torture/etc. could be qualitatively different from tons of pinpricks because torture exceeds the threshold, while pinpricks don't. Say  $\text{value}(\text{pinprick}) = -1$  and  $\text{value}(\text{torture}) = -\infty$ . This is an argument about how bad objective stimuli are. Hence, no number of pinpricks can outweigh torture (for this hypothetical bifurcated mind gives a hypothetical lexical difference in

hypothetical bifurcated mind given a hypothetical lexical difference in suffering/disvalue between its two equilibrium states). This could answer torture vs. dust specks and most other practical cases where these arguments come up, because the practical cases are about objective stimuli.

1 ^ | v • Reply • Share ›



**Jess Riedel** • 5 months ago

You quote Matt Ball and link to an essay at [utilitarian-essays.com](http://utilitarian-essays.com), but following that source back leads to a broken link: <http://www.veganoutreach.org/a...>

1 ^ | v • Reply • Share ›



**Brian Tomasik** → Jess Riedel • 4 months ago

Ah, thanks, Jess. :) I've now replaced that link with its Internet Archive version: <https://web.archive.org/web/20...>

^ | v • Reply • Share ›

ALSO ON EVERYDAY UTILITARIAN

WHAT'S THIS?

## Personal Review for December 2013 - February 2014

3 comments • 9 months ago



**Pablo Stafforini** — I don't have strong views, except that diversification may be justified for learning purposes. True, in ...

## The Making of the Personal Review: Behind the Scenes

1 comment • 5 months ago



**Pablo Stafforini** — I also track my time, though I use Toggl rather than Google Calendar. One useful rule I've adopted ...

## Sunday Links #26

2 comments • 7 months ago



**peterhurford** — I think the discomfort comes from the implicit assumption that I am a rapist, or that I am more ...

## Career Advice for Phyllis Schmidt: Choosing Between Masters, ...

2 comments • 8 months ago



**Holly Morgan** — "American-based The Life You Can Save might be of particular interest to Phyllis. Though I don't ...

[Subscribe](#)

[Add Disqus to your site](#)

[Privacy](#)

---

Everyday Utilitarian by [Peter Hurford](#). Powered by [Jekyll](#) and hosted on [GitHub Pages](#).

