Anna Folland

Harm

Essays on Its Nature and Normative Significance

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Humanistiska teatern, Engelska Parken, Thunbergsvägen 3C, Uppsala, Thursday, 6 February 2025 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Associate Professor Travis Timmerman (Seton Hall University, Department of Philosophy).

Abstract
Folland, A. 2025. Harm. Essays on Its Nature and Normative Significance. 38 pp. Uppsala: Department of Philosophy. ISBN 978-91-513-2328-2.

This thesis examines how we should understand the concept of harm, and its moral and prudential importance. It discusses various analyses of harm and normative principles that appeal to harm. In broad terms, it offers a defense of the view that harm is normatively important and useful for philosophical theorizing. Further it proposes a novel analysis of harm, which aligns with that view.

The first paper, "The Harm Principle and the Nature of Harm", defends John Stuart Mill's *Harm Principle* against the criticism that the principle has unacceptable implications regardless of which analysis of harm we plug into it. I argue that the criticism is built on mistaken assumptions – most importantly, the assumption that the Harm Principle is plausible only if there exists an unproblematic analysis of harm.

The second paper, "Feit on the Normative Importance of Harm", criticizes Neil Feit's suggested solution to the so-called *Failing to Benefit Problem* for the Counterfactual Comparative Account (CCA). Feit argues that CCA's inability to align with some commonsense views about harm's moral importance is no flaw since those views are false. I object to that argument, in part by showing that the cases that Feit appeals to are not genuine counterexamples.

The third paper, "Doing Away with Skepticism about Harm", scrutinizes *the elimination thesis*, which states that we should do away with the concept of harm in philosophical theorizing. I examine various claims in support of that thesis – for instance that the concept is defective – but conclude that we lack good reasons to accept it.

The fourth paper, "Misfortune and Missing Out", focuses on Kaila Draper's famous challenge for deprivationism – the view that death harms a subject in so far as it deprives her of life's goods. Since not winning the lottery is also a deprivation, the challenge is to explain why only death is a misfortune in the sense that it merits negative emotional responses. I argue that the challenge is serious, in part by criticizing some prominent suggested solutions, and identify a parallel challenge for CCA.

The fifth paper, "A Fitting Attitudes Analysis of Harm", puts forward a novel analysis of harm. Roughly, this analysis says that an event harms me if, and only if, it is fitting for me to disfavor the event for my own sake.

Keywords: harm, benefit, the counterfactual comparative account, omission, pre-emption, normative reasons, well-being, skepticism, fitting attitudes, misfortune, The Harm Principle

Anna Folland, Department of Philosophy, Ethics and Social Philosophy, Box 627, Uppsala University, SE-75126 Uppsala, Sweden.

ISBN 978-91-513-2328-2
URN urn:nbn:se:uu:diva-543566 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-543566)

To Måns and Jack

List of Papers

This thesis consists of a general introduction and the following papers.

I. Folland, A. (2022) The Harm Principle and the Nature of Harm. *Utilitas*, 34(2): 139–153.

II. Folland, A. (2023) Feit on the Normative Importance of Harm. *Theoria*, 89(2): 176–187.

III. Folland, A. (accepted) Doing Away with Skepticism about Harm. *Ethical Theory and Moral Practice*.

IV. Folland, A. (submitted) Misfortune and Missing Out.

V. Folland, A. (manuscript) A Fitting Attitudes Analysis of Harm.

Reprints were made with permission from the respective publishers.

Acknowledgments

My heart is full. I am worried that I will omit mentioning someone who deserves it. But here we go.

I am extremely grateful to my supervisors Jens Johansson and Erik Carlson. Throughout my doctoral studies, they have given me the most heart-warming guidance in the form of expert knowledge, professional advice and encouragement, constructive criticism, and insightful feedback – invariably delivered with illuminating precision. Know that your generosity is appreciated and that your conscientious approach to philosophy is inspiring.

There is a long list of people who made my doctoral studies at the Department of Philosophy at Uppsala University so memorable and rewarding. On it are the former Head of Department Matti Eklund and the current Head of Department Elisabeth Schellekens; I thank them for their help with both practical and philosophical matters. The list also includes Alexander Stöpfgeshoff, Andreas Stokke, Andrew Reisner, Anna Nyman, Axel Rudolphi, Carl Montan, Edit Karlsson, Ekrem Çetinkaya, Elena Prats, Emil Andersson, Guy Dammann, Hallvard Stette, Henrik Rydéhn, Irene Martínez Marín, Jeremy Page, Jessica Pepp, Jonathan Shaheen, Karl Bergman, Karl Ekendahl, Kasper Kristensen, Katharina Felka, Maarten Steenhagen, Magnus Jedenheim Edling, Mahdiyeh Mousavi, Maria Svedberg, Nicholas Wiltsher, Nils Franzén, Nils-Hennes Stear, Oda Tvedt, Olle Risberg, Per Algander, Rebecca Wallbank, Sebastian Lutz, Sebastián Reyes Molina, Sharon Rider, Silvana Hultsch, Simon Rosenqvist, Sofia Bokros, Tobias Alexius, Tommaso Braida, Zehao Lyu, and Åke Gafvelin. Thanks to Susanne Gauffin, Ulrika Valdeson, Johan Löfström, and Rysiek Sliwinski for assistance with various practical matters – and to Susanne and Ulrika especially for the much-needed coffee breaks. I am also grateful to Pauliina Reemes, Jasmina Nedevska, and Folke Tersman for helpful comments on a draft of this thesis, in connection to my final seminar.

I am thankful to Björn Petersson, who was the opponent of my final seminar, for valuable feedback and fruitful discussion.

Thanks to the staff and graduate student community of the Philosophy Department at CU Boulder; my visit there in the spring of 2022 was both philosophically and socially stimulating. Special thanks to Chris Heathwood, David Boonin, Alastair Norcross, and my fellow Swede Henrik Andersson for inspiring discussions and valuable feedback on my paper drafts.

Among the philosophers in Gothenburg that I would like to thank is Karl de Fine Licht, with whom I have been fortunate enough to co-author two research papers. The experience of working with Karl has been very valuable.

I was kindly offered to work from the Department of Philosophy, Linguistics and Theory of Science at Gothenburg University during the last stages of writing this dissertation. I am grateful to everyone there who helped make that intense period productive and enjoyable. For valuable comments on a draft of the general introduction, I am very grateful to Alexander Andersson, Richard Endörfer, John Eriksson, Ragnar Francén, Simon Rosenqvist, M. Hadi Fazeli, Olle Blomberg, Mattias Gunnemyr, and Erik Malmqvist.

I would like to acknowledge the financial support that I have received from Gothenburg Nation in Uppsala (the O. Andrén Foundation), the H. F. Sederholm Foundation for the Faculty of Arts, the J. Håkansson Travel Grant, the Erik and Gurli Hultengren Foundation for Philosophy, Kungl. Humanistiska Vetenskaps-Samfundet, and the Jubelfeststipendierna Foundation for the Faculty of Arts.

Finally, I would like to express my gratitude to my family and friends. By offering me unwavering love and support, they have been my guardians and my foothold through these years. I treasure them immensely.

Contents

General Introduction

1. Introduction

During the course of our lives, we all feel pain, we get disrespected, our loved ones hurt us, and we fall sick. We are all harmed at various points in our lives. And whether or not we intend to, we all harm others. Broadly, this thesis explores two themes, both involving harm. The first theme is *the nature of harm,* or what harm is. According to one central family of theories, harming someone is a matter of causing something that is bad for them. According to another, harming someone is a matter of making them worse off compared with some counterfactual scenario – for instance compared with if the agent had not performed the relevant act. The second theme of this thesis concerns when we are morally permitted to harm others and when we are not. Questions about when morality permits us to do harm belong to the study of harm's *normative significance*.

The purpose of this general introduction is to summarize the five essays that make up the thesis, put them into context, explain the connections between them, discuss methodological choices that I have made when writing them, and say what I think we can learn from them.

The concept of harm is familiar from everyday thinking, when we evaluate our own and others' behavior in various ways. Consider behavior that we typically find morally wrong: using violence, stealing, lying, discriminating, etc. One natural explanation of why such behavior is wrong is that it *harms* the recipients of that behavior. Generally speaking, we blame, punish, and react with negative emotions toward those who harm others.

Moreover, we tend to think that harmless behavior should be allowed. Suppose you leave a dinner party putting on someone else's jacket by mistake. The jackets are similar enough that the owner of the jacket just takes yours

and never notices the mistake. In this and similar cases, we commonly think something like "no harm, no foul". [1]

With those associations between harming and wronging in mind, it is not surprising that many policies and principles aimed at promoting welfare and justice employ the concept of harm. For instance, in the United Nations *Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power*, UN defines victimhood by appealing to the concept of harm: "'Victim' means persons who, individually or collectively, have suffered harm […]" (The United Nations, n.d.) The first of the ten core principles of ethical AI, proposed by UNESCO, is the following: "Proportionality and Do No Harm: The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses" (UNESCO, n.d.).

The concept of harm is also frequently employed in philosophical arguments and principles. John Stuart Mill's (1859/1977, pp. 223–224) so-called *Harm Principle* says that the state cannot justifiably intervene against an individual unless it thereby prevents harm to others. Many versions of the doctrines of double effect and doing and allowing appeal to harm (Foot, 1967; Quinn, 1989; Woollard, 2015). There is a debate concerning whether harmfulness can explain wrongful risk imposition (Finkelstein, 2003; Oberdiek, 2017; Perry 1997; Rowe, 2021; Stefánsson, 2024). Tom Beauchamp and James Childress' (1979) influential principles in biomedical ethics include the principle of non-maleficence: avoid imposing harm on others. One of W.D. Ross's (1933/2002, pp. 21–22) prima facie moral duties is the duty of non-maleficence: the duty not to harm others. These are just a few examples of philosophical arguments and principles that appeal to harm.

In light of the role that the concept of harm plays in everyday thinking, in practices and policies, as well as in academic research, philosophers have found it crucial to explore the nature of this concept.

[1] This phrase is thought to have been coined in the 1950's in a more violent form of basketball, called 'streetball'. In order to get a better flow in the games and reduce the number of interferences, the referee should only call a contact foul if it led to harm.

2. Desiderata

What is a good analysis of harm? What conditions must or should an analysis satisfy in order to be adequate? This section presents a list of desiderata for an analysis of the concept of harm. This list is based on the desiderata formulated by Ben Bradley (2012, pp. 394–396).[2] In the papers in this thesis, I make use of these desiderata when evaluating different proposals of what harm is. I am not alone in doing so. Many philosophers in the debate use desiderata borrowed from or inspired by Bradley to evaluate analyses of harm.[3]

Aiming to illustrate how the desiderata can be applied, I will use them to evaluate the following toy account of harm:

> **The Physical Injury Account (PIA)**: An event harms a subject if, and only if, the event causes the subject physical injury.

As we shall see, PIA satisfies some of the desiderata, but far from all.

Before presenting the desiderata, I should point out that the list is not intended to be exhaustive and the desiderata are intended to be desirable features rather than absolute requirements.[4] This means that there may be further desiderata which are not listed and that it is possible that we should accept an analysis that fails to satisfy some of the desiderata, especially if no analysis satisfies them all.

> *Desideratum 1: Extensional Adequacy*
> The analysis of harm should "fit the data". The analysis should for instance imply that killing someone ordinarily harms that person and be compatible with the idea that harm comes in degrees (meaning that events can be more or less harmful for someone).

I think that it is safe to say that the most common objection against analyses of harm is that there are counterexamples – which is another way of saying that they fail to satisfy *Extensional Adequacy*. Here is how such an objection,

[2] I do not use Bradley's (2012) exact formulations.
[3] See for instance Feit (2023), Johansson & Risberg (2023), Purves (2019), Rabenberg (2015), and Unruh (2023).
[4] This aligns with Bradley's intentions: "This list is not meant to be exhaustive. Nor do I claim that any of these is an absolute requirement for an acceptable theory of harm; they are desirable features […]" (2012, p. 396).

targeting the toy account, PIA, would go. PIA fails to satisfy *Extensional Adequacy*, since it is clear that we can be harmed without being physically injured.[5] I have mentioned some examples already: being (non-physically) hurt by loved ones, or being disrespected by someone whose opinion we care about. Simply getting a headache may be another example. If I get a headache from a long and intense day, I have hardly been injured. But it seems clear that, at least in normal circumstances, getting a headache harms me. The just mentioned examples illustrate that PIA fails to satisfy *Extensional Adequacy* by failing to capture the harm in cases that arguably are harmful. Another common way of phrasing this failure is to say that PIA *undergenerates* harm in these cases. But PIA plausibly *overgenerates* harm too. For not all physical injuries harm us – at least not overall.[6] If I scrape my knee because someone pushes me just to stop me from getting hit by a truck, I am plausibly not harmed overall – but rather benefited – despite the minor physical injury to my knee.

Depending on the interpretation of *Extensional Adequacy*, one may wonder whether this condition plausibly is an absolute condition. As already indicated, Bradley claims that no desideratum "is an absolute requirement for an acceptable theory of harm" (2012, p. 396). But consider the interpretation that being an extensionally adequate analysis means having true implications about particular cases. It then seems doubtful that an analysis that fails to satisfy *Extensional Adequacy* – i.e., has false implications – is "acceptable". Suppose that it is true that torturing a kitten harms the kitten. Consider an analysis of harm that manages to fit the data perfectly, except with respect to this "data point": the analysis implies that torturing a kitten is harmless. This analysis may deserve our attention or it may be onto something; but it must nevertheless be false.

However, our epistemic situation is arguably not such that we know for every particular event whether it is harmful or not. In other words, what the "data" are is itself a matter of dispute. Judgments about whether particular cases are cases of harming or not are often built on intuition. But philosophers have different intuitions and make different judgments about these cases. This invites the interpretation that *Extensional Adequacy* requires that an analysis is in line with our intuitions (perhaps not all, but those that are generally shared and thoughtfully reflected upon). Given this interpretation of *Extensional Adequacy* there seems to be no reason to think that this is an absolute condition. It is possible that we use the best of our abilities, but still end up with an inconsistent or otherwise faulty understanding of the concept; our intuitions may not correspond to the actual extension, due to limitations of our abilities.

[5] For a discussion of the differences between harming and injuring, see Feinberg (1984, Chapter 3, sec. 2).

[6] In the next section, I return to the distinction between *overall* (or *all-things-considered*) and *pro tanto* harm.

I think that the former interpretation is preferable: being an extensionally adequate analysis means having true implications. The desideratum so understood makes for more straightforward discussions. Given that our intuitions can be faulty, the desideratum so understood is more in line with what the term 'extensional' adequacy plausibly is supposed to capture. After all, on the alternative interpretation an analysis that captures the concept's extension perfectly can fail to satisfy the desideratum, simply because it fails to align with our intuitions.

> *Desideratum 2: Axiological Neutrality*
> The analysis should not presuppose any substantive theory of well-being.

A common view in the harm literature is that harm should be understood in terms of negatively affecting the subject's *well-being*. (Exactly how to spell that out is of course controversial.) However, Bradley claims, "[p]roponents of different axiologies should be able to agree—at some suitable level of abstraction—about what it takes for someone to be harmed" (2012, p. 394). The analysis should thus be compatible with different theories of well-being.

Given that PIA is a view about harm and includes no claims about well-being it does not by itself presuppose any theory of well-being. But given the assumption, say, that something lowers a subject's well-being only if it harms her, PIA is clearly problematic with regard to *Axiological Neutrality*. For combined with that assumption, it presupposes that only events that cause physical injury can lower our well-being level – desire frustration, suffering, and pain unaccompanied by physical injury cannot.

> *Desideratum 3: Ontological Neutrality*
> The analysis should be compatible with different sorts of (plausible) entities being both the "agents" and subjects of harm. The analysis should for instance allow that events besides intentional acts can harm – like earthquakes and attacks by animals. Similarly, the analysis should allow that beings besides human adults can be harmed – like non-human animals and human babies.

PIA does not have any obvious issue with fulfilling *Ontological Neutrality*, since it is possible for different sorts of beings to be the cause of and the victim of physical injury. PIA does well with regard to all of the examples listed in the description of the desideratum. But there may be other possible subjects of harm, which PIA rules out. So, whether PIA satisfies *Ontological Neutrality* depends on what subjects plausibly can be harmed. For instance, maybe it is plausible that *collectives* can be the subject of harm; but even if one or several members of a collective are physically injured it is not obvious that the *collective* is thereby physically injured.

Desideratum 4: Amorality
The analysis should be compatible with different (plausible) views on the moral significance of harm. The analysis should for instance not presuppose that an event necessarily is morally wrong because it harms or that the agent of harm necessarily *intends* to harm.

PIA seems to do well with regard to *Amorality*, since it is silent on the moral importance of physical injury.

Bradley (2012) claims that *Amorality* follows from *Ontological Neutrality* but treats it as a separate desideratum to emphasize it. If an earthquake can harm, then events can be harmful without being morally wrong and without involving an agent with the intention to harm. He suggests that it is best to test intuitions about harm using cases that do not involve bad intentions, in order to avoid moralistic fallacies.

Desideratum 5: Unity
The analysis should locate a common core to harm – i.e., align with the idea that all harms have something in common by specifying that thing. The analysis should for instance not simply be a list of bad things or include ad hoc features which have the sole purpose of accounting for some type of cases.

PIA satisfies *Unity*, since it specifies a non-ad hoc, common core to all harms: they cause physical injury.

Pluralist or disjunctive analyses of harm are, due to their nature, vulnerable to the objection that they fail to satisfy *Unity*. Disjunctive analyses usually have the following structure: An event harms a subject if, and only if, the event is *F* or the event is *G*.[7] On this analysis, two harmful events need neither share the property of being *F* nor the property of being *G*. One of the harmful events can be *F* (and not *G*), while the other is *G* (and not *F*). Since both are harms, but they have neither *F* nor *G* in common, it seems that this analysis fails to locate a common core to all harms.

However, one natural idea is that a disjunctive view can satisfy *Unity* if *F* and *G* are similar enough – i.e., there is a common core to *F* and *G*.[8] Consider this hybrid variant of PIA: An event harms a subject if, and only if, the event causes the subject physical injury or the event causes the subject physical pain. One could argue that this analysis satisfies *Unity* because there is a common core to physical injury and physical pain: they are both negative bodily states. That particular account is perhaps not plausible. But my point is that

[7] For examples of proposed disjunctive analyses, see Unruh (2023, 2024) and Woollard (2012) – also, cf. McMahan: "Although I cannot argue for this here, I suspect that a pluralist or disjunctive account of harm, which includes both comparative and noncomparative dimensions, is unavoidable" (2013, p. 8, note 3).
[8] For a similar defense of a disjunctive view, see Unruh (2023, sec. 4.1).

proponents of disjunctive analyses can argue that their view is unified, in the way just demonstrated.

> *Desideratum 6: Normative Importance*
> The analysis should accommodate the idea that harm is prudentially and morally significant. For instance, it should make sense for me to care about events described in the analysans if they happen to me. And when plugged into reasonable moral principles, the analysis should not make those principles absurd.[9]

Philosophers in the debate generally agree that harm is prudentially and morally important, in that it is in our interest to avoid things that harm us and that we should avoid harming others when we can. Exactly *how* prudentially and morally significant harm is and what that means in detail are questions open for debate. However, a common view is that we have a pro tanto prudential reason to avoid harm to ourselves and a pro tanto moral reason not to harm others.[10] The idea is that those reasons exist at least in part in virtue of the harm involved, meaning that they are *harm-based* prudential and moral reasons. To clarify further, if there is no moral reason in favor of performing a harmful act, the act's harmfulness makes it wrong to perform it. But if there is moral reason to perform a harmful act, it can outweigh the harm-based moral reason against performing it. If the reason for performing the act outweighs the reason against performing it, then it is morally acceptable to perform it.

Michael Rabenberg (2015, p. 2) points out that *Normative Importance* can be problematic if understood as requiring something stronger than the view described above (that harm gives rise to pro tanto moral and prudential reasons). In order to square *Normative Importance* with *Amorality,* the former cannot require that the analysis entails that harmful events are always wrong. Moreover, *Normative Importance* should arguably not tell against theories that allow for the existence of minor harms (perhaps a very short and mild episode of pain). It would do so if it required the analysis to accommodate the idea that harms are always hard to justify morally, because justifying a minor harm presumably is not hard. Similar remarks apply to the prudential importance of harm. If there are minor harms, they arguably play a more limited role in prudential deliberation. Thus, for *Amorality* and *Normative Importance* to be compatible, the latter cannot require that an analysis aligns with a view of harm's normative importance that is as strong as those just mentioned.

---

[9] Bradley has the two following separate desiderata: *Prudential Importance* and *Normative Importance*. I choose only to use the "Normative Importance" desideratum and include prudential considerations, since both prudential and moral considerations standardly are considered normative.

[10] Joseph Raz goes as far as claiming that "[…]'causing harm' entails by its very meaning that the action is prima facie wrong […]" (1988, p. 414).

PIA does not seem to satisfy *Normative Importance*. The problem is not obvious at first glance, because it does make sense to care about physical injuries in prudential and moral deliberations. I should try to avoid that they happen to me and I should avoid causing others to endure them. The problem is that we cannot plug PIA into reasonable moral principles without making them absurd. Recall the aforementioned *Harm Principle* attributed to Mill: the state cannot justifiably intervene against an individual unless it thereby prevents harm to others. Many find this principle plausible, but it is not with PIA plugged into it. Since PIA counts acts of discrimination, fraud, threat, and blackmailing (unaccompanied by physical injuries) as harmless, PIA and the Harm Principle imply that individuals performing such acts are never eligible for punishment or other types of interventions.

As I mentioned in the introduction, harm's normative importance is a central theme in this thesis. For instance, Paper II, "Feit on the Normative Importance of Harm", discusses an objection to a prominent counterfactual analysis of harm, saying that this account fails to accommodate the idea that harm is morally significant. In Paper I, "The Harm Principle and the Nature of Harm", I address a certain type of criticism against Mill's *Harm Principle*, which draws on objections similar to the just mentioned objection to PIA. The critics argue that since there is *no* analysis of harm such that Mill's *Harm Principle* is plausible when that account is plugged into it, the principle is implausible. In Paper III, "Doing Away with Skepticism About Harm", I discuss the idea that *Normative Importance* is an implausible desideratum. More specifically, I challenge the view that the fact that an event is harmful never constitutes a moral reason to avoid or prevent it.

In Paper IV, "Misfortune and Missing Out", I suggest a novel desideratum for an analysis of harm:

> *Desideratum 7: The Fitting Emotions Desideratum*
> The analysis should accommodate the idea that it is fitting for a subject of harm to have negative emotional responses toward the harmful event for her own sake.

This desideratum highlights the tight connection between harmfulness and fitting negative attitudes. Examples of negative emotions that can be fitting to have toward harmful events are dread, sadness, despair, disappointment and fear.

PIA does not satisfy this desideratum. Since a successful hernia surgery causes you physical injury, PIA entails that the surgery is harmful for you. But clearly, since the surgery repairs the hernia, it is fitting for you to have overall positive emotional responses to it for your own sake. Hence, PIA fails to make sense of the idea that it is fitting for us to have negative emotional responses for our own sake toward harms.

In sum, the toy account, PIA, seems to clearly satisfy *Amorality* and *Unity*. With regard to the rest of the desiderata, PIA is unsuccessful or its success depends on some further questions that I have raised. As can be expected, many other analyses discussed in this thesis do better with regard to the desiderata.

I hope this section has made it clear that giving an analysis of harm is not solely trying to capture ordinary language use. In other words, the mission is not merely giving an analysis that would suit a dictionary. Rather, philosophers try to capture a concept of harm that is philosophically fruitful.[11] Most notably, the concept must be morally significant in order to play the role assigned to it by many philosophical arguments and principles. Some of the distinctions presented in the upcoming section further explicate what giving an analysis of the nature of harm is meant to achieve.

[11] Feinberg claims that since the word 'harm' is both ambiguous and vague, we should for the purposes of normative theorizing select the relevant sense and make it sufficiently precise: "[I]nsofar as it is ambiguous, we must select among its normal senses the one or ones relevant for our normative purposes, and insofar as it is vague in those senses, it should be made more precise—a task that requires some degree of stipulation, not simply a more accurate reporting of current usage" (1984, p. 32).

3. Distinctions

There are some helpful and commonly used distinctions between different *types of harm* and different *types of analyses* of harm. There are also distinctions that help explicate more exactly what philosophers in the debate take an analysis of harm to be. As will become clear, it is controversial how to best understand some of the distinctions in this section. I will do my best to explain the most important disagreements and motivate my own views. I will begin the discussion by presenting some distinctions between different types of harm. The two most important distinctions of that kind are *pro tanto* vs. *overall* (or *all-things-considered*) harm and *intrinsic* vs. *extrinsic* harm.

Pro Tanto and Overall Harm

The distinction between pro tanto and overall harm is useful in relation to the idea that an event with both positive and negative effects on a subject can on balance be either positive, negative or neutral on the whole. Again, consider the event of going through surgery. A surgery can be overall beneficial in virtue of curing a serious illness, although it is also pro tanto harmful in virtue of the pain it leads to and what it makes you miss out on when you lay in bed recovering from it. If unsuccessful, a surgery can be overall harmful for you, since the pro tanto harm outweighs the pro tanto benefit (if any).

Intrinsic and Extrinsic Harm

Here is a common understanding of the distinction between intrinsic and extrinsic harm: roughly, intrinsic harms are harmful in virtue of their intrinsic properties (or in virtue of themselves) and extrinsic harms are harmful in virtue of their extrinsic properties (Bradley, 2012; Feit, 2023, Chapter 1; Klocksiem, 2012, p. 13; Purves, 2016, p. 90). Assuming hedonism about well-being, Klocksiem gives the following example:

> [S]uppose that Archie is diagnosed with cancer. If it is left untreated, the tumor is very likely to harm him, but this harm occurs only because there exists a mechanism by which the tumor will cause pain, suffering, and an early death, and is therefore merely extrinsic; whereas the event consisting of Archie's experience of the pain itself might plausibly be regarded as an intrinsic harm. (2012, p. 13)

Common examples of extrinsic properties in virtue of which extrinsically harmful events are harmful are *causes* of the events and what the events *prevent*. In the example above, Archie's getting a tumor is considered extrinsically harmful in virtue of causing pain, suffering, and early death.

Some make the further and more controversial claim that for something to be *intrinsically harmful* for a subject is for it to be *intrinsically bad* for the subject (Bradley, 2012; Feit, 2023, Chapter 1). What thing or things are intrinsically bad for us is a controversial question, but pain and desire frustration are two of the main candidates. I take no stand on this further claim that the relation between intrinsic harm and intrinsic prudential badness is identity.

Philosophers are not always explicit about what type of harm they discuss; those who are explicit mainly focus on *overall extrinsic* harm (see for instance Bradley, 2012; Feit, 2023). They also tend to identify intrinsic harm with intrinsic badness (i.e., they subscribe to the claim I mentioned in the previous paragraph). Given that assumption, it is natural to think that discussions of the harm concept should center around extrinsic harm, since discussions of intrinsic harm essentially regard well-being. (And the discussion of what well-being amounts to is a separate philosophical debate.) The focus on overall harm may at least partly be explained by the simple fact that we often care about how an event affects a subject on balance, counting both the negative and the positive. While overall harm takes into account pro tanto harms and benefits, the converse is not true.

Comparative and Non-Comparative Analyses

Other distinctions – distinctions between different types of analyses (or accounts) – are useful to better understand the various analyses that philosophers have proposed and the crucial differences between them. A frequently employed distinction is that between *comparative* and *non-comparative* analyses of harm.

Some philosophers understand the distinction as one between accounts that appeal to a comparative notion (such as comparisons between possible worlds or comparisons over time) and accounts that do not appeal to such a notion (Bradley, 2012; Gardner, 2021, p. 389, note 13; Rabenberg, 2015).[12] They provide no explanation of what a comparative notion is. Intuitively speaking, there are many ways for something to be comparatively negative for a subject.

[12] Some philosophers use the comparative/non-comparative distinction to describe different types of harms, rather than different accounts of harm (McMahan, 2022; Woollard, 2012). According to Fiona Woollard, this is how I harm someone comparatively: "I harm someone if and only if my behaviour leads to him being worse off overall than he would have been if I had acted differently" (2012, p. 685). This is how I harm someone non-comparatively: "I harm someone if and only if I bring it about, in a sufficiently direct manner, that he suffers a harm" (ibid). Woollard takes this distinction to pick out two senses of the term 'harming'. That is different from how it is commonly used, namely, as mentioned before, to pick out two different categories of accounts.

To mention some examples, an event can make a subject worse off compared to a neutral level, compared to how well off some relevant comparison subject is, compared to how well off she otherwise would have been, compared to how well off she was before, compared to how well off she could have been, or compared to how well off she should have been.

It is not obvious that we can spell out the distinction more precisely and still keep the traditional grouping of which analyses belong to the comparative and non-comparative side. For instance, the view that harming consists of *causing something that is intrinsically bad for the subject* traditionally counts as a non-comparative account. However, given the not uncommon view that 'badness' should be understood in terms of 'worseness', this account seems to appeal to a comparative notion.

I have not ruled out that the distinction between comparative and non-comparative accounts can be specified in a way that makes this distinction useful. However, it is noteworthy that many prominent accounts of harm appeal to either counterfactual dependence or causation – where causation is not understood as mere counterfactual dependence. Therefore, I will use the categorization of *counterfactual* and *causal* analyses in the next section of this general introduction, where I present central accounts of harm.

The Currency and the Structure of Harm

There are some distinctions that help explicate what philosophers in the debate take an analysis of harm to be. The *currency* of harm is commonly distinguished from the *structure* of harm – and both are part of an analysis of harm.[13] The currency tells us what about the subject a harmful event interferes with. As I mentioned earlier (without using the term 'currency'), the most common suggestion is that the currency of harm is the subject's *well-being* (Feit, 2015; Gardner, 2021; Johansson & Risberg, 2023). Other suggestions are the subject's *rational will* (Shiffrin, 1999), her *interests* (Feinberg, 1984), or her *opportunities* for future well-being or to exercise autonomy (Raz, 1988, p. 414; Simester, 2011, Chapter 3). A proponent of the toy account, PIA, would say that the currency of harm is *physical injury*.

Roughly put, the structure of harm is the answer to *how* the harmful event interferes with the currency: is it by making the subject worse off in some sense, is it by causation, or something else? PIA is an example of an analysis with a *causal* structure. As we shall see in the next section, a popular counterfactual analysis is that an event harms a subject if, and only if, the subject

[13] Gardner (2021, p. 381) uses the terms 'formal analysis' and 'substantive component' to refer to the structure and the currency. Tadros (2014, p. 172) uses the term 'measure' of harm to refer to the structure. I think that using 'measure' for this purpose is misleading, since measuring something is connected to there being more or less of something.

would have been better off in terms of well-being had the event not occurred. On that analysis, the currency is well-being and the structure is *counterfactual*.

Linguistic, Conceptual, and Metaphysical Analyses

Another useful distinction is that between analyses of the concept HARM (conceptual analyses), the meaning of the word 'harm' (linguistic analyses), and the nature of the phenomenon harm (metaphysical analyses) (cf. Johansson & Risberg, 2023, p. 511). Philosophers in the debate seldom say explicitly which category the analysis they discuss belongs to. This is unfortunate. Consider the fact that 'the morning star' and 'the evening star' do not have the same meaning, although they have the same extension (the planet Venus). This means that an argument for, or objection to, a particular analysis may be plausible when the analysis is conceived of as metaphysical, but implausible when conceived of as linguistic.

That being said, this distinction does not generally play an important role in the papers in the thesis. For instance, while I consider the novel analysis that I propose in Paper V, "A Fitting Attitudes Analysis of Harm", an attractive *conceptual* analysis, I note that many of my arguments in that paper are unaffected by whether that analysis and its competitors are conceived of as linguistic, conceptual or metaphysical.

Harming and Harm

The final distinction I will present is that between *harming* and *harm.* Some philosophers find this distinction widely overlooked and important for giving a correct analysis; they give explicit and separate conditions for harming (when an event harms a subject) and harm (when a state of affairs is a harm for a subject) (Gardner, 2015, 2019, 2021; Unruh, 2023, 2024, p. 297). As will become clear in the next section, given this distinction, many analyses labelled an analysis of 'harm' should strictly speaking be labelled an analysis of 'harming'. For they give an analysis of when an event harms a subject.

4. Central Accounts of Harm

Before presenting prominent accounts of harm along with their virtues and vices, a comment about *benefit* is in order. The most common view in the literature is that harm and benefit are mirror opposites. That is, benefit is positive for its subject in the same way that harm is negative for its subject. For instance, if harming someone consists in causing them to be in a *bad* state, then benefiting them is causing them to be in a *good* state. Many of the accounts below are explicitly accounts of both harm and benefit, where benefit is the mirror opposite of harm.

Counterfactual Accounts

As previously mentioned, there are two main categories among the prominent accounts of harm: counterfactual accounts and causal accounts. Counterfactual accounts align with this general formula:

> An event harms (benefits) a subject if, and only if, the subject is better (worse) off *in counterfactual scenario X*.

There are different versions of counterfactual accounts, since counterfactual scenario X is spelled out in different ways. Here follows the most popular counterfactual account (Boonin, 2014, Chapter 3; Feit, 2015, 2016, 2019, 2023; Klocksiem, 2012, 2022; Parfit, 1984, p. 69):

> **The Counterfactual Comparative Account of Harm (Benefit) (CCA):** An event harms (benefits) a subject if, and only if, the subject would have been better (worse) off in terms of well-being if the event had not occurred.

CCA thus compares the subject's well-being in the scenario where an event occurs with her well-being in the closest possible counterfactual scenario where that event does not occur. (CCA proponents typically hold that it is the subject's well-being over her *lifetime* that matters for overall harm.) If a subject is better off in terms of well-being in the counterfactual scenario where the event does not occur, then CCA implies that the event harms the subject.

This account is considered to have many virtues. For example, it is considered a natural and intuitive understanding of harm, to have explanatory power, to capture deprivational harms (like death), and to align with the idea that harm is prudentially and morally important (Bradley, 2012; Feit, 2015, 2019; Klocksiem, 2012).

Among the most discussed problems for CCA are the so-called *pre-emption, non-identity* and *omission* problems. I will present those problems one by one, before presenting some commonly suggested solutions. The following cases illustrate the *pre-emption* problem for CCA.[14]

> *Batman's Heart Attack*: Suppose Batman drops dead of a heart attack. A millisecond after his death, his body is hit by a flaming cannonball. The cannonball would have killed Batman if he had still been alive. (Bradley, 2012, p. 397)

> *Bobby Knight*: Bobby Knight chokes a philosopher, injuring her windpipe; if he hadn't choked her, he would have torn her arms off, which would have been much worse for her. (Bradley, 2012, p. 407)[15]

CCA counterintuitively implies that Batman's dying of a heart attack does not harm him, because he would not have been better off in the absence of that event. In its absence, he would have died by a flaming cannonball a millisecond later.[16] Even worse, one may think, CCA implies that Bobby Knight's choking the philosopher benefits her rather than harms her. For had Bobby Knight not done so, he would have torn her arms off – which would have been even worse for her.

Another problem for CCA stems from the discussion of our moral duties to future generations. More particularly, can acts that are performed in the present and that affect which people later exist, harm and wrong the people that later exist? Many think that they can, for instance by deciding on environmental policies which make future people suffer. CCA is thought to be inconsistent with that idea. The following case illustrates the *non-identity* problem for CCA (for discussion of the non-identity problem, see for instance Boonin, 2008, 2014, 2019; Gardner, 2015; Harman 2004, 2009; Parfit, 1984, ch. 16; Roberts, 2007, 2009).

---

[14] The *overdetermination* problem for CCA is similar (see for instance Norcross, 2005, p. 152; Parfit, 1984, p. 70; and Petersson, 2018).

[15] This case was originally formulated by Alastair Norcross (2005, pp. 165–166).

[16] To avoid the complication that dying from the heart attack is harmful (to a very minor degree) because living a good life for a millisecond longer would have been better for Batman, we can assume that Batman's well-being level would have been neutral during the extra millisecond that he would have lived had he not died of the heart attack.

> *Mary's Baby*: Suppose Mary is contemplating pregnancy. If she becomes pregnant now, she will conceive a child, Jane, who will have a painful disease. If she waits a few months to conceive, she will conceive a different child, John, who will not have that disease. In that case, Jane would never come into existence at all. Mary chooses to conceive Jane. Jane lives a good life on the whole, despite the pain she endures from her disease; but due to all that pain, her life is much worse than the relatively pain-free life John would have had if she had waited. (Bradley, 2012, p. 398).

This is called a non-identity case since Jane appears to be wronged by Mary's decision although it is a condition for her existence and although her life is worth living. Moreover, many think that Mary acts wrongly because her act harms Jane (Bontly, 2016; Gardner, 2015; Harman, 2004, 2009; Roberts, 2007, 2009; Unruh, 2019; Woollard, 2012). However, given the assumption that Jane would not have had a well-being level had she not existed, CCA entails that Mary's decision to conceive when she does is not harmful for Jane. For given that assumption, Jan is not better off in the relevant comparison scenario, as she lacks a well-being level in that scenario. This does not rule out that Mary acts wrongly for some other reason – but CCA seems incompatible with the claim that Mary's decision is harmful for Jane.

The *omission* (or *failing to benefit*) problem for CCA is thought to bring out a slightly different kind of flaw than the two previous problems. The preemption and non-identity problems are thought to show that CCA fails to satisfy *Extensional Adequacy* (see section 2. Desiderata) by counting harmful events as harmless. The omission problem for CCA is that it fails to satisfy that desideratum by counting harmless events as harmful. The following case illustrates the problem:

> *Golf Clubs*: Suppose Batman purchases a set of golf clubs, which Batman intends to give to Robin and which Robin would be happy to receive. Batman tells The Joker about his intentions. The Joker says to Batman, ''why not keep them for yourself?'' Batman is persuaded. He keeps the golf clubs. (Bradley, 2012, p. 397)

CCA counts Batman's decision to keep the clubs as harmful for Robin, since he would have been better off had Batman decided otherwise. This is counterintuitive since his decision seems to be a mere (non-harmful) failure to benefit.[17]

[17] It is easy to design cases to illustrate the analogous problem, which we may call the *failing to harm* problem: Batman contemplates stealing a set of golf clubs from Robin. He decides to not steal them. Had he not decided that, he would have stolen them – which would have made Robin worse off in terms of well-being. Intuitively, Batman's decision is a "failure to harm": it

In Paper IV, "Misfortune and Missing Out", I formulate a further problem for CCA. I argue that it fails to satisfy *The Fitting Emotions Desideratum* – and thereby fails to align with the plausible idea that it is fitting for a subject of harm to have negative emotional responses (such as dread, sadness, despair, disappointment and fear) toward the harmful event. For one thing, consider that it does not seem fitting for Robin to have such negative emotional responses toward Batman's decision. That is especially clear if we stipulate that Batman's decision does not result in Robin's well-being level being lower than it was before; for instance, it is not the case that Robin learns about Batman's decision and feels insulted and sad because he really cares about golf clubs.

There are two main responses that friends of CCA give to the pre-emption, non-identity, and omission problems. Some argue that despite first appearances, they are not genuine problems. Neil Feit (2019) argues that Batman's decision in *Golf Clubs* is a genuine form of harming: harming by failing to benefit.[18] In Paper II, "Feit on the Normative Importance of Harm", I reply to Feit's arguments for that claim. David Boonin (2008, 2014) argues that we should accept that Mary's decision in *Mary's Baby* is both harmless and morally permissible. Justin Klocksiem (2012) argues that since harm judgments, as well as counterfactuals, are highly sensitive to context, CCA can handle the pre-emption problem.

The other main response amounts to proposing revised versions of CCA, or additions to it, which are supposed to handle (some of) the problems. One proposal is Feit's (2015, 2023) account of *plural harm*. Here follows a slightly simplified version of it, where E is a plurality of events (Feit, 2023, sec. 4.5).

> **Plural Harm:** *E* harms a subject, if and only if, (1) if none of the events of *E* had occurred, *S* would have been better off; and (2) each event *e* of *E* is such that *S* would not have been better off if *e*, but none of the other events of *E*, had occurred.

This account is thought to capture the harm in *Batman's Heart Attack* by pointing out that had neither the heart attack happened nor the cannonball approached him, Batman would have been fine. On Feit's account, what harms Batman is a plurality consisting of two events: *Batman's having a heart attack* and *the cannonball's approaching his body*. Had neither of those events occurred, Batman would not have died and thereby been better off. Feit gives an analogous explanation of the harm in *Bobby Knight*: "In this case, certain of Knight's token psychological states (perhaps feelings of rage) ground the

seems to be neither harmful nor beneficial to Robin. On CCA however, Batman's decision not to steal Robin's golf clubs benefits Robins.

[18] Nathan Hanna (2016) gives another argument for the claim that CCA's verdicts in omission cases are correct. See also Tanya de Villiers-Botha's (2018) criticism of Hanna's argument.

counterfactual, and so these events, with the choking, constitute the harm" (2015, p. 381).

This account has been accused of failing to make sense of the idea that making a change for the worse is essential for harming (Petersson, 2018). Some argue that this account fails with regard to some paradigmatic examples of events that together harm a subject, for instance a case in which two agents act such that the subject dies, where none of the individual acts would have been lethal by itself (Carlson et al., 2023, sec. 7).

Daniel Immerman (2022, forthcoming) proposes an alternative counterfactual account that he argues escapes the pre-emption and omission problems. Here is the account spelled out:

> **The Worse than Nothing Account:** An agent harms a subject, if and only if, the subject would have been better off had the agent done nothing at all.

Consider what this account says about *Golf Clubs*. Compare the scenario in which Batman decides to keep the clubs for himself to the scenario in which Batman does nothing at all. Since Robin is equally well off in those two scenarios, Immerman claims, the Worse than Nothing Account implies that Robin is not harmed by Batman.

Since this account is mainly motivated by pre-emption cases like *Bobby Knight* it is no wonder that it provides a straightforward explanation of the harm involved in that case. The idea is that since the philosopher would have been much better off had Bobby Knight done nothing at all (including not tearing her arms off, of course) Bobby Knight's choking her constitutes a harm to her. Although Immerman presents his account mainly as a solution to the pre-emption problem, it is not obvious that it can handle *Batman's Heart Attack*. Since *Batman's Heart Attack* involves no relevant agent and no single event such that had it not occurred Batman would have been better off, this account needs to be expanded to cover events – and more precisely pluralities of events – in order to deal with this particular case. Immerman (2022, secs. 2.1 and 2.3) discusses and embraces the possibility of expanding the account to cover single events and plural agents – but not pluralities of events.[19]

The final counterfactual account that I will present is the following contrastive account (Norcross, 2005; for a variant, see Gunnemyr, 2023).

> **Contrastive Harm:** Performing act A rather than act B harms (benefits) a subject if, and only if, the subject would have been better (worse) off had B been performed.

[19] For discussion of problems with The Worse than Nothing Account, see Johansson and Risberg (2022, forthcoming).

Unlike on the other counterfactual accounts, on this account there are no true harm claims of the following sort: "my shooting that person harmed her". But contrastive harm claims can be true: "my shooting that person *rather than eating an apple* harmed her". However, it is also the case that my shooting that person *rather than torturing and then giving her deadly poison* is *beneficial* for her. This means that my shooting a person is harmful contrasted with some alternative acts and beneficial contrasted with others. Therefore, and since there is no non-contrastive fact about whether my act harms the person, this account is accused of failing to satisfy *Normative Importance*. Bradley claims that "[…] no deontological principle prohibiting harm will be remotely plausible if the contrastive account of harm is true" (2012, p. 408).[20]

Causal Accounts

Note that the counterfactual accounts presented above are difficult to square with the idea that Mary's decision harms Jane in *Mary's Baby*. For instance, Jane would not have been better off had Mary done nothing at all. And there seems to be no suitable plurality of events such that those events together harm Jane. Hence, the non-identity problem is a significant motivation for the second family of analyses: *causal analyses* that do not identify causation with counterfactual dependence.

Causal analyses generally align with this formula:

> An event harms (benefits) a subject if, and only if, the event causes a harm for the subject.

This formula makes use of the previously introduced distinction between harming and harm. In slogan form, causal accounts say that *harming is causing harm*. Due to the fact that causal accounts give separate conditions for harming and harm, I will in the rest of the discussion of these accounts be careful to distinguish between 'harming' and 'harm'.

One of the most discussed causal accounts of harming is the following (cf. Harman, 2009; Smuts, 2012):

> **The Causal-Intrinsic Badness Account:** An event harms (benefits) a subject if, and only if, it causes a state of affairs that is intrinsically bad (good) for the subject.

On this account, a *harm* is a state of affairs that is intrinsically bad for the subject; and *harming* occurs just in case an event causes such a state of affairs to obtain. Let us suppose that pain is intrinsically bad for us. This account then

[20] Cf. Norcross's (2005, pp. 171–172) discussion about his own contrastive view, concerning harm's (limited) role in ethical theorizing.

appears to correctly categorize many of the problem cases. Since Bobby Knight's choking the philosopher causes the philosopher to be in pain, that amounts to harming on this account. Since Mary's decision causes Jane to be in pain, her decision is (at least pro tanto) harmful for Jane on this account.[21] Since Batman's decision to keep the golf clubs instead of giving them to Robin does not cause Robin any pain, this account entails that it is harmless.

This account is considered problematic primarily due to its inability to capture the harm of death (Bradley, 2012, pp. 400–401).[22] If death is the end of our existence, then it arguably does not cause anything intrinsically bad for us – for instance, we cannot be in pain after our death.

Another prominent causal account of harming is Molly Gardner's (2015, 2017, 2019, 2021) account.[23]

> **Gardner's Causal-Counterfactual Account:** An event harms a subject, if and only if, it causes a state of affairs such that had the subject existed and the state of affairs not obtained, then the subject would have been better off.[24]

Gardner argues that the non-identity problem should be solved by appealing to this account; this account entails that Mary's decision harms Jane. The idea is that Mary's decision causes the state of affairs of Jane having a painful disease. And had Jane existed and that state of affairs not obtained, then Jane would have been better off. This account can capture the harm in pre-emption cases too, Gardner claims. Consider the state of affairs of, say, the philosopher being in pain. Had (the philosopher existed and) that state of affairs not obtained, then the philosopher would have been better off. And since Bobby Knight's choking the philosopher arguably causes that state of affairs to obtain, this account entails that Bobby Knight's choking the philosopher harms her. This account also seems to deliver the intuitively plausible result in *Golf Clubs*; a proponent of a causal view can arguably say that Batman's decision *allows* but does not *cause* the state of affairs of, say, Robin's owning no golf clubs.

---

[21] Since Mary's decision also causes states of affairs that are intrinsically good for Jane, this account plausibly entails that Mary's decision both pro tanto harms and pro tanto benefits Jane. To solve the non-identity problem, one has to claim that Mary's decision is wrong because it is pro tanto harmful for Jane – despite the fact that it is also pro tanto beneficial for her. (Similar remarks apply to Gardner's suggested solution, which I present shortly.) Elizabeth Harman advances such an argument, according to which the pro tanto benefits cannot morally outweigh the pro tanto harms: "Reasons against harm are morally serious reasons that are difficult to outweigh; the mere presence of benefits more beneficial than the harms are harmful is not sufficient to render harming permissible" (2004, p. 108).

[22] For discussion of other problems for the Causal-Intrinsic Badness Account and some modifications of it, see for instance Gardner (2015), Rabenberg (2015), and Carlson et al. (2022).

[23] Bontly (2016) and Northcott (2015) propose similar accounts, which we may also call 'causal-counterfactual' accounts.

[24] I use a simplified formulation. For her own formulation, see Gardner (2021, pp. 390–391).

Boonin (2019) argues that Gardner's account fails with regard to *Normative Importance* and *Extensional Adequacy*; for instance, he argues that some acts that are harmful on this account are insignificant, such that we have no good reason to feel entitled to an apology or compensation, feel resentment toward or disrespected by the agent, wish that the act had not been performed, etc. Carlson et al. (2022, sec. 5) suggest that a causal-counterfactual account is the most promising type of causal account of harming; but they argue that it fails to satisfy *Extensional Adequacy* by overgenerating harming in various sorts of cases. For instance, they present a version of *Golf Clubs* where Batman has bought two golf clubs, each of which will give its owner 10 units of pleasure. Batman gives Robin exactly one of them, which gives Robin exactly 10 units of pleasure. Had Batman not done so, he would have given Robin both clubs, which would have given Robin 20 units of pleasure (Carlson et al., 2022, p. 433). According to Carlson et al., since the event of Batman's giving Robin exactly one golf club causes the state of affairs of Robin's owning exactly one golf club (and Robin would have been better off had that state of affairs not obtained), Gardner's account counterintuitively implies that the event harms Robin.

The last causal account that I will mention can be called The Causal-'Temporal' Account, since it appeals to a comparison between how well the subject does before and after the occurrence of an event (Foddy, 2014; Perry, 2003).

> **The Causal-Temporal Account (CTA)**: An event, occurring at time *t*, harms a subject if and only if it causes the subject to be worse off after *t* than the subject was before *t*.[25]

This account captures the harm in *Bobby Knight*, since Bobby Knight's choking the philosopher causes the philosopher to be worse off after that event than she was before it. Given that Batman's decision in *Golf Clubs* does not cause Robin to be worse off after the decision than he was before it, CTA implies that it is harmless. CTA's implications in non-identity cases and cases involving death are unclear. For they depend on further claims about when subjects start and stop existing and whether we can make sense of comparisons between how well a subject does when existing compared to before and after.

One of the main objections to CTA is that it undergenerates harming in so-called 'preventive harm' cases: "Suppose a person's pain would have gone

[25] Note that the account is highly implausible if spelled out without a causal (or similar) element, in the following manner: an event harms a subject if, and only if, she is worse off after the event than she was before the event. This account must be false, since it implies that all events that occur at the same time as a harmful event are themselves harmful (Carlson et al., 2022, p. 427). For instance, say that breaking my leg at *t* harms me, since I am worse off after *t* than I was before. This account implies that *all* events (with the same duration) that occur at *t* are harmful for me, because it is true of all of them that I am worse off after their occurrence than I was before. This includes events that have no relevant connection to me; a deep-sea anglerfish's flapping its fins at *t* harms me.

away, had I not acted to ensure that it continues. Clearly, I harm him, despite the fact that I leave him in no more pain than he was prior to my intervention" (Holtug, 2002, p. 368). CTA thus counts events that prevent someone from getting better as harmless.

Other Accounts

There are some analyses that do not neatly fit in either the counterfactual or the causal category. They do not employ either counterfactual dependence or causation to spell out the relation between the event and the subject.

One of those analyses is the following, recently proposed by Jens Johansson and Olle Risberg (2023):

> **The Negative Influence on Well-Being Account (NIWA):** An event harms a subject if, and only if, the event adversely affects the subject's well-being.

Since 'to affect' is not intended to be analyzable further in purely counterfactual or causal terms, this account falls outside the counterfactual/causal categorization. As Johansson and Risberg note, this account may strike one as unsatisfactory due to being uninformative. (That complaint plausibly targets the proposed structure of harm and not the currency. On NIWA the currency of harm is well-being, which is the standard view as I have mentioned before.) Although they themselves agree that NIWA is less informative in some respect than for instance CCA, they suggest that expecting an adequate analysis to be more informative may be expecting too much.

Another example of an analysis that fits in neither category is the novel analysis that I propose in Paper V, "A Fitting Attitudes Analysis of Harm".

> **The Fitting Attitudes Analysis of Harm (Benefit) (FAAH):** An event harms (benefits) a subject if, and only if, it is fitting for the subject to have negative (positive) emotional responses toward the event for the subject's own sake.

Along with straightforwardly satisfying *The Fitting Emotions Desideratum*, I argue in Paper V that this account satisfies all of the other previously presented desiderata. I argue that the account is extensionally adequate, by showing how it handles problem cases like those discussed in this section. For instance, it is arguably fitting for the philosopher to have negative emotions toward Bobby Knight's choking her for her own sake and for Batman to have negative emotions toward his heart attack for his own sake. So, FAAH's verdicts align with the intuitive judgment: those events are harmful.

5. Paper Summaries

In this section I will provide brief summaries of the papers in this thesis. A thread of argumentation running through all the papers is a defense of the concept of harm. In light of skeptical arguments, I defend its place in philosophical arguments and principles, its normative importance, and its connection to fitting negative emotions. I also defend certain normative principles that invoke it and the idea that continuing to theorize about its nature is a worthwhile project. Another reoccurring theme is objections to the popular Counterfactual Comparative Account (CCA). Objections to CCA is a common topic of discussion in the debate; and I add reasons to doubt that CCA is plausible. Finally, another issue that reoccurs in the papers is how harmfulness and prudential badness relate to each other. While it is an intuitive idea that being *harmful for* a subject is closely related to – or even the same thing as –being *bad for* the subject, the idea that harm should not be identified with prudential badness becomes relevant in several of the discussions in the papers.

Paper I: The Harm Principle and the Nature of Harm

This paper defends John Stuart Mill's Harm Principle against recent criticism. This principle says that the state can justifiably intervene against an individual only if it thereby prevents harm to others. Some critics argue that we should reject the principle because there is no unproblematic account of harm to support it. I examine this type of criticism, focusing on various accounts of harm, such as CCA and the Temporal Comparative Account (TCA). I show that the objections do not refute the Harm Principle. I suggest that the critics' assumptions – such as the need for an unproblematic theory of harm – are misguided. I conclude by suggesting ways in which the concept of harm is useful for philosophical theorizing even if it is true that no unproblematic account of harm has been formulated.

Paper II: Feit on the Normative Importance of Harm

This paper argues against the view that some harms lack moral significance. In light of the objection that CCA is incompatible with common views about harm's moral importance, some CCA proponents argue that harm is not as morally important as philosophers often think. The most elaborate argument

of that kind is given by Neil Feit. He rejects the so-called "Strong View", which roughly is that there are strong moral reasons against harming. Moreover, he argues that CCA can respect the commonsense asymmetry between the moral significance of harms and benefits: harming someone is more morally problematic than failing to provide someone with a (proportionately sized) benefit. I argue that Feit's defense of CCA is unsuccessful. I show that his attempts to provide counterexamples to the Strong View do not succeed. With regard to the asymmetry, I argue that Feit at best has shown that CCA can respect a version of it that is not part of common sense. But I also point out that his arguments for that version of the asymmetry are at odds with his objection to the Strong View. I conclude that the challenge for CCA remains.

Paper III: Doing Away with Skepticism about Harm

In this paper, I address the idea that we should do away with the concept of harm in philosophical theorizing. Some philosophers claim that the harm concept is not fit to play the central role in philosophical theorizing that it does: a vast number of theories, principles and arguments appeal to it. Ben Bradley argues that the harm concept is so problematic that we should eliminate it from philosophical theorizing. I call this *the elimination thesis*. Other philosophers propose versions of the elimination thesis. I address several claims in support of the elimination thesis, for instance that the harm concept is defective (or a "Frankenstein" concept) and that it lacks moral significance. I argue, contrary to Bradley's assertions, that even if various analyses of harm suffer from the serious problems that Bradley claims they do, that is a poor reason for thinking that we are dealing with a "Frankenstein", disunified and disjunctive, concept. I also argue that convincing objections to the intuitively plausible view that the harm concept is morally significant are lacking. I conclude that the challenges from skepticism can be met: they do not show that harm's central role in philosophical theorizing is illegitimate.

Paper IV: Misfortune and Missing Out

This paper discusses Kaila Draper's famous challenge to deprivationism about the badness of death. Proponents of deprivationism argue that when death is bad – and supposedly a misfortune – for the person who dies, this is because death *deprives* her of future intrinsic goods. Draper challenges this view by arguing that some such deprivations, like missing out on winning the lottery, are not misfortunes in the sense that they are worthy of fear, sadness or other negative emotions. Death's being a deprivation thus cannot explain why it is a misfortune. In the first part of this paper, I critique different strategies of addressing the challenge, for instance Ben Bradley's attempt to explain the apparent unfittingness of negative emotions toward certain deprivations by distinguishing between *merited* and *rational* emotions. In the second part of

the paper, I explore an analogous challenge involving fitting emotions toward *harm*. I argue that a version of Draper's challenge affects CCA.

Paper V: A Fitting Attitudes Analysis of Harm

This paper proposes a novel analysis of harm (and benefit), which I call the Fitting Attitudes Analysis of Harm (Benefit) (FAAH): *an event harms (benefits) a subject if, and only if, it is fitting for the subject to disfavor (favor) the event for the subject's own sake.* Despite the fact that fitting attitudes analyses are popular concerning closely related concepts, such as prudential badness, there is to my knowledge no similar analysis of harm proposed in the literature. I argue that FAAH has several merits: it is equipped to handle cases that are thought to pose serious problems for competing analyses; it satisfies commonly accepted desiderata for an analysis of harm. For instance, it satisfies the desideratum that an analysis of harm should accommodate the idea that harm is normatively significant; and it amounts to a unified understanding of intrinsic and extrinsic harm and benefit. I address two potential objections to FAAH, appealing to cases that allegedly show that fitting negative attitudes and harm can come apart. I argue that, given the correct understanding of FAAH and plausible claims about the relevant cases, the objections are not forceful.

References

Beauchamp, T. L., & Childress, J. F. (1979). *Principles of Biomedical Ethics*. Oxford University Press.

Bontly, T. D. (2016). Causes, Contrasts, and the Non-Identity Problem. *Philosophical Studies*, *173*(5), 1233–1251.

Boonin, D. (2008). How to Solve the Non-Identity Problem. *Public Affairs Quarterly*, *22*(2), 129–159.

Boonin, D. (2014). *The Non-Identity Problem and the Ethics of Future People*. Oxford University Press.

Boonin, D. (2019). Solving the Non-Identity Problem: A Reply to Gardner, Kumar, Malek, Mulgan, Roberts and Wasserman. *Law, Ethics and Philosophy*, *7*, 127–156.

Bradley, B. (2012). Doing Away with Harm. *Philosophy and Phenomenological Research*, *85*(2), 390–412.

Carlson, E., Johansson, J., & Risberg, O. (2022). Causal Accounts of Harming. *Pacific Philosophical Quarterly*, *103*(2), 420–445.

Carlson, E., Johansson, J., & Risberg, O. (2023). Plural Harm: Plural Problems. *Philosophical Studies*, *180*(2), 553–565.

Feinberg, J. (1984). *The Moral Limits of the Criminal Law. Vol. 1: Harm to others*. Oxford University Press.

Feit, N. (2015). Plural Harm. *Philosophy and Phenomenological Research*, *90*(2), 361–388.

Feit, N. (2016). Comparative Harm, Creation and Death. *Utilitas*, *28*(2), 136–163.

Feit, N. (2019). Harming by Failing to Benefit. *Ethical Theory and Moral Practice*, *22*(4), 809–823.

Feit, N. (2023). *Bad Things: The Nature and Normative Role of Harm* (1st ed.). Oxford University Press.

Finkelstein, C. (2003). Is Risk a Harm? *University of Pennsylvania Law Review*, *151*(3), 963.

Foddy, B. (2014). In Defense of a Temporal Account of Harm and Benefit. *American Philosophical Quarterly*, *51*(2), 155–165.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, *5*, 5–15.

Gardner, M. (2015). A Harm-Based Solution to the Non-Identity Problem. *Ergo, an Open Access Journal of Philosophy*, *2*, 427–444.

Gardner, M. (2017). On the Strength of the Reason Against Harming. *Journal of Moral Philosophy*, *14*(1), 73–87.

Gardner, M. (2019). When Good Things Happen to Harmed People. *Ethical Theory and Moral Practice*, *22*(4), 893–908.

Gardner, M. (2021). What Is Harming? In J. McMahan, T. Campbell, J. Goodrich, & K. Ramakrishnan (Eds.), *Principles and Persons*. Oxford University Press.

Gunnemyr, M. (2023). Harming Others. In A. G. Garcia, M. Gunnemyr, & J. Werkmäster (Eds.), *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen*. Department of Philosophy, Lund University.

Hanna, N. (2016). Harm: Omission, Preemption, Freedom. *Philosophy and Phenomenological Research*, *93*(2), 251–273.

Harman, E. (2004). Can We Harm and Benefit in Creating? *Philosophical Perspectives*, *18*(1), 89–113.

Harman, E. (2009). Harming as Causing Harm. In M. A. Roberts & D. T. Wasserman (Eds.), *Harming Future Persons* (pp. 137–154). Springer.

Holtug, N. (2002). The Harm Principle. *Ethical Theory and Moral Practice*, *5*(4), 357–389.

Immerman, D. (2022). The Worse than Nothing Account of Harm and the Preemption Problem. *Journal of Moral Philosophy*, *19*(1), 25–48.

Immerman, D. (forthcoming). Harm, Baselines, and the Worse than Nothing Account. *The Philosophical Quarterly*.

Johansson, J., & Risberg, O. (2022). Against the Worse Than Nothing Account of Harm: A Reply to Immerman. *Journal of Moral Philosophy*, *20*(3–4), 233–242.

Johansson, J., & Risberg, O. (2023). A Simple Analysis of Harm. *Ergo, an Open Access Journal of Philosophy*, *9*(19), 509–536.

Johansson, J., & Risberg, O. (forthcoming). The Worse than Nothing Account of Harm: A Fallen Hero. *Utilitas*.

Klocksiem, J. (2012). A Defense of the Counterfactual Comparative Account of Harm. *American Philosophical Quarterly*, *49*(4), 285–300.

Klocksiem, J. (2022). Harm, Failing to Benefit, and the Counterfactual Comparative Account. *Utilitas*, *34*(4), 428–444.

McMahan, J. (2013). Causing People to Exist and Saving People's Lives. *The Journal of Ethics*, *17*(1–2), 5–35.

McMahan, J. (2022). Creating People and Saving People. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (p. 0). Oxford University Press.

Mill, John Stuart. (1977). *On Liberty, in Collected Works of John Stuart Mill, vol. XVIII*, ed. by John M. Robson (Toronto: University of Toronto Press), pp. 213–310. (Originally published in 1859)

Norcross, A. (2005). Harming In Context. *Philosophical Studies*, *123*(1–2), 149–173.

Northcott, R. (2015). Harm and Causation. *Utilitas*, *27*(2), 147–164.

Oberdiek, J. (2017). *Imposing Risk: A Normative Framework*. Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Clarendon Press.

Perry, S. R. (1997). Risk, Harm, and Responsibility. In D. G. Owen (Ed.), *The Philosophical Foundations of Tort Law*. Oxford University Press.

Perry, S. R. (2003). Harm, History, and Counterfactuals. *San Diego Law Review*, *40*, 1283–1314.

Petersson, B. (2018). Over-Determined Harms and Harmless Pluralities. *Ethical Theory and Moral Practice*, *21*(4), 841–850.

Purves, D. (2016). Accounting for the Harm of Death. *Pacific Philosophical Quarterly*, *97*(1), 89–112.

Purves, D. (2019). Harming as Making Worse Off. *Philosophical Studies*, *176*, 2629–2656.

Quinn, W. S. (1989). Actions, Intentions, and Consequences: The Doctrine of Double Effect. *Philosophy & Public Affairs*, *18*(4), 334–351.

Rabenberg, M. (2015). Harm. *Journal of Ethics and Social Philosophy*, *8*(3), 1–32.
Raz, J. (1988). *The Morality of Freedom*. Oxford University Press.
Roberts, M. A. (2007). The Non-identity Fallacy: Harm, Probability and Another Look at Parfit's Depletion Example. *Utilitas*, *19*(3), 267–311.
Roberts, M. A. (2009). The Nonidentity Problem and the Two Envelope Problem: When is One Act Better for a Person than Another? In M. A. Roberts & D. T. Wasserman (Eds.), *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem* (pp. 201–228). Springer Netherlands.
Ross, W. D. (2002). *The Right and the Good* (D. Ross & P. Stratton-Lake, Eds.). Oxford University Press. (Original work published 1933)
Rowe, T. (2021). Can a Risk of Harm Itself be a Harm? *Analysis*, *81*(4), 694–701.
Shiffrin, S. V. (1999). Wrongful Life, Procreative Responsibility, and the Significance of Harm. *Legal Theory*, *5*(2), 117–148.
Simester, A. P. (with Von Hirsch, A.). (2011). *Crimes, harms, and wrongs: On the principles of criminalisation* (1st ed.). Hart Publishing.
Smuts, A. (2012). Less Good but Not Bad: In Defense of Epicureanism About Death. *Pacific Philosophical Quarterly*, *93*(2), 197–227.
Stefánsson, H. O. (2024). How a Pure Risk of Harm Can Itself be a Harm: A Reply to Rowe. *Analysis*, *84*(1), 112–116.
Tadros, V. (2014). What Might Have Been. In Oberdiek, John (Ed.), *Philosophical Foundations of the Law of Torts* (pp. 171–192). Oxford University Press.
The United Nations. (n.d.) *Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power.* UN Human Rights Office. https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-basic-principles-justice-victims-crime-and-abuse
UNESCO. (n.d.). *Ethics of Artificial Intelligence*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
Unruh, C. (2019). Can We Benefit in Non-Identity Cases? *Intergenerational Justice Review*, *5*(2).
Unruh, C. F. (2023). A Hybrid Account of Harm. *Australasian Journal of Philosophy*, *101*(4), 890–903.
Unruh, C. (2024). More on the Hybrid Account of Harm. *Journal of Ethics and Social Philosophy*, *28*(2), 291–298.
Villiers-Botha, T. de. (2018). Harm: The counterfactual comparative account, the omission and pre-emption problems, and well-being. *South African Journal of Philosophy*, *37*(1), 1–17.
Woollard, F. (2012). Have We Solved the Non-Identity Problem? *Ethical Theory and Moral Practice*, *15*(5), 677–690.
Woollard, F. (2015). *Doing and Allowing Harm*. Oxford University Press.