Cambridge Quarterly of Healthcare Ethics (2025), 34: 4, 536–546
doi:10.1017/S0963180126100164

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The Wrong Motives for Potentially Harming a Being

Simon Knutsson[1,2]

[1]Stockholm University, Sweden and [2]Center for Reducing Suffering, USA
Emails: simon.knutsson@philosophy.su.se; simonknutsson@gmail.com

Abstract

I aim to make progress on the following question: When is it morally wrong to risk harming another being? I will pay special attention to the use of the roundworm *Caenorhabditis elegans* (*C. elegans*), but my points are relevant to other situations and beings. I focus on the *motive* or *purpose* for exposing a being to risk of harm. I argue that it is morally wrong to potentially harm a being for the sake of others' positive well-being or for a purported good such as knowledge (that is, knowledge for its own sake). The practical ramifications include that there is a moral hurdle of justification for potentially harming another being, and the justification cannot be others' positive well-being or a good such as knowledge. Essentially, if the use of a being can be morally justified, it needs to be justified based on reducing enough ill-being or unpleasantness. One should recognise the creation and use of beings as a moral issue and view any warranted potential harm as a regrettable lesser evil. If feasible, it is desirable to use alternative methods that carry less risk and do not involve potentially sentient beings.

Keywords: *Caenorhabditis elegans*; *C. elegans*; counterbalancing suffering; ethics; intentions; motives; painience; risk; sentience

Introduction

Consider a being that might have the capacity to be badly off. In other words, the being can perhaps, as far as we know, be harmed. When is it morally wrong to risk harming such a being? I will usually take the roundworm *Caenorhabditis elegans* (*C. elegans*) as the example being, and I will pay special attention to the use of non-human animals in research. That said, my moral points are relevant to other beings and situations.

When an individual is exposed to the risk of harm, there needs to be some chance of a benefit that justifies that risk. I will focus on the intended benefit. That is, I will focus on the motive or purpose for exposing the being to risk of harm. I will argue that it is morally wrong to potentially harm a being (including *C. elegans*) for the sake of others' positive well-being or for a purported good such as knowledge.

The structure of this paper is as follows: Section "Preliminaries" contains preliminaries, mainly some conceptual and terminological clarifications. In Section "There are ethical concerns with using and creating a being such as *C. elegans*", I set the stage further by noting that also for a being such as *C. elegans*, there are ethical concerns with using that being. Section "The moral relevance of motives" is about the moral relevance of motives. In Section "My proposal: Do no harm for goods", I put forth a principle about when it is morally wrong to potentially harm a being. I present arguments for that principle in Section "Arguments for the principle" and reply to objections against it in Section "Replies to objections against the principle". Section "Practical ramifications" suggests practical ramifications, and Section "Conclusions" concludes.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

Preliminaries

For moral purposes, a key feature of an individual is their capacity for ill-being. In other words, the individual can be badly off. I take the term "ill-being" to be interchangeable with the terms "negative welfare", "negative well-being", and "negative quality of life".

There are different theories of ill-being. For example, one theory is that unpleasant experiences are bad for an individual, while another theory is that frustrated preferences are bad for an individual.[1] In my view, unpleasant experiences, broadly understood, are the most obvious and important constituents of ill-being, as well as the core feature for moral purposes, so I will often speak about unpleasantness. I use the term "unpleasant" in a wide sense for any aversive, bothersome, disturbing, or negative experience. An unpleasant experience may be mild, but I intend to cover the whole range of severity. For example, experiencing the worst form of torture is extremely unpleasant.

In this paper, the possibility of harming a sentient being is important. One definition of sentience is capacity for valenced experience.[2] I take negatively valenced experiences to be the same as unpleasant experiences, and I gather that it is the possibility of *negatively* valenced experiences that gives rise to the moral issue of how to deal with the possibility that a being might be sentient. The moral concern is not that a being such as *C. elegans* can have pleasant experiences. So, sentience is important in this paper, but, more specifically, it is *painience* in the sense of capacity for unpleasant experience that is important.[3]

When I talk about "harm", I have in mind ill-being and unpleasantness. I do not have in mind, for example, having fewer goods than one could have had.

Positive well-being is often thought of as the opposite of ill-being. Positive well-being concerns what has positive value for an individual for its own sake. For example, some think that experiencing pleasure is good as an end for an individual (aside from any reduced ill-being or other beneficial effects of the experiences). To be clear, getting well after an illness or feeling better does not imply having positive well-being because one may have merely reached an improved state that is not positive in itself. Besides positive well-being, one can also hold that there are goods such as beauty, biodiversity, and knowledge that are good for their own sake above and beyond any individual's well-being, experiences, and interests. As mentioned, I will argue that it is wrong to potentially harm a being for the sake of someone else's positive well-being or such other goods. Among other purported goods, I will focus on knowledge rather than, say, beauty because knowledge seems more relevant to the use of animals in research, but it is also wrong to potentially harm a being for the sake of beauty.[4]

There are ethical concerns with using and creating a being such as *C. elegans*

In the literature on *C. elegans*, there are statements of the kind that this roundworm can be used without ethical concerns.[5,6,7] One can read such statements in at least two ways: First, they can be read as describing existing regulations or other practical obstacles to the use of the being in question (e.g. there is no existing ethics regulation hindering the use of *C. elegans* in research). That is not my focus here. My focus is on a second reading, namely to read the statements as saying something normative about ethics (e.g. there are no ethical concerns or issues with using *C. elegans*).

As a claim about ethics, it sounds very strong to claim that a being such as *C. elegans* can be used without ethical concerns. We can compare this to someone saying that an activity has absolutely no risk, to which we might react by questioning whether there really is no risk at all of any kind.

Here are a few examples of ethical concerns with damaging a being such as *C. elegans* (or creating such a being and thereby exposing the being to the potential harms that occur in life). Of course, there is the risk that the being might have the capacity for ill-being or unpleasant experience. Another ethical consideration is opportunity cost. That is, perhaps the agent should do something more useful instead. If one could reduce more suffering by doing something else, that plausibly gives one a moral reason to do that instead. Other potential ethical considerations pertain to what the acts in question express about the agent. For example, using a being might express a lack of moral attention, caution, compassion, consideration, or responsibility. There is also the question of which attitude is morally appropriate. For example, it may be morally appropriate to view one's practice as regrettable and to be bothered by

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

aspects of it. Another example of a potential ethical concern is the wider effect of how an agent is and behaves. For instance, perhaps acts that deliberately expose beings to potential harm contribute to the continuation of harmful attitudes, ideas, or norms. For any of our characteristics and behaviours, including any treatment of non-human animals, we can ethically assess whether we contribute to a practice or norm that has negative effects. If we do contribute to that, why would that be ethically justified? Finally, one can ethically assess the motives of the agent, which is the focus of this paper and something I will talk more about in the next section.

In sum, as a starting point for this paper on the ethics of potentially harming a being such as *C. elegans*, it seems safe to say that doing so raises plenty of ethical issues. It is a substantive area for ethical inquiry with many angles tying into different strands in moral philosophy.

The moral relevance of motives

Here is an example to convey that motives matter morally at least sometimes: It seems morally nice or at least morally permissible to give coloured pencils to a child to cheer them up or contribute to their development. In contrast, to give the pencils to the child in order to make another child jealous and sad seems morally wrong.

The philosophical literature contains other examples of wrong-making motives. Sverdlik brings up various acts motivated by the hope of being paid, cruelty, or racism.[8] One of the examples is that it is wrong to refuse to sell one's house to someone because of their skin colour while it is permissible to refuse to sell it due to doubts about their creditworthiness.

There are different ways to think about the moral role of motives.[9,10,11] One option, which I will not endorse, is to hold that an act's rightness or wrongness does *not* depend on the motive with which it was performed.[12] Neither will I analyse the moral role of motives in terms of the consequences of the motives. For example, I will neither hold that the moral qualities of an *act* depend on the consequences of the motive of the act, nor that the moral qualities of a *motive* depend on the consequences of the motive.[13] That said, one can probably get a position similar to the one I propose in this paper even if one only considers the consequences of acts and motives, as long as one gives enough weight to reducing suffering.[14]

Similarly, I will not think about the moral role of motives in terms of the final (non-instrumental) or intrinsic value of the motive or of an action taken from the motive.[15,16,17]

I mean to leave it somewhat open where to theoretically locate the moral role of motives. My concern is not so much theoretical normative ethics about the exact moral role of motives in general. Rather, I wish to focus on the more substantive issue of whether some specific motives are morally wrong or problematic. The principle I propose in this paper (see Section "My proposal: Do no harm for goods") is about the moral wrongness of potentially harming beings for the sake of various ends. The most straightforward understanding of this principle might be that it is about the wrongness of acts where the motive is sufficient to make the act morally wrong. Still, I do not have a firm stance on which of the following hold: the motive is sufficient to make the act wrong; it is wrong or morally bad to have a certain motive or intention (even if no act is carried out); or having the motive or being disposed to act with it is a moral character flaw. Probably all of them hold, but regardless, as long as one of them holds, having the motive or acting with it is, in a broad sense, morally undesirable.

My proposal: Do no harm for goods

Here is a rough formulation of my ethical proposal: let us avoid potentially harming an individual for others' pleasure or preference satisfaction, or for a purported good that goes beyond what matters to individuals.[18] More precisely, I propose the following principle:

> **Do no harm for goods:** It is morally wrong to potentially harm a being for the sake of someone else's positive well-being or for a good such as knowledge.

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

In the rest of this section, I will make some comments and clarifications about the principle. I say "potentially" harm because it might be uncertain whether the individual can have ill-being or experience unpleasantness, and it might be uncertain whether an action would result in unpleasantness even if the individual can experience unpleasantness. There are different ways to potentially harm an individual. One way is to directly damage an individual, by, for example, cutting their body, but I also wish to cover, for example, creating an individual and thereby exposing the individual to tissue damage, starvation, death, and so on (even if no one is actively inflicting those things on the individual).

The principle is general in that it is not restricted to beings that are, for example, sufficiently likely to be sentient. It is not a matter of balancing the likelihood of the negative with the benefit of the positive, but rather a general moral prohibition of this risk-imposition and tradeoff. The principle morally prohibits potentially harming any being for the sake of someone else's positive well-being or for a good such as knowledge. Of course, the principle covers beings that are likely to be sentient, but there is no likelihood threshold for inclusion. The principle's prescription does not hinge on, for example, the scientific evidence for the ability to feel unpleasantness or which theory of phenomenal consciousness is correct. Essentially, if there is any risk that a being can be harmed, then the principle prescribes that we should not risk harming the being for the sake of someone else's positive well-being or for a good such as knowledge.

There are ways to adjust the principle so that it is less general or more modest. I will not do so, but I will mention a few ways in which it could be done. One way is to specify that it is *non-consensual* acts of potentially harming that are morally wrong. A second way to narrow it is to say that it is wrong to potentially *severely* harm a being for goods. A third way is to not claim that it is morally *wrong* but rather just "morally *problematic*" or "an ethical *concern* that *speaks against* the action". A fourth way is to add the qualification that it is, *in general*, morally wrong to potentially harm a being for the purposes in question, although perhaps not in all cases. Having mentioned those four ways to adjust the principle, I wish to note that when it comes to the use of non-human animals, it seems that an adjusted principle would still be relevant even if all these changes were made. They seemingly do not consent, the potential harm is often severe, the potential harming would still be morally problematic (or the like), and even if there are exceptions, the direct, intentional damaging of a being seems to be a core case.

The part of the principle that reads "for the sake of" could be understood in terms of motive, intention, purpose, end, or reason for the action. I usually speak of motives, but the other options may be at least as viable. The part of the principle that concerns someone else's positive well-being refers to anyone besides the victim (e.g. the agent doing the potential harm or a third-party beneficiary).

The principle says that it is morally wrong to potentially harm a being when someone else's positive well-being or a good such as knowledge is the *only* motive. What does the principle say when an agent also has other motives? We can compare to the example I mentioned in Section "The moral relevance of motives" in which it was wrong to refuse to sell one's house to someone because of their skin colour but permissible to refuse to sell it due to concerns about their finances. What if the seller's concerns about the buyer's finances were sufficient reasons for the seller to refuse to sell the house, but the seller also had some racist motives that contributed (but did not tip the scales)? It seems that we want to say that the act was morally justified based on the financial concerns, but that the racist motives are still morally relevant. As I see it, acting with those racist motives is morally wrong even if the financial concerns already settled the matter, and the morally right thing is to not have and not act with racist motives. Along the same lines, I would specify the "Do no harm for goods" principle as follows: It is morally wrong to potentially harm a being when at least one of the motives is to increase someone else's positive well-being or create a good such as knowledge.[19]

I am not granting that there is such a thing as positive well-being or that knowledge, the beauty of a mountain, or the like is good for its own sake. (And I do not grant the existence of positively pleasant experiences in the sense of experiences with a positive hedonic level.) Whenever I speak of such things, I could preface such talk with "purported" and say "purported positive well-being" and "purported good", but I often omit "purported" for brevity and because I am concerned with the agent's motives (e.g. an agent thinks that positive well-being exists and potentially harms a being for the sake of someone else's positive well-being).[20,21]

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

My ethical proposal (the principle) is not new when it comes to the priority of avoiding unpleasantness, suffering, or ill-being over increasing positive well-being or other goods. There are many works where authors emphasise the importance of reducing suffering or the like over the promotion of what is purportedly positive.[22,23,24,25,26,27] Relatedly, there are writings on the ethics of risk when risk is imposed for a pure benefit,[28,29,30,31] as well as on the moral weight of harm, badness, or suffering compared to goodness.[32]

As far as I know, my ethical principle is different from existing ideas mainly in its focus on motives. The risk or potentiality component in my principle is perhaps unusual, but I would say that the role of motives is the main new feature of my principle.

Arguments for the principle

This section is about two positive arguments for the "Do no harm for goods" principle, as opposed to responses to objections against it (which is the topic of the next section). I do not attempt to give an exhaustive list of arguments.[33]

The first argument is to appeal to the intuitive plausibility of the principle, in general. This is similar to appealing to the intuitive plausibility of general ideas such as the following: one has a moral reason to repay one's debts, and one should help someone in distress if it costs nothing for oneself or anyone else. When it comes to the principle I advanced above, it sounds to me like simple, basic morality. It sounds very reasonable that one should not potentially harm one individual who would bear the cost while someone else would merely get, for example, more intense pleasure, which is not needed and does not reduce any ill (or for the pure benefit of no one, in the case of harming for knowledge or beauty which is purportedly good above and beyond its significance to individuals).

The second type of argument for the principle is support from examples or cases. In the philosophical literature, there are plenty of hypothetical cases where the pleasure or positive well-being of some is risked or realised at the expense of someone else's suffering.[34,35,36,37,38,39] For example, one example is about driving an injured person faster to the hospital and thereby risking severe harm to innocent pedestrians, which seems permissible if the risk is low enough and driving faster has a great chance of preventing lifelong suffering. In contrast, it seems wrong to drive like that and impose that risk on pedestrians if the benefit is only others' more intense pleasure.[40]

When considering cases to assess the principle (both when arguing for and against it), it seems that we want to try to contemplate cases where we *isolate* the motive of increasing someone else's positive well-being or creating a good such as knowledge. Perhaps someone has several motives such as reducing suffering and increasing others' positive well-being. Considering a case where someone acts from both of those motives might make it hard to morally assess the motive of increasing others' positive well-being. For example, if someone aims to reduce others' suffering and this motive is enough to justify and motivate the act, but the agent also hopes to increase others' positive well-being, then our judgment might be muddled by the mix of motives. If we want to morally assess the motive of increasing someone else's positive well-being or creating a good such as knowledge, it seems we should hone in on that motive and try to avoid potentially confounding motives. We seemingly want cases where the potential harm to the being stands clearly against someone else's positive well-being or good such as knowledge, so that we can assess whether that tradeoff is morally wrong.[41]

The following is a new case about potentially harming invertebrates: Suppose that you are outdoors and encounter a stranger. It could be an adult, teenager, or younger child. They are deliberately damaging a group of worms. You ask why they are doing that. What kind of reply would be morally satisfactory? I submit that the following replies would be unsatisfactory: my friend wanted this (my friend would be perfectly fine either way but has a preference for this activity); I am making something that will entertain some people I know (they would be perfectly fine regardless but will now become more amused); I will gain knowledge from this; no one knows what happens if one performs the very specific test that I am doing so I am testing merely to fill that knowledge gap. And if we try to probe the last two replies further and ask the follow-up question "Would this knowledge prevent more ill-being than the potential harm

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

that you are causing the worms in front of you?" If the answer is "I have no reason to believe that it would; I am just aiming for the knowledge", then that is, in my view, a morally unsatisfactory reply.

Replies to objections against the principle

In this section, I reply to possible objections to the "Do no harm for goods" principle. As with the positive arguments, I do not aim to present an exhaustive list of possible objections and replies.[42] The objections I will deal with involve providing putative counterexamples of the kind "this act of potentially harming a being is done for the sake of others' positive well-being or good such as knowledge, but it is not wrong." Presenting counterexamples is a standard way of objecting to a principle or general claim. The formulation of counterexamples can range from brief, general descriptions to elaborate descriptions of a specific situation for a specific agent. I will not try to formulate specific counterexamples that an objector might bring up. Rather, I will reply more generally and thereby try to cover a broad range of putative counterexamples that may be formulated.

The following is a generally described putative counterexample against the "Do no harm for goods" principle: Suppose that an act of potentially harming actually and in expectation reduces the most suffering, but the agent performs the act not to reduce suffering but for the sake of someone else's positive well-being or a good such as knowledge. According to this objection, it is not morally wrong to perform the act that reduces the most suffering, even if the motive is positive well-being or knowledge. One can elaborate on this objection and, for instance, stipulate that the alternative is to perform another act for the sake of reducing suffering, but that this alternative act results in more suffering (that is the motive would be better but the results in terms of suffering would be worse). Again, the objection to the principle would be that it is not morally wrong to reduce the most suffering. Essentially, this putative counterexample says that reducing suffering is more important than the motive.

For this sort of putative counterexample, the question arises of what the alternatives are and what the agent can do. Can the agent reduce the most suffering but with another motive than for the sake of positive well-being or knowledge (e.g. the motive to reduce suffering)?[43] It seems to me that, in this simple case, the agent should ideally reduce the most suffering with the right motive. But if the agent cannot act with that motive, one may consider it dubious that the agent should do what the agent cannot do. With that in mind, I will mention five points related to this sort of putative counterexample. The first two points are possible replies to the counterexample, and the essence of the remaining three points is that whether or not the agent can now act with the right motives, there are still other ideas similar to the "Do no harm for goods" principle that merit consideration.

The first point is that the agent in the thought experiment could have performed the action that reduces the most suffering without doing it for the sake of someone else's positive well-being or for a good such as knowledge. In other words, there is also the morally right alternative that the agent reduces the most suffering with a better motive such as relieving distress. Given the presence of that morally superior alternative, reducing the most suffering for the sake of others' positive well-being or knowledge would be wrong.

The second point is that if the thought experiment is stipulated in such a way that the agent can only perform the act that reduces the most suffering if they do it for the sake of others' positive well-being or a good such as knowledge (the agent cannot perform that act with better motives), then that stipulation is unrealistic. One can argue that unrealistic putative counterexamples are not convincing.[44]

The third point is that regardless of whether the agent can now act with the right motives, the agent should at least strive to do so in the future. This point would reflect a principle of the following kind: we should strive to not potentially harm a being for the sake of someone else's positive well-being or for a good such as knowledge. This principle is not exactly the "Do no harm for goods" principle, but it is similar.

The fourth point is that regardless of whether the agent can now act with the right motives, the fact that the agent now potentially harms a being for the sake of someone else's positive well-being or for knowledge indicates that the agent did something wrong in the past. For example, the agent earlier failed to develop the right motives. Perhaps the fact that the agent now acts with certain motives does not *entail* that the agent did something wrong in the past, but one could argue on a case-by-case basis (about

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

specific putative counterexamples) that the agent acted wrongly in the past. Again, this is not exactly the "Do no harm for goods" principle, but rather the idea that the agent in a specific counterexample who potentially harms a being for the sake of others' positive well-being or for knowledge at least acted wrongly in the past.

The fifth point is that even if the agent cannot act with the right motives now, cannot strive to do so in the future, and could not have developed the right motives in the past, it still seems that there is a moral problem. Consider someone who could never develop, strive for, or act with empathy, compassion, benevolence, or the like. The agent can only act out of self-interest and sadism. It seems unreasonable to say that this person has no moral flaw and that there is no moral problem. So, even if the right motives are unavailable to the agent, it still seems that there is a moral problem. This amounts to a defence of at least a principle such as "there is a moral problem with potentially harming a being for the sake of someone else's positive well-being or for a good such as knowledge."

I will spend the rest of this section on another reply to putative counterexamples (in general). The reply goes as follows: Someone presents a putative counterexample, but the counterexample does not work because the act of potentially harming is not, on closer inspection, done purely for the sake of positive well-being or knowledge. The potential harming is in fact partly or wholly done for the sake of reducing bads such as ill-being, ignorance, or the like. If we make a favourable moral judgment about the case, that judgment may very well be driven by the act partly or wholly being done for the sake of reducing bads. In contrast, when we isolate the tradeoff between, on the one hand, potentially harming a being and, on the other hand, increasing others' positive well-being or creating knowledge, then we see more clearly that intentionally making the tradeoff is morally wrong.

What we might think of as increasing someone's pleasure or positive well-being might reduce disappointment, dissatisfaction, boredom, health problems, and so on. Similarly, knowledge may reduce an ill such as being unaware, ill-informed, or having false beliefs. Regarding the motive of making others feel good or happy, we could ask questions such as what would have happened otherwise. If the beneficiaries were not made happy, would they have experienced (more) boredom, discomfort, discontentment, frustration, or the like? And if so, was the act of potentially harming a being not done at least in part or perhaps largely or only to avoid those ills? Even an act such as giving someone an unnecessary surprise gift might be done for the sake of reducing bads. The giver may hope that the recipient will feel less unappreciated or less unstimulated, even though the state might commonly be described in more positive terms.

In daily-life cases where we putatively think that it is morally right to potentially harm a being for the sake of someone else's positive well-being or for a good such as knowledge, it is plausible that the potential harming is at least in part done for the sake of reducing ills. In some examples, the motive might not be the reduction of ills, such as if the example involves potentially harming to create a new being (e.g. a brain in a vat) living in isolation and experiencing pleasure. But I submit that such a counterexample would not be convincing. That is not an end that justifies potentially harming another being. When the tradeoff is so clear and isolated, it seems morally wrong.

Even if it is difficult to sort out and determine what the motive is, that still is a weakness for putative counterexamples because it is at least *doubtful* whether it really is the case that the potential harming is done for the sake of someone else's positive well-being or a good such as knowledge. Perhaps the reduction of bads is a motive that is driving our moral judgments about putative counterexamples.

My conclusion is that there are significant challenges for putative counterexamples against the "Do no harm for goods" principle. Moreover, even if an objection against the principle were successful, there are other, similar principles and ideas that would still be practically relevant, such as the ideas that we *should strive to* not potentially harm a being for the sake of someone else's positive well-being or for a good such as knowledge, and that there is at least a *moral problem* with such potential harming.

Practical ramifications

The practical ramifications include that there are ethical concerns to recognise and attend to. One example is to recognise creation as an ethical issue that exposes beings to potential harm. Regardless of

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

whether it is creation or another act that exposes a being to the risk of harm, there is a moral hurdle of justification for potentially harming another being. The justification cannot be others' positive well-being or a good such as knowledge for its own sake.

In practice, an agent might have several motives for potentially harming a being. The agent can pay attention to their motives and assess whether a motive is to increase their own or others' positive well-being or create knowledge for its own sake. If such a motive is present, the agent can try to discount or rule out that motive from their deliberation and decision-making. Ultimately, the same action might or might not be taken, but at least the wrong motives are sorted out and discounted.

This paper is not about what, if anything, could justify potentially harming a being, but I will now say a bit about it. Let us assume that it is granted that others' positive well-being and a good such as knowledge are not acceptable justifications. That leaves some remaining candidates such as the being's own positive well-being and reducing something bad. For the use of non-human animals in research, the being's own well-being seems dubious as a justification. One can argue that it is dubious because creation cannot be in the being's own interest, but also because, in practice, the use of non-human animals in research is for some other purported benefit (not for the sake of the animals being used). So what about reducing the negative? Negative values beyond what matters to individuals could be held to include injustice, inequality, the ugliness of a pile of trash, and so on. Without arguing for it here, it seems morally problematic to potentially harm a being, which potentially matters greatly to that individual, for a purported disvalue that is supposedly disvaluable above and beyond any effects on individuals. Arguably, only the reduction of suffering or ill-being can justify acts that potentially harm a being. Essentially, if the use of a being can be morally justified, it needs to be justified on the basis of reducing enough ill-being or unpleasantness.

Yet this is not to say that such use can necessarily be justified in many cases, or even in any cases. Moreover, even if the use of beings for the sake of reducing ill-being can be justified, it would still be important to consider broader effects in terms of the attitudes, norms, and practices that one might contribute to. In addition, it is worth being aware of potential biases and distortions that might make someone allow suffering for some group of beings for the sake of others (e.g. self-serving motivated reasoning, and undue neglect of the beings' interests).

For some tradeoffs, such as tradeoffs between two individuals' potential suffering, it might be important to have more evidence of the individuals' ability to experience unpleasantness, the severity of the unpleasantness, and so on. But that is not central to the application of the "Do no harm for goods" principle. We can draw ethical conclusions given the current state of knowledge about whether the being is sentient. As long as there is any risk, we should not potentially harm for the sake of someone else's positive well-being or a good such as knowledge. We do not need to settle exactly how likely it is that a being can be harmed to apply that principle. We should generally avoid potentially harming beings for those purposes.

Fundamental research that involves potentially harming a being is not necessarily ruled out by the principle. Even if it is unclear exactly which applications fundamental research will have, it could be argued that some fundamental research will, in expectation, eventually contribute to a reduction of ill-being. Yet if there is no reasonable way that some fundamental research would reduce ills, then it seems wrong to potentially harm a being to do that fundamental research. As mentioned above, there needs to be sufficient justification in terms of reducing enough ill-being or unpleasantness.

Finally, given that there is a risk of harming a being, it is desirable to use alternative methods that carry less risk and do not involve potentially sentient beings. This might sound obvious, but perhaps it is still worth highlighting that there is a risk that should be taken into account (the risk is not zero), and so a lower risk option is preferable, provided that the alternative does not come with at least as great costs in terms of ill-being and unpleasantness.

Conclusions

The potential harming of a being, including a worm, raises many ethical issues. I have focused on the moral relevance of the motive or purpose for potentially harming a being. I have put forth and argued for

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

the principle "Do no harm for goods". The principle essentially says that it is morally wrong to potentially harm a being for the sake of others' pleasure or preference satisfaction, or for the creation of knowledge for its own sake.

Besides that specific principle, there are other, similar ethical ideas that seem reasonable. For example, it seems reasonable that when it comes to our future behaviour and the personal development of our motives over time, we should at least strive to not potentially harm a being for those purposes.

There are plenty of practical ramifications of the ethical ideas in this paper. Broadly speaking, it is morally desirable to pay more attention to the risk of harm, and to think of an act that potentially harms a being as a regrettable lesser evil (if it is indeed the lesser evil). When making decisions about whether to potentially harm a being, we should try to discount or rule out motives aimed at others' positive well-being or a good such as knowledge for its own sake. Instead, we should seek to reduce the risk of ill-being and unpleasantness.

Acknowledgements. I am grateful to Magnus Vinding for comments on earlier versions of this text. I also thank everyone working for the journal.

Funding statement. This research received no funding of any kind.

Competing interests. The author declares no conflict of interest.

Notes

1. For more on well-being, see e.g. Crisp R. Well-Being. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University; 2021. When I talk about unpleasant experiences or frustrated preferences being bad for an individual, the badness in question is final (non-instrumental) badness. The idea is not that unpleasant experiences tend to have unfavourable consequences for an individual, but rather that the unpleasant experiences are, by themselves, bad for the individual who has them.
2. Birch J. *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford: Oxford University Press; 2024:2.
3. On painience, see Ryder RD. *Painism: A Modern Morality*. London: Centaur Press; 2001. Ryder writes that "the only morally relevant quality is painience—the capacity to suffer pain or distress" (p. 120).
4. A final terminological clarification is that I use "moral" and "ethical" as synonyms throughout the paper.
5. Greener M. What has *Caenorhabditis elegans* ever done for us? *Prescriber* 2021;32(11–12):29–32, at 30. "'In addition, we can use *C. elegans* without ethical concerns and there are numerous mutant and transgenic strains,' adds Dr Saul."
6. Poupet C, Chassard C, Nivoliez A, Bornes S. *Caenorhabditis elegans*, a host to investigate the probiotic properties of beneficial microorganisms. *Frontiers in Nutrition* 2020;7:135, at 5. "In recent years, the *C. elegans* nematode has become a powerful animal model for studying host-probiotic interactions. Its advantages include ease of handling, transparency of the body, short lifespan, and the absence of ethical issues."
7. Zhang S, Li F, Zhou T, Wang G, Li Z. *Caenorhabditis elegans* as a useful model for studying aging mutations. *Frontiers in Endocrinology* 2020;11:2. "[E]xperiments with *C. elegans* are free of ethical concerns."
8. Sverdlik S. *Motive and Rightness*. Oxford: Oxford University Press; 2011:13–5.
9. McDougal A. *The Dualism of Being for and Bringing About*; 2023; available at https://drive.google.com/file/d/107abbqDS8bT2tH0yYUQU2IqagA2ceYe_/view.
10. Vasiliou I. Introduction: Moral motivation and its history. In: Vasiliou I, ed. *Moral Motivation: A History*. New York: Oxford University Press; 2016:3–14.
11. See note 8, Sverdlik 2011.

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

12. A similar phrase comes from Sandberg J. Utilitarianism. In: Runehov ALC, Oviedo L, eds. *Encyclopedia of Sciences and Religions.* Dordrecht: Springer; 2013:2324–4, at 2326. "[U]tilitarianism holds that whether a certain action is right or wrong does not depend on what kind of action it is (whether, e.g., it is a case of lying or killing) or with what intention it is performed (whether, e.g., it is performed with good or bad intentions) but the only interesting issue is what outcomes different actions produce."
13. Sinnott-Armstrong W. Consequentialism. In: Zalta EN, Nodelman U, eds. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University; 2023, at 5.
14. See e.g. Knutsson S, Vinding M. Introduction to suffering-focused ethics; 2024; available at https://centerforreducingsuffering.org/research/introduction-to-suffering-focused-ethics/ (last accessed 29 Nov 2024). See sections 2.1 "Consequentialist views" and 6.4 "Reducing suffering in everyday life" on suffering-focused consequentialist views.
15. See note 8, Sverdlik 2011, at 5.
16. Ewing AC. Utilitarianism. *Ethics* 1948;58(2):100–11, at 108. "Another objection that may be advanced against the utilitarian is that the rightness or wrongness of an action does not depend on its results but on its motives. That it depends only on its motives is obviously false but that it should depend partly on its motives and partly on its consequences is not necessarily incompatible with utilitarianism, though it is with the hedonistic form of utilitarianism. If pleasure is not the only good, among the things which are intrinsically good may well be actions done from certain motives".
17. For a recent text on the final disvalue of some relevant acts and traits, see Knutsson S. Pessimism about the value of the future and the welfare of present and future beings based on their acts and traits; 2022; available at https://www.simonknutsson.com/pessimism-value-future-welfare-acts-traits/.
18. By this formulation, I do not mean that others would sadistically take pleasure in or prefer the harming, but rather that the potential harming would be done for the sake of increasing others' pleasure or preference satisfaction.
19. There might be a moral difference between when potentially harming for the sake of someone else's positive well-being or a good such as knowledge is (i) the only motive; (ii) a sufficient motive (the motive is alone enough for the agent to potentially harm a being); (iii) a deciding motive that tips the scale in favour of the potential harming; or (iv) an existing but motivationally and causally redundant motive (the same act would be favoured and carried out even without the motive). It seems to me that when the motive has one of the first three roles, it is morally wrong or bad, and when it has role (iv) it is perhaps less morally wrong.
20. See note 14, Knutsson, Vinding 2024, at 3.3.
21. Knutsson S, Undisturbedness as the hedonic ceiling; 2022; available at https://www.simonknutsson.com/undisturbedness-as-the-hedonic-ceiling/.
22. See note 14, Knutsson, Vinding 2024.
23. Vinding M. *Suffering-Focused Ethics: Defense and Implications*. Copenhagen: Ratio Ethica; 2020, at I.
24. Hofmann B. Moral obligations towards human persons' wellbeing versus their suffering: An analysis of perspectives of moral philosophy. *Health Policy* 2024;**142**:105031.
25. Tranöy KE. Asymmetries in ethics. *Inquiry* 1967;**10**(1–4):351–72, at 8.
26. Findlay JN. *Values and Intentions: A Study in Value-theory and Philosophy of Mind*. London: Routledge; 2014, at 358.
27. Examples of principles similar to mine include: The Principle of Unacceptable Tradeoffs in Ohlsson R. *The Moral Import of Evil: On Counterbalancing Death, Suffering, and Degradation.* Stockholm: Akademilitteratur; 1979, at 76. The Misery Principle in Wolf C. O repugnance, where is thy sting? In: Ryberg J, Tännsjö T, eds. *The Repugnant Conclusion.* Dordrecht: Kluwer Academic Publishers; 2004:61–80, at 63. See note 23, Vinding 2020, at 62. In particular, see Vinding's principle of sympathy for intense suffering.
28. Shiffrin SV. Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory* 1999;**5**(2):117–48.
29. Shiffrin SV. Harm and its moral significance. *Legal Theory* 2012;**18**(3):357–98.

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press

30. Magnusson E. On risk-based arguments for anti-natalism. *The Journal of Value Inquiry* 2022;**56**:101–17.
31. Rulli T. Effective altruists need not be pronatalist longtermists. *Public Affairs Quarterly* 2024;**38**(1):22–44.
32. Munthe C. *The Price of Precaution and the Ethics of Risk*. Dordrecht/London/New York: Springer; 2011, at 5.5–5.7.
33. Arguments for similar ideas can be found in the sources mentioned by Knutsson and Vinding. See note 14, Knutsson, Vinding 2024, at 3. Magnus Vinding formulated an argument for the 'Do no harm for goods' principle in personal correspondence. I do not pursue that argument in this paper, but I will state it here. I rephrase it slightly and state it in terms of two premises and a conclusion. Premise 1: It is morally wrong to potentially harm a being for the sake of something that has no value. Premise 2: Someone else's positive well-being and a purported good such as knowledge have no value. These two premises imply the 'Do no harm for goods' principle. See note 14, Knutsson, Vinding 2024, at 3.3. There and in the sources mentioned there, one can find support for the second premise.
34. See e.g. the cases "The game" and "Christians and lions" in Mulgan T. *Understanding utilitarianism.* Stocksfield: Acumen; 2007, at 94–5.
35. See note 23, Vinding 2020, at 63. That is, see the discussion of roller coaster rides.
36. Vinding M, Comparing repugnant conclusions: Response to the 'near-perfect paradise versus small hell' objection; 2021; available at https://centerforreducingsuffering.org/comparing-repugnant-conclusions/ (last accessed 18 Oct 2023). This source formulated Creating Hell to Please the Blissful.
37. The Ultra Repugnant Conclusion in Tännsjö T. Why we ought to accept the repugnant conclusion. In: Ryberg J, Tännsjö T, eds. *The Repugnant Conclusion.* Dordrecht: Kluwer Academic Publishers; 2004:219–37, at 220.
38. Wolf C. Social choice and normative population theory: A person affecting solution to Parfit's Mere Addition Paradox, *Philosophical Studies* 1996;**81**(2–3):263–82, at 274.
39. Wolf C. Person-affecting utilitarianism and population policy; or, Sissy Jupe's theory of social choice. In: Fotion N, Heller JC, eds. *Contingent Future Persons.* Dordrecht: Springer; 1997:99–122.
40. Knutsson S. Ethical analysis of purported risks and disasters involving suffering, extinction, or a lack of positive value. *Journal of Ethics and Emerging Technologies* 2025;**35**(2):1–20, at 6.
41. If one wants to consider cases with several motives, one can construct cases such as the following: The agent potentially harms two worms for the sake of saving another worm from the same potential harm and for the sake of increasing someone else's positive well-being. In this case, potential harm stands relatively clearly against positive well-being because the potential harm being reduced for one worm is the same potential harm being inflicted on another worm, so the remaining potential harm can be compared to the increased positive well-being. My judgment about this case is that it is an example of moral wrongdoing.
42. See note 14, Knutsson, Vinding 2024, at 5. Additional relevant objections and replies can be found there and in the sources cited there.
43. See note 8, Sverdlik 2011, at 8. (The topic is the accessibility of motives.)
44. See e.g. Elster J. How outlandish can imaginary cases be? *Journal of Applied Philosophy* 2011;**28**(3):241–58.

Cite this article: Knutsson S (2025). The Wrong Motives for Potentially Harming a Being. *Cambridge Quarterly of Healthcare Ethics* 34: 536–546, doi:10.1017/S0963180126100164

https://doi.org/10.1017/S0963180126100164 Published online by Cambridge University Press