

Ethical Puzzles of Time Travel

Sara Bernstein, University of Notre Dame

From pop culture to science fiction to undergraduate teaching, one often comes across the idea that if time travel were possible, one could go back in time and kill Hitler, and thus presumably save millions of lives. But little serious philosophical thought has been given to this claim, and to the background issues that surround it: if time travel were possible, what sorts of ethical puzzles, dilemmas, and obligations would time travel introduce? For example, would one be morally permitted or even morally obligated to go back in time and kill Hitler or remove him from power? Would less dramatic interventions, such as travelling back in time to prevent a single car accident, also be subject to moral duties and obligations? What moral risks does time travel introduce?

In this paper I take a more careful look at these questions. I articulate several ethical puzzles of time travel and divide them into three different categories: permissibility puzzles, obligation puzzles, and conflicts between past and future selves. In each category, I suggest that ethical problems involving time travel are not as dissimilar to parallel “normal” ethical puzzles as one might think. The intent will not be to establish an ambitious overarching point about the ethics of time travel, but rather to begin a serious philosophical conversation about such topics—conversations that often happen in passing, but that rarely play a central role in serious philosophical debate. The hope is that the questions raised and the parallels drawn will be useful for further metaphysical and ethical investigation.

Since the focus of this discussion will be ethical puzzles, I will assume for the sake of argument that time travel is possible, including controversial “second time around” time travel which changes the past.¹ I will assume that we can at least conceive of such travel and changes to the past, and draw interesting ethical lessons from these scenarios. I will try to remain as neutral as possible on particular models of time travel.

Roadmap: in Section 1, I discuss the case for the moral permissibility of changing the past, and argue that apparent unique moral risks of time travel are not so different

¹ See Goddu (2003), van Inwagen (2010), Hudson and Wasserman (2010), van Inwagen (2010), and Bernstein (2017) for models of “second time around” time travel that changes the past.

from normal moral risks. I show how a particular time travel case connects with the non-identity problem, and poses a challenge for actualist responses to the non-identity problem. In Section 2, I discuss the case for and against moral obligations to change the past, and argue that apparent differences between time travel-involving moral obligations and “normal” moral obligations are illusory. In Section 3, I introduce puzzles involving conflicts between present and future time slices of the same person, and argue that the puzzles are not so different from contemporary problems of peer disagreement, time bias, and the primacy of present consent.

1. The Moral Permissibility of Changing the Past

1.1 The Argument from Moral Risk

Intuitively, it seems morally permissible to remove a present-day dictator from power in order to save millions of lives.² Thus it also seems morally permissible to send a time-traveler into the past to remove Hitler from power. Many lives could be saved, and much suffering prevented, by a single change to the past. Changing the past to dramatically alter the world for the better is, on the face of it, morally allowed.

But there are multiple grounds for holding that it is *never* morally permissible to change the past. The first is what I will call the *argument from moral risk*. Assuming that even small changes (including the time traveler’s mere arrival in the past) can lead to big changes in the present day (for example, preventing the conception of a presently-existing person),³ one line of thought holds that it is causally and morally risky for a time traveller to change the past at all. For example, some historians theorize that the casualties of World War II would have been even greater had Hitler been removed from power.⁴ Though it seems clear that saving millions of lives is preferable to the suspected historical alternatives, we still do not know for certain what the alternatives would have been. Thus one might take time travel to pose an unacceptable moral risk.

² According to some, but not all, leading moral theories.

³ This problem was explored in the film *Back to the Future*, in which Marty McFly inadvertently threatens his own existence by interfering with his conception.

⁴ This scenario has also been explored in science fiction. In Stephen Fry’s *Making History* (1996), time travelers make the world such that Hitler never existed, only to unintentionally replace him with a more competent Nazi leader who then wins the war.

Consider an example less dramatic than Hitler's assassination:

(Loud Arrival) David's time machine arrives with a bang, causing Suzy to look towards the source of the loud noise. Had David's time machine not arrived, Suzy would have noticed handsome Billy sitting in the coffee shop, leading to their eventual marriage and conception of Jane, a presently-existing person.

Here, the mere arrival of David's time machine initiates a causal chain that results in presently-existing Jane's removal from existence.

It might seem that this is a unique causal and ethical risk posed by time travel: only in these situations can one remove a person and her future existence—a stretch of human existence that has “already” occurred-- from the temporal manifold. Such a possibility suggests that any sort of time travel poses not just a moral risk, but a unique one: any change to the past, no matter how minute, might cause numerous downstream effects, including preventing the existence of a person or people who “already” existed later in the timeline. We have knowledge of how the past turned out, but lack knowledge about the possible effects of interventions into the past.

Lacking knowledge about how the future will evolve given changes to the past is no different than lacking full knowledge about the future consequences of any “normal” intervention in the present day. We lack knowledge of how the past would have evolved given Hitler's early removal from power just as we lack knowledge of how current geopolitical situations will evolve given removal of present-day dictators. Given the uncertainty in both cases, removing Hitler from power is not significantly morally different than removing a present-day dictator from power. Doing so would be morally permissible were time travel possible.

I suggest that most changes to the past are no more morally risky than normal actions in everyday life. Every time we leave the house or pass someone in traffic, we risk preventing a meeting of two people that might later conceive another human. Such possible downstream effects are part of the messy causal stuff of everyday life. Just as we are not morally forbidden from leaving the house on the grounds that we might prevent a meeting between two people, we should not be morally forbidden from time travel simply because it is causally risky in this most basic sense.

One might worry that the fact that Jane “already” existed in the future poses a special moral obligation to ensure that Jane is conceived. According to this view, there is a moral difference between preventing a person’s conception in the present day (by, say, making a person late for a date by cutting her off in traffic—a date that would have led to the conception of a person had she been on time for it, or more simply, failing to have the romantic night that leads to the conception) and time-travelling to prevent the initial conception of Jane, who already exists in the present day. The latter is akin to murder, the worry goes, whereas the former is not. For the time traveler is not only *ending* Jane’s life (we can imagine that she would have lived a long life were it not for the time traveller), but removing Jane’s stretch of existence from the temporal manifold altogether.

But consider that removing Jane from the temporal manifold isn’t *murdering* her, in the typical sense of the term. The time traveller doesn’t cause her physical or emotional pain; her loved ones do not grieve. Her loss is not felt by the world. The time traveller does not “hole punch” Jane out of existence, leaving her traces and influences intact. For the time traveler changes the world so that Jane never existed in the first place. If one holds that there are not duties to bring nonexistent people into existence, one should agree that there is no duty to ensure that Jane is conceived upon arrival in the past. Consider a non-existent (i.e. not-yet-conceived) person Jim. If Billy and Suzy have a romantic evening, they will conceive Jim. But presumably, Billy and Suzy are not morally required to conceive Jim.⁵ Similarly, the time traveller does not have a duty to ensure that Jane is conceived. From the temporal vantage point of the time traveler, her duty to Jane is no different than Billy and Suzy’s duty to Jim. Assuming there is no obligation to conceive future persons, the Argument from Moral Risk does not render time travel morally impermissible.⁶

A defense of the Argument from Moral Risk holds that there is a moral difference between the world in which Jane exists and Jane is removed. Even if the time traveller

⁵ For a dissenting argument that people are required to bring people into existence if they can have good lives, see Gardner (2016).

⁶ One might be concerned that such a result introduces *moral chaos*. If everyone is permitted to go back in time to right past wrongs, then the whole of reality would be chaotic and ever-changing, responding to countless minute causal differences brought about by multiple time travellers. But this is a practical problem rather than one about moral permissibility of individual time-travelling episodes.

makes it such that Jane never existed in the first place, there is something morally questionable about doing so. Intuitively, the world containing Jane's removal seems somehow different than the one without it. Setting aside time travel for a moment, suppose that one could push a button and make it so that Jane never existed. *Ceteris paribus*, this seems worse than not doing so, even if one is not technically doing harm to Jane.

This argument leads to an interesting and unexpected result: the existence or nonexistence of hypertime has ethical consequences. *Hypertime*, roughly speaking, is an extra temporal manifold against which the temporal manifold is measured. The idea is that the passage of time must be measured against something, and the most natural "something" is another dimension of time. When we say that time passes at a rate of one second per second, for example, we are measuring time against another dimension of time.

In the context of time travel, hypertime provides a natural model for understanding changes to the past.⁷ For example, if our time traveller ventures to the past and prevents Jane's conception, then Jane's conception hyper-occurs "the first time around" but does not hyper-occur "the second time around" after the removal. Hypertime provides theoretical resources within which to compare the worlds in which Jane exists and the time traveller removes Jane from existence. With hypertime, both worlds exist in diachronic order and can be examined for metaphysical and moral differences.

But the nonexistence of hypertime prevents moral comparison of the removal world and the non-removal world, since her existence has been entirely erased from the temporal manifold. There is just one occurrence of the event in question--the prevented conception of Jane. The time traveller makes it so that Jane *never existed in the first place*, not the case that she hyperexisted first and failed to hyperexist later. There is no basis for a comparison of the world in which Jane exists and which she doesn't exist.

The interesting result here is that ethical comparisons between worlds depend on the existence of hypertime. Only a theory that commits to the existence of hypertime has the resources to compare the worlds. It is thus *prima facie* preferable to have a

⁷ See Goddu (2003), van Inwagen (2010), Hudson and Wasserman (2010), and Bernstein (2017) for models of time travel to the past that involve hypertime. Van Inwagen (2010) explicitly mentions the moral problem of erasing past people who have already existed.

metaphysical theory that posits hypertime in order to have the resources to make such comparisons.

1.2 Time Travel and the Non-Identity Problem

Originating with Parfit (1984), the *non-identity problem* arises when the existence of a person is caused (*ceteris paribus*, a good moral act), but causing a person to exist also causes that person to have a feature or features that negatively affects quality of life. When the relevant alternatives to bringing the negatively-featured person into existence are (i) not bringing anyone into existence at all, or (ii) bringing a different, non-flawed, but *non-identical* person into existence, it is morally unclear why (ii) is morally preferable to causing the flawed person's existence.

The moral permissibility of time travel connects with the non-identity problem in several ways. First, note that from the vantage point of the time traveller in Loud Arrival, there is no future person, Jane, to which the time traveller is beholden. For Jane does not yet exist in the time traveller's present. Thus it seems that the time traveller commits some sort of harm, but not a harm directed towards a particular existent person.

Second, we can easily vary Loud Arrival to introduce a time-travelling variant of the non-identity problem. *Consider*:

(Conception-Causing Arrival) David's time machine arrives with a bang, causing Suzy to look towards the source of the loud noise. In looking towards the source of the noise, Suzy spots Billy, with whom she falls in love and conceives Jane, a child with developmental impairments. Had David's time machine not arrived loudly and with a distracting bang, Jane would not have been conceived.

The non-identity problem arises when making a choice that results in someone's conception but that negatively effects that person's long-term well-being.⁸

Ethicists differ on the moral status of this sort of action. Here, David causes Jane's life with developmental impairments to occur. The relevant contrast class is Jane's non-occurrence. Some philosophers hold that causing Jane's existence is morally preferable to her nonexistence. According to one such view, *moral actualism*, the moral

⁸ Here I do not assume that all developmental impairments effect well-being; I stipulate that that the particular developmental impairment effects well-being in this token case.

status of an action is evaluated on the basis of its effects on only actual past, present and future people. Interests of, and outcomes for, merely possible people are not morally relevant. According to Jackson and Pargetter's (1986) formulation, roughly, for every act-type *A*, you *ought* to do *A* if and only if your conduct would be (in the relevant way) *better* if you did *A* than if you did not. According to actualism about the non-identity problem, the disabled person's existence is preferable to her nonexistence, since the morally relevant features of the choice are given by the impact on her actual existence rather than her unactualized counterparts.

The possibility of time travel reveals an ambiguity in what constitutes "the actual world" for the moral actualist. In normal scenarios, evaluating whether a particular action obeys actualist doctrine involves examining its effect on past, present, and future people in the actual world. But in time travel scenarios, it is unclear how to evaluate which world is the actual one, since there are multiple actualities. One view takes the actual world to be the world as it is at the moment of the time traveller's arrival—i.e., the world that does not include Jane, since Jane has not yet been conceived. Another view takes the actual world to be the world as it is after the time traveller's intervention—i.e., the world that include Jane's existence, which is a causal result of the time traveler's arrival.

Time travel poses another intuitive problem for actualism. Consider a time traveller like the one in *Loud Arrival*. As in the original case, let us imagine that Jane is a presently-existing person at the temporal location of the time-traveller's initial departure. Suppose that the time traveller arrives with a bang, delaying Jane's conception by a mere one second. (And suppose, further, that this delay causes no untoward effects, genetic or otherwise.) Intuitively, the time traveller does Jane no harm by delaying her conception by one second.

But according to the actualist, the time traveller has *taken Jane out of existence*. Most actualists believe in what Parfit (1984) called *the Time-Dependence Claim*: if any particular person had not been conceived when they were in fact conceived, they would never have existed. Since Jane's origins are essential to her, any change in the circumstances of her conception effectively removes her from existence. Thus any change in origin changes the actual person to whom moral obligations might be owed. Call non-delayed Jane "Jane1", and delayed Jane "Jane2". According to the actualist,

Jane2 is not identical to Jane1 because her circumstances of conception have changed. To take Jane1 out of existence is to harm her. There are countless other similar harms that would occur in a time travel scenario-- numerous people harmed by being removed from existence in virtue of the circumstances of their creations being slightly altered. Yet it is counterintuitive to hold that the time traveller has really harmed countless people in this way. Without the theoretical resources to account for slight differences in origin, the actualist is left with a large theoretical cost.

2.0 The Moral Obligation to Change the Past

If it turns out that changing the past is morally permissible, is changing the past ever morally *obligatory*? *Puzzles of obligation* involve apparent moral obligations introduced by the possibility of time travel. I suggest that the possibility of time travel would introduce rampant moral obligation to prevent suffering and harm. Suppose that the following moral principle is true:

(Obligation) *Ceterus paribus*, if one can save a life at little cost to oneself, one should.⁹

For the moment, let us set aside controversial interpretations of this principle such as, e.g., Peter Singer's rampant obligation-inducing principle about giving to charity. Rather, let consider a simple example:

(Disease) Athena, who lives in 1880, will die of a disease that is easily treatable today with one dose of antibiotics. Dr. Smith could press a button on a time machine, travel back in time, deliver one antibiotic pill to Athena, and return home within ten minutes.

Given the ease with which Dr. Smith could save Athena's life, is such a trip morally obligatory?

I say: yes. If Dr. Smith knows about Athena's condition, and if the cost to him is low, he is morally obligated to make the trip and deliver the pill. So many lives could be

⁹ See Singer (1972) and Unger (1996) for canonical formulations of this principle.

easily saved; so many interventions could be performed. Consider the following similar case:

(Accident) On your way to work, you strike and kill a child with your car. But, like many, you have an iTime, an easily usable personal time machine. You could activate the device and change the near past so that you do not hit the child.

Intuitively, you are morally obligated to activate the device and save the child: at very little cost to you, you can save the life of a child whose life you would have otherwise ended.

The possibility of time travel would introduce rampant moral obligation to prevent easily preventable past deaths, and to right wrongs that initially resulted in death, suffering, and harm. This profligate moral obligation would change the terrain of moral theory. Given such a possibility, it is not the case that we would say “It is too bad that the innocent deaths occurred; we will try better next time.” Rather, all the wrongs of the past have the possibility of being righted; all the past suffering can be prevented. Overriding moral obligations would extend to the past as well as the present and future.

This result is not merely theoretical; it extends to many tendrils of applied ethics. The ethics of reproduction and biotechnology to some extent involve the righting of wrongs: preventing two people who conceive a baby that they don’t want to have or correcting genetic arrangement that lead to suffering and harm. The terrain of these problem spaces would change given the possibility of time travel: rather than be concerned with preventing harm given non-ideal situations, the concern would be changing the harm that already occurred: preventing the conception of the baby rather than questioning the moral permissibility of abortion; retroactively changing genetic dispositions to conditions that negatively impact well-being, and so on.

3.0 Conflicts between Past and Future Selves

Time travel introduces the possibility of conflicts in consent between concurrently existing past and future selves. Suppose that a time-travelling future version of you shows up in the present and insists on (present) you doing something you don’t want to do—getting a tattoo, for example, or quitting your job. (And suppose that it is 100% certain

that this time traveller is in fact a future version of you.) The future version of you plans to physically force you to get the tattoo if you do not consent. Do the wishes of present you override the wishes of future you?

Here, there is a conflict of consent between past and future selves: the preferences of present you conflict with the preferences of future you. Conflicts of consent between different synchronous temporal stages of the same person seem *prima facie* unsolvable: principles of personal autonomy do not apply, since both person-stages are you. The puzzle is whether the consent of one person-stage trumps the consent of another person-stage, and if so, which one does.

There are several possible resolutions of the puzzle. First, one might judge that the primacy of consent lays with the future self rather than the present one. One motivation for this reading is that the future self might have information that the present self does not have. For example, the future self might know that the tattoo will lead to great pleasure and serve to initiate conversations that lead to friendships with like-minded people. If this is the case, then it seems wise to give the future slice primacy of consent.

However, there is a counterexample to this principle in *auto-erasure*, suicide by eliminating one's past self. Is it permissible to travel back in time and remove yourself from existence entirely?

Consider the right to suicide under normal circumstances. Suppose that Joe is a fully rational agent not suffering from clinical depression or any other perspective-altering mental illness. And suppose, further, that Joe seeks to commit suicide. Joe's suicide seems morally permissible. *Ceteris paribus*, suicide is a fundamental human right: it invokes fundamental autonomy and control over one's own life and body.

But now consider a time-travelling variant of the case:

(Suicide) Future Joe seeks to kill himself. Future Joe travels back in time to remove his past self from existence.

This case uncovers a tension between a commitment to the moral permissibility of suicide and the moral permissibility of auto-erasure. For Joe is entitled to remove himself from existence. But the primacy of present consent seems to trump Future Joe's desire, no matter what Future Joe knows. Even if Future Joe knows that the apocalypse is near

and that Present Joe's life will be filled with suffering and struggle, it seems wrong for Future Joe to forcibly remove Present Joe from existence. Even in less dramatic cases, such as the case of the tattoo, hindsight knowledge possessed by the future time slice shouldn't necessarily trump the desires of the presently-existing person stage. This line of thinking casts doubt on the primacy of future consent over present consent.

One might read *both* selves—the “present” self and the time-travelling self—as equally laying claim to being present, and thus to primacy of consent. For the past *feels* like the present to the time-travelling self. In a sense, both are present once the time-travelling slice completes her journey. Read this way, the case might be seen as evidence against *time bias*, which grounds preferences for painful events to be in one's past and pleasurable events to be in one's future.

If this reading of the case is correct, then the preferences of neither time slice trumps the other. We might view the case as a time-travelling variant on peer disagreement. Here, the “peer” is oneself. As in epistemic peer disagreement, the logical space of options for the two time slices is: (i) to give equal weight to the peer, (ii) give no independent weight to the peer, (iii) give more weight to the peer. I leave this particular puzzle to experts on peer disagreement, largely because it is not substantively different from normal instances of this problem. Again, we can note the strong parallels between an apparently bizarre time travel puzzle and a more normal philosophical problem.

Finally, one might hold that the present person stage has primacy of consent. There is independent motivation for this reading. According to Dougherty (2014), present consent trumps past consent in cases of conflicting consent at different times. For example, suppose that you tell your friend on Tuesday that she must tattoo you on Thursday no matter what. (You want a tattoo, but you always get scared at the last minute, darnit.) But on Thursday, as the time for the tattoo draws near, you recant your consent, begging your friend not to tattoo you. Intuitively, the Thursday (present-time) consent trumps your past consent: no matter what you consented to in the past, what you say in the present time transmits the morally binding message. Dougherty holds that this lesson generalizes: present consent seems to trump past projections of consent into the future, and extends *inter alia* to similar cases involving, for example, sexual consent.

It is natural to generalize Dougherty's lesson further: present consent trumps

consent projected from the backwards-travelling time traveler. According to this view, time travelling scenarios involving conflicts between past and future selves are special instances of the priority of present consent: the consent of whoever is *not* the time traveller trumps the consent of the time traveller. However strange-seeming time-travelling conflicts of consent are, they are parallel to Dougherty's less strange cases involving past and future consent: instances of a more general problem of disagreements between past and future time slices of the same person. Again, the seemingly bizarre problem introduced by time travel is not so different from the more ordinary case.

4.0 Conclusion

Time travel seems to introduce fanciful and unique ethical problems involving moral risk, moral obligation, and conflicts of consent. However, these problems can often be assimilated to more familiar problems in ethics and in metaphysics, shedding new light on the structure of problems in both areas. While time travel cases might seem impossibly theoretical, they are useful for eliding the shape of ethical puzzles, illuminating their most morally relevant features, and opening up new avenues of investigation.

References

- Bernstein, S. (2017) "Time travel and the movable present" In J. Keller (Ed.), *Being, freedom, and method: Themes from the philosophy of Peter van Inwagen* (pp. 80–94). Oxford: Oxford University Press.
- Dougherty, T. (2014) "Fickle Consent," *Philosophical Studies*, 167 1: 25-40.
- Fry, S. (1996) *Making History*. Auckland: Random House New Zealand.
- Gardner, M. (2016) "Beneficence and Procreation," *Philosophical Studies*, 173 (2):321-336.
- Goddu, G.C. (2003) "Time Travel and Changing the Past: (Or How to Kill Yourself and Live to Tell the Tale)" *Ratio* 16: 16-32.
- Hudson, H., & Wasserman, R. (2010). "Van Inwagen on time travel and changing the past" In D. Zimmerman (Ed.), *Oxford studies in metaphysics* (Vol. 5). Oxford: Oxford University Press.
- Jackson, F. & Pargetter, R. (1986). "Oughts, options, and actualism," *Philosophical Review* 95 (2):233-255.
- Parfit, D. (1984) *Reasons and Persons*. Oxford: Oxford University Press.
- Singer, P. (1972) "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1, 3, 229- 243.
- Unger, P. (1996) *Living High And Letting Die*, New York: Oxford University Press.
- van Inwagen, P. (2010) "Changing the Past" *Oxford Studies in Metaphysics* 5:3-40.