



Moral Facts Do Not Supervene on Non-moral Qualitative Facts

Frank Hong¹

Received: 6 May 2023 / Accepted: 29 May 2024
© The Author(s) 2024

Abstract

It is very natural to think that if two people, x and y , are qualitatively identical and have committed qualitatively identical actions, then it cannot be the case that one has committed something wrong whereas the other did not. That is to say, if x and y differ in their moral status, then it must be because x and y are qualitatively different, and not simply because x is identical to x and not identical to y . In this fictional dialogue between Socrates and Cantor involving infinitely many qualitatively identical agents, this assumption is challenged.

1 Introduction

CANTOR: Oh, hi Socrates! Why do you look so self-righteously indignant today?

SOCRATES: Oh Cantor, if you knew what I knew (which admittedly is very little), you would be indignant too, and not self-righteously so.

C: Do tell, Socrates. What has gotten you riled up this time?

S: Well, if you watched the news recently, you would have seen how Political Party X has been haranguing y , nominee of Political Party Y , for months; but as soon as their *own* nominee, x , was caught doing the exact same thing as y , they have been completely silent!¹ The whole situation is most unjust.

C: Ah, so you say the situation is unjust. You must know what Justice is then.

¹ For the sake of protecting the privacy of the individuals mentioned, but mostly for the purpose of achieving maximum generality, all names have been replaced by variables.

✉ Frank Hong
franksho@usc.edu

¹ University of Hong Kong, Pokfulam, Hong Kong

S: Don't you start that with me, kid. I am the one who invented the Socratic method.

C: But surely you must know what justice is if you have claimed the whole situation unjust.

S: I don't need to know the exact nature of justice to say whether this is a case of injustice. For example, I didn't say whether it was *right* for Political Party X to harangue *y*. For all that I said, *y* may have done something completely wrong and Political Party X may have been completely justified for criticizing *y*, or perhaps *x* has done nothing wrong and Political Party X is completely justified for *not* criticizing *x*. But one thing is for sure, if *x* and *y* did the exact same thing, then it cannot be the case that *y* did something wrong while *x* did not.

C: You overstate your case, Socrates. It can't be possible that *x* and *y* did the *exact same* thing. The two actions can't be the same because they have different properties. One is an action done by *x*, and the other is an action done by *y*.

S: Okay, I suppose that if certain special circumstances applied to *x* but not to *y*, then I suppose *x* would be justified while *y* would not. Maybe *x* has a certain history, or a certain mental state, or was located in a sufficiently different circumstance that makes his action relevantly different from *y*'s.

C: No, no. I just meant to say that one is an action committed by *x* and another an action committed by *y*. For all I care, they may be identical in every other way.

S: What?! How could their moral statuses be different just because one is identical to *x* and another identical to *y*? If they differ morally, there must at least be *some* qualitative difference between them.

C: Ah, so you accept the principle:

Wrongness Supervenes on Non-moral Qualitative Facts: For any set, *X*, of non-morally qualitatively identical agents,² either all members in *X* are

² Here, "qualitative properties" are contrasted with "haecceitistic properties", and two different objects are qualitatively identical iff they share all the same qualitative properties but not necessarily the same haecceitistic properties. Qualitative properties are like those properties that Max Black's two iron spheres share (e.g. having a certain mass, chemical composition, being a certain size, being beside an iron sphere, etc.) (Black, 1952). If any properties differentiate the spheres, they would be haecceitistic, where, roughly, a property is haecceitistic if the linguistic expressions that stand for them makes mention to particular individuals (e.g. being identical to *x* and not to *y*, being co-located to *x*, being five feet apart from *y*, etc). More rigorous analyses of the qualitative/haecceitistic distinction are available, but they are controversial. Cowling (2023) gives a great overview of the different attitudes philosophers have taken towards this distinction. For example, Adams (1979) ties the distinction to certain linguistic terms, and Lewis (1986) takes qualitative properties to supervene upon a distinct class of "perfectly natural properties". Some philosophers like Diekemper (2009), Cowling (2015), and Bacon (2019) take the distinction to be primitive. But even if we leave this distinction unanalyzed, we can connect the concept to related ideas (e.g. see Fine, 1977; Fritz & Goodman, 2016 who connect it to ideas related to higher-order contingentism). For our purposes, we can similarly take this distinction as unanalyzed and rely on our rough characterization in order to get a hold of the thought experiment to be given below.

doing something wrong or none of the members in X are doing something wrong.

S: Woah. Okay, I guess that's true. That seems like an overly formal way of stating my point, though.

C: It helps with what I'm about to say later.

S: Okay, so what is it you want to say?

C: I want to say that that principle is false.

S: That's the most ridiculous thing I've ever heard, and I've talked to Parmenides.

C: Hear me out. Let's start with some preliminaries so we're on the same page.³

2 Preliminaries

C: Let's start with some examples. Suppose you have two people, Alice and Bob, and Alice attempts to punch Bob who himself has done nothing wrong. Would you say that Alice has done something wrong?

S: Well, I suppose that would be a prime example of wrongdoing, barring certain cases where a person is punching an "innocent threat" in self defense.⁴

C: That's fine. In the cases I will be describing, self-defense won't be involved. Instead, imagine now that Bob is attempting to punch Carl, and Carl has done nothing wrong. Would you say that Alice has not done something wrong in attempting to punch Bob with the intention of saving Carl from being punched?

S: I suppose.

C: Would you still say it's not wrong if Alice will in fact be unsuccessful in her attempts to save Carl?

³ Elsewhere, Dreher (2002) has also argued that moral facts do not supervene on qualitative facts. But his argument requires moral facts to be contingent, and so two qualitatively identical people can differ morally in two different possible worlds where the moral facts are different. The following argument can be run even if we suppose moral facts to be necessary as we will only look at qualitatively identical agents in the same world.

⁴ The term "innocent threat" comes from Judith Thomson's paper, "Self-Defense" where Thomson imagines a case where one must kill an innocent fat man who happens to be falling off a bridge in order to avoid being squashed and killed by the fat man Thomson (1991) thinks killing the innocent fat man in self-defense would be permissible, though she suspects that others find it only excusable. But when it comes to *other-defense*, even Thomson thinks that killing the innocent fat man to save *another* is impermissible.

S: Well, if Alice has good reason to think that she can punch Bob to save Carl, then it wouldn't be wrong for her to at least try to do what's right. I suppose that sometimes it can even be admirable to intervene on behalf of a victim even if one doesn't know for sure that one would be successful in helping.

C: Great. So supposing that Carl has done nothing wrong, Bob is attempting to punch Carl, and Alice is attempting to punch Bob because she justifiably believes that punching Bob will stop Bob from punching Carl, then everything you agreed to would suggest that Bob is doing something wrong while Alice is not.

S: That's correct.

C: And supposing that Alice never did anything wrong in the past and Alice is doing nothing wrong in attempting to punch Bob, then Alice would be as innocent as Carl.

S: Sounds right.

C: So if Alice is as innocent as Carl, and it's wrong for Bob to punch Carl, then it would similarly be wrong for anyone to punch Alice.

S: Sure...

C: Ok, in order to get our puzzle off the ground, let's consider some first-pass principles that generalize these intuitions, and we can worry about finessing them later. For now, let's accept these two principles:

Punching the Innocent is Wrong: If y has done nothing wrong and x is attempting to punch y , then x is doing something wrong in attempting to punch y .

Punching an Aggressor is Not Wrong: If x is doing something wrong in attempting to punch y , then it is not wrong for z to attempt to punch x to save y from being punched.

S: These two principles seem to be exactly what someone like me would argue. If x punches y , then whether or not x is innocent depends on whether y is innocent or whether y is wrong in attempting to punch someone else. In each case, whether or not x is doing something wrong depends on some qualitative difference.

C: You're right to say that whether x does something wrong depends on whether y is innocent, but that's different from saying that x does something wrong because of some qualitative fact about y . If the innocence of y depends only on the fact that y is identical to y , then facts about y 's innocence would not supervene on qualitative facts about y .

S: Well, how could the matter of being innocent simply depend on the identity of y ? Surely whether or not you are innocent depends on what you do.

3 Infinite Punchers

C: Well, suppose we have an infinite set, X , of qualitatively⁵ identical agents. We'll say that there is an agent $x_i \in X$ for each integer i such that $-\infty < i < \infty$.

Let's stipulate that for each x_i , x_i has done nothing wrong up till this moment. But let's also stipulate that each x_i finds themselves attempting to punch x_{i+1} in order to save x_{i+2} from being punched by x_{i+1} .

S: So we have an infinite row of people, who up till now have been totally innocent, where each person is attempting to punch the next person in order to save the next next person?

C: That's right. To get an illustration, you can imagine that each person looks exactly the same, has the exact same mental state, and they are each standing right behind the next person with their right fists raised.

S: Ok. I can see why you want an infinite chain where there is no first member so that you can't distinguish one member as punching but not being punched (or as being punched but not punching). But couldn't you achieve the same result by considering a circle of punchers?

C: Circles add an extra layer of difficulty I would like to avoid here. For example, when considering "circles" containing only one person punching themselves, we may have to consider whether anyone has an absolute right to harm themselves due to their autonomy, or in considering circles containing two people or more, issues about self-defense may arise, and this is another thing we said we wished to avoid.

S: Fair enough.

C: Now let's just look at x_0 . Either x_0 is doing something wrong or x_0 is not doing anything wrong. If x_0 is not doing anything wrong, then since x_0 was innocent until now, x_0 remains innocent even now. But if x_0 is innocent, then by *Punching the Innocent is Wrong*, x_{-1} must be doing something wrong in attempting to punch x_0 since x_{-1} is attempting to punch an innocent person. And since x_{-2} is attempting to punch x_{-1} in order to save x_0 , then by *Punching an Aggressor is Not Wrong*, x_{-2} is not doing anything wrong. And since x_{-2} has been innocent until now, then they remain innocent even now. We can then repeat the argument for x_{-3} to show that x_{-3} is doing something wrong and thereby show that x_{-4} is innocent, and so on. Thus, every even numbered person before x_0 is innocent and every odd numbered person before x_0 is doing something wrong.

⁵ For economy of presentation, whenever we say two agents are "qualitatively" identical, we mean to say that they are qualitatively identical in all non-moral respects. If moral facts are themselves qualitative, then it will just be trivially true that all qualitatively identical agents have the same moral status. What we are interested in, however, is whether moral facts supervene on all the *non-moral* qualitative facts.

Similarly, since x_0 is innocent, then by *Punching the Innocent is Wrong*, x_0 cannot be attempting to punch someone who is innocent, and so x_1 is not innocent. Furthermore, since x_1 is not innocent, then x_2 must be innocent. For suppose x_2 were not innocent. If x_2 were not innocent, then it can only be because x_2 is doing something wrong in attempting to punch x_3 . But by *Punching an Aggressor is Not Wrong*, that would mean x_1 would not be wrong in punching x_2 . But we already established that x_1 is doing something wrong, and so x_2 must be innocent. And now that we have established that x_2 is innocent, we can repeat our reasoning to establish that x_3 is not innocent. By following this reasoning, we can show that every odd-indexed person after x_0 is doing something wrong while every even-indexed person after x_0 is not doing anything wrong.

Since some members of X are innocent while others are not, and all members of X are qualitatively identical, *Morality Supervenes on Qualitative Facts* would be false in this case.

S: Okay, but what if x_0 is doing something wrong?

C: Well, if x_0 is doing something wrong in attempting to punch x_1 , then by *Punching an Aggressor is Not Wrong*, it is not wrong for x_{-1} to attempt to punch x_0 in order to save x_1 . And so long as x_{-1} is not doing something wrong, then by the same argument we gave above, we can show that every other odd numbered person before x_{-1} is innocent while every even numbered person before x_{-1} is not.

Similarly, if x_0 is doing something wrong in attempting to punch x_1 , then by *Punching an Aggressor is Not Wrong*, x_1 must *not* be doing anything wrong in attempting to punch x_2 . By the same kind of reasoning above, we can then show that every *odd-indexed* person after x_0 is innocent while every *even-indexed* person after x_0 is not.⁶

So again, *Morality Supervenes on Qualitative Facts* is false.

S: Now all you've done is make me suspicious about the two principles you proposed.

C: Alright, which one would you reject?

S: I'm suspicious of *Punching the Innocent is Wrong*. Conspicuously missing from this principle is any mention of the beliefs or intentions of the agent. Surely this would matter when it comes to whether any of them is doing something wrong? For example, if each is punching another other while believing the other is not innocent, then perhaps none of them are doing anything wrong.

C: I agree that one's beliefs may be morally relevant. But I don't think one's beliefs can make an action *right* or *wrong*.⁷ One's beliefs only makes one *praiseworthy* or

⁶ For an independently developed, but structurally similar, argument, see Jeremy Goodman's dissertation (Goodman, 2017) who uses a reverse ω - sequence.

⁷ Thomson (1991) explicitly argues that beliefs and intentions are not relevant to whether an action is right or wrong.

blameworthy for the action. For example, a jury may wrongfully convict an innocent person, even though the jury justifiably believes the person is guilty based on some misleading evidence. In this case, the jury may be blameless, since they are doing the best they can given the evidence they have, but this does not make their conviction right. They simply had the false belief that they were doing the right thing when in fact they did the wrong thing. Similarly, if each person in the infinite chain justifiably believed that the person in front of them is not innocent, that would at best show that each person (like the jury) is blameless. It doesn't show that their action is not wrong.⁸

But if you insist that beliefs and intentions can make an action right or wrong, *Punching the Innocent Is Wrong* is not the principle you want to attack. After all, no one will argue that you do *nothing wrong* by punching innocent people on the street so long as you believe they are not innocent. Perhaps you want to attack *Punching an Aggressor is Not Wrong* on the basis that it is sometimes wrong to punch aggressors if you believe they are innocent. However, such an objection can be avoided in our context by assuming that everyone in the chain of punchers justifiably believes the person they are punching is *not* innocent.

S: In that case, perhaps we can find another reason why *Punching an Aggressor is Not Wrong* might be false. Perhaps you might think vigilantism is wrong, and so if a vigilante is about to wrongly punch a runaway criminal, it would still be wrong for you to punch the vigilante lest you engage in impermissible vigilantism yourself.

C: So perhaps I should weaken the principle to only apply to those who are not vigilantes. Even then, we can still generate the puzzle by supposing that this all takes place in the state of nature where no laws are being contravened.

S: The principle is still too strong. It would be wrong to punch aggressors for pleasure, for example.

C: Again, this at best shows that an agent would not be *praiseworthy* in punching someone to save another. It doesn't quite show that the punching is itself wrong. Besides, in this case, we can just imagine that the agents here have no such perverse desire.

S: Okay, I have a more general worry for these two principles. These principles are clearly not *fundamental*. They tell you whether an act of punching is wrong, but they only do so by linking normative properties with normative properties. For example, these principles don't link facts about punching to facts about

⁸ For more on the distinction between blameworthiness and wrongness, see Hawthorne and Srinivasan (2013). They argue that divergences between blameworthiness and wrongness are inevitable. So even on moral theories where the "right" thing to do is determined by what you believe the right thing to do is, the fact that we are not always in a position to know what we believe will tell us that there can be cases where we blamelessly do the wrong thing because we simply do not know what our beliefs are.

wrongful punching; instead, they link facts about *wrongful punching* with other facts about wrongful punching.

C: Fair. I can see why these two principles are not fundamental. But why should we care about whether these principles are fundamental?

S: Well, if you concede that these principles are undergirded by more fundamental principles, then we cannot really evaluate the case of infinite punchers unless we know what those fundamental principles are. For example, a fundamental principle linking normative facts with non-normative facts might say: "If you're punching someone who's punching someone... who's punching someone who's not punching anyone, and the number of people between you and the punched-but-not-punching is even, then you're wrong. If it's odd, you're ok". Of course, this doesn't tell you what happens in the infinite case, and the fundamental principle that applies in that weird case might just say everyone (or no one) is wrong in punching. So until we have that fundamental principle, then for all we know, everyone in the chain may have the same moral status, and so we won't have a counter-example to *Moral Facts Supervene on Non-Moral Qualitative Facts*.

C: I see. Perhaps the fundamental principle may say that everyone (or no one) in the chain of infinite punchers is wrong. But if that's the case, then that would just mean that either *Punching the Innocent is Wrong* or *Punching an Aggressor is Not Wrong* is simply *false*. Nothing in my argument requires that these principles be *fundamental* in order to get the result that qualitatively identical agents can differ in their moral properties. As long as they are *true*, then it follows from our case that moral facts do not supervene on qualitative facts. Furthermore, if you are really looking for fundamental principles that *undergird* our two principles, then it cannot be the case that these fundamental principles turn out to contradict what they undergird. So if you are really looking for the fundamental principles that undergird our two principles, then we can know that, *whatever these principles may be*, they will tell you that some people in the infinite chain are wrong and others are not.

S: Okay, I didn't want to resort to this, but a utilitarian would say that everyone is doing something wrong because everyone will get punched in this scenario and so we have a case where disutility is maximized.

C: Indeed, utilitarians would say that the wrongness of an action only depends on how much net utility it produces relative to all alternative actions, and it doesn't matter who produces the utility or who enjoys the utility. So it's no surprise that utilitarians would resist this argument. In this case, they might simply say that everyone is an aggressor but that it is still wrong to punch the aggressor.

However, utilitarians would also deny *Punching the Innocent is Wrong* for similar reasons. So long as an act of punching produces more utility, it doesn't matter if the person being punched is innocent or not.

S: I think your argument is more an argument against non-utilitarian ethical theories than it is an argument against *Moral Facts Supervene on Qualitative Facts*.

C: Hardly. Utilitarianism is implausible precisely because they deny *Punching the Innocent is Wrong*. I think my argument is better off interpreted as an argument against *Moral Facts Supervene on Qualitative Facts*.

That being said, there may be a more utilitarian friendly version of this argument that tells against the supervenience of moral facts on qualitative facts. For example, suppose there are two qualitatively identical possible worlds, w and w' where the same infinitely many people live in both worlds. Suppose further that for every integer, i , there is exactly one person in each world with wellbeing level i .

S: Okay, so in both w and w' , we have the same people, where the distribution of wellbeing is the same.

C: That's right. However, let's suppose that in w' , *everyone* is better off by *one unit* of wellbeing. This means that the person with -1 unit of wellbeing now has 0 , the person with 0 now has 1 , the person with 1 now has 2 , etc... This, of course, doesn't change the qualitative distribution of wellbeing. In both worlds, there is exactly one person with wellbeing i , for every integer i . Nonetheless, if *everyone* is better off in w than in w' , I think that's a strong reason to think w is better than w' , even if they are qualitatively identical. If so, then at least *goodness* would not supervene on qualitative facts. That being said, this conclusion is itself controversial, so I won't pursue it any further.⁹ But if you think goodness doesn't supervene on qualitative facts, then I think you can cook up a case where wrongness, which the utilitarian thinks supervenes on goodness, also fails to supervene on qualitative facts.

S: Okay, but this is all quite fanciful. You still have to admit that Political Party X did something wrong, right?

C: Oh yea, of course. They are Political Party X after all.

⁹ This example comes from Hamkins and Montero (2000) to show that the *Pareto Principle* (which says that a world w is better than w' if everyone in w is better off in w') is inconsistent with an *Anonymity Principle* (which says that any two worlds where you simply permute the wellbeing levels of the individuals is equally as good). However, Hamkins and Montero themselves opt for *Anonymity*, and so they would deny that, in this case, w is better than w' .

Appendix

Stipulation Let X be a countably infinite set of qualitatively identical agents, with each x_i punching x_{i+1} (for every integer i).

1. If y has done nothing wrong and x is attempting to punch y , then x is doing something wrong in attempting to punch y . [*Punching the Innocent is Wrong*]
 2. If x is doing something wrong in attempting to punch y , then it is not wrong for z to attempt to punch x to save y from being punched. [*Punching an Aggressor is Not Wrong*]
 3. Either x_0 is wrong in punching x_1 , or x_0 is not wrong in punching x_1
 4. If x_0 is wrong in punching x_1 , then for every even integer i , x_i is wrong and x_{i+1} is not wrong. [from premise 1 and 2]
 5. If x_0 is not wrong in punching x_1 , then for every even integer i , x_i is not wrong and x_{i+1} is wrong. [from premise 1 and 2]
 6. Therefore, either every even x_i is wrong and every x_{i+1} is not wrong, or every even x_i is not wrong and every x_{i+1} is wrong. [from premise 1, 2, and 3]
- ∴ X is a set of qualitatively identical agents where some are doing something wrong and others are not. [*Moral Facts Do Not Supervene on Non-moral Qualitative Facts*]

Author Contributions Frank Hong is single author.

Funding Written while receiving funding from Longview Philanthropy.

Data Availability Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of interest The author has no conflict of interest to declare that are relevant to the content of this article.

Ethics Approval This research did not contain any studies involving animal or human participants, nor did it take place on any private or protected areas. No specific permissions were required for corresponding locations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, R. M. (1979). Primitive thisness and primitive identity. *Journal of Philosophy*, 76(1), 5–26.
- Bacon, A. (2019). Is reality fundamentally qualitative? *Philosophical Studies*, 176(1), 259–295.
- Black, M. (1952). The identity of indiscernibles. *Mind*, 61(242), 153–164.
- Cowling, S. (2015). Non-qualitative properties. *Erkenntnis*, 80(2), 275–301.
- Cowling, S. (2023). Haecceitism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (2023rd ed.). Stanford: Metaphysics Research Lab, Stanford University.
- Diekemper, J. (2009). Thisness and events. *Journal of Philosophy*, 106(5), 255–276.
- Dreher, J. H. (2002). Can there be brute, contingent moral facts. *Philosophical Studies*, 108(1–2), 23–30.
- Fine, K. (1977). Properties, propositions and sets. *Journal of Philosophical Logic*, 6(1), 135–191.
- Fritz, P., & Goodman, J. (2016). Higher-order contingentism, part 1: Closure and generation. *Journal of Philosophical Logic*, 45(6), 645–695.
- Goodman, J. (2017). *Distinctions in reality*. PhD thesis, New York University.
- Hamkins, J. D., & Montero, B. (2000). With infinite utility, more needn't be better. *Australasian Journal of Philosophy*, 78(2), 231–240.
- Hawthorne, J., & Srinivasan, A. (2013). Disagreement without transparency: Some bleak thoughts. In D. Christensen & J. Lackey (Eds.), *The epistemology of disagreement: New essays* (pp. 9–30). Oxford: Oxford University Press.
- Lewis, D. K. (1986). *On the plurality of worlds*. Malden: Wiley-Blackwell.
- Thomson, J. J. (1991). Self-defense. *Philosophy and Public Affairs*, 20(4), 283–310.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.