**Title**

A Unified Framework for Moral Appraisal

**Permalink**

https://escholarship.org/uc/item/5tk7d9dr

**Author**

Conklin, Sherri Lynn

**Publication Date**

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

A Unified Framework for Moral Appraisal

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Philosophy

by

Sherri Lynn Conklin

Committee in charge:

Professor Matthew Hanser, Chair

Professor Aaron Zimmerman

Professor E. Sonny Elizondo

December 2020

The dissertation of Sherri Lynn Conklin is approved.


_____

Aaron Zimmerman



_____

E. Sonny Elizondo



_____

Matthew Hanser, Committee Chair



September 2020

A Unified Framework for Moral Appraisal

Copyright © 2020

by

Sherri Lynn Conklin

iii

ACKNOWLEDGEMENTS

First and foremost, I begin by acknowledging the importance of this enormous achievement for me. Surrounded by academics who have earned or who will soon earn PhDs, it becomes easy to forget the significance of the undertaking. The PhD is afterall just the minimum requirement for obtaining full time academic employment. At some point, it feels like everybody has a doctoral degree and that there is nothing all that special about mine or the work I put in to earn it. And then I remember that I am the first person in my direct family to graduate from high school, to go to college and earn a degree, to spend over 25 cumulative years in the pursuit of knowledge. I have never experienced such a deep well of pride in my accomplishments as I do in these final hours of preparing this manuscript for submission.

With me on this journey went my life-partner, Michael, who taught me more about what it means to be a good researcher and a good person than any other of my mentors. I hope I get to spend the rest of my life learning with him. I cannot express my joy at having someone in my life who is so completely dedicated to ensuring my success, while simultaneously racking up his own set of admirable achievements, including beating me to the finish line on earning the PhD. I'm especially thankful for the last several years of "coffee fellowships". Can you believe that Michael wakes up before me every morning, clears my workspace of kipple, situates my computer, and then wakes me up with a fresh cup of hand ground coffee? He did this and so much more. Without him I would not have had the psychological fortitude necessary for meeting each day. He is my morning light.

# ACKNOWLEDGEMENTS

This research would have been impossible without my many mentors. Laurie Clarcq and Damian Grzeskowiak helped me reclaim my humanity in some of the darkest moments of my life. Fierce defenders of knowledge in a wilderness of ignorance, they triaged the worst of my wounds and push me forward before I had a chance to look back. Teresa Celada and Rolf Nelson, my undergraduate advisors, were the gardeners who cultivated the first crop of ideas blossoming at my intellectual well spring. They taught me the fundamentals of academic research and guided me in the decision to pursue a PhD. Elinor Mason, Nicole Hassoun, and Kelly Anne Brown, former and ongoing research advisors, were there for me when I was floundering in the early days of graduate school. As mentors and colleagues, these women helped me take the initial steps in crafting my professional identity as an academic and philosopher. Aaron Zimmerman and E. Sonny Elizondo, my dissertation committee members, asked many hard questions. I'm still not sure I have any of the answers.

The last but not the least. Matthew Hanser, my dissertation advisor, provided many years of reserved-yet-optimistic support while I embarked on this journey. He encouraged me to swing for the fences, and he never made me feel embarrassed when I was off the mark. He trusted me to make things work, and I trusted him to tell me when they didn't. This research was only possibly because I had the freedom to explore my outlandish ideas, which nonetheless eventually developed into something plausible and cogent. It wasn't always clear that I was making progress, but, when I was, he let me know. He was the compass that guided me on my course but never told me which way to go. Thanks Matt.

# CURRICULUM VITAE

## SHERRI LYNN CONKLIN

September 2020

## EDUCATION

PhD           Philosophy, University of California, Santa Barbara, 2020

MA           Philosophy, University of California Santa Barbara, 2016

MSc           by Research, Philosophy, University of Edinburgh, Scotland, 2011

BA           Philosophy [Hons] & Psychology, Wheaton College, *cum laude*, 2010

**Specialization**    Moral Theory (Metaethics, Moral Psychology, Normative Ethics)

**Competencies**    Applied Ethics (esp. Computers and Technology), Action Theory, Epistemology, Metaphilosophy, Writing Studies

## PEER REVIEWED PUBLICATIONS

2019           **Conklin, S.**, Artamonova, I., & Hassoun, N. "The State of the Discipline: New Data on Women Faculty in Philosophy." *Ergo* (Forthcoming).

2018           Wilhelm, I., **Conklin, S.**, & Hassoun, N. "New Data on the Representation of Women in Philosophy Journals. *Philosophical Studies*, 175(6): 1441–1464.

2014           Nelson, R., Reiss J. E., Gong, X., **Conklin, S.**, & Parker, L., & Palmer, S. E. "The Shape of a Hole is Perceived as the Shape of its Interior." *Perception*, 43: 1033–1048.

## HONORS & AWARDS

2020           Residential Research Group Fellowship, UC Humanities Research Institute

2020, 2019    NHSEB|MAP Scholarship Program, Marc Sanders Foundation

2019           Charlotte Stough Memorial Prize in Philosophy, UCSB

2019           Humanities and Social Sciences Research Grant, UCSB

2019           Engaging Humanities Grant, UC Humanities Research Institute

2019, 2016    Graduate Research Fellowship, UCSB

2018           Weinpahl Award for Teaching Excellence in Philosophy, UCSB

2017 – 2015    Multi-Campus Graduate Research Working Group Award, UC Humanities Research Institute (x2)

2016           Humanists@Work Graduate Advisory Committee Fellowship, UC Humanities Research Institute

2015           Dixon-Levy GSA Service Award, UCSB

ABSTRACT

A Unified Framework for Moral Appraisal

Sherri Lynn Conklin

Most moral theorists subscribe to a vindicatory account of moral justification, according to which an action is morally justified if and only if the action is not morally wrong. Vindicatory accounts of moral justification are incompatible with genuine moral dilemmas, since any action the agent performs in a dilemma will fail to conform with at least one operative duty and therefore render morally wrong. Yet, in moral dilemmas, it typically seems as though there is a morally preferred solution – an action the agent is justified in performing. So, if agents are sometimes morally justified in performing the preferred act in dilemma cases, then the only available account of justification is a non-vindicatory account, according to which agents are sometimes justified in performing morally wrong actions. While such views receive little attention in the existing philosophical literature, I argue that rejecting vindicatory accounts of moral justification positions me to unify a range of apparently contradictory patterns of moral assessment in a single moral framework, including in moral dilemmas.

Because moral justification is closely linked to moral permissibility, by conceding the possibility of justified wrongdoing, I concede the possibility of a *prima facie* incoherent act-assessment: Permissible-Wrong Action [PWA] – morally permissible actions that are morally wrong. To make sense of PWA, I distinguish between wrongness (or rightness) and moral permissibility. Wrongness is a first-order act-assessment, an evaluation of the morally relevant features of the act alone. Moral permissibility is a second-order (hybrid) act-assessment, an evaluation of the relationship between the action and the motivations of the agent performing it. This distinction sheds light on the Doctrine of Double Effect [DDE] debate. The central

question for the DDE is whether agent motivations inform assessments of moral permissibility. I argue that we can offer a principled explanation for our patterns of judgment in DDE cases by reframing the debate as disagreement about whether permissibility is a first- or second-order act-assessment. PWA is possible considering the distinction between first- and second-order act-assessments and my account of the relationship between moral justification and permissibility.

Specifically, I argue that genuine moral dilemmas and PWA are compatible with an adequacy account of moral justification, according to which an action sometimes overcomes concerns about its wrongness when it conforms with other important moral demands. PWA helps to explain DDE cases, where two acts comprising all of the same morally significant agent-independent features receive different permissibility assessments. While both actions are morally wrong, only one is an instance of PWA. Because the only difference in pairs of DDE cases involve differences in the agents' motivations, I follow DDE proponents in suggesting that one act is rendered impermissible by the offending agent's criticizeable motivations in acting. I propose that the agent's motivations pick out the features of the action that get to do the justifying or unjustifying. If the agent is motivated by the action's wrongness, then the action's wrongness unjustifies the action, rendering it impermissible. If not, then it could turn out to be a case of PWA. I argue that the proper understanding of the conceptual relationships between these and other puzzling moral phenomena (including moral residue, suberogation, and praise-blame asymmetry in Side-Effect Effect cases) and their corresponding moral assessments reveals an underlying unity in our patterns of assessment, providing evidence for a new, unifying moral framework, the backbone of which I develop throughout my dissertation.

# TABLE OF CONTENTS

# Chapter 1: The Complexity of Moral Assessments

These wonderful narrations inspired me with strange feelings. Was man, indeed, at once so powerful, so virtuous and magnificent, yet so vicious and base? He appeared at one time a mere scion of the evil principle, and at another as all that can be conceived of noble and godlike.

-- Mary Shelley, *Frankenstein*.

**Introduction**

As moral judges, we often find it difficult to accurately morally evaluate other agents and their actions.[1] We might think of the very same agent that she is simultaneously vicious yet praiseworthy or think of the very same action that it is wrong yet permissible. On the face of it, such judgments are contradictory, but we cannot help but feel the pull of these incongruent evaluations when faced with the complex moral dimensions of human lives. Consider how one might regard a person like Mary Shelley's Dr. Frankenstein.[2]

> I created a rational creature, and was bound towards him, to assure, as far as was in my power, his happiness and well-being. This was my duty; but there was another still paramount to that. My duties towards the beings of my own species had greater claims to my attention, because they included a greater proportion of happiness or misery. Urged by this view, I refused, and I did right in refusing, to create a companion for the first creature. He showed unparalleled malignity and selfishness, in evil: he destroyed my friends; he devoted to destruction beings who possessed exquisite sensations, happiness, and wisdom; nor do I know where this thirst for vengeance may end. Miserable himself, that he may render no other wretched he ought to die. The task of his destruction was mine, but I have failed.

---

[1] Throughout this dissertation, I use "moral evaluation," "moral assessment," "moral judgment," and "moral appraisal" interchangeably to denote a moral judge's reaction to an agent.

[2] Shelley, M. (1994). *Frankenstein*. Eds. S. Applebaum, C. Ward. (Dover thrift Editions, first published 1818): 161-2.

As academics, we might find it easy to admire a person who doggedly pursues profound intellectual discoveries. Nonetheless, we forbid the creation of monstrosities, such as the daemon. Frankenstein acts contrary to duty, and he is not due any sympathy for his self-inflicted woe – after creating, spurning, and sabotaging a new life. His failure to cultivate the daemon's moral sensibilities is his own moral failure, and the daemon's lethal comportment is, to some extent, his responsibility. Even so, we pity Frankenstein for his predicament, we are sensitive to the tragedy of his lost loved ones, and we believe he is ultimately justified for failing to keep his promise to the daemon and for his ardent pursuit of the creature's demise.

Frankenstein's creation is similarly puzzling for the moral judge. At first, he is unequipped for the cruelty of the world he is thrust into, and he is therefore not responsible for the bad environment that served as his model for humanity. As a result, we might think his maliciousness is partially excused. Even so, the daemon's monstrous appearance indicates a seed of evil within, and we are unsurprised when he embraces it. He is a vicious murderer, yet we are more sympathetic to the daemon's suffering than we are to his creator's. We admire the daemon's attempts to reason with his foe, and we sympathize with his contempt for Frankenstein. Finally, while we do not believe his lust for revenge against Frankenstein is justified, we cannot help but think Frankenstein has earned some portion of it.

While our assessments of these characters, which compound and intensify as events unfold, are not mysterious to us, we should be struck by their incongruities. We justly sympathize with a murderer and scorn his victim. Supposing the lens through which I assess Frankenstein and his daemon is the correct one, we get the troubling result that not all moral judgments naturally align along the lines of good and evil. Instead, we assess agents and their

actions using a variety of different criteria, and we sometimes arrive at surprising and complicated moral evaluations.

More often than not, our moral evaluations feel torn in two directions rather than unified. As, more often than not, we find ourselves evaluating people like Frankenstein than people like Shakespeare's Aaron.[3]

> If there be devils, I would that I were a devil,
>
> To live and burn in everlasting fire,
>
> So, I might have your company in hell,
>
> But to torment you with my bitter tongue!

We can easily assess the evilness of an evil man. In Aaron's case, he fantasizes about becoming a devil, so that, even in death, he can continue to do evil. However, people are more complicated, and philosophers inevitably confront difficulty accounting for our puzzling moral judgments.

A good theory of moral judgments will be able to systematize such *prima facie* incongruent assessments into a unified framework if doing so is at all possible. This motivates the task set out for me in this dissertation. I am writing in order to provide the building blocks for a maximally inclusive unified account of our moral assessments. The framework should be maximally inclusive in order to accommodate a rich assortment of commonly articulated moral evaluations, as well as to create conceptual space for some controversial or entirely new

---

[3] Shakespeare, W. (2005). *Titus Andronicus*. Eds. B. Mowat, P. Werstine (Eds.). (The Folger Shakespeare Library, first published 1594): 5.1.2282-85.

4

ones. The framework should be unifying in that it can help to reveal the coherence of well-known asymmetries and inconsistencies, which I will discuss throughout the remainder of this chapter.

Before proceeding, however, I will clarify my task. The philosophical literature is riddled with puzzling moral assessments. In this dissertation, I identify five well-known but relatively controversial moral phenomena for which any good theory of moral judgments should provide an account. Either that theory conforms with common judgments about these moral phenomena or provides an error theory to explain how judgments about these moral phenomena are misguided. The moral phenomena under consideration in this dissertation are: (1) genuine moral dilemmas; (2) asymmetrical assessments of permissibility in Doctrine of Double Effect cases; (3) asymmetrical assessments of praiseworthiness and blameworthiness in Side-Effect Effect cases; (4) moral residue; and (5) the suberogatory.

Each of these phenomena stem from or give rise to intuitively plausible yet *prima facie* inconsistent moral assessments – as I will soon show. Generally, theories of moral judgments take on some of these phenomena but deny others. However, I believe that the apparent inconsistencies in our patterns of moral assessment reveal an underlying unity to these phenomena, which provide evidence for a new moral framework. My job is to show just how this might be. Throughout this dissertation, I will assume that, for the most part, we are correctly making assessments about genuine moral phenomena, even when those assessments stem from or give rise to apparent inconsistencies. I therefore reject any view according to which any of the moral judgments under consideration turn out to be false, since the correctness of these moral judgments is a constraint on the sort of views I can adopt. As it turns out, I identify several previously unconsidered solutions to well-known philosophical

puzzles, which reveal a unifying structure to our moral judgments. What follows is an overview of the moral phenomena I explain in the course of this dissertation.

## 2. The Dilemma Trilemma

The first such puzzling moral phenomena are moral dilemmas, which have plagued philosophers for as long we have recorded our moral lives. Consider a case, by Aeschylus.

> **ORESTES:** Clytemnestra, Queen of Argos, killed her husband King Agamemnon in cold blood. According to the Greek gods, Orestes, their son, faced conflicting duties in the aftermath. One duty requires that Orestes avenge his father's murder. Another duty requires that he not commit matricide.

In this scenario, suppose duties are dictated to humans by the gods, rendering each of our moral duties absolute or exceptionless, such that failing to conform with a duty is always morally wrong.[4] If Orestes conforms with the duty prohibiting matricide, he will fail to avenge his father because his mother was the murderer, and if he avenges his father, he will kill his mother and thereby violate the duty prohibiting matricide.[5] If Orestes fails to avenge his father, his father's Erinyes, deities of vengeance, will torment him, and if he commits matricide, then his mother's Erinyes will torment him. Orestes will be punished for violating one of his conflicting duties whatever action he chooses. Sure enough, when Orestes avenges his father

---

[4] Of course, we might also think that duties issued to Orestes' by the gods were not absolute or exceptionless and that the commands of one god could countermand the commands of another. However, given Orestes tragic circumstance, my reading of the scenario seems plausible enough.

[5] Presently, I use duty generically to refer to that which we are ordinarily morally required to do. Absolute duties are duties that cannot be overridden or outweighed by other duties or other morally relevant considerations. Such duties are therefore exceptionless in a way that makes failing to conform with them always morally wrong. Most philosophers suppose further that this makes the action therefore morally impermissible. Whether such duties exist is not my concern in this paper. For discussions of moral absolutism, see J. Rachel's (1970) "On Moral Absolutism" (*Australasian Journal of Philosophy*, 48.3); F. Jackson and M. Smith's (2006) "Absolutist Moral Theories and Uncertainty" (*Journal of Philosophy*, 103.6); P. Hawley's (2008) "Moral Absolutism Defended" (*Journal of Philosophy*, 105.5).

by committing matricide, his mother's Erinyes torment him. Because his mother killed his father, conforming with either of his duties would result in nonconformance with the other and nonconformance with either absolute duty is morally wrong.

Orestes faces a classically depicted moral dilemma. Moral dilemmas have two central characteristics. First, the conflicting duties are such that conforming with either duty would result in nonconformance with the other, thereby rendering any action the agent performs morally wrong.[6] Second, dilemmas nonetheless typically have a morally preferred solution – dilemmas are resolvable.[7] In cases where an agent faces a genuine moral dilemma, the agent can typically permissibly perform at least one of the two actions. In Orestes' case, Apollo reveals that avenging his father and killing his mother is the morally preferable solution in the eyes of the gods, even though doing so is morally wrong. This case gives rise to the dilemma trilemma, which arises from the following three independently plausible, yet *prima facie* jointly inconsistent, claims and subsequent line of argument.[8]

(1)     The two duties binding Orestes conflict

---

[6] As Hursthouse explains it, the nature of a dilemma (of any sort) is such that each alternative is equally unfavorable for the choosing agent. For moral dilemmas, the unfavorable consequences spring from conflicting duties where conforming with either duty would result in nonconformance with the other, thereby making any action the agent performs morally wrong. As I analyse moral dilemmas, I treat both actions as equally morally wrong, since neither absolute duty yields to the other. In Chapter 2, I will define the moral preferability of an action in terms of the normative moral reasons, separate from the duties, that favor one action over another. On my account, the normative reasons favoring one action over another never affect the "weight" of an action's wrongness in any way. See R. Hursthouse's (2002) *On Virtue Ethics* (Oxford University Press), esp. Chapter 2.

[7] Although philosophers, such as Hursthouse (ibid. Chapter 3), do occasionally defend the possibility of unresolvable dilemmas.

[8] For other thoughtful discussions of moral dilemmas, see B. Williams' (1965) "Ethical Consistency" (*Proceedings of the Aristotelian Society*, 39); T. McConnell's (1978) "Moral Dilemmas and Consistency in Ethics" (*Canadian Journal of Philosophy*, 8.2); R. Hursthouse's (1995) "Fallacies and Moral Dilemmas" (*Argumentation*, 9); J. De Haan's (2001) "The Definition of Moral Dilemmas: A Logical Problem" (*Ethical Theory and Moral Practice*, 4.3).

(2)     The two duties binding Orestes are absolute

(3)     Orestes can resolve his dilemma

*Suppose (1) and (2).*

If Orestes is facing a conflict between two absolute duties, then nonconformance with either conflicting duty is morally wrong. If so, then Orestes is morally prohibited from performing either of the two actions. If performing either action is morally prohibited, then we cannot say morality prefers that Orestes perform one action rather than the other. Morality requires that Orestes perform neither. Thus, (3) is false. Conflicts between absolute duties are unresolvable.

*Suppose (1) and (3).*

If morality prefers that Orestes perform one action rather than the other when two duties conflict, then Orestes is not prohibited from performing the morally preferred action. If Orestes is not prohibited from performing the morally preferred action, then performing the morally preferred action is not morally wrong. If performing the morally preferred action is not morally wrong, then the duty prohibiting Orestes from performing the morally preferred action is not absolute, and (2) is therefore false. Orestes cannot be bound by two absolute duties when (1) and (3) are true.

*Suppose (2) and (3).*

If Orestes' dilemma is resolvable, then morality prefers that Orestes perform one action rather than the other. If so, then Orestes is not morally prohibited from performing the morally preferred action. Performing the morally preferred action is, therefore, not morally wrong. But failing to conform with an absolute duty is morally wrong. In failing to perform the morally preferred action, then, Orestes does not fail to conform with an absolute duty. Yet, by

hypothesis, both of the duties binding Orestes are absolute. It follows that these duties do not genuinely conflict in Orestes' case – the conflict is only apparent and (1) is therefore false.

Because of the dilemma trilemma, we have difficulty making sense of the possibility of moral dilemmas. Inconsistency could be resolved by giving up any one of the three claims. Claim (2) is a likely candidate. Afterall, many philosophers do not think moral laws are handed to humans by the gods. If not, then perhaps moral duties are never absolute such that failing to conform with the duty is always morally wrong and there are no real moral dilemmas. Yet, some duties do seem to be unyielding in this way.

Consider yet another example of a classic moral dilemma – the Trolley Problem, which is a family of cases used to discern intuitions about moral permissibility.

> **SIDE-TRACK:** Imagine a runaway trolley traveling along a track. Ahead of the trolley, five people are tied down to the track, and they are unable to flee from the oncoming trolley. Medea is standing next to a lever that can switch the trolley to a different set of tracks. If she pulls the lever, the trolley will switch tracks and miss the five people. If Medea pulls the lever, she will save the lives of five people. Unfortunately, a person, who I will call "Jason", is on the other track. If Medea pulls the lever and switches the tracks to divert the trolley, the trolley will kill Jason.

In SIDE-TRACK, Medea has two choices. Medea can refuse to pull the lever, which will result in the deaths of the five people, or Medea can pull the lever, which will result in Jason's death. So, Medea can let five people die while preserving Jason's life, or she can kill Jason while preserving five lives. We wonder: Is pulling the lever morally permissible?

As it turns out, pretty much everyone agrees that pulling the lever is permissible in SIDE-TRACK.[9] Our judgments about the case are settled, yet we should be puzzled over them. On the one hand, we are ordinarily required to save a person who is in danger. Likewise, we are ordinarily prohibited from killing. Moreover, we are normally prohibited from killing one person to save another.[10] Even so, according to common intuitions we are nonetheless permitted to kill the one in order to save the five in SIDE-TRACK.

Consider just one way of articulating the source of the dilemma. Although I am sometimes permitted to kill in defense of myself or another person, I am justified in killing in other-defense when the person I kill will inevitably kill others and when the people I am trying to save have a right that that person not kill them. Thomson argues that the assailant forfeits her right not to be killed because unless she is killed, she will violate another person's right not to be killed. Thus, I have no duty prohibiting me from killing an assailant in what Thomson regards as justifiable cases of killing in defense of another.[11]

However, even if I am sometimes permitted to kill in other-defense, I cannot take just any action to defend others from death. In particular, I am prohibited from killing innocent bystanders.[12] I am prohibited from killing bystanders in order to save myself or others because people ordinarily have a right against me that I not kill them, and I have a corresponding duty

---

[9] Green, J. "The Cognitive Neuroscience of Moral Judgment." The Cognitive Neurosciences IV Ed. M.S. Gazzaniga (MIT Press, 2009): 13-14.

[10] One justification for the theoretical claim that we are usually prohibited from killing one person to save another is based on philosophers' intuitions that killing is sometimes worse than letting die. See, for example, W.S. Quinn's (1989) "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing" (*The Philosophical Review*, 98.3). Although some philosophers take issue with this intuition and the resulting doing/allowing distinction, see J. Bennett's (1980) "Killing and Letting Die" (The Tanner Lectures in Human Values, Delivered at Brasenose College, Oxford University May 9,16, and 23, 19).

[11] Thomson, J.J. (1991). "Self Defense." *Philosophy and Public Affairs*, 20(4).

[12] *Ibid.*, 289-292.

prohibiting me from killing them. An agent's right not to be killed is only forfeited when the agent poses a lethal threat and will imminently violate another's right.[13] Thomson claims that bystanders are individuals who are not causally involved in the current lethal threat, so the duty prohibiting me from killing a bystander is operative because her right not to be killed has not been forfeited.[14]

So, one might think, when a trolley is barreling down a track towards five immobile people and Medea is the only person around to save them, that she has a duty to save them. However, because Jason is a bystander and so has a right against Medea that she not kill him, Medea has a corresponding duty prohibiting her from killing him. Redirecting the trolley onto the side-track, thereby saving the five and killing Jason, is bystander substitution – an unacceptable form of self- or other-defense because Jason has not forfeited his right not to be killed.[15]

My interest here is not in the correct interpretation of the Trolley Problem, and I am not committed to the above interpretation of SIDE-TRACK. As it turns out, neither is Thompson. While bystander substitution is wrong in one-for-one cases, she does not think

---

[13] According to Thomson, the primary difference between attackers and bystanders is that attackers pose a lethal threat and will imminently violate a right. If we are justified in defending ourselves or others from attackers but not bystanders, then the fact that attackers pose a lethal threat and will imminently violate a right (where bystanders do not and will not) must be the reason that attackers forfeit their rights not to be killed (*Ibid.*, p300-302). Some would argue against Thomson by claiming that the attacker only forfeits her right not to be killed if she is an unjust or villainous attacker. See, for example, M. Otsuka's (1994) "Killing the Innocent in Self-Defense" (*Philosophy & Public Affairs*, 23.1).

[14] Thomson, op. cit. 298.

[15] Bystander substitution occurs when an agent prevents a death by redirecting the lethal blow to a bystander. Normally we have difficulty explaining why the agent acts impermissibly in contrast scenarios for Trolley Problem & other Doctrine of Double-Effect cases. See Section 4 below. Here, I am claiming that we have the opposite problem; we have difficulty explaining why switching the trolley onto the sidetrack in the standard Trolley scenario is permissible in the first place.

bystander substitution is wrong in one-for-five cases. Thomson holds that merely pulling the lever in standard Trolley Problem scenarios does not itself constitute a violation of any duty or an infringement on any rights.[16] Nonetheless, the interpretation is plausible – grasping something deep about the nature of the scenario. Thomson surely does not think Jason has forfeited his right not to be killed, which entails that Medea has a duty to not kill Jason. Thus, killing Jason certainly seems wrong. If so, then the judgment about the permissibility of Medea's action is subject to the dilemma trilemma. At least *prima facie*, Medea faces a resolvable conflict between two absolute duties. If this is so, we might question the mechanism by which Medea permissibly kills Jason while failing to conform with the duty to not kill him.

If the relevant duty is present and active (whatever that might mean), then the duty maintains its significance to our moral assessment of the action – specifically in that failing to conform with one's duty renders one's action wrong. We need a way of explaining the permissibility of performing ordinarily wrong actions in dilemma cases that does not undermine our assessment of the action's wrongness. Otherwise, we have difficulty making sense of our competing claims about the cases. Of course, we can always give up one of the competing claims – namely that failing to conform with one's duty, in the above dilemma cases, is wrong, but I will propose that we do not have to. And why would we want to? Certainly, one must act in dilemma cases. This is the bind all agents face. But we need not

---

[16] Afterall, the mere pulling of the lever does not hurt anyone, and we leave open the possibility that the person on the side-track can still get away. See J. J. Thomson's (1985) "The Trolley Problem" (*The Yale Law Journal*, 94.6). However, in most hypothetical cases, the death of the one is a fact of the scenario, so deflecting the harm the trolley will inevitably lead to his death. In such cases, we have no reason to believe the one has given up his right not to be killed, so we might nonetheless think the bystander at the lever yet has a concomitant duty prohibiting acting so as to kill the one, which is violated when the one dies as a result of pulling the level.

suppose moral standards are compromised by our failures as practical deliberators or by the whims of fate. Instead, we must make sense of the ways in which we sometimes simultaneously succeed and fail from the moral standpoint – allowing for permissible actions that are nonetheless deeply morally marred. Just as otherwise virtuous agents sometimes have tragic flaws, so too can actions. In later chapters, I argue that genuine moral dilemmas are possible and, I suspect, quite common.

## 3. Asymmetrical Assessments of Permissibility

Once the possibility of genuine moral dilemmas enters our moral theory, we face another puzzling moral phenomenon. When two agents perform the same action, we should – *ceterus paribus* – generate the same moral judgments about that action. Judgments about moral permissibility are judgments about actions, so it seems we should assess agents who perform the same action symmetrically with regard to the action's permissibility. Otherwise, we violate the principle of parity. Yet, we do not do this for all cases. Sometimes, we assess agents who perform the same action differently with regard the action's permissibility. We make asymmetrical assessments of permissibility [AAP].

AAP is a *prima facie* inconsistent but, as I will show, intuitively plausible pattern of assessment especially present in Doctrine of Double Effect [DDE] cases. Consider a standard DDE case. Suppose there is a war going on. Bombing a munitions factory will help end the war early. However, bombing the munitions factory will cause many civilian deaths.

> TACTICAL BOMBER: The tactical bomber believes the killing of civilians during war is morally objectionable but that the promise of ending the war early, thereby saving lives on both sides, is a good enough reason for bombing the factory. So, the tactical bomber drops the bomb with the intention of destroying the munitions factory, thereby destroying the munitions factory and ending the war early. He foresees that his action will kill civilians, but he does not intend this result.

The TACTICAL BOMBER is contrasted with the TERROR BOMBER.

**TERROR BOMBER:** The terror bomber believes the killing of civilians is important to the bombing because the death of civilians will demoralize the enemy government, thereby ending the war early and saving lives. So, the terror bomber drops the bomb with the intention of killing civilians, thereby demoralizing the enemy and ending the war early.

Most think the Terror Bomber acts impermissibly and that the Tactical Bomber acts permissibly.[17] But these moral judgments leave us with a puzzle. Both the Tactical Bomber and the Terror Bomber attempt to end the war early, and they both kill civilians by bombing the munitions factory. The actions in both cases seem to have the same morally relevant features, even as the agents' motivations differ. So how do we justify our AAP about the actions in these two cases?

Some philosophers deploy the DDE to handle the above sorts of cases. According to the DDE, we sometimes permissibly cause a foreseen harm in the course of intentionally bringing about a good when intentionally bringing about the same harm as a means of bringing about the good would be impermissible. Philosophers generally agree that a thoroughgoing account of the DDE provides a principled justification for many of our moral assessments but question whether the DDE is the right principle to use. They worry that the permissibility of a person's action sometimes depends on the intentions with which the person acts according to the DDE. In particular, one can read the DDE as suggesting that the intentions with which

---

[17] See, for example, W.S. Quinn's (1989) "Actions, Intentions, and Consequences: The Doctrine of Double Effect" (*Philosophy and Public Affairs*, 18.4); J. McMahan's (1994) "Revising the Doctrine of Double Effect" (*Journal of Applied Philosophy*, 11.2); D.K. Nelkin and S.C. Rickless' (2014) "Three Cheers for Double Effect" (*Philosophy and Phenomenological Research*, 89.1).

the person acts sometimes count as reasons against performing an action.[18] So, while applying

the DDE, one might advise the Tactical Bomber to perform an action if he does not intend the

harmful consequences and advise the Terror Bomber against performing the action if he does

intend the harmful consequences. Yet, one might think that when the target of an assessment

is an action, as it is with assessments of moral permissibility, then the considerations relevant

to the assessment should surely have to do with the action itself and not the agent, since the

agent is not the target of the assessment.

Thus, one way of framing the long-standing debate surrounding the DDE is as a

disagreement about the appropriate application of our concept 'permissible.' In general

(although not always), proponents of the DDE argue that claims about moral permissibility

sometimes appeal to agent motivations, while opponents of the DDE argue that claims about

moral permissibility instead appeal only to agent-independent features of actions.

If proponents of the DDE are correct, then they can offer a principle for supporting the

patterns of moral assessment resulting in AAP. Yet, if judgments about moral permissibility

are judgments about actions and not agents, then our asymmetrical judgments about

permissibility in DDE cases are puzzling. It seems like we should assess the agent's action

symmetrically in each case, while assessing the agent's motivations asymmetrically. But this

is not what we appear to be doing in the bomber cases.[19] When we assess that the Tactical

---

[18] See T.M. Scanlon's (2008) *Moral Dimension: Permissibility, Meaning, and Blame* (Harvard University Press), esp. Chapter 1; Thomson's "Self Defense," op. cit.

[19] Bennet argues that we can identify no morally relevant differences between the two actions. If so, then perhaps we should judge the TACTICAL BOMBER and TERROR BOMBER symmetrically. While this gets the correct result with regard to assessing the action symmetrically, it gets the wrong result with regard to assessing the action asymmetrically on the basis of the agent's motivations. See J. Bennett's (1995) *The Act Itself* (Oxford University Press), esp. Chapter 11.

Bomber acts permissibly while the Terror Bomber acts impermissibly, we seem to judge the action asymmetrically.

To resolve the puzzle surrounding AAP, we need a framework that can simultaneously help us assess the actions in pairs of DDE cases asymmetrically and symmetrically. This would allow us to preserve the patterns of assessment indicated by AAP while conforming to the principle of parity. Since we cannot render the very same assessment of an action's permissibility both asymmetrically and symmetrically, it seems that we need a framework generating at least two distinct categories of action assessment – only one of which is an assessment of permissibility. If one way of framing the long-standing debate surrounding the DDE is as a disagreement about the appropriate applications of our concept 'permissible,' then such a framework might also help us understand the DDE and the role of agent motivations in assessments of permissibility, which is often cited to justify AAP.

In the end, I hold that some version of the DDE accurately captures a description of the moral considerations relevant to assessments of permissibility in pairs of cases where AAP are present. However, the DDE, like AAP, only makes sense in the context of the framework I develop in Chapter 2. Because I am developing a framework to justify AAP only, my approach is conciliatory between proponents and opponents of the DDE, including opponents who deny AAP because the DDE does not suffice to justify it. So, while I ultimately argue that some version of the DDE is probably correct, I set out to vindicate both the proponents and opponents of the DDE. As a result, neither is likely to be satisfied with my conclusion even as I make concessions to each.

## 4. Praise-Blame Asymmetry

In AAP cases, two agents perform the same action but act on differently valanced motivations. Although permissibility is characterized as an act-assessment, and the actions are the same, permissibility is assigned asymmetrically in AAP cases. Presumably, the justification for AAP involves the two agents' differing motivations, since appeal to their qualitatively identical actions alone should yield symmetrical assessments. Praise-blame asymmetry [PBA] is a complementary phenomenon where two agents perform differently valanced actions but act on the same motivations. Although praise and blame are characterized as agent-assessments, and the agent's motivations are the same, praise and blame are assigned asymmetrically in PBA cases. Presumably, the justification for PBA involves the qualitative differences between the two agents' actions, since appeal to the agents' motivations alone should yield symmetrical assessments.

The complementary nature of AAP and PBA suggests to me that they are closely related moral phenomena, and I think understanding AAP requires fully understanding PBA. As illustrated below, PBA, like AAP, is an intuitively plausible pattern of assessment. Unlike AAP, PBA does not generate the same level of controversy. While many philosophers will happily allow that assessments of praise and blame appeal to actions, the view that permissibility appeals to agent motivations is more immediately puzzling. However, we might wonder why that is, since the view that praise and blame are solely determined by agent motivations is a live option. For example, we could reasonably (and correctly) think the TERROR BOMBER is blameworthy while the TACTICAL BOMBER is not. Since their actions are the same, we might generate these assessments by sole appeal to their differently valanced

motivations. If so, PBA is a clear violation of the principle of parity, and the problem is just as deep as the problem surrounding AAP.[20]

The issue is exacerbated by the literature surrounding PBA, which suggests that PBA is a pattern of assessment ultimately stemming from *de facto* inconsistent assessments about agent intentions in Side-Effect Effect cases. Because the patterns of assessment justifying PBA are *de facto* inconsistent, PBA is *prima facie* problematic without further justification. Consider a standard pair of Side-Effect Effect cases, which originate with a series of experiments conducted by Joshua Knobe.[21] Some participants read the following vignette.

> **HARM**: The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.' The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed.[22]

---

[20] Further, Kantian philosophers might reasonably deny that PBA is ever appropriate when the agent's motivations are the same because the quality of the agent's will and not the outcome of the action dictates our assessments of moral worth. Afterall, what actually happens is often a matter of luck and does not fully attribute to the agent, while the motivations in acting do attribute fully to the agent. Kant famously stated that morality is immune from luck, and anyone with similar anti-luck intuitions might similarly find PBA cases puzzling. See I. Kant's (2002) *Groundwork for the Metaphysics of Morals* (Ed. Allen Wood, Yale University Press, first published 1785), as well as B. Williams and T. Nagel's (1976) "Moral Luck" in the *Proceedings of the Aristotelian Society (Supplementary Volumes*, 50: 115-135; 137-151).
[21] Knobe, J. (2003). "Intentional Action and Side Effects in Ordinary Language," *Analysis*, 64: 81 – 87.
[22] Ibid.

After reading HARM, participants rated the chairman on how much blame he deserved for his actions, and they were asked to say whether they believed that the chairman intentionally harmed the environment. Meanwhile, other participants read a different vignette.

> HELP: The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.' The chairman of the board answered, 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped.[23]

After reading HELP, participants rated the chairman on how much praise he deserved for his actions, and they were asked to say whether they believed that the chairman intentionally helped the environment.

The two different conditions generated two different responses from the participants in the experiment. Participants evaluating Chairman HARM judged that he harmed the environment intentionally, while participants evaluating Chairman HELP judged that he did not help the environment intentionally. These asymmetrical judgments about intentions appear to yield asymmetrical judgments about an agent's praiseworthiness and blameworthiness. Participants evaluating Chairman HARM judged he was blameworthy for harming the environment intentionally, while participants evaluating Chairman HELP judged

---

[23] Ibid.

he was not praiseworthy for helping the environment.[24] But these patterns of judgment leave us with a puzzle.

These results should be surprising, since Knobe stipulates that both Chairman HELP and Chairman HARM are motivated by the same thing – they both implement the program in order to earn profit. Their intentions seem to have the same morally relevant features, while the outcomes of the agents' actions differ. Yet, assessments of the agents' intentions and corresponding assessments of praiseworthiness and blameworthiness were assigned differently in light of the outcome. The asymmetry in the assessments of intentional action in the two cases is what is known as the Side-Effect Effect. Asymmetrical judgments about praiseworthiness and blameworthiness in HARM and HELP are PBA.[25]

Knobe argued that laypeople are more willing to say that harmful side-effects are brought about intentionally and that this attribution will be made regardless of whether or not the side-effect was actually brought about intentionally. Given lay intuitions about the chairmen's intentions in HELP and HARM, PBA seems to follow.

While intuitively plausible, PBA justified by lay judgments about intentional action is *prima facie* problematic. Laypeople appear to apply inconsistent criteria for assessing whether an agent acts intentionally, thereby violating the principle of parity. They judge that Chairman

---

[24] These are not fine-grained cases, so one might dispute whether it is appropriate to describe the chairmen as performing right or wrong actions when the chairmen only agreed to implement some program. Though, this is somewhat due to Knobe's original description of the experiment. He asked participants how much praise or blame the chairman deserved for what he did. In the scenario, what the chairman *did* was 1) consent to implementing the program and 2) help or harm the environment as a side-effect of consenting to the program. My characterization is based on how participants responded to the cases. However, I would expect to see similar patterns of assessment when examining cases where the action is more clearly defined.

[25] Hindriks, F. (2008). "Intentional Action and Praise-Blame Asymmetry," *The Philosophical Quarterly*, 58.233: 630 – 641.

HARM acted intentionally because he performed an action with harmful consequences. Although an agent may very well intend to perform a bad action for its harmful consequences, the mere presence of such consequences need not entail that the agent intended them. Meanwhile, this is precisely the sort of reasoning used to exclude Chairman HELP from assessments of acting intentionally while helping the environment. Although an agent may very well intend to perform a good action for its beneficial consequences, the mere presence of beneficial consequences need not entail that Chairman HELP intended them. PBA is at least *prima facie* problematic even if only partially justified by the *de facto* inconsistent assessment criteria employed by laypeople used to evaluate agent intentions.

Of course, philosophers have access to more theoretical tools than laypeople assessing these scenarios, and plausible theoretical justifications for PBA are on offer. I briefly address these in Chapter 5. However, existing solutions are inadequate, I argue, because they fail to eliminate or vindicate the underlying inconsistencies undermining the justification for PBA.

More importantly, for my purposes, no other solutions connect PBA to AAP in their attempt to explain the phenomenon. We need a framework that accounts for both while conforming to the principle of parity. I propose that contextualizing PBA as the complement to AAP, using my framework, provides a new theoretical route through the problem. Specifically, I argue that AAP and PBA, taken together, reveal a simple, symmetrical, and unified organization to landscape of second-order moral assessments, which comprises judgments about moral worth (praiseworthiness and blameworthiness) and moral permissibility.

## 5. Moral Residue

> In our nature, however, there is a provision, alike marvelous and merciful, that the sufferer should never know the intensity of what he endures by its present torture, but chiefly by the pang that rankles after it.
>
> -- Nathaniel Hawthorne, *The Scarlet Letter*.

Sometimes I permissibly fail to conform with a duty because the duty appears to be outweighed by other morally relevant considerations, yet it seems as though I continue to be bound by the defeated duty in some way. Gardner discusses *Vincent v. Lake Erie Transportation Co.* – a famous case in tort law, which can be used to illustrate this point.[26]

> VINCENT: A ship captain ties his ship to a dock without the owner's permission in order to prevent a violent storm from damaging the ship. The storm caused the ship to collide with the dock while the ship is tied to it. The collision damaged the dock, and the dock's owner sued the ship captain for damages.[27]

Everyone apparently agrees that the ship captain in VINCENT was justified in tying his ship to the dock amidst the storm. Otherwise, the storm might have damaged the ship and injured the crew. However, the courts ruled that the ship captain owed reparations to the dock's owner. This case is puzzling because if the ship captain was justified in tying his ship to the dock, then he did nothing wrong. In which case, we might find it difficult to explain why he would

---

[26] Gardner, J. (2005). "Wrongs and Faults." *The Review of Metaphysics*, 59.1: 95-132.
[27] Vincent v. Lake Erie Transp. Co., 109 Minn. 456, 124 N.W. 221, 1910 Minn. LEXIS 588 (Minn. 1910)

owe anything to the dock owner. If he owes something to the dock owner, then his action was not justified.

When we are bound by an apparently defeated duty, we experience moral residue.[28] When we owe reparations for failing to conform with a defeated duty, we are subject to residual compensatory obligations. The ship captain has a compensatory obligation to reimburse the dock owner for repairs. In addition to compensatory obligations, we sometimes are subject to residual compunctory obligations. Compunctory obligations are residual obligations to apologize, explain oneself, or possibly to feel guilt and remorse about what one has done. The latter might count as an obligation insofar as we appropriately express remorse when failing to conform with a defeated duty (and inappropriately fail to express remorse when failing to conform with a defeated duty).

The experience of guilt and remorse, upon realizing that one has failed to conform with a duty, is one of the oldest themes in literature. Consider the case of OEDIPUS.

> **OEDIPUS:** Oedipus was adopted. Upon traveling to Thebes, he kills his father Laius, King of Thebes, by mistake. Later, also by mistake, he marries his mother Jocasta. He knew not that either was his parent. Upon discovering that he had killed his father and married his mother, Oedipus' guilt led him to blind himself.

---

[28] See P.S. Greenspan's (1983) "Moral Dilemmas and Guilt" (*Philosophical Studies*, 43: 117-125), as well as discussions in P.S. Greenspan's (1995) *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*, (Oxford University Press), H.E. Mason's (Ed) (1996) *Moral Dilemmas and Moral Theory* (Oxford University Press), and C.W. Gowans' (Ed) (1987) *Moral Dilemmas* (Oxford University Press).

Recall that the Greek gods generally prohibit incest and patricide. So, when Oedipus realizes that he has failed on both counts, he is literally blinded by his guilt because he has no other recourse for making reparations.

If moral residue helps to explain why Oedipus blinded himself, then it has a more general flavor than we usually taste in the philosophical literature. OEDIPUS is unlike VINCENT in that Oedipus is not justified in killing his father and marrying his mother, while the captain in VINCENT is justified in mooring his boat to the dock.[29] Nonetheless, we might think residual compensatory or compunctory obligations derive from both. I posit moral residue as a general phenomenon typically present whenever an agent fails to conform with a duty, regardless of whether the agent was justified in doing so. Agent regret and dirty hands, discussed in Chapter 5, are other instances of this general phenomenon, which is only especially puzzling in cases like VINCENT. The existence of moral residue, upon failing to conform with a defeated duty, is in tension with any claims that the agent was justified in so acting – generating yet another *prima facie* inconsistency.

Some philosophers posit the existence of genuine moral dilemmas in light of agents who are subject to residual obligations following apparent conflicts between duties. Afterall, we find it difficult to explain residual obligations derived from apparently defeated duties without positing the existence of that duty. I do not use the existence of moral residue to argue

---

[29] As a clarificatory point, I am claiming that Oedipus was not justified in failing to conform with the duties prohibiting patricide and incest, which is a separate question from whether he was otherwise justified in killing his father in self-defense or marrying Jocasta and taking the throne of Thebes after defeating the Sphinx. Ordinarily, both are justified, yet both are tainted with miasma in Oedipus' case. Despite all the good he did and all the good that befell him, his successes were undercut by moral transgressions so deep that his children were forced to pay for Oedipus' sins, with his sons killing each other over the throne and his daughter, Antigone, eventually sharing the same fate as Jocasta.

for the existence of dilemmas as some have, nor need I do so. Instead, I argue that the existence of moral residue imposes constraints on the correct account of moral justification in Chapter 5.

## 6. The Suberogatory

The final moral phenomena I will consider is suberogation. Even the most reasonable ethicist is likely to raise an eyebrow at the possibility of permissible yet "patently bad" actions.[30] I am typically required to perform the morally best action available (according to some measure or other), so when I perform an action with decisive moral reason weighing against it, I have almost certainly done something wrong.[31] Regardless, many philosophers defend the possibility of the suberogatory – the logical counterpart of the supererogatory.[32]

While supererogatory actions are non-obligatory good actions, suberogatory actions are non-prohibited bad actions. Consider an example of suberogatory action (Ramirez 2014).[33]

> **PERRY ZOSO:** Suppose it is lunch hour at a popular diner. Perry Zoso was fortunate to arrive early, get a seat, and finish his lunch. He is now lazily sipping his coffee and reading the newspaper, despite the very long line of hungry guests amassing at the entrance.

Perry Zoso is under no obligation to vacate his seat to another guest. Yet, we might also think Perry Zoso has more moral reason for giving his seat to another guest than he has for continuing to sip his coffee. When Perry Zoso continues sipping his coffee, he performs the

---

[30] Heyd, D. (2002). "Supererogation." *The Stanford Encyclopedia of Philosophy*. www.plato.stanford.edu

[31] Michael Zimmerman gives a nice summary of different accounts of obligation, all of which guide us to take the best action according to some measure or other. See M. Zimmerman's (2014) *Ignorance and Moral Obligation* (Oxford University Press), esp. Chapter 1.

[32] Chisholm, R. (1963). "Supererogation and Offence: A Conceptual Scheme for Ethics," *Ratio*, 5: 1–14; Driver, J. (1992). "The Suberogatory." *Australasian Journal of Philosophy*, 70: 286–295; Massoud, A. (2016). "Moral Worth and Supererogation." *Ethics*, 126: 690–710.

[33] Example extends on R. Chisholm and E. Sosa's (1966). "Intrinsic Preferability and the Problem of Supererogation" (*Synthese*, 16:321-331). See Ramirez, A. J. (2014). "Righteous Villainy: A Defense of Offence" (PhD thesis, PDF retrieved from personal correspondence: June 01, 2015).

action he has decisive moral reason to not perform but no duty not to perform, so we might think Perry Zoso's action is suberogatory.

Supposing suberogation exists as a legitimate category of action, we might yet wonder whether the suberogatory helps us to analyse any morally significant situations. While staying too long at a restaurant, driving too slowly during rush hour, or writing a check at the cash register are all, as Chisholm might put it, offensive behaviors, they are all morally trivial. If suberogation is useful for moral theorizing, then it should help us to analyse morally significant situations.

In which case, I need a clear case of suberogatory non-trivial moral offence. Consider the case of Virgil's AENEAS.

> **AENEAS:** Aeneas was destined to found the city of Rome. During his journey, he encounters Dido, Queen of Carthage, and falls in love with her. While lingering in Carthage, the god Mercury reminds Aeneas of his glorious destiny, and Aeneas chooses to proceed with his journey. Dido promises to commit suicide when Aeneas leaves. Even so, Aeneas leaves Carthage, and, sure enough, Dido commits suicide.

On the above description, we may well think Aeneas faces a moral dilemma, yet his case is unlike that of Orestes. While Orestes faces a conflict between two duties handed to him by the gods, Aeneas faces a conflict between the pursuit of love and the pursuit of glory, which, as fortune would have it, were both inflicted on him by the gods. Aeneas's love of glory outweighed his love for Dido, so he ventured forth for glory rather than lingering with Dido.

The death of Dido is a tragedy, which we may very well think Aeneas had decisive moral reason to prevent. He could have sacrificed his own glory in order to ensure that Dido

lived, and one could plausibly hold that morality favors this option. Afterall, the gods forced her to fall in love with him, making her an innocent bystander wrapped-up in his fortunes. Even so, many of us (though not all) think people are under no obligation to stay with romantic partners once we are ready to move on. This is true even if we are aware that leaving that person will cause her a great deal of suffering. Afterall, very few romantic relationships end painlessly. So, although staying with Dido might have been the best thing for Aeneas to do morally speaking, he was nonetheless permitted to fulfill his destiny and found Rome. Still Aeneas is troubling. One might wonder how morality makes room for manifestly bad actions and whether such actions serve any function in our moral theories, besides filling out the logical space.

So, we must account for the suberogatory because it is needed for making sense not just of offensive action but also tragic ones. I argue that the suberogatory is a legitimate category of actions resulting from our commitments to other moral phenomena – specifically the possibility of conflicting moral duties. I further argue that the existence of the suberogatory helps to set limits on other commitments we can take on in our moral theories. So, our existing moral commitments make room for manifestly bad suberogatory actions and these actions do serve a function in our moral theories.

**7. Methods**

This research does not aim at proving any particular thesis. Nor does it aim at vindicating any existing philosophical orientations. Nor does it aim at impugning them. This research aims at making sense of the aforementioned puzzles within a single unified framework, and I will use any tools available to me to achieve this task. In one respect, this project is purely exploratory as I am exploring the question of what such a framework might look like because, right now, no such framework exists. So, if this research aims at proving anything, it aims at proving that such a framework is possible if we reject certain widely held assumptions in moral philosophy. I argue that when we give up these assumptions, the making of the framework is relatively straightforward.

The issue at hand is that philosophers have foresworn the most direct route through these puzzles due to concerns about apparent contradictions. I believe these philosophers have given up too easily. These puzzles are about potential worrisome contradictions. This dissertation answers the question of what might have to be true to resolve the contradictions. Even so, the project is complicated by the various ways philosophers have attempted to solve these puzzles. I will not survey all the various solutions on offer, since most solutions come with commitments incompatible with the aim of this dissertation. Instead, I primarily take up issue with several of these core commitments, since the specifics of the various solutions are of little help when based in a view with deep incompatibilities with the framework I aim to develop.

In another respect, I think what follows, in rough outline, is a step closer to the truth of things. As the reader will come to see, the framework on offer is explanatorily powerful. It also has the virtue of elegance in its simplicity and symmetry. The nitty gritty details are, of

31

course, complex, and spelling them out is onerous. So, what follows is undoubtedly underdeveloped. Yet, I hope that this dissertation demonstrates proof of concept – a term used in the sciences when some evidence of a project's feasibility is required. I am proposing to outline a new moral framework with a new way of thinking about issues in moral philosophy – a tall order indeed.

This research begins with the assumption that I have accurately characterized the above puzzling patterns of moral assessment and that the patterns of moral assessment under consideration are all correct (i.e. they are accurate assessments of real moral phenomena). I will treat (1) – (5) as datapoints that my own account must accommodate. The moral framework I develop in this dissertation will therefore be constrained by these patterns of moral assessment.

One central assumption of this research is that a unifying theory is *ceterus paribus* more desirable from the theoretical standpoint than a jerrymandered one. I have proposed that a unifying theory is one that can explain and justify the patterns of assessment observed in the previous sections. This constraint on my framework renders it unfriendly with any account taken to be incompatible, for example, with genuine moral dilemmas, such as straightforward Kantianism or Rossianism. However, my project relies on many assumptions a deontologist might welcome. For example, any moral theory consistent with genuine moral dilemmas must be friendly to moral absolutism, according to which some of our moral duties are exceptionless. On my account, the property of an action that makes it wrong always makes it wrong without exception. Relatedly, my account aims to be consistent with the general deontic prohibitions on harm indicated by (1) – (5), where whatever property of an action that makes it wrong is usually concomitant with something harmful to the agent in question.

This constraint on my framework also demands that I develop a particular kind of moral framework – one that justifies our *post hoc* moral assessments from the standpoint of the moral judge rather than the moral deliberator. The first kind of moral theory provides us with the standards of morality, which allow a critical observer to judge whether the agent has, in fact, acted, intended, etc. morally. The second kind of moral theory provides us with a moral agent's manual, which instructs agents on how to be moral. For those acquainted with Scanlon's two uses of moral principles, this sort of distinction should be familiar. The *theory of morality* instructs us on the *critical use* of moral principles. Moral principles, used as standards of criticism, help a critical observer to determine whether an agent took the reasons for or against performing the action into proper consideration. The *moral agent's manual* instructs us on the *deliberative use* of moral principles. Moral principles, used as guides to deliberation, pick out the features of an action that count as reason for or against doing it for the purpose of determining whether the action is permissible.

On Scanlon's sort of view, moral principles serve both as guides for action and as standards against which we assess actions. However, Scanlon holds that the deliberative use of moral principles is the standpoint from which the permissibility of an action is assessed, while the critical use of moral principles is the standpoint from which the moral worth of an agent is assessed. On this account, morality does not have a standpoint from which the moral judge can critically assess the permissibility of an action. However, I hold that the critical assessment of an action's permissibility can be achieved using the sort of "non-deliberative principles" that make up a theory of morality but not the moral agent's manual. As a result, the view on offer does not engage extensively with concerns about action guiding principles, a concern I will address at the end of this dissertation, especially as it relates to concerns about

the possibility of genuine moral dilemmas. Moral judges are positioned to assess any feature of a morally charged event *ex post facto*, including the agent or her action, by appealing to considerations agents cannot use in their deliberations about what to do, such as the unforeseen actual, rather than expected, consequences of an action.

Beginning from a critical or objective, rather than deliberative, standpoint places additional constraints on my view. For example, the moral framework on offer is most consistent with an account of negative and therefore non-optimizing duties, making my account incompatible standard consequentialist views. I will say more about this concern in Chapter 2. Because the highlighted aspects of the moral framework in this dissertation are only really consistent with an account of non-optimizing duties, I consider it to be an adequacy account – an account of good-enough morality rather than an account of moral excellence (moral perfectionism or virtue). In that vein, I assume that we often fail from the moral standpoint in a wide variety of different ways while nonetheless succeeding at the same time.

We agents and our actions are often morally imperfect, yet we seem to do well-enough from the moral standpoint to go about our daily lives. Moral adequacy is a measure of the ways in which we succeed from the moral standpoint, despite our moral failures. A morally adequate agent or action has succeeded sufficiently from the moral standpoint in the relevant realm of moral assessment. Moral excellence relates more directly to moral worth considerations, which I do discuss in Chapter 3. However, the aim of that chapter is to relate an account of moral worth compatible with my framework and my account of permissibility, rather than dealing with concerns relating to moral excellence. I reserve this work for future projects, but I flag it here because I want to make it clear that the framework as it stands in this dissertation is set-up to make way for the ethicist's interests in the right and the good.

In another respect, a theory is unifying when the same underlying principles employed for one moral assessment are the same as those employed for assessing other similar phenomena. This use of a unifying theory underlies much of the argumentative strategy throughout this dissertation. Here, I am appealing to supervenience considerations, or what I call the principle of parity above and throughout. I suggest that if we always assume that similar things should be assessed similarly and different things should be assessed differently, we can identify a cohesive framework underlying our apparently contradictory moral assessments. However, parity is not the only unifying consideration here.

I am also appealing to a general need to provide overarching explanations for our various moral assessments in a way that accounts for the trivial as much as the non-trivial. I think a moral theory should offer criteria for rendering assessments on all morally relevant considerations, even if philosophers do not normally care about such considerations. For example, questions about moral justification only ever seem to arise when an agent performs an ordinarily wrong action in the course of performing an ordinarily right action, such as when killing in self-defense. Yet, all ordinarily permissible actions are presumably justified, since such actions would not be permissible if they were not morally justified. Philosophers rarely talk about moral justification for unproblematically permissible actions, but I propose that an account of moral justification should explain why any action is morally justified using the same set of criteria, regardless of whether we rarely talk about a large swath of possible cases. Ordinary permissible actions are not excepted from this principled explanation. A unifying theory is complete in this respect.

I achieve my aims starting in Chapter 2, where I outline the bare structure of the Unified Moral Framework [UMF]. In the first third of this chapter, I lay out the conceptual

space utilized in the remainder of the dissertation, beginning with a pair of distinctions. In general, moral assessments have two possible targets – actions and agents. I call this the Fundamental Moral Distinction. Assessments are made by appeal to certain considerations. Some assessments involve only the target of the assessment, such as the agent-independent features of the action alone or the agent's motivations alone. Some assessments are hybrids, appealing to interactions between both sorts of considerations. I call the former first-order moral assessments and the latter second-order moral assessments. This dissertation is primarily about the structure of second-order moral assessments, the most important of which are moral worth (praiseworthiness and blameworthiness) and, I will argue, permissibility.

However, first-order moral assessments, especially first-order act-assessments play a significant role in second-order moral assessments, so the second two-thirds of Chapter 2 focuses on my account of first-order moral-assessments. Most importantly, I argue that the possibility of resolvable genuine moral dilemmas suggests that we sometimes have moral reasons to do things we have duties not to do and duties to do things we have moral reasons not to do. I suggest that we can therefore distinguish between what we have moral duty to do and moral reason to do. I argue that the Realm of the Erogatory involves first-order act-assessments about what we have moral reason to do but no duty to do. Suberogation is a subclass of erogation assessment, which follows straightforwardly from my framework.

Chapter 3 deals exclusively with my account of moral worth as a second-order agent-assessment. Although I do not directly require this account to explain any of the above puzzles, it does carve out some of the conceptual space reserved for my account of permissibility and provides a model for permissibility as a second-order act-assessment. For this reason, Chapter

3 is crucial scaffolding for my framework. I start with the assumption that moral worth assessments are clear candidates for second-order moral assessments. Moral worth is an agent-assessment rendered by appeal to the relationship between the agent's motivations and her action. If moral worth assessments are agent-assessments, one might wonder what role the agent's action serves in the assessment of her. I propose that the agent-independent features of the action provide the background against which the agent's motivations are assessed by specifying which of the agent's motivations are relevant to a particular moral worth assessment. The relation between the agent and her action is the reasonably well-established responsibility relation. Moral worth is an assessment of an agent for the quality of her responsibility relation to her action.

Although primary focus of this chapter is the actual account of praiseworthiness and blameworthiness on offer, this chapter brings in the responsibility relation to showcase that an agent's responsibility for her action is distinct from an assessment of the agent for the quality of her responsibility relation. This allows us to make sense of phenomena like responsibility without blame – a moral assessment that agents are sometimes fully morally responsible for their wrong actions but not blameworthy. This category of moral worth assessment creates space for a category of permissibility assessment I heavily utilize in making sense of (1) – (5).

Chapter 4 is the real meat of this dissertation and deals primarily with my account of permissibility as a second-order act-assessment. I begin by rejecting what I call ordinary permissibility accounts, which typically subscribe to the following claim: (A) an action is morally permissible if and only if the action is not morally wrong. I argue that any theory of permissibility compatible with genuine moral dilemmas is incompatible with claim (A)

because the duties relevant to genuine moral dilemmas are absolute duties. Any action failing to conform with an absolute duty is morally wrong because there are no exceptions to such duties. In genuine moral dilemmas, any action the agent performs is morally wrong because any action the agent performs fails to conform with one such duty. Yet, in genuine moral dilemmas, it typically seems as though there is a morally preferred solution – an action the agent permissibly performs.

The straightforward solution to genuine moral dilemmas is therefore to allow for the possibility of permissible-wrong actions [PWA] (the permissibility counterpart to responsibility without blame). PWA is possible on the UMF because wrongness is a first-order act-assessment while permissibility, I argue, is a second-order act-assessment, meaning that they appeal to different assessment criteria. I also argue that AAP in DDE cases can be explained by PWA. While both actions are morally wrong in standard pairs of DDE cases, I suggest that AAP is best explained by appeal to the agent's motivations, which establish one action as an instance of PWA.

If PWA provides a worthwhile solution to AAP and genuine moral dilemmas, then I require an account of permissibility as a second-order act-assessment because PWA is only possible if an assessment of the action's wrongness is different from the assessment of the action's permissibility, and if the agent's motivations are directly relevant to the permissibility of the agent's action. On the latter point, if permissibility assessments are act-assessments, one might wonder what role the agent's motivations serve in the assessment of her action.

To address this concern, I argue that just as an agent's action is directly relevant to assessing her moral worth, the agent's motivations are directly relevant to assessing the permissibility of the agent's action. The agent's motivations establish the background against

38

which we assess the agent-independent features of the action by specifying which agent-independent features of the action are relevant to our second-order act-assessments. Put very roughly, the difference between the two agents in the bomber cases (a central example of the application of the AAP) is that the Tactical Bomber's motivations highlight the action's dutifulness as the first-order act-assessment relevant to the action's permissibility, while the Terror Bomber's motivations highlight the action's wrongness as the first-order act-assessment relevant to the action's permissibility. The agent's motivations are relevant to the permissibility assessment because they settle the question of the description under which the action should be assessed for its permissibility.

Chapter 5 develops my account of moral justification, specifically my account of the function of moral justification and what actually does the justifying. Importantly, PWA is incompatible with ordinary permissibility accounts, which typically subscribe to claim (A). I believe that (A) is a corollary of another claim: (A*) an action is morally justified if and only if the action is not morally wrong. (A) follows from (A*) where an action's permissibility is dependent on its justification and an action is morally permissible if and only if the action is morally justified. So, to secure a theoretical basis for PWA, I defend a non-vindicatory account of moral justification – an account of moral justification according to which the action is not wholly vindicated or shown to be non-wrong such that morally wrong actions are sometimes morally justified.

Chapter 5 therefore takes issue with vindicatory accounts of moral justification and the commitment to (A*). I argue we have good reason to think vindication is not the function of moral justification because it rules out moral dilemmas and fails to make sense of moral residue. I then offer an alternative to vindicatory accounts of moral justification, suggesting

that the purpose of justification is to overcome the problem of impermissibility and not the wrongness objection. To overcome the problem of impermissibility the action must be justifiable, so I develop an account of what makes an action, especially a wrong action, justifiable. Next, I propose an alternative model to justificatory vindication, which I call an adequacy account of moral justification [AAJ]. The AAJ employs, what I call, a proving method, the aim of which is to show an action is good-enough despite its flaws. With that, I conclude in Chapter 6, where I give an overall, summary description of my framework.

**Conclusion**

The aim of this dissertation is to develop a moral framework constrained by the puzzles detailed above. Now that I have outlined my dissertation and highlighted some central assumptions in this introduction, I take the first steps by outlining the UMF, which carves out the conceptual space utilized throughout the remainder of the dissertation in Chapter 2.

Chapter 2: The Unified Moral Framework

and

First-Order Moral Assessments

**Introduction**

Chapter 1 established the overarching aim of this dissertation, which is to construct a moral framework capable of accommodating five relatively controversial moral phenomena: (1) genuine moral dilemmas; (2) asymmetrical assessments of permissibility [AAP] in Doctrine of Double Effect [DDE] cases; (3) Praise-Blame Asymmetry in Side-Effect Effect cases; (4) moral residue; and (5) the suberogatory. The task set out for me is straightforward, beginning with the basic framework in which I lay out the conceptual space utilized in the remainder of the dissertation.

In the first third of this chapter, I outline the bare structure of the Unified Moral Framework [UMF]. The UMF is based on a pair of distinctions between types of moral assessments. In general, moral assessments have two possible objects – actions and agents. I call this the Fundamental Moral Distinction. Assessments are made by appeal to certain considerations. Some assessments involve only the object of the assessment, such as the morally significant agent-independent features of the action alone or the agent's motivations alone. Some assessments are hybrids, appealing to interactions between both sorts of considerations. I call the former first-order moral assessments and the latter second-order moral assessments. This dissertation is primarily about the structure of second-order moral assessments, the most important of which are moral worth (praiseworthiness and blameworthiness) and, I will argue, moral permissibility.

However, first-order moral assessments, especially first-order act-assessments, play a significant role in second-order moral assessments, so the second two-thirds of Chapter 2 focuses on my account of first-order moral-assessments, especially as they relate to (1) and (5) above. In what follows, I define the conceptual niche used to distinguish the obligatory

from the erogatory by distinguishing between what we have moral duty to do and moral reason to do. I then present the account of duty I think best fits the UMF – one according to which duties are negative, absolute, and consistent with moral objectivism. Next, I briefly explain how such an account suggests that the UMF is an account of mere moral adequacy before briefly addressing how we arrive at first-order act-assessments about suberogation. The remainder of the chapter details a pair of first-order agent-assessments, namely moral sensitivity and moral judgment, which involve an appeal to the agent's motivating reasons *de re* and the agent's motivating reasons *de dicto*. I then conclude.

## 1. The Simplified Moral Framework

In Chapter 1, I described the case of Oedipus, who murdered his father and married his mother. I described both as actions in violation of duty, neither of which afforded Oedipus opportunities for recompense until he eventually blinded himself and abdicated the throne. The appropriate assessment of the case is complex. On the one hand, Oedipus' incest and patricide are repugnant. On the other, Oedipus is the victim of a family curse where the outcomes of his actions were fated yet unforeseen, and he was racked with guilt and made wretched upon becoming aware of his abomination. In our judgment of Oedipus, we distinguish the moral assessment of the actions and of the agent performing them. Oedipus the agent is prideful and wretched, while his actions are repugnant and abominable – though perhaps not entirely his fault. I call the distinction between agent-assessments and act-assessments the fundamental moral distinction [FMD].[34]

The FMD is a plausible to the point where philosophers sometimes align entire moral theories along the primacy of one type of assessment or the other.[35] Judgments about good

---

[34] Not to be confused with the elemental moral distinction, offered by Gardener, who opposes what I call the fundamental moral distinction. According to Gardener, the fundamental objects of moral assessment are people and their lives. On this account, we are unable to make act-assessments without regard for the agent's motivations because the agent's motivations are manifested in the agent's action, and we are unable to make agent-assessments without regard for the agent's action because the agent's actions partially constitutes how we regard agents. For example, we cannot say of Kant's Good Willed Grocer that she is honest with her customers unless she has demonstrated a history of acting honestly towards her customers, nor can we make sense of her action's being an honest one without her being motivated by the action's honesty. See J. Gardner's (2005) "Wrongs and Faults" (*The Review of Metaphysics*, 59.1).

[35] While the FMD appears to be a relatively common way of carving up the moral landscape, many philosophers suppose one side of the division is more central to ethical theory. Kant, holds that the good will is central to our moral theorizing, while Mill holds that the consequences of our actions are. Both philosophers argue that the other side of the division, while important, holds a subsidiary position in ethical theory. For Kant, attempting to save a drowning swimmer on the basis of my correct judgment that I ought to save the drowning

will and ill will, praiseworthiness and blameworthiness, and virtuousness and viciousness are assessments relating to the agent, while judgments about rightness and wrongness, permissibility and impermissibility, and moral justification are moral assessments relating to an action. While I uphold the FMD throughout this dissertation, I do not hold that either sort of moral assessment is more fundamental – more important or conceptually central to ethical theory. Instead, I propose that the FMD is not the only principal moral distinction, and that we can carve out yet another division in the moral space. The second distinction I discuss is a distinction between First-Order Moral Assessments and Second-Order Moral Assessments.

Consider a central difference between types of agent-assessments. Sometimes we assess the quality of the agent's will without regard for her action. We might say of someone that she means well or that she is cruel at heart, even if she never succeeds in benefitting or harming anyone. We might think this agent is motivated by good will or ill will. Yet, such assessments are not enough for taking the full measure of a moral agent. Sometimes we assess the interaction between the agent's motivations and her action. For example, one might think, as does Nomy Arpaly, that an agent is praiseworthy when she performs the right action while motivated by her good will and that an agent is blameworthy when she performs the wrong action while motivated by her ill will.[36]

---

swimmer is good regardless of whether I actually save the save the swimmer. Even so, actually saving the swimmer is also a moral good – just not a good in itself. In the other direction, Mill would think that actually saving the swimmer is likely more morally valuable than not saving the swimmer, but my correct judgment that I ought to save the swimmer is morally good insofar as making such judgments is the sort of thing that ultimately motivates people to do good things.

[36] In speaking of an agent's good or ill will, I mean only to point to the agent's motivations and not the agent's actions. In doing so, I deviate from philosophers like Kant and align myself more with contemporary philosophers like N. Arpaly (2003) in *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford University Press).

Both sorts of assessments are agent-assessments, yet judgments about good will and ill will are non-hybrid moral assessments in the sense that they appeal only to the morally relevant features of the agent's motivations. These judgments only ever indirectly involve consideration of the agent's action, such as when she desires to act in a way that will bring about good or bad consequences or judges that so acting is good. Her judgments and desires are morally assessible even if she never actually acts on them. Such judgments do not appeal to the agent's action. Judgments about praiseworthiness and blameworthiness, on the other hand, are hybrid moral assessments appealing, in addition, to the agent's action. In order to assess whether an agent is praiseworthy or blameworthy, we need to know something about what the agent actually did. Otherwise, we have nothing to praise or blame that agent for.[37]

Thus, within the category of agent-assessments, we distinguish non-hybrid and hybrid moral assessments. I propose to call the former first-order moral assessments.

**First-Order Moral Assessments:** Moral assessments the objects of which are evaluated by appeal only to the morally relevant features of the object itself.

When making claims about the morally relevant features of the object itself, what we appeal to when making first-order moral assessments, I will refer to the assessment properties.

Plausibly, some act-assessments are first-order moral assessments as well. Meaning, we sometimes morally judge actions without regard for the agent performing the action. I

---

[37] One might wonder whether we are ever praiseworthy or blameworthy for having certain virtues or vices. This is a complicated question. Attributing a virtue, such as honesty, to a person is itself a kind of propitious moral assessment of the agent. If we think all propitious moral assessments are a kind of praise, then I am here talking only about the specific sort of praise warranted when an agent's good-will motivates her right action. To say that an agent is virtuous is to talk about a different sort of praise about the agent's overall disposition to do the right thing at the right time for the right reasons. I do not, herein, address the role of virtue in my framework, but I do think my approach can accommodate it.

propose that judgments about rightness and wrongness are first-order act-assessments. Typically, to assess whether an action is right or wrong, we need only know whether the action conformed with the relevant duty, maximized aggregate happiness, was the action that the virtuous person would perform, etc. Such judgments do not appeal to the agent's motivation in acting.

When we make hybrid moral assessments appealing to the interactions between the assessment properties in first-order moral assessments with different sorts of objects, we issue what I call second-order moral assessments.

> **Second-Order Moral Assessments:** Moral assessments the objects of which
>
> are evaluated by appeal to interactions between the assessment properties in
>
> first-order moral assessments.

As already observed, we commonly hold that assessments of praiseworthiness and blameworthiness are second-order agent-assessments. I propose, in addition, the possibility of second-order act-assessments.[38] Figure 1 illustrates the framework:[39]

---

[38] Hitherto, second-order act-assessments, as a category of moral assessment distinct from first-order act-assessments and second-order agent-assessments, have been neglected in the existing philosophical literature. The failure to recognize this category of assessment leads to verbal disputes, as indicated (but not made explicit) in Section 3. Another reason to recognize second-order act-assessments concerns parity. My view yields a more symmetrical and elegant taxonomy of moral assessments than one that excludes it. While not all theorists will be moved by these virtues in a theory, they may well be moved by the simplicity with which I handle nigh intractable debates in ethics with this minimalist machinery.

[39] The actual moral framework under development comprises more than two orders of assessment. Third-order moral assessments involve the assessment of interactions between second-order moral assessments and corresponds more directly to Gardener's elemental moral distinction, according to which the objects of moral assessment are people and their lives.

**Figure 1: The Simplified Moral Framework**

|  | Act-Assessment | Agent-Assessment |
|---:|:---:|:---:|
| **First-Order** | Act Only<br>(1) | Agent Only<br>(2) |
| **Second-Order** | Act & Agent<br>(4) | Agent & Act<br>(3) |

One might initially worry that second-order moral assessments fail to uphold the FMD. Afterall, second-order moral assessments are hybrid moral assessments, which coadunate evaluations of agent motivations and their actions. Nevertheless, second-order moral assessments uphold the FMD in another respect. Sometimes we make second-order assessments, such as praiseworthiness and blameworthiness, the objects of which are agents, and, I propose, sometimes we make second-order assessments the objects of which are actions. I, therefore, propose that a plausible moral theory should yield at least two different act-assessments – (Box 1) one involving only the morally relevant features of the action, and (Box 4) one involving, in addition, the agent's motivations.

The distinctions identified for the SMF are plausible even if we disagree about what type of moral assessment fits into each box. Second-order act-assessment, labeled Box 4 above, is ultimately the category of act-assessment into which I fit judgments about the moral permissibility of an action. I propose moral judgments about the permissibility of an action appeal to the agent's motivations in acting, much like moral judgments about praiseworthiness and blameworthiness.

**Figure 2: The Simplified Moral Framework [Completed]**

|  | Act-Assessment | Agent-Assessment |
|---:|:---:|:---:|

| First-Order | Right & Wrong (1) | Good Will & Ill Will (2) |
|---|---|---|
| Second-Order | Permissible & Impermissible (4) | Praiseworthy & Blameworthy (3) |

Figure 2 illustrates the SMF – complete with examples of the specific moral assessments I propose for each box. My reader can agree to the content of Figure 1 without agreeing to the content of Figure 2. If so, I would fail to convince my reader of the possibility of permissible-wrong action, but I might yet convince my reader that the relevant category of act-assessment exists.[40]

The SMF consists in a pair of plausible moral distinctions yet yields a relatively controversial result. In this section, I proposed that moral judgments about the permissibility of an action at least sometimes appeal to the agent's motivations in acting. Here, I need not defend the claim that permissibility is a second-order act assessment – I do that work in Chapter 4. Here, I only aim to define the conceptual niche used in distinguishing rightness and wrongness from permissibility (or its like). My characterization of permissibility is consistent with the sort of characterization offered by proponents of the Doctrine of Double

---

[40] Ultimately, I do not mind losing custody of the term 'permissibility' in the divorce, so long as I get to keep the house into which I propose to settle it. However, I believe we have good reason to use the term 'permissible' in the way I suggest. As I show in Chapter 3, some philosophers already deploy a version of my usage, so my suggestion is plausible if not uncontroversial. Moreover, I will argue that current debates suggest the likelihood of two distinct categories of act-assessment. If so, we need two distinct terms for making those assessments. If judgments about rightness and wrongness already account for one sort of act-assessment, then judgments about permissibility are a likely candidate for the second. Otherwise, we have two sets of terms for making the very same act-assessment, and we are in need of identifying yet another. These are questions for conceptual ethics. Conceptual ethics is an emerging sub-field in normative theory, which deals with questions including: "Which concepts should we use to think and talk about the world and to do all of the other things that mental and linguistic representation facilitates?" See A. Burgess and D. Plunkett's (2013) "Conceptual Ethics I" and "Conceptual Ethics II" (*Philosophy Compass*, 8.12).

Effect, which I address also in Chapter 4. So, although the view that judgments about permissibility appeal to agent motivations is relatively controversial, the view is not unheard of, and I hope readers will set questions about this issue aside for now. I now turn to a discussion of first-order moral-assessments, starting with first-order act-assessments.

## 2. Reasons, Duties, and Dilemmas

The core characteristic of a genuine moral dilemma involves conflicting moral duties. Recall SIDE-TRACK from Chapter 1. Although Medea has a duty to save the five, she has a simultaneous conflicting duty to not kill the one. If the dilemma in SIDE-TRACK is a genuine moral dilemma, Medea is bound equally by each – she faces two countervailing conflicting duties. If we take such claims about duties seriously, then we are stuck making sense of how conflicts between duties are resolvable. In Chapter 4, I will propose a novel solution to the puzzle of moral dilemmas and provide a way to resolve conflicts between moral duties using a hybrid account of moral permissibility. To develop this account of permissibility, as a second-order act-assessment, I require additional distinctions at the level of first-order moral assessments.

In this section, I propose the first of a pair of plausible distinctions between types of first-order moral assessments, which make up the UMF. I discuss the distinction between what we have moral duties to do and what we have moral reasons to do. With this distinction in place, I examine how it helps us to resolve genuine moral dilemmas. My solution reveals conceptual space for a somewhat controversial category of moral assessment – suberogation, which, I propose, is a straightforward consequence of allowing genuine moral dilemmas into one's moral theory.

Genuine moral dilemmas are not possible on views according to which an agent only ever has a single operative duty.[41] On such a view, when Medea pulls the lever, she conforms to the single operative duty to save the five. Her action is morally right, since she violates no

---

[41] This sort of view will not always be the same as one according to which our duties only ever generate a uniquely right action. One might think that an agent sometimes has more than one operative duty but that conflicts ultimately resolve in favor of performing one action.

duty to not kill Jason. On this view, Medea never had a duty to not kill Jason. She therefore never faced a moral dilemma. Because I aim to develop a framework compatible with the existence of genuine moral dilemmas, such views are not under consideration herein. However, such views do reveal two important features of the moral landscape.

When we assess the rightness or wrongness of an action, we assess the agent's conformance with an operative duty. According to most moral theories, our duties to perform a specific action are derivative of the normative moral reasons for or against performing that action. As with the case of *pro tanto* duties, we often say that an agent has a duty to do only what she has the most and mightiest moral reasons to do. So, if Medea had more and mightier moral reasons to save the five than to not kill Jason, then she has only a single operative duty to save the five and no duty to not kill Jason. Of course, these theorists might think Medea would have a duty to not kill Jason had she not had a weightier duty to save the five.

On these views, first-order act-assessments appeal to two distinct features of actions – what we have normative moral reasons to do and what we have duties to do. Recall that I define first-order moral assessments as follows.

> **First-Order Moral Assessments:** Moral assessments the objects of which are
> evaluated by appeal only to the morally relevant features of the object itself.

A first-order act-assessment is a moral assessment of an action, which is evaluated only by appeal to the morally relevant features of that action. Following my predecessors, I believe a good account of first-order act-assessments should accommodate these two distinct assessment properties of actions.

**Figure 3: Normative Moral Reasons & Duties**

| | Object of Act-Assessment |
|---|---|
| | Normative Moral Reason (1a) |
| **First-Order** | Duty (1b) |

Figure 3 depicts the distinction between Normative Moral Reasons (Box 1a) and Duties (Box 1b) categorized as the morally relevant features of action evaluated by first-order act-assessments. From here on, I will typically refer to normative moral reasons as moral reasons.

With this distinction in hand, I examine how it helps us to resolve genuine moral dilemmas. If we suppose duties sometimes enter into conflicts, then we should probably reject the view that an agent has a duty to do only what she has the most and mightiest moral reasons to do. Duties are derived differently, but this does not mean we should reject the view that moral reasons play a significant role in our moral assessments. We might think the duty with the most and mightiest moral reasons in its favor inevitably defeats any putative countervailing duties. This explanation for how we might resolve conflicts between duties has important implications for our moral theorizing. Specifically, we get the result that what we have a moral duty to do is sometimes different from what we have moral reason to do.

Suppose, as I have throughout, that SIDE-TRACK is an instance of a genuine moral dilemma. Medea has a duty to save the five, and she has a simultaneous conflicting duty to not kill the one. Moreover, she has moral reasons that count against killing the one at the very same time as she has more and mightier moral reasons favoring saving the five. As a result, Medea has moral reasons against performing the action that she has a decisive duty to perform, and she has moral reasons for performing an action that her defeated duty prohibits. If what

54

we have moral reasons to do and what we have duties to do sometimes point towards different actions, then what we have a moral duty to do and what we have moral reason to do are possibly different sorts of moral considerations.

I propose that first-order act-assessments appealing to moral reasons versus duties generate two distinct kinds of act-assessments. The first are judgments about the erogatory.

**Erogation:** Moral assessments of actions evaluated by appeal only to the normative moral reasons regulating the action.

In this category of assessment, we find the supererogatory and the suberogatory, the latter of which was introduced in Chapter 1.[42] I further discuss this category of first-order act-assessment later in this chapter. The second are judgments about the obligatory.

**Obligation:** Moral assessments of actions evaluated by appeal only to the duties regulating the action.

In this category of assessment, we find rightness and wrongness. The distinction between the erogatory and the obligatory can be mapped follows.

**Figure 4: Erogation and Obligation**

|  | **Act-Assessment** |
| --- | --- |
|  | Suberogatory, Erogatory, & Supererogatory (1a) |
| **First-Order** | Right & Wrong |

---

[42] Harman's (2016) "Morally Permissible Moral Mistakes" (*Ethics*, 126) deals with phenomena that most likely fall into this category of moral assessment. As does G. Mellema's (1987) "Quasi-Supererogation" (*Philosophical Studies*, 52.1).

| | (1b) |
|---|---|

Figure 4 depicts the distinction between the erogatory (Box 1a) and the obligatory (Box 1b) categorized as first-order act-assessments.

Before proceeding with this discussion, let me first make three notes about the erogatory. First, I recognize that my use of erogation is non-standard. Erogation assessments, as they appear in acts of supererogation or suberogation, refer to duty. The former is typically formulated as acting above and beyond the call of duty while the latter might be characterized as acting somehow "below" the call of duty. I am interpreting these assessments to instead refer to acts beyond the scope of duty but not beyond the scope of morality, which is not a wildly deviant interpretation of what we might mean by "beyond the call of duty". So, I am stipulating a new but definition for this term that already has uses in our moral language. Further, the sort of moral phenomena we typically assess using erogation fit neatly into the category of action I assign to erogation assessments, so I am analyzing and assessing typical cases using different criteria, but I am arriving at the same sorts of judgments about the same sorts of phenomena.

Second, I do typically define specific erogation assessments with reference to duty. For example, I define supererogation is a moral assessments of an action that we have the more moral reason to do than against doing but no duty to do. Defined in this way, supererogation looks like a second-order act-assessment, since it involves an assessment of reasons and duties. However, supererogation is an assessment only of the normative reasons regulating the action, since, in cases where supererogation is an appropriate assessment, there is no duty to be assessed, so we cannot be assessing the reasons in relation to any duty.

56

Third, we can assess the normative moral reasons for or against performing an action regulated by duty without any regard for the duty. Such assessments are erogation assessments, achieved in the same way we make more familiar erogation assessments. Many such assessments will have the same flavor as supererogation or suberogation, except they will be assessible alongside a duty. I refrain, here and throughout, from applying the familiar terms to the sort of erogation assessments under consideration primarily because of the unintuitive usage, but it makes no difference to me. The assessments of particular relevance to this dissertation are those that determine the preferability of an action in dilemma cases.

**Morally Preferred:** An action is morally preferred IFF the action is one those with the most and mightiest moral reasons in its favor.

In the event of a tie, the actions share in moral preference. Morally preferred actions are the best available, morally speaking, but need not always be characterized as the actions with the most morally desirable outcome. I think some considerations, such as the risk of bringing about a bad outcome, typically weigh heavily against performing actions with good outcomes. This idea will be developed later in this chapter and brought up many times throughout this dissertation. Lastly, moral preference is an assessment we can make of an action regardless of whether the action is regulated by duty, but moral preference is central to sorting out the way in which a dilemma should resolve. This idea will be developed in greater detail in Chapters 4 and 5.

I now turn to a short discussion about the realm of the obligatory. The possibility of genuine moral dilemmas forces us to rethink how we might categorize the rightness and wrongness of action. Ordinarily, an action renders morally wrong when the agent performing the action fails to conform with her duties, and an action renders morally right when it

conforms with the agent's duties at the time of acting. Moral dilemmas are troublesome because the action conforms to some of the agent's duties and fails to conform with others at the time of action, where, given the externalities of the case, no action the agent performs will conform with all of her operative duties at the same time. This seems to make whatever action she selects from her set of alternatives both *prima facie* right (as the discharge of a duty) and prima facie wrong (as a failure to live up to his or her obligations at the time). However, an action cannot be both right and wrong, so we need a different way of categorizing these actions along the relevant dimension. Because both actions in dilemma cases are morally wrong, it should turn out that failing to conform with one's duty suffices for making an action wrong, while conforming with one's duty does not suffice to make an action right.

> **Morally Wrong:** An action is morally wrong IFF the action fails to conform
> with an operative duty.

Even if an agent shoulders her most pressing obligations (whatever those amount to), failing to conform with a countervailing conflicting duty is sufficient to render her actions wrong, including in dilemma cases. Although rightness is impossible in dilemma cases, an action can conform with at least one of the conflicting duties in such cases. I call such actions dutiful.

> **Dutiful:** An action is dutiful IFF the action conforms with an operative duty.

I suggest actions only ever render as right when the action conforms with all of the agent's duties at the time and does not fail to conform with a countervailing conflicting duty. Alternatively, we might say the following.

**Morally Right:** An action is morally right IFF the action conforms with an

operative duty in a non-dilemma case.

A non-dilemma case is a case where the agent is in a position to conform with at least one

operative duty and the action is non-wrong. As a point of clarification, to say that an action is

non-wrong, on my account, is to offer a description and not an evaluation of an action on my

account, since non-wrongness cannot refer to any existing assessment properties (i.e. there is

no duty appeal to in the assessment).

I use the latter formulation of rightness in my characterization of rightness because

rightness does not involve an assessment of anything other than the action's conformance with

a single operative duty, and the initial formulation makes rightness seem like a second-order

duty in the same way erogation assessments might. All morally right actions are dutiful, but

not all dutiful actions are morally right. Dutiful actions are morally wrong in dilemma cases.

## 3. Absolute versus Pro Tanto Duties

In the previous section, I proposed that first-order act-assessments appeal to at least two plausible sets of morally significant considerations. I defined the conceptual niche used to distinguish the obligatory from the erogatory by distinguishing between what we have moral duty to do and moral reason to do. Over this and the next sections, I characterize the notion of duty utilized throughout this dissertation. Recall the dilemma trilemma, which arises from the following three independently plausible, yet *prima facie* jointly inconsistent, claims.

(1)     The two duties binding Orestes conflict

(2)     The two duties binding Orestes are absolute

(3)     Orestes can resolve his dilemma

For my purposes, a dilemma is resolvable when the agent is morally permitted to perform at least one of the actions. The inconsistency could be resolved by giving up any one of the three claims. However, I believe rejecting any of the three claims would leave us with unsatisfactory assessments of actions in this dilemma case. We cannot reject (1) because we are attempting to explain the possibility of a genuine moral dilemma and moral dilemmas are only possible when duties conflict, so we would need to reject either (2) or (3).[43] We cannot reject (3) because doing so would give us the result that the dilemma is unresolvable, but we want to allow the possibility that agents like Orestes are sometimes permitted to fail to conform with a duty in dilemma cases. This leaves us with (2) only.

---

[43] Of course, philosophers like Kant might reject (1), since the categorical imperative issues only one duty. As might a utilitarian who holds that we have only one moral duty – to maximize utility.

So, suppose we were to give up (2), that the duties binding ORESTES are absolute. By rejecting (2), one could accommodate the intuition that we sometimes permissibly fail to conform with our duties. Rather than holding the view that duties are always absolute, one might suggest instead that some duties are *pro tanto*. According this view, duties admit of exceptions. Exceptions to duties arise when conflicting duties provide conflicting reasons for action, where these reasons outweigh the reasons for conforming with the initial duties. Thus, the conflict about which duty one ought to conform to is resolvable within rational moral thinking and nonconformance with one of one's two *pro tanto* duties is not morally wrong, even though nonconformance with that duty alone is ordinarily morally wrong.

However, the idea that all duties are absolute has long held appeal. Intuitively, nonconformance with one's duty is morally wrong because ordinarily prohibited actions are the sorts of actions that manifest a criticizeable moral failure – whatever that may be for a given theory.[44] A moral absolutist holds that ordinarily wrong actions possess some property P, which is precisely the feature of the action that makes it wrong. That the action possesses P grounds the duty to not act so as to instantiate P in acting. Individual duties to do or not do something, on this account, are therefore derivative of the action's actually possessing P and serve to delimit the moral boundaries for an action.

Consider a classic example called DROWNING, where a child is drowning in a pond. I am the only one around to save her. If I ignore the child in DROWNING, then, presumably, I have done something wrong. I have done something wrong not only because I failed to conform with my duty. I have a duty to save the drowning child because otherwise, in allowing

---

[44] Because we are discussing action assessments, we ordinarily think the action and not the agent is the source of the moral failure and the object of criticism.

the child to drown, I will manifest some moral failure, such as violating one of her rights or failing to treat her as an end-in-herself.[45] (I leave it to my readers to decide what precisely makes ignoring the child in DROWNING morally wrong.) I have a duty in such cases, with which nonconformance is morally wrong, because of the particular type of moral failure manifested by my inaction – the instantiation of the wrong-making property (whatever it is). And, plausibly, the death of the child becomes no less of a failure when the duty to prevent the child's death enters into conflict with some other, apparently more pressing duty. We have a name for that failure – wrongness. The motivation to claim that duties are sometimes absolute, supporting (1), stems from the view that acting so as to manifest certain kinds of criticizeable moral failures is what makes those actions wrong under all circumstances.

Note that my characterization of moral absolutism is nonstandard. Typically, philosophers characterize absolutist theories as prohibiting actions because the action is an instantiation of a particular act type, such as committing murder.[46] According to my characterization, it may very well turn out that certain act types always possess P. In which case, performing some act types is always morally wrong. Even so, I suggest many absolutists hold that regardless of the description under which we classify the action, the action is wrong if and only if it possesses P. Depending on the details of the particular moral theory then, it may yet turn out that some token actions of ordinarily prohibited act types are not wrong because they lack P, while some token actions of ordinarily permitted act types are wrong

---

[45] The action is wrong because of something about the action. The failure that would be manifested by the action is a feature of the action. The action fails to conform with duty to prevent the child's death. The action morally fails, presumably, because of the impact of the action on the child – namely, the child's death.

[46] As with Anscombe, perhaps, who states that intending some act types is unconditionally wrong. See G.E.M. Anscombe's (1958) "Modern Moral Philosophy" (*Philosophy*, 33.124).

because they possess P. This helps to explain why lying is so morally complicated, since lying sometimes instantiates P and sometimes does not.

This general characterization of moral absolutism, as a prohibition on certain features of actions rather than act types, better accommodates a wide variety of non-consequentialist absolutist views. For example, this account better captures the correct characterization of Kantian-style absolutism. While Kant might think lying is always morally wrong, the explanation for why Kant might think lying is wrong stems from the view that all lying is grounded in a rationally unwillable maxim or a maxim to which no rational agent can consent. This is the feature of the action that makes it wrong, that it derives from an unwillable maxim, and the source of the duty prohibiting lying. Kant surely does not think that lying is wrong because all actions sharing the description of lying ought to be prohibited, even if he thinks that all actions sharing the description of lying will inevitably be prohibited. If I were to identify a rationally willable maxim involving lying, then Kant would have to concede that some lying is non-wrong. For Kant, this would not be forswearing his moral absolutism, since the correctness of his view relies not on the prohibition of certain act types. In fact, many philosophers argue Kant does not think all lying is wrong, yet Kant is nonetheless a moral absolutist about the wrongness of acting on rationally unwillable maxims.[47]

*Pro tanto* duties are not like this. Instead, proponents of *pro tanto* duties must hold that what makes actions wrong all-things-considered is a kind of relation. In general,

---

[47] For more on the complexities of lying see, for example, J. Saul (2011) "Just Go Ahead and Lie" *(Analysis*, 72.1). Regarding Kant on lying, see I. Kant's (1799) "On a Supposed Right to Lie on Altruistic Motives" (*Ethical Philosophy*, (trans. J.W. Ellington, Hackett Publishing, 1994) and, for example, H. Varden's (2010) "Kant and Lying to the Murderer at the Door... One More Time: Kant's Legal Philosophy and Lies to Murderers and Nazis" (*Journal of Social Philosophy*, 41.4).

proponents of *pro tanto* duties hold that what we have a decisive moral duty to do depends on what we have the most and mightiest moral reason to do. A *pro tanto* wrong action is an action with wrong-making features, whatever they may be, ordinarily sufficient for making the action wrong. This means the action ordinarily has, on-balance, more and mightier moral reasons against performing the action than for. [48] Yet, when two *pro tanto* duties enter into conflict, the balance of moral reasons might sometimes favor that same ordinarily prohibited action. In such cases, the action is no longer wrong, since wrongness has to do with what we have moral reasons to do and not just what we have *pro tanto* duties to do. Instead the action is right, since rightness and wrongness are a relation the action has to the relevant moral reasons.[49] Alternatively, the action is a non-wrong permissible action. If so, we can accommodate the view that Orestes faces a resolvable conflict between two duties by rejecting (2) and supposing instead that the duties Orestes faces are *pro tanto* duties.[50]

---

[48] Individual duties to do or not do something, on this account, must therefore delimit the moral boundaries of actions that we ordinarily have decisive moral reason to perform or not perform. This is consistent with the view that duties just are certain kinds of protected reasons, which are nonetheless overridable. See J. Raz's (1975) *Practical Reason and Norms* (Oxford University Press), esp. Chapter 2 and J. Gardner's (2004) "The Wrong-Doing That Gets Results" (*Philosophical Perspectives*, 18.1, Ethics).

[49] On this account, it is possible for us to judge that ordinarily wrong actions are sometimes morally right. Yet, it seems theoretically problematic to suppose the very same action could buoy two contradictory moral values without any mechanism for morally assessing the value of the two actions in the same way. This is exactly the sort of puzzle that motivated concerns about the DDE, in Chapter 1. When two agents perform the same action, a moral theory should have a mechanism for assessing the two actions in the same way, even if they are assessed differently on some other measure.

[50] Supposing all duties are *pro tanto* comes with problems, even for proponents of *pro tanto* duties themselves. For example, some proponents of *pro tanto* duties hold rights-based moral theories. If no rights are absolute rights with concomitant absolute duties, and if the wrongness of nonconformance with one's rights-based duty is a relation that the action has to what the agent has moral reason to do or not do, then nothing is intrinsically wrong with nonconformance with one's rights-based duty. Put more bluntly, violating someone's right is not intrinsically wrong according to this account. Violating someone's right is wrong only insofar as we have decisive moral reason to not violate that person's right. When someone has

While *pro tanto* duties can explain Orestes' dilemma, proponents of *pro tanto* duties cannot deliver symmetrical assessments about the value of the agent's action in the two DDE cases and therefore cannot accommodate the view that two actions that are structurally the same should be judged in the same way. When a *pro tanto* duty is defeated, the action is shown to be non-wrong as stated above, since the action has the correct relationship to the relevant moral reasons. So, proponents of *pro tanto* duties are forced to say that while the Terror Bomber performs a morally wrong action, the Tactical Bomber performs a non-wrong morally permissible action or morally right action – making only asymmetrical act-assessments about the two DDE cases.[51]

---

a right that we permissibly fail to conform with, *pro tanto* theorists, like Thomson, call that an infringement on the right. Rights are infringed rather than violated when nonconformance with the right's concomitant duty is non-wrong. Yet, one might plausibly hold that failing to uphold a right is intrinsically wrong, even when doing so is a necessary evil. We might think of Thomson's language of infringement as an attempt to hold onto the ugliness of failing to uphold a right despite any apparent permissibility of doing so, thereby appeasing absolutist intuitions. However, if the moral value of an action – an action's wrongness – is this aforementioned relation, then proponents of *pro tanto* duties are not in a position to accommodate absolutist intuitions in such cases. The relevant object of unpropitious moral assessment for the proponent of *pro tanto* duties is not present in cases of rights infringement, since the action has the correct relationship to the relevant moral reasons for performing that action.

[51] Of course, proponents of pro tanto duties may not agree that DDE cases are structurally the same and that they need not, therefore, accommodate the view that we should have any mechanism for assessing the cases symmetrically. Such a view cannot accommodate my aims for this project.

## 4. Objectivism about Negative Duties

In the previous section, I contrasted my preferred notion of absolute duties with *pro tanto* duties while explaining how one might go about resolving the dilemma trilemma. I hold that moral absolutism is the only account of duty consistent with (1) – (5), even if proponents of *pro tanto* duties can make sense of moral dilemmas. In this section, I motivate the need for endorsing an account of negative duties, which is especially important to my account of moral worth in Chapter 3.

My account of first-order act-assessments aims at consistency with moral objectivism – an account according to which rightness and wrongness are assessed only by appeal to the morally significant agent-independent features of action. The agent-independent features of actions that make them right or wrong are only ever instantiated after the action has been performed (or not performed, as the case may be). Genuine moral objectivism about duties is therefore best suited for a retrospective stance where duties correspond to requirements for or prohibitions against doing the thing that was actually done. On a genuine account of moral objectivism, act-assessments are only possible *post hoc* because the actual agent independent features of action determine the action's rightness or wrongness, even if these features of the action were unforeseen and impossible for the agent to use in her practical deliberations.

As a result, it could turn out that an agent acts wrongly or impermissibly even if she perfectly deliberated about what to do, given the available information. For example, suppose Orestes was unaware that he had any duty to avenge his father by killing his mother. Suppose he was only aware of the duty prohibiting matricide and he chose to conform with this duty. Orestes would have failed to kill his mother and avenge his father, and his father's Eumenides would have chased him to the ends of the earth. Because the action was not the morally

preferred solution to the dilemma, the gods likely would not have cleansed his miasma. Orestes' would have tragically lingered in suffering because of his ignorance about the second duty, despite his best efforts at deliberating about what to do.

Some philosophers take this to be an undesirable result because one might think agents should always be able to reason their way out of performing a morally wrong action. Of course, I aim to accommodate genuine moral dilemmas where the agent will perform a morally wrong action come what may. This sort of result is therefore not particularly objectionable to me. Moreover, I think we should not always be able to predict whether things will go well for us in the moral domain. We are fallible. Other agents and events can interrupt and affect our plans. I think we often hold our breath and hope we land in the right place. Even so, I am sympathetic to the concern that moral objectivism runs the risk of seeming needlessly harsh on the poor moral deliberator.

The most significant problem with moral objectivism has nothing to do with moral objectivism *per se*. As I indicated toward the end of Chapter 1, philosophical views tend to come with clusters of commitments, and one commitment common to most accounts of moral objectivism is what Zimmerman calls The Objective View, an account, most commonly associated with positive duties, according to which an agent ought to perform an action if and only if the action is in fact the morally best option available.[52] Moral objectivism, when coupled with The Objective View, has difficulties handling cases where the agent is ignorant of which action is in fact the morally best option available. It asks the agent to act on the morally best option available even when the agent is ignorant of which action is the morally best option available. Yet, this requirement seems problematic from the deliberative

---

[52] See Zimmerman, op. cit., especially Chapter 1 (pg. 2).

standpoint, since it sometimes seems like agents should choose the morally best option that the agent can be reasonably expected to know about (the prospectively best option).

He illustrates this point using Jackson-style cases which generally have the following form. Agent X has a minor illness, which agent Y aims to treat. Agent Y has four treatment options available. Option A is in fact the morally best option available (e.g. it will completely cure the minor illness). Option B is a morally adequate option (e.g. it will partially treat the illness). Option C is a morally unacceptable option (e.g. it will kill agent X). Option D is morally undesirable (e.g. it will render the illness untreatable). Unfortunately, because of some accident (e.g. missing labels on the medication bottles), agent Y cannot distinguish between options A and C. She knows that one is the morally best option available while the other is morally unacceptable, but she does not know which is which.

Zimmerman contends that the stakes are too high for agent Y to treat agent X with either of options A or C because the impact on agent X is morally unacceptable if Y gets things wrong. So, it seems like agent Y ought to treat agent X with option B. Option B is not the morally best option, nor is it the worst; option B is the best of the available options that she knows about. However, moral objectivism, when coupled with The Objective View, would probably indicate that agent Y ought to administer option A regardless of whether she can distinguish between options A and C. If correct, then it seems like moral objectivism has a problem in that it requires agents to take significant moral risks when we have otherwise morally acceptable options available.[53]

My aim here is not to argue about whether Zimmerman has the right view of things. Instead, I aim to argue that the moral objectivist need not take on the baggage of The Objective

---

[53] Ibid., Chapter 2.

View. An account of negative duties can circumvent Zimmerman's worry. If duty only prohibits the morally unacceptable option C, then administering any of options A, B, and D is morally permissible (assuming option D is not also prohibited and is instead simply suberogatory). This result is consistent with Zimmerman's claim that the agent ought only to administer option B.

Now, I recognize that this solution is dissatisfying for my typical reader. An account of negative duties does not seem to indicate which of the three available options the agent should administer all-things-considered, even if she's not morally obligated to administer any specific option. Duties, on my account, are not action guiding in the sense that they do not typically guide an agent toward performing a particular token action.[54] This is the role of normative moral reasons, the realm of the erogatory and not the realm of the obligatory – what she has moral reason to do and not moral duty to do. Because duty points her away from option C, the relevant normative reasons would probably also point her away from option A when pursuing option A poses a significant risk of ending up with option C. She would also be pointed away from option D, since, between the two, she has more and mightier moral reason to administer option B. So, if Y deliberates correctly about what to do, then she would straightforwardly choose to administer option B.

In any case, an account of negative duties can help make sense of how a genuine moral objectivist might hold that an agent can permissibly perform the prospectively best action available when she is ignorant of the morally best option available. An account of negative duties is useful for other reasons as well. Specifically, I believe it rules out the existence of

---

[54] For a discussion on action guidance in objective versus prospective accounts of rightness, see H.E. Mason's (2013) "Objectivism and Prospectivism About Rightness" *(Journal of Ethics and Social Philosophy*, 7.2).

derivative duties to perform any token act, so it will turn out that an agent permissibly performs any action conforming with duty (and not also failing to conform with duty). And I believe this can help make sense of the permissibility of trying and failing, which I will address in the next section.

However, the real purpose of what follows is to highlight the kind of moral framework an account of negative duties is consistent with. The UMF, as presented in this dissertation, must be what I call an adequacy account or an account of "good enough" morality. We can already see this with the above conclusions. When presented with a range of options, the agent is not required to administer the morally best option available, but she is prohibited from administering the morally worst option. All of the other options are good enough, even if one option (e.g. option B) is morally preferred given the information the agent has available. Administering the morally best option available or the prospectively best option available satisfies what I call requirements of virtue, which pick out what is virtuous about some act and tell us to pursue that action in order to promote the good – to be virtuous. I do not discuss requirements of virtue much in this dissertation because they are not necessary for making sense of (1) – (5). Instead I focus on minimal moral adequacy requirements, which are in part comprised of negative duties. I propose that the permissibility of trying and failing is another way in which an account of negative duties can help us make sense of minimal moral adequacy in acting, as does an account of suberogation, which I will also discuss in the next section.

## 5. Minimal Moral Adequacy (a.k.a. Good Enough Morality)

I previously suggested that a moral absolutist holds that ordinarily wrong actions possess some property P, which is precisely the feature of the action that makes it wrong. On an account of negative absolute duties, agents are prohibited, without exception, from acting so as to instantiate P. Any action not instantiating P is either consistent with or conforms with duty and is therefore permissible, and sometimes we have more than one permissible way to conform with duty, such as when choosing between throwing a drowning child a life preserver versus swimming out to rescue the child. Duties therefore cannot specify the performance of any particular token action. An agent's token action is a mere means of conforming with duty.[55] I will sometimes, for ease of locution, carelessly speak of the duty to perform some token action, but I do not suppose any such duty exists. So, for example, Medea has no duty to pull the lever (or to save the five).[56]

Because an action's wrongness only ever renders after the action is actually performed and because whatever duty operating on the action prohibits the doing of the thing that was actually done, we should take this to mean that duty regulates the instantiation of certain substantive wrong-making features that are relativized to the actual action and that the prohibition should be read in the *de re* sense (even if it turns out that acting so as to instantiate the relevant wrong-making feature is universally prohibited). We can reasonably speak of an

---

[55] As a result, there can be many right acts. I do not typically talk about *the* right act or the act that one ought to perform because there are often many. I recognize this is contrary to the standard way of talking about right action.

[56] Strictly speaking, I am not certain Medea faces a conflict between two absolute duties. Instead, Medea faces a conflict in choosing between two actions, either of which simultaneously conforms and fails to conform with the duty prohibiting whatever makes her contemplated action wrong. The outcome is the same, but the structure of a dilemma is a probably bit different on my account. However, throughout this dissertation I will continue to characterize moral dilemmas as a conflict between two separate duties.

agent's nonconformance with duty in one respect and conformance with duty in another respect, since the prohibition is a prohibition on acting so as to instantiate the specific substantive features of the contemplated actions.

An action renders morally right when conforming with an operative duty by not instantiating P in a non-dilemma case. Performing a morally right action is morally required when doing so is the only permissible means of conforming with a single operative duty. It could turn out that the only permissible means of conforming with duty is to act so as to instantiate another property – G, which is a morally desirable or good-making property. Thus, acting so as to bring about G is required. Even so, the action is not morally right because of its G-ness. Any goods the agent brought about in acting are merely a part of the means of satisfying the duty and not the features of the action that make it morally right. Instead, the action renders morally right because it conforms with an operative duty while failing to instantiate P (i.e. the action does not also fail to conform with duty).[57]

These observations have several theoretical implications for the relationship between duty and what duty can require of us. For example, my reader might note that if the doing-allowing distinction is relevant on an account of negative duties, it does not track the distinction between positive and negative duties. As Foot initially framed the issue, the typical prohibition against *doing* harm tracks a negative duty and the typical prohibition against *allowing* harm tracks a positive duty.[58] She characterizes the former as duties to refrain from

_____

[57] In Chapter 3, I will suggest that if we want to assess the praiseworthiness of an agent for performing a morally right action, we should not appeal to any desire (or whatever) she might have to bring about G, since the action's G-ness has nothing to do with the action's rightness. As a result, I tend to think that wrongness is related to the substantive features of an action in a way that rightness is not.

[58] Foot, Phillippa. (1977). "The Problem of Abortion and the Doctrine of Double Effect." *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford University Press).

acting in certain ways, and she characterizes the latter as duties to act in certain ways. In other words, negative duties tell us what we ought not do, and positive duties tell us what we ought to do. We have negative duties to refrain from killing people and positive duties to help people in need, so we satisfy a positive duty when we confer a preventive benefit as a result of some action we take.[59] The prohibitions against allowing harm are paired with positive duties because the positive duty to provide aid is taken to be the flipside of the prohibition against allowing harm. If we are prohibited from allowing a child to drown in a pond, then we are thereby saddled with a duty to save the person from drowning.

First off, conforming with the duty prohibiting me from allowing a child to drown in a pond does not always involve demonstrating positive agency to act so as to save the drowning child. An agent employs positive agency, generally speaking, when the agent's most direct contribution to an outcome is an action.[60] Consider an alternative version of DROWNING. If I am traveling with a friend, and we both spot the drowning person, then we both have a duty prohibiting us from allowing the person to die. Suppose however that there is only one life ring, which can only be thrown by one person, and my friend reaches the life ring first. If I attempt to help, then I will interfere, and the person will drown. I can best conform with the duty prohibiting me from allowing the person to drown by not interfering. I can best conform with the duty prohibiting me from allowing the person to drown by not acting at all. So, we need not hold that the duties prohibiting allowings of the instantiation of a wrong-making property corresponds to a positive duty to do anything at all.

---

[59] A preventive benefit is given to prevent or alleviate harm. For more on preventive benefits see S. Shiffrin's (1999) "Wrongful Life, Procreative Responsibility, and the Significance of Harm" (*Legal Theory*, 5.2).

[60] Quinn, op. cit.

Yet, even on an account of purely negative duties, we could still be required to achieve dutifulness by taking a preventive action, which would involve taking an action aiming at conferring a preventive benefit.[61] Moreover, we are sometimes required to act so as to actually confer the preventative benefit. As a result, we would not need positive duties to make sense of why someone would be required to do something (and succeed) to prevent the instantiation of the wrong-making property. For example, I am sometimes required to save the drowning child although I have no duty to do so – when saving the child is the only permissible means of conforming with the duty prohibiting harm to the child (assuming, for the sake of discussion, that the harm to the child is what makes the action wrong). For this, I imagine a Singer-style example where I can easily save the child and the only thing standing in the way of successfully saving the child is the worry that I will get my clothes dirty. In such cases, trying and failing to save the child would probably not suffice to achieve dutifulness. Typically, philosophers would explain this in one of two ways. First, one might suggest that the contents of our duties include a success condition, according to which the agent is required to prevent the instantiation of the wrong-making property. For example, I might have a straightforward duty to not instantiate P in acting. In which case, if I tried and failed to save the child, then I instantiate P and fail to conform with my duty thereby. Second, one might suggest that the contents of our duties include an attempt condition, according to which the agent is required to attempt or try to prevent the instantiation of the wrong-making property. Afterall, one way of trying to do something is to succeed at doing something. I might have a duty to try to not instantiate P in acting. In which case, if I failed to save the child in my

---

[61] This idea of preventive action is follows Shiffrin's use of preventive benefits.

apparent attempt, then one might suggest I never really tried at all given the relative ease of success.

I think neither is a fruitful analysis of the content of our duties, though my reader need not follow me on this point to appreciate my overall framework. Here, I am only highlighting possible merits of an account of negative duties (and my idiosyncratic views on them). However, I will use this way of talking about things to make some progress on the sort of view I do endorse – though, I will not articulate that full view here, and the view I do articulate is not one I fully endorse.

I propose that one can endorse a view with something like a success condition that nonetheless makes room for failed attempts. To make sense of this, we must be careful to distinguish between the different possible objects of these conditions. In particular, we should distinguish between the wrong-making feature of an action and the action itself. As formulated above, the success condition operates on the wrong-making feature of the action. If the child's death is what makes the action wrong and the child dies, then the action fails to conform with duty because the feature of the action that makes it wrong is instantiated. Alternatively, the condition can operate on the action itself, but the remaining content of the duty matters.

My current preferred formulation of a negative duty is "do not act in such a way that the action itself ensures the instantiation of P". This formulation of the duty is meant to accommodate views according to which P is a consequence of the action and views according to which P is instantiated in the very act. I am open to other formulations where the gist of what follows still applies. This duty only ever arises when an action instantiating P is available and imminent. The action might be imminent because the agent is considering such an action, because the agent has no other alternatives, or because the agent is at risk of doing so without

76

her knowledge. An agent does not fail to conform with this duty by intending to instantiate P in acting if the action is not the sort that ever instantiates P. I should probably also add the caveat that this duty applies to the action at the point of initiation, since agent-independent factors can impact our actions and their consequences as we execute them. On my understanding of this duty, the scope of the success condition operates on the action itself and not the instantiation of P. To assess whether the act succeeds, we must assess whether the objective facts are such that the action itself ensured the instantiation of P and not whether P was instantiated.

Returning to my thoughts on this issue, an agent who attempts, but with little effort, to save the drowning child in a singer-style example will almost certainly find himself in a scenario where he ensures the child's death and fails to conform with duty thereby. The action ensures the child's death because the facts are such that, upon initiating the action, it was one of relatively few actions, among available actions, with the consequence of the child dying. In this case, the agent is therefore required to succeed in preventing P but only because the success condition operates on his not acting in any of the ways that fail to prevent P when more promising actions (or action implementations), ones that achieve the aim of the duty, are available.

However, sometimes the agent is not required to succeed in preventing P, even when only a single required action is available. So, sometimes trying, but ultimately failing, to prevent P will conform with duty. This scenario is only likely when the agent attempts one of the preventive actions in such a way that the objective facts reveal that the action was a part of the process of preventing P and that the failure to prevent P is attributable to the conjunct of the following: (a) the facts are such that, upon initiating the action, it was one of relatively

many actions, among available actions, that would inevitably instantiate P and, for those relatively few actions that would prevent P, (b) preventing P would require a level of motor control or skill (or whatever relevant capability) that is ordinarily achievable but only with great difficulty for the agent. When the available options are (a) – limited – and (b) – difficult to achieve, one might reasonably hold that doing one's damnedest counts as acting so as to not ensure the instantiation of P, even if P was instantiated.

This is really just to say that the initiated action was an appropriate means of conforming with duty but that the action was also reasonably likely to fail to achieve the aims of the duty – preventing P. Only the former and not the latter is regulable by duty, since duty cannot regulate facts external to the action itself. That is, duty cannot regulate how likely the action is to prevent P, rather duty might instead regulate action's progress towards successful prevention in such unfortunate circumstances. In this sort of case, it will not be true that the agent should not have done what they did or that the agent should have done something else, since the action was the only one or was only one of the relatively few that could achieve the aim of duty at the point of initiation. If, in the execution of the action, P is instantiated, P instantiates despite the action and not because of it and is probably not properly attributed as a property of the action.

One might think that this view really just amounts to a kind of attempt condition. But I think we should conceptualize the view as more of a success-unless condition. The action is required to successfully prevent P unless achieving success is at the farthest margins of our locus-of-control over the situation – the threshold after which duties cease to operate on us at all. When no preventive actions are available to the agent at all, the duty no longer operates on the agent, since P cannot be prevented. When the only preventive actions available fail to

conform with some other duty, then the agent faces a dilemma and the action will be wrong regardless, and it might still turn out that the agent must nonetheless do her damnedest in now tragic circumstances.

To make this more concrete, keep in mind that many morally significant scenarios call for a sequence of actions and not the performance of a single action, such as swimming in tumultuous waters to save a drowning person where swimming is the only way to save the person. In these sorts of circumstances, successfully saving the drowning person requires relatively high swimming skills and high precision in act-execution over the course of the entire sequence (i.e. one false move will lead to failure). Failure to attempt the swim would lead to a failure to conform with duty, since the facts were such that, at the time of initiating the saving sequence, the drowning person could be saved, no one else could intervene, and the swimmer could not save himself. So, the only way for the agent to conform with duty is by initiating the swim, thereby taking the appropriate means of conforming with duty.

However, due to external circumstances, relating to, for example, the swell, the riptide, or the emergence of sharks, the saving sequence suffers from a number of act-execution failures. We can suppose many acts in the sequence were successfully executed, and that the failed actions are still actions objectively speaking, properly attributable to an agent, but not actions as intended. And, as a result of these act-execution failures, the saving sequence results in the victim drowning. Now, one might think the agent probably has an excuse for failing to achieve the aim of the duty, since her saving sequence was thwarted by an angry ocean, but my question is whether the act failed to conform with duty by failing to achieve the aim of duty. If, at each point, the action was the appropriate means of conforming with duty at act-initiation while nonetheless suffering from act-execution failure attributable to external

79

factors, the agent was at each point doing what duty asked – to not act so as to ensure the instantiation of P in acting. The agent satisfied the success condition on the action.

Again, we are considering a success condition and not an attempt condition on the action. Consider what it would mean to analyse the action in terms of an attempt condition. We can understand an attempt in several ways. One might attempt to choose the correct action but get things wrong. One might attempt to initiate the action before discovering oneself incapable of movement. One might initiate the action in an attempt to execute but find one's action aborted by, for example, becoming stunned. And one might initiate an action in an attempt to execute in a particular way and find oneself executing in a different way. This could be an attempt in the sense that the agent attempted to do one thing and ended up doing something a little different. The first three forms of attempts would not count as conforming with duty. The first is a failure, the duty probably does not apply in the second case, and the agent has an excuse in the third case. Only the fourth form of attempt is useful for our purposes. The swimmer finds herself in this situation.

However, the fourth form of attempt is not really an attempt at all. The agent successfully acts, even if things did not go as planned. The duty regulates the action itself, as executed. What the agent planned is irrelevant to the assessment of the action. The agent-independent facts settle the question of how to assess it, and the question is simply whether the action was an appropriate means of conforming with duty. The agent could not refrain from acting, since she would fail to conform with duty. So, the agent had to do something. And the something she did made progress towards achieving the aim of duty – presumably the swimmer was actually progressing towards the drowning person rather than treading water the entire time. And we should not think that she would fail come what may because duties

cease to operate on us at that point. So, if she was bound by a single operative duty, and she performed the action making progress toward achieving the aim of duty, then she conformed with duty. But please note that this only applies in cases where preventing P is unlikely. When the agent has more promising options, even if she does not know about them, it will no longer be true that trying and failing to prevent P will conform with duty because the facts will be such that the action actually preventing P is the one that makes progress towards preventing P. Whereas the attempt will only make progress away from actually preventing P, via the other action, when it fails.

Negative duties thus help us to make sense of how an agent sometimes adequately performs from the moral standpoint despite any apparent flaws in her action, such as a person's drowning as a consequence. The agent's failure is still a moral failure because she did not manage to prevent the morally undesirable feature of the action, but she did not fail to conform with duty. Her action was good enough. However, it might fall somewhere in the realm of the suberogatory (or related assessments). So, minimal moral adequacy considerations extend to the realm of other first-order act-assessments as well. Before characterizing first-order agent-assessments, I will briefly explain how the UMF delivers judgments about the suberogatory.

I like to think most of our everyday actions are not regulated by the realm of the obligatory. That is – most of our everyday actions are neither prohibited by some duty nor required by some duty. Even so, moral considerations continue to delimit and guide our actions. For example, if my niece and nephew both ask me for the last cookie in the cookie jar, I might have no duty to give the cookie to one rather than the other, but considerations of fairness might prompt me to give the cookie to my niece rather than my nephew, if he has

already eaten one while she none. The realm of the erogatory is thus inveterate to our daily lives.

In such a situation, I am vulnerable to the possibility of committing a criticizeable moral failure even if I am not bound by any duties. If I give the cookie to my nephew rather than my niece, then I fail to appropriately respond to considerations of fairness. While I do nothing wrong, I nonetheless do something I have decisive moral reason not to do, and a critical onlooker would be entitled to make a negative moral assessment of my action. The judgment that my action is suberogatory – that my action is a non-prohibited bad action – is just this sort of unpropitious moral assessment.

**Suberogation:** An action is merely suberogatory IFF the action is non-wrong
and has, on balance, more and mightier moral reasons against than for.

Suberogation is, thus, an unpropitious first-order act-assessment, which is evaluated by appeal to the moral reasons regulating the action. Yet, it might turn out that we sometimes have all-things-considered reason to perform suberogatory actions (due to prudential concerns, for example). Furthermore, suberogatory actions might sometimes turn out to be the morally preferred action when no better alternatives exist. Now consider first-order agent-assessment.

## 6. Motivating Reasons

Two agents can perform two actions with the same structure and the same morally significant features from two very different underlying motivations (hereafter – referred to as performing the same action from different motives). This phenomenon sometimes leads to difficulties in morally assessing agents and their actions, as evidenced in the cases of the Tactical Bomber and the Terror Bomber. Although they perform the same action, they are motivated differently, and, somewhat controversially, we assess their actions differently as a result. A great deal of philosophical disagreement hangs on what we want to say about these and other cases typically handled by the DDE. So, regardless of where one stands on the DDE, we should agree that any good moral framework should help us make sense of the moral significance of an agent's motivations in acting. In this section, I propose a second, reasonably well-known, pair of plausible distinctions between types of first-order moral assessments, which make up the UMF. I refer to the distinction between motivating reasons *de re* and *de dicto*.

Again, consider the bomber cases from Chapter 1. Plausibly, the Tactical Bomber and Terror Bomber agree that bombing the munitions factory is the right thing to do. Moreover, bombing the munitions factory probably is the all-things-considered right thing to do because bombing the munitions factory will end the war early thereby saving lives on both sides. Further, we might think neither the Tactical Bomber nor the Terror Bomber would bomb the munitions factory were it not the right thing to do. Plausibly then, both the Tactical Bomber and Terror Bomber bomb the munitions factory because doing so is right.

In one respect, we might therefore think the Tactical Bomber and Terror Bomber act from the same motivation – the motivation to do the right thing because it is right. They are motivated by the action's rightness *de dicto*. However, we also think the differences in their

84

motivations distinguish them. The Terror Bomber desires to kill the nearby civilians or judges that killing civilians would be a good thing (or whatever), while the Tactical Bomber does not.

As I have characterized DDE cases, they involve genuine moral dilemmas. Plausibly, the Tactical Bomber and Terror Bomber face simultaneous conflicting duties to save those endangered by the war and prohibiting the killing of innocent civilians (or something of the sort). Yet, the Terror Bomber acts with the motive of killing the civilians. The Terror Bomber, and not the Tactical Bomber, is motivated to act by the actual morally significant features of the action that count precisely against acting – he is motivated by the action's wrongness *de re*.

We appear to make different second-order act-assessments about the two bombers' actions because of this difference in motivation. Thus, we can potentially assess two distinct features of an agent's motivations – her motivating reasons *de dicto* and her motivating reasons *de re*. I propose either can feature in our first-order agent-assessments, which are moral assessments of agents evaluated only by appeal to the morally relevant features of the agent herself.

**Figure 5: Motivating Reasons *de re* & *de dicto* in the UMF**

|  | **Object of Agent-Assessment** |
| --- | --- |
|  | Motivating Reason *de re* (2a) |
| **First-Order** | Motivating Reason *de dicto* (2b) |

Figure 5 depicts the distinction between motivating reasons *de re* (Box 2a) and motivating reasons *de dicto* (Box 2b) categorized as the assessment properties of agent motivations.

I propose that an agent's motivating reasons *de re* and her motivating reasons *de dicto* are central to two distinct kinds of first-order agent-assessments. The former are judgments about the agent's sensitivity or responsiveness to the actual morally significant features of an action. The latter are assessments of the agent's moral deliberations on the morally significant features of an action. I will not say much more on this because these assessments are not all that important to developing the aspects of my framework under consideration in this dissertation. But I should note that philosophers typically use the same terms to pick out assessments about each type of motivation and argue about which set of motivations should be privileged in our moral assessments. The contentious terms appeal to the agent's quality of will. In this category of assessment, we find good will, ill will, and moral indifference. I leave it to my reader to decide which assessment applies to which category of motivation just as long as my reader grants that the innominate category of motivation is independently morally assessible.

While the above distinction between motivating reasons *de re* and *de dicto* is generally acknowledged, ongoing debates in contemporary ethics focus on the question of which motivations are central to our second-order moral assessments about an agent's moral worth.[62] I discuss this question in detail in Chapter 3, but the answer ultimately hinges on facts about the action for which we are assessing the agent. The same is true in making second-order act-assessments about an action's permissibility. Neither motivation type is more central to ethical

---

[62] See, for example, Arpaly (op. cit. Chapter 3) and Z. Johnson King's (2019). "Praiseworthy Motivations" (*Nous*, 54.2).

theory or privileged in our moral assessments in general, but different motivations matter more in making different types of moral assessments in different situations.

Although motivating reasons receive less attention in this chapter about first-order moral assessments than normative reasons and duties, they are no less important to my view, and much of my discussion on second-order moral assessments makes appeal to them. Because my views about motivating reasons are a little more standard, they require less development to show how they fit into my overall framework. However, before proceeding into Chapter 3, I should say a bit more about motivating reasons generally.

So far, I have been discussing motivating reasons *de re* and *de dicto* primarily in the context of the actions they motivate. I framed wrong-making reasons *de re* as the sort of motivations underlying the Terror Bomber's wrong action. We can only make sense of the agent's *de re* motivations in the context of the action, since the agent only acts on de re motivations when the reasons for which the agent acted link up to a feature manifested in the action. But recall that I made the distinction between first- and second-order moral assessments to highlight the ways in which we can assess an agent's motivations without regard for the action they motivate. The distinction between *de re* and *de dicto* motivations is the distinction I require for explaining the role of an agent's motivations in our moral worth assessments, but agents can be motivated to perform actions from a range of motivating reasons, many of which are morally significant, and many of which we can assess without knowing anything at all about what the agent actually did.

Before getting into this, I say a bit more on what I mean by motivating reasons. Our motivating reasons are the reasons agents utilize in their practical deliberations. In general, I use deliberation extremely loosely throughout this dissertation – perhaps to the point of

abusing the term. Something can serve as a motivating reason in any number of irrational, arational, or rational non-deliberative processes, which nonetheless lead to action. Facts about a ball's trajectory can partially serve as an expert tennis player's motivating reason for hitting it just so, even if the player never reasons about what action to take. That the rat is a white fuzzy thing can partially serve as a motivating reason in Little Albert's conditioned aversion to the animal.[63] Or that a vase is present can partially serve as my motivating reason for throwing it on the ground in a fit of rage. I say partially in these cases because we often act for more than a single reason. Some motivating reasons are morally significant, such as those discussed in DDE cases, and some motivating reasons are morally irrelevant, such as those mentioned in the paragraph above.

These facts, often in combination with various desires (e.g. to win the tennis match), judgments (e.g. that white fuzzy things are bad), and impulses (e.g. for violent catharsis), help to move and agent to action and help to explain why they did what they did. However, while we might talk about an agent's motivating reasons as making her action rational or as rationalizing her action (when the action only renders rational in a certain light), I am using a more permissive sense of motivating reason, since they can also serve to highlight the agent's rational and moral failings without any reference at all to what the agent actually did.

Acting on sinister motivations is always criticizeable, even when the agent's motivations bore no fruit in action. For example, a woman poking a voodoo doll with a pin, in order to kill her wife, invariably fails to succeed using that method, yet the desire to kill

---

[63] The Little Albert experiment was an early behavioral study demonstrating classical conditioning and stimulus generalization in humans. The researchers induced a phobia of a rat and similarly fuzzy items in a psychologically normal infant. See J.B. R. Rayner's (1920) "Conditioned Emotional Reactions" (*Journal of Experimental Psychology*, 3.1).

one's wife is a moral failure on the part of the agent. I need look no further to find something

morally objectionable. So, not all morally relevant motivations will map onto a similarly

relevant feature of the action, yet the motivations themselves are still appropriate objects of

assessment. Having said that, I will conclude this chapter so that I can develop my account of

second-order agent-assessments in Chapter 3.

**Conclusion**

In this chapter, I outlined the bare structure of the UMF. Here is the final result.

**Figure 8: The Unified Moral Framework [Completed]**

|  | **Act-Assessment** | **Agent-Assessment** |
|---|---|---|
| **First-Order** | Erogation (1a) | Quality of Will *de re* (2a) |
|  | Obligation (1b) | Quality of Will *de dicto* (2b) |
| **Second-Order** | Permissibility (4) | Moral Worth (3) |

Figure 8 illustrates the UMF – complete with examples of the specific moral assessments I propose for each box. My reader can agree to the structure of Figure 8 without agreeing to the content. The relationship between boxes (1) and (2) will be examined over the next two chapters, beginning with second-order agent-assessments of Moral Worth in Chapter 3.

# Chapter 3: Moral Worth as a Second-Order Moral Assessment

**Introduction**

In Chapter 2, I developed the first half of the moral framework required for unifying assessments about the puzzling moral phenomena introduced in Chapter 1. To do this, I proposed several plausible distinctions between types of moral assessments, which together make up what I call the Unified Moral Framework. Central to the Unified Moral Framework is the distinction between first-order moral assessments and second-order moral assessments. In Chapter 2, I focused on developing the account of first-order moral assessments – moral assessments the objects of which are evaluated by appeal only to the morally relevant features of the object itself.

**Figure 1: The Unified Moral Framework**

| | Act-Assessment | Agent-Assessment |
|---|---|---|
| **First-Order** | Erogation (1a) | Motivating Reasons *de re* (2a) |
| | Obligation (1b) | Motivating Reasons *de dicto* (2b) |
| **Second-Order** | Permissibility (4) | Moral Worth (3) |

Figure 1 depicts the Unified Moral Framework filled out with the corresponding categories of moral assessment (not the assessments themselves.).

However, sometimes we want to assess the interactions between the assessment properties in first-order moral assessments with different sorts of objects, resulting in hybrid moral assessments, or, what I call second-order moral assessments.

> **Second-Order Moral Assessments:** Moral assessments the objects of which
> are evaluated by appeal to interactions between the assessment properties in
> two first-order moral assessments.

This chapter deals exclusively with my account of moral worth as a second-order agent-assessment. I start with the assumption that moral worth assessments are clear candidates for second-order agent-assessments. Moral worth is an agent-assessment rendered by appeal to the relationship between the agent's motivations and her action. When we make such assessments, we want to know how to view the agent in light of her morally significant actions. Following Nomy Arpaly, I refer to this category of moral assessment as assessments of the agents about whom we issue judgments of praiseworthiness or blameworthiness.

Although I do not directly require this account to explain any of the puzzles articulated in Chapter 1, it does carve out some of the conceptual space reserved for my account of permissibility and provides a model for permissibility as a second-order act-assessment. For this reason, Chapter 3 is crucial scaffolding for my framework. The central theoretical concern for this chapter is as follows. If moral worth assessments are agent-assessments, one might wonder what role the agent's action serves in the assessment of her. I propose that the morally significant agent-independent features of the action provide the background against which the agent's motivations are assessed by specifying which of the agent's motivations are relevant to a particular moral worth assessment. The relation between the agent and her action is the reasonably well-established responsibility relation. Moral worth is an assessment of an agent for the quality of her responsibility relation to her action.

> **Moral Worth:** Second-order moral assessment the subject of which is an agent
> who is evaluated by appeal to the quality of the relationship between the

reasons for which she acted and the morally significant features of the action constraining these reasons.

While primary focus of this chapter is my account of what makes an agent praiseworthy or blameworthy, this chapter brings in the responsibility relation to showcase the underlying structure of moral worth assessment as hybrid moral assessment. This allows us to make sense of phenomena like responsibility without blame – a moral assessment that agents are sometimes fully morally responsible for their wrong actions but not blameworthy. This category of moral worth assessment creates space for a category of permissibility assessment I heavily utilize in making sense of (1) – (5).

This chapter proceeds as follows. Section 1 covers my account of praiseworthiness, according to which the features of the actions that constrain our assessments of praiseworthiness depend on the relevant first-order act-assessment. Section 2 handles the controversial case of Huckleberry Finn. In Section 3, I develop my account of blameworthiness. In Section 4, I explain my account of responsibility without blame. I then conclude.

## 1. Praiseworthiness

Sometimes agents perform the morally right act and are morally praiseworthy for doing so. Consider the case of the GOOD-WILLED GROCER.

> **GOOD-WILLED GROCER:** The good-willed grocer always treats her customers fairly. She cares about profit as much as the next grocer, but she treats her customers fairly because it is the right thing to do. If treating her customers fairly were unprofitable compared with cheating them, this good-willed grocer would treat them fairly anyway.

Like Kant, many of us are inclined to think the Good-Willed Grocer is morally praiseworthy for treating her customers fairly. We think she performs the right act for the reasons for which it is right and that her decision to perform this act is morally commendable.[64] While the gist of this assessment does not vex, explaining the assessment and how it relates to others like it can.

While we might describe any sort of positive moral assessment as moral praise, praiseworthiness, as I use it here, refers to a kind of second-order moral assessment about the extent to which the agent was properly motivated by the morally significant agent-independent features of the action. The difficulty in assessing praiseworthiness lies in the different ways first-order moral assessments interact.

The Good-Willed Grocer performs a morally right action, presumably because she conforms with her single operative duty to not cheat her customers, for which she is a candidate for praiseworthiness. She achieves candidacy for praiseworthiness because we

---

[64] Kant, op. cit.

cannot assess her actual praiseworthiness without further consideration of her motivations. A morally desirable action, such as a morally right one, is the first requirement for an agent's candidacy.

However, not all morally desirable actions are morally right, since not all of our actions are regulated by duties. Some actions are supererogatory – actions we have on-balance the most and mightiest moral reason to perform and no duty to perform in a non-dilemma case. Others have argued that supererogatory actions are actions for which agents performing them are candidates for praise, and I see no reason to object for the most part. Generally, we think promoting the moral good is morally commendable, especially if the agent is not bound to do so by any duty. So, on my account, an agent who performs a morally right action or who performs a supererogatory action is a typical candidate for praiseworthiness.

Once again, the purpose of this discussion is not to identify, in general, which category of motivation is central to our second-order moral assessments about an agent's praiseworthiness for performing a morally desirable action. Different motivations matter more or less for a particular type of moral assessment at different times. When assessing an agent's praiseworthiness for performing some action, the motivations privileged in our assessments are only those dealing with the corresponding first-order act-assessment. The morally significant normative features of an action, which are the objects of a corresponding first-order act-assessment, specify which of the agent's motivations in acting are relevant to assessing her moral worth. Only this can explain why we examine whether an action is right or wrong or whatever to determine the agent's candidacy for praise or blame. The action establishes the background against which we assess the agent.

So, when an action is morally right, motivations responsive to the action's rightness or

dutifulness tend to make the agent praiseworthy. Meanwhile, when an action is supererogatory, motivations responsive to the action's supererogation tend to make the agent praiseworthy. I will address complications to this picture later in the chapter. Otherwise, when the agent performs the morally desirable action, the agent fails to be properly motivated by morally significant features of the action for which we are assessing her praiseworthiness. On this account, while the action establishes the background against which we assess agent's motivations, these motivations do not in any way modify the agent's motivations or their moral significance; they only bear on the quality of the relationship between the agent and her actions.

As stated in Chapter 2, an action renders morally right when conforming with an operative duty by not instantiating P in a non-dilemma case. If the Good-Willed Grocer conforms with her duty to not cheat her customers by treating them fairly, her action is morally right when her action does not fail to conform with any countervailing conflicting duty. If she is praiseworthy for performing a morally right action, then she is likely motivated by the action's rightness. Because the action's rightness does not itself correspond to any substantive moral goods, the Good-Willed Grocer is most likely praiseworthy only when she is motivated by its rightness *de dicto*. She is likely praiseworthy when she is motivated to perform the action only because she takes it to be right action. The substantive features of the action for which the Good-Willed Grocer could be motivated *de re*, such as the action's fairness or its contribution to the customer's well-being, do not correspond to the features of the action that make it right.[65]

---

[65] On Arpaly's view (op. cit. Chapter 3, esp. 70-75), an agent who acts "from the motive of duty" may not always be praiseworthy. Performing an act for duty's sake or because the act is in conformity with what one takes to be one's duty could lead one to perform acts conflicting

Recall that while right actions often manifest substantive moral goods, the substantive moral goods themselves never make the action right on my account. For why this is so, recall the discussion on minimal moral adequacy in Chapter 2 (section 5). When it turns out that the available permissible means of conforming with duty in non-dilemma cases involve acting so as to instantiate another property – G, which is a morally desirable or good-making property, acting so as to bring about G may well be required. Even so, the action is not morally right because of its G-ness or any other substantive good brought about in performing the action. A morally right action need not instantiate any goods at all. Instead, a morally right action may well have no moral reasons in favor and moral reasons against (as with Perry Zoso's action). Any goods the agent brings about in acting are merely a part of the means of satisfying the duty and not the features of the action that make it morally right.

This is, in part, because substantive moral goods often correspond to moral successes, such as successfully saving a drowning swimmer, yet, as Kant correctly pointed out, an agent performs a morally right action even if she tries and fails to save the drowning swimmer and brings about no substantive moral goods thereby. Further, even morally wrong actions sometimes instantiate substantive goods, making them unreliable motivators for performing praiseworthy morally right actions. For example, Zoe Johnson-King correctly notes that an obsession with fairness, at the expense of other sorts of moral considerations, could be precisely the reason that an agent performs a wrong action, such as when a parent insists on sending all of her children to bed at the same time, when one is an infant and another is nearly an adult.[66] Although the policy of sending all children in a household to bed at the same time

with morality.
[66] Johnson King, op. cit.

may well be fair (perhaps in the a distorted sense, since none of the children will get the extra benefit of an additional hour of playing video games), doing so is nonetheless inappropriate and possibly wrong.

Instead, substantive moral goods correspond to the features of the action regulating the Realm of the Erogatory. Such considerations are more relevant to assessing the praiseworthiness of an agent performing a supererogatory action. Unlike the case of right action, acting from the motive of the action's supererogation *de dicto* does not typically suffice to make the agent praiseworthy.

Supererogation, like any assessment of erogation, is an assessment of an action for its normative moral reasons, specifically of the substantive moral goods manifested by the action – the action's Gs. So, Gs are the relevant assessment properties of the action, and these assessment properties specify which of the agent's motivations are evaluated in the relevant second-order agent-assessments. Only assessment properties of first-order moral assessments are considered in second-order moral assessments, so only the action's Gs are relevant to the moral worth assessment. I get into the details of how exactly this view works in section 5, but I can say how it applies to supererogation here.

Specifically, Gs in acting pick out corresponding G-related motivations. An agent is never praiseworthy when motivated by an action's supererogatoriness in the *de dicto* sense because supererogatoriness is not an assessment property of first-order act-assessments. Meaning, there is no G with the type of supererogatory that can pick out supererogation-related motivations for assessments of praiseworthiness. Otherwise, we would have to think an action's supererogatoriness is a normative feature of the action we can appeal to when making the assessment that the action is supererogatory, and this does not seem right.

Similarly, an agent is never praiseworthy when motivated by "the action with the most and mightiest moral reasons to perform and no duty to perform" (again, in this general sense) because there is no G of this type to specify a corresponding motivating reason in the agent. Although the former is an appropriate evaluation of the action and the latter is an appropriate description of the action, neither are the normatively significant features of the action appealed to in the first-order act-assessment. So, neither play a role in our moral worth assessment on my view. An agent is therefore only praiseworthy for performing a specific supererogatory action when she is sufficiently sensitive to the specific normative reasons that make that action supererogatory.

Moreover, performing some particular supererogatory action only because it is fair generally or only because it promotes the customer's well-being generally inadequately captures the action's supererogatoriness. An action's fairness or virtuousness as such is not a normative moral reason in favor of performing some action, so it cannot serve as an assessment property of an action. Instead, I think that we have moral reasons to act in some way because of the impact it will have on other people and that the explanation for why that reason counts as a moral reason could be something like our value of fairness or virtuousness. Fairness is really just a kind of first-order act-assessment that assesses an action along a particular value dimension.

Performing some action only because it promotes the customer's well-being generally has a similar issue. Well-being promotion is a value that plays a role in our assessment of the Good-Willed Grocer's action. And just to clarify, I do think fairness and well-being promotion as such can serve as motivating reasons, meaning they can function as assessment properties of agents. As a result, they can play a key role in our permissibility assessments, so I am not

saying that such motivations are entirely irrelevant to our second-order moral assessments. However, the concepts of various goods have nothing to do with supererogation generally, so acting with those as one's motivations cannot make one praiseworthy for performing a supererogatory action. Only sensitivity to the specific features of the action that make that action fair and ultimately supererogatory are the sorts of motivations from which praiseworthy supererogatory actions arise (or rather, they are the sorts of motivations from which praiseworthy agents performing supererogatory actions arise – I use the former, rather than the latter, sort of locution for the sake of expediency throughout).

## 2. *Prima Facie* Mismatched Cases

One problem with this account is that it appears to get praiseworthiness in the case of HUCK wrong. Consider a case introduced by Bennett and made popular by Arpaly.[67]

> **HUCK:** Huckleberry Finn is traveling with his friend Jim – a runaway slave. Huck is faced with the decision of whether to turn Jim in to the authorities. Huck deliberates about what he ought to do and concludes that he ought to turn Jim in to the authorities because helping a slave escape is tantamount to stealing, and stealing is wrong. Despite his deliberations, Huckleberry Finn helps Jim escape the authorities, and he concludes that he is just a bad boy.

Huck is clearly akratic – he acts contrary to his best judgments or suffers from weakness of will. Akratic agents are criticizeably irrational because they fail to do what they judge they ought to do. Although Huck judges that he ought to turn Jim in, he helps Jim escape instead. Weakness of will is hard enough to make any sense of philosophically, but Huck is not just weak willed. Something else is going on in this case as well. Helping Jim escape, so we critical onlookers might think, is the thing that Huck all-things-considered ought to have done. So, although Huck is akratic, we might yet think Huck is more rational in helping Jim escape than he would have been if he had acted in accordance with his best judgment. This is what Arpaly calls "Inverse Akrasia," and she argues that not only is Huck more rational for helping Jim escape, he is praiseworthy for doing it as well.[68] Huck performs a morally right action, and he

---

[67] Arpaly, op. cit. Chapter 3; Bennett, J. (1974). "The Conscience of Huckleberry Finn." *Philosophy*, 49(188).
[68] Arpaly, op. cit. 75

does it for the right sorts of reasons, such as his compassion for Jim's plight or his appreciation for Jim's humanity.

Pamela Hieronymi provides resources for thinking Arpaly's characterization of the case is correct. Hieronymi argues that we can explain acting contrary to one's best judgment by distinguishing between practical reason, on the one hand, and theoretical reasoning about practical subject matter, on the other.[69] According to Hieronymi, the output of practical reasoning is an intention to act, and the agent forms this intention upon "taking certain considerations to settle the question of whether to act." Judgments about what one ought to do derive from pieces of theoretical reasoning about the reasons one has for acting. They are two different kinds of reasoning.

Ideally, the considerations an agent takes to settle the question of whether to act would be the very same considerations, weighed in the same way, that the agent judged to count in favor of acting, but sometimes this does not happen, as with Huck. Huck incorrectly judges that he ought to turn Jim in based on certain salient cultural considerations. Fortunately for Jim, the considerations Huck took to settle the question of whether to turn Jim in, such as his compassion for Jim's plight or his appreciation for Jim's humanity, were those that he had deemed insufficient in the course of his deliberations. Although Huck deemed such considerations insufficient in the course of his deliberations, Hieronymi gives us the tools to say that he nonetheless acted for a reason when he helped Jim escape and perhaps we can say that he acted for the right reasons in doing so.

---

[69] Hieronymi, P. (2009). "The Will as Reason." *Philosophical Perspectives*, 23, Ethics: 207.

Arpaly's assessment of this sort of case gave us richer notions of the types of agents and actions that have moral worth, while Hieronymi's account of reasons supply us with more complex notions of the varieties of rationality and irrationality necessary for making any sense of the case at all. However, the reader may well have difficulty seeing how my account is compatible with the assessment that Huck is praiseworthy for performing a morally right action, since, as Arpaly stipulates, he is motivated by the action's rightness *de re*. Meaning, Huck is motivated by the specific normative reasons in favor of performing the action and not the action's rightness *de dicto*. If I propose that assessments of an agent's motivating reasons *de re* only ever appropriately feature in assessments of praiseworthiness when the agent performs a supererogatory action (and if we think no right actions are supererogatory), then Huck's motivations do not make him praiseworthy when he performs a right action by helping Jim escape.

However, much depends on which specific considerations settle the question, for HUCK, of whether to help Jim escape. If, for example, Huck is motivated, in part, by his compassion for Jim, then his praiseworthiness depends on what his compassion is responsive to. If Huck's compassion is responsive to the unjust suffering Jim has experienced at the hands of his captors or to the anticipated unjust suffering Jim would experience in future captivity, then Huck is almost certainly appropriately responsive to the normative reasons against turning Jim in. Furthermore, such features of the action are likely the very same features Huck has a duty to avoid in acting. Meaning, Huck is responsive to the would-be *de re* wrongness of turning Jim in, and he is motivated to perform some other action due to these considerations. If so, Huck performs a morally right action while responsive to the *de re* rightness of the action only insofar as the action's *de re* rightness is determined by facts about the substantive wrong-

making features the action does not instantiate (the suffering Jim would have experienced if he had turned Jim in).[70]

If something along these lines accurately captures Huck's motivations in acting, that he is in some way performing the action because the action is not wrong *de re*, then I can account for Huck's praiseworthiness. To do so, I must first explain my account of blameworthiness. Specifically, I must say a bit about why the substantive features of an action's wrongness are relevant to assessments of moral worth in a way that the substantive features of an action's rightness are not.

---

[70] My reader my worry about how I can account for the praiseworthiness of Huck's taking positive steps to help Jim escape, when I have only an account of negative duties. It might seem like I can only account for why Jim must refrain from turning Jim in. First, recall from Chapter 2, that negative duties can require agents to take positive action when the alternatives would amount to a failure to conform with duty. If the risk of Jim getting caught, if left entirely on his own, was high enough, then Huck would be required to intervene on his behalf. But even if he was not strictly speaking required to help, any moral reasons in favor of helping Jim escape to which he is responsive might nonetheless point him in that direction. Furthermore, many right actions are also the actions the agent has the most and mightiest moral reasons to perform. Although such actions are not, strictly speaking, supererogatory, I am open to assigning praiseworthiness to agents motivated by these considerations in the same way I would in the case of supererogatory actions. If Huck was motivated to perform this sort of action in this way, then I would generally agree with Arpaly that Huck is praiseworthy for performing the action for its de re goods, but not because the action is morally right. However, I am setting aside these sorts of complications for the sake of simplified presentation. Once my reader understands how things work in the case of supererogation, then the view can easily be applied to other relevantly similar scenarios.

### 3. Blameworthiness

Sometimes agents perform the morally wrong act and are morally blameworthy for doing so. Consider another Kantian-variety grocer: the MALEVOLENT MERCHANT.

> **MALEVOLENT MERCHANT:** The malevolent merchant always cheats her customers. If cheating her customers were less profitable than treating them fairly, this malevolent merchant would cheat them anyway. She does not care at all about profit and instead works as a grocer because she enjoys victimizing unsuspecting customers.

Plausibly, the Malevolent Merchant is blameworthy for cheating her customers. We think she performs the wrong act for the reasons for which it is wrong and that her decision to perform this act is morally criticizeable. While we might describe any sort of negative moral assessment as moral blame, "blameworthiness", as I use it here, refers to a kind of second-order moral assessment about the way in which the agent was improperly motivated by the morally significant agent-independent features of the action. Specifically, we are assessing how the moral undesirability of a particular action featured in the agent's motivation for acting.

The Malevolent Merchant performs a morally wrong action by failing to conform with her single operative duty to not cheat her customers, for which she is a candidate for blameworthiness. She achieves candidacy for blameworthiness because we cannot assess her actual blameworthiness without further consideration of her motivations. Having performed a morally undesirable action, such as a morally wrong one, is the first requirement for an agent's candidacy. Not all morally undesirable actions are morally wrong, since not all of our actions are regulated by duties. Some actions are suberogatory. Generally, we think promoting

106

moral ill is morally criticizeable, yet agents who perform suberogatory actions are not candidates for blame, as I will demonstrate shortly.

Merely performing a morally undesirable action does not suffice to make an agent blameworthy. An agent's specific motivations play a role in assessing her blameworthiness. The morally significant normative features of an action specify which of the agent's motivations in acting are relevant to assessing her moral worth. When assessing an agent's blameworthiness for performing some action, the motivations privileged in our assessments are only those that deal with the corresponding first-order act-assessment. So, when an action is morally wrong, an agent is blameworthy if and only if she is inappropriately motivated with regard to the action's wrongness. Largely in agreement with Arpaly, I think this amounts to the following.

In Chapter 2, I proposed that our duties are formulated as prohibitions on acting so as to bring about distinctively moral harms – however that pans out on any particular view. I have been assuming the Malevolent Merchant has a duty to not cheat her customers, perhaps because cheating one's customers is unfair, or disrespectful, or harmful in some other way. Regardless, the Malevolent Merchant fails to conform with her duty to not cheat her customers, making her action morally wrong. If she is blameworthy for performing a morally wrong action, then she is insufficiently motivated to treat her customers fairly. An agent is insufficiently responsive to the relevant moral considerations under two conditions.[71] First, an agent can be motivated to perform an action from the motive of ill will, which is performing a wrong action for the reasons for which it is wrong. In this case, the agent is insufficiently responsive in the sense that she is responding to the relevant moral reasons in the wrong way.

---

[71] Arpaly, op. cit. Chapter 3

Second, an agent can be motivated to perform a wrong action from moral indifference or without regard for the wrongness. In this case, the agent is motivated by morally irrelevant reasons or while apathetic to moral considerations (or both). All can serve as a motivating reason *de dicto* or *de re*.

The Malevolent Merchant appears to instantiate a version of the former. She is not motivated by the action's wrongness *de dicto* – she is not like Shakespeare's Aaron who does evil for the sake of evilness. Instead, she is motivated by the substantive features of the action for which the action is regulated by duty – the action's wrongness *de re*. The Malevolent Merchant delights in opportunities to abuse her customers, and she motivated to act by these opportunities for abuse.

Plausibly, Chairman HARM, from Chapter 1, instantiates a version of the latter. Chairman HARM is the CEO of a company who desires to earn profits. Because he does not care at all about the environment, he implements a program that will earn a great deal of profits – destroying the environment in the process. Destroying the environment is morally wrong, which makes Chairman HARM a candidate for blame. He is blameworthy because he is insufficiently sensitive to the normative moral reasons or the specific substantive ills that count against performing the action.

This analysis initially appears at odds with what I say about praiseworthiness for right action above. However, I tend to think an action's wrongness is grounded directly in the substantive features of an action in a way that an action's rightness is not, so the blameworthiness of an agent performing a morally wrong action can be assessed both in terms of her responsiveness to the morally significant features of the action that make it wrong and to the wrongness itself. I have proposed that a judgment of wrongness is an unpropitious

assessment of actions for failing to conform with the duty against acting so as to bring about a particular property – P – in acting. Failing to conform with one's duty or acting so as to bring about P in acting both make an action wrong because the latter will always be an instance of the former, unless the agent has a certain kind of excuse where we would say that the instantiation of P is not properly said to be a property of the action, and a blameworthy agent might be motivated to perform an action either for its wrongness or its P-ness.

If Huck is motivated to conform with his duty, thereby helping Jim escape, because he is motivated to act so as to not bring about P in acting (where we might suppose P is Jim's suffering or some other morally undesirable feature of the action), then Huck is almost certainly praiseworthy for doing so. Because actions render morally right only when some likely alternative is morally wrong, then Huck's responsiveness to the substantive features of the alternative action's wrongness, which are relevant to assessments of moral worth, are potentially the kinds of motivations from which praiseworthy right actions arise. Strictly speaking on my account, Huck is not praiseworthy for performing a morally right action for the action's rightness *de re* – contrary to Arpaly. Instead, Huck is praiseworthy for performing a morally right action while appropriately responsive to the *de re* reasons for which the alternative action is wrong. These are the morally significant substantive features of Huck's action that make his action right.[72] These sorts of considerations do not, however, bear on

---

[72] I do not mean for any of this to imply that performing a right act because the alternative action is wrong in the *de dicto* sense is a potentially praiseworthy motivation for someone like Huck. Although acting for the *de re* reasons for which the alternative action is wrong corresponds to the reasons for which the right action is right, the alternative action's *de dicto* wrongness has nothing to do with the right action's rightness *per se*. Such motivations could similarly correspond to my performing a merely dutiful action, which could turn out to be either of the morally preferred or the morally undesirable actions in a dilemma scenario. I can apply the same reasoning as used for Huck to assess the Good-Willed Grocer's praiseworthiness. She might be praiseworthy because she registered the *de re* wrongness of

questions about assessments of moral worth for suberogatory actions. Suberogatory actions

do result in specific substantive ills that count against performing them. If an agent were

blameworthy for performing a suberogatory action, then she would be blameworthy for

insufficient responsiveness to these morally relevant considerations. The problem with such

an account is that we have no principled way of distinguishing between specific substantive

ills manifested in suberogatory actions and those manifested in merely erogatory actions –

actions we have an equal balance of moral reasons for and against. So, we would have to hold

that agents performing merely erogatory actions are also blameworthy when acting with

insufficient regard for the morally relevant considerations, and this seems problematic, since

it could result in agent blameworthiness while participating in a number of morally ambiguous

but relatively mundane activities, such as drinking coffee or using electricity (depending on

their source).

One might propose that this concern applies equally to assessments of praiseworthy

supererogation, since we have no principled way of distinguishing between specific

substantive goods manifested in supererogatory actions and those manifested in merely

erogatory actions. However, the motivations from which praiseworthy supererogation arises

are those corresponding to the specific substantive moral goods that make a particular

supererogatory action supererogatory, and this is the only time that an agent's motivation to

manifest the substantive goods of an action are morally assessible. Meaning, the only context

in which it would be appropriate to assess the quality of the agent's motivations *de re*, for the

---

cheating one's customers but not because she was motivated by the *de dicto* wrongness of
cheating them, since the former and not the latter helps to make sense of why the particular
right action she performed was right and can pick out a corresponding set of motivating
reasons.

sake of assessing her praiseworthiness, is when the agent performs a supererogatory action. If we thought the motivations from which blameworthy suberogation arises are those corresponding to the specific substantive moral ills that make a particular suberogatory action suberogatory, then we should likewise think that this is the only time an agent's motivation to manifest the substantive ills of an action are morally assessible. We know this is incorrect because we can assess an agent's motivation to manifest the substantive ills in the performance of morally wrong actions.

The alternative would be to say that an action makes an agent a candidate for blame whenever there are more and mightier moral reasons against the action than for it. Surely, one might think, all wrong actions and all suberogatory actions share this feature, which can rule out erogatory actions as candidates for blame. However, such an account would rule out the blameworthiness of agents who perform wrong actions for their wrongness *de dicto* – the likes of Shakespeare's Aaron – and, furthermore, might include the blameworthiness of agents who perform morally right actions, since my account is compatible with the possibility of an action's being right and nonetheless the action the agent has decisive moral reason against performing. Moreover, we simply cannot hold that all wrong actions are the actions the agent has decisive moral reason against. Typically, one wrong action in a genuine moral dilemma has more going for it morally than the other, which partially explains the resolvability of moral dilemmas. And, if DDE cases show us anything, an agent is a potential candidate for blame even when performing the morally preferred solution to a dilemma, since the Terror Bomber's desire to kill civilians almost certainly makes him blameworthy for killing them.

So, while it makes sense to say that an agent's motivating reasons *de dicto* are privileged in one kind of praiseworthiness assessment while her motivating reasons *de re* are

111

privileged in another, no such distinction exists for blameworthiness, since both play a role in assessing an agent's blameworthiness for performing morally wrong actions. Furthermore, we can identify no principled way of distinguishing between the potential blameworthiness of agents performing suberogatory actions and those performing erogatory action that correctly handles agents performing right or wrong actions. Better to reject blameworthiness for suberogatory actions and, instead, hold that such actions and their agents satisfy moral adequacy requirements, which I discuss in more detail in Chapter 5.

## 4. Responsibility without Blame

The final section of this chapter serves two purposes. Most importantly, this section steps away from the focus of the previous sections, where I developed an account of what makes an agent praiseworthy or blameworthy. This section identifies key structural features of the responsibility relation, which I intend to use as a model for the justification relation. These relations help to make sense of the structure of second-order moral assessments as hybrid assessments. This section also carves out some of the conceptual space reserved for my account of permissibility. Specifically, I explain responsibility without blame – a moral assessment that agents are sometimes fully morally responsible for their wrong actions but not blameworthy. This category of moral worth assessment creates space for a category of permissibility assessment I heavily utilize in making sense of (1) – (5).

Beginning with the structural underpinnings of my account of moral worth, recall that, in Chapter 1, I observed a previously overlooked parallel between the asymmetrical assessments of permissibility [AAP] observed in DDE cases and another set of cases – those resulting in asymmetrical assessments of praise and blame, otherwise known as praise-blame asymmetry [PBA], stemming from side-effect effect cases.

I constructed the parallel between the two puzzles as follows. Put roughly, two agents perform two actions with the same agent-independent features, yet they perform their actions for countervailing motivating reasons in AAP cases. The asymmetries in the agent's motivations appear to generate asymmetries in permissibility assessments about the actions, but this is puzzling because we have difficulty explain how one's motivations are relevant to the assessment of one's actions. PBA cases are often considered less troubling, but, I contend, they provide a good model for understanding the asymmetries AAP cases. Put roughly, two

agents perform two actions with different agent-independent features, yet they perform their actions for the same motivating reasons. The asymmetries in the agents' actions appear to generate asymmetries in moral worth assessments of the agents, but this should be just as puzzling as AAP since one might wonder how the action is relevant to the assessment of the agent who performed it.

For AAP, the actions are the same, while the agent motivations are the same for PBA. For PBA, the actions are different, while the agent motivations are different in AAP. For AAP, the assessment targets the actions, while the assessment targets the agent for PBA, so we might expect the pairs of assessment for each pair of cases to render symmetrically – doing otherwise violates the principle of parity. Yet, the agent-assessments render asymmetrically for PBA, while the act-assessments render asymmetrically for AAP. The asymmetries in moral assessment, in both sets of cases, can be explained by proposing the relevant assessment is a second-order or hybrid moral assessment. However, the parallels between the asymmetries observed in the two cases indicate much more. AAP and PBA are mirror phenomenon, which suggests that the asymmetrical moral assessments in the pairs of cases render on the basis of similar mechanisms. I believe analyzing PBA helps to understand and to explain the mechanism contributing to AAP.

The cases of the two chairmen represent the paradigm for PBA cases. The two chairmen act from the same desire for profit, without any regard for moral considerations. If we were to assess the agents on the basis of their motives alone, we would be forced to assess them symmetrically. Because the profit motive is largely morally irrelevant, any assessment of praiseworthiness or blameworthiness would be hard to justify. Such motivations are neither. Because the agent's motivations alone do not settle the question of whether the agent should

114

be assessed positively or negatively, the agent's action provides the background against which the agent should be assessed.

I therefore propose that an agent's action is directly relevant to assessing her moral worth, in the same way I will later propose that an agent's motivations are directly relevant to assessing the action's permissibility. More specifically, I am proposing that (i) the morally significant normative features of the action specify which of the agent's motivations are relevant to assessing the agent's moral worth. I call the motivating reasons specified in (i) the oneratory set. Onerate, in this context, means to burden with responsibility. The contents of the oneratory set establish the way in which an agent is morally responsible for the moral quality of some action. I call the relationship between the oneratory set and the features of the action constraining them the responsibility relation, but it could equally be called the oneration relation.

To assess the difference between the two agents in PBA cases (e.g. the chairmen cases), the agent-independent features of the action relevant to the assessment of the agent's moral worth are the assessment properties considered in the appropriate first-order act-assessments of the action for which the agent is being assessed. To assess Chairman HELP, we look to the right-making features of the action, and to assess Chairman HARM we look to the wrong-making features of the action.

The assessment properties of the action pick out any of the agent's corresponding morally significant motivating reasons, and only these motivating reasons can onerate or exonerate (i.e. burden or unburden with responsibility) the agent. So, the difference between the two chairmen's actions highlight the primary difference between how we assess the set of motivating reasons playing a role in onerating them. Because both chairmen are motivated by

profit, nothing relating to the morally significant features of the action are present in the oneratory set, except, perhaps, their genuine indifference to the relevant moral considerations and their intentions to perform the action despite these considerations.

The moral worth of an agent is determined by the way in which the action's assessment properties align with the agent's motivating reasons. The action's assessment properties align in the right way when they pick out motivating reasons indicating good will without picking out motivating reasons indicating ill will. This amounts to (ii) an assessment of the alignment between the action's assessment properties and the oneratory set and (iii) an assessment of whether any misalignment is due to the inclusion of an exonerating motivation in the content of the oneratory set or the exclusion of an onerating motivation in the content of the oneratory set.

Regarding (ii), if things go well, the agent's good will alone is a member of the oneratory set and the agent performed a morally right or supererogatory action that specifies those motivations as relevant to the moral worth assessment, as with the Good-Willed Grocer. Regarding (iii), exonerating and onerating motivations for praiseworthiness and blameworthiness are inversed. The good will serves as an onerating motivation for praise and an exonerating motivation for blame, while ill will serves as an onerating motivation for blame and an exonerating motivation for praise. Although the two chairmen both act with moral indifference, their motivations play different roles in the assessment.

Chairman HELP is not praiseworthy because her right action fails to pick out anything indicating that she acted with a good will in the oneratory set. Meaning, the onerating motivation, the motivation that burdens Chairman HELP with responsibility for the rightness of her action and makes her praiseworthy, is excluded from the oneratory set. Chairman HARM

is blameworthy because his wrong action fails to pick out anything indicating that he acted with a good will in the oneratory set. Meaning, the exonerating motivation, the motivation that unburdens Chairman HARM with responsibility for the wrongness of his action and makes him not blameworthy, is excluded from the oneratory set. The good will is relevant to the action's wrongness insofar as it deals with the agent's appropriate responsiveness to the wrong-making features of the action. Instead, we find only Chairman HARM's indifference to the action's harmfulness and his intention to perform the action despite its wrongness in the content of the oneratory set, which clearly saddles him with the burden of responsibility for the action's wrongness.

This model of the underlying structure of moral worth assessments helps to explain the direct relevance of an action to these assessments. In the next chapter, I propose something parallel for the direct relevance of motivations to permissibility assessments. An analysis of PBA cases provides a model for analyzing AAP cases.

With this model in hand, I will now address a different concern for my view. Chairman HARM is straightforwardly blameworthy for his wrong action. However, we problematically get the same result for Medea in SIDE-TRACK. While we might plausibly think Medea did not intend the morally objectionable outcomes of her actions, she nonetheless intentionally performed the relevant action – pulling the lever and killing the one thereby. If she performed a morally wrong action for which she is morally responsible, then we may well think she is blameworthy for doing so. Since, according to a long-standing view on moral responsibility, an agent is blameworthy if and only if she is morally responsible for performing a wrong action. The underlying idea is that if the agent is morally responsible for performing a wrong action, then she was improperly motivated when performing that action.

If Medea is blameworthy, then she cannot have acted permissibly, since we cannot say that an agent acts permissibly but that we will blame her. Blame characteristically involves experiencing hostile emotions to the subject of the blame, forming negative judgments about the subject's character, taking punitive actions towards the subject, and feeling entitled to participate in the three aforementioned regarding the subject.[73] Permissibility entails that we not take punitive actions towards, or interfere with the agent performing the action. So, permissibility and blameworthiness are incompatible. In which case, I must show that an agent's moral responsibility for performing a morally wrong action can separate from her blameworthiness for performing a morally wrong action if I aim to make sense of permissibly taking the morally preferred action in genuine moral dilemmas.

Again, consider Medea. Medea performs a morally wrong action in killing Jason, so she is a candidate for blame. Only motivating reasons relating to the action's wrongness are therefore selected for the contents of the oneratory set. The motivating reasons privileged in the assessment of blame relate to ill will, so first we would need to establish that Medea did not act from ill will. Medea pulls the lever out of ill will if hitting Jason with the trolley was taken to be a good-making feature of pulling the lever in SIDE-TRACK, since hitting Jason and killing him thereby is the feature of the action that makes it wrong. However, we stipulate that Medea is not motivated to pull the lever by the fact that the train will hit Jason and hitting Jason with the trolley is not a means of saving the five. So, we need not judge that Medea is acted from the motive of ill will. So, no such motivation is featured in the oneratory set.

---

[73] Pickard, H. (2011). "Responsibility Without Blame: Empathy and the Effective Treatment of Personality Disorder." *Philosophy, Psychiatry, and Psychology*, 18(3); Pickard, H. (2014). "Responsibility without Blame: Therapy, Philosophy, Law." *Prison Service Journal*, 213.

Now, consider whether Medea pulls the lever out of a deficiency of good will. Medea pulls the lever out of a deficiency of good will if Medea is insufficiently responsive to the relevant moral reasons against hitting Jason with the Trolley. This seems plausible, since Medea has an operative duty to not kill Jason, and she is failing to conform with that duty and the associated relevant moral reasons. However, throughout I have assumed that one's duty does not generate decisive moral reasons for conforming with it. The reasons for conforming with the duty to not kill Jason can be outweighed. If Medea judges that the killing Jason is an unfortunate consequence of pulling the lever and doing the thing she has most and mightiest moral reason to do, hopes he will escape, and feels guilty at the prospect of hurting him, then she is not insufficiently responsive to the relevant moral reasons. The reasons against killing Jason were outweighed by the reasons in favor of saving the five, and Medea accurately judged this to be the case. Thus, Medea may be appropriately responsive to the reasons against pulling the lever. So, instead of moral indifference, we would expect to see exonerating motivating reasons, reasons suggesting that Medea acted from the good will, in the oneratory set.

Medea was not improperly motivated, and we observe no misalignment between the reasons for which she acted and the morally significant features of the action. So, one might reasonably hold that Medea is not blameworthy for pulling the lever. If Medea and her dilemma counterparts are not blameworthy, then I have conceptual space for suggesting that they permissibly perform the morally preferred action in dilemma scenarios, and I can proceed with my account of permissibility in Chapter 4.

**Conclusion**

In Section 4, I noted that some theorists believe that agents who are morally responsible for their wrong actions are always blameworthy. If I want to say that agent's like Medea permissibly perform the morally preferred action in a dilemma scenario, such a view would commit me to saying that Medea is nonetheless blameworthy for performing a morally wrong action. Yet, when an agent is blameworthy for acting, then we cannot coherently say that the agent acts permissibly. So, I argued for responsibility without blame. If we can coherently say that Medea is morally responsible for her wrong action but not blameworthy, then it is possible for us to say that Medea acts permissibly. Responsibility without blame establishes a little discussed category of act-assessment I call permissible-wrong action, a prima facie inconsistent category of action that is not inconsistent on my account. I heavily utilize this category of permissibility assessment in making sense of (1) – (5). Over the next chapter, I will make sense of this moral assessment and explain how an account of moral permissibility can support such an assessment.

# Chapter 4: Permissibility as a Second-Order Moral Assessment

**Introduction**

This chapter is about permissibility as a second-order act-assessment. On my account, if permissibility is a second-order act-assessment, then permissibility assessments are made by appeal to interactions between the targets of at least two first-order moral assessments.

> **Permissibility:** Second-order moral assessment the object of which is an action evaluated by appeal to relationship between the reasons for which the agent acted and the morally significant features of the action constraining these reasons.

As formulated above, this account of permissibility does not rule out permissibility as a second-order act-assessment about interactions between the targets of first-order act-assessments only. Thus, the view that the permissibility of some action is ultimately settled by appeal to the morally significant features of the action alone is consistent with my account thus far. A proponent of this sort of view can adopt Unified Moral Framework [UMF] without taking on the views I develop in this chapter. However, they would have difficulty making sense of the possibility of genuine moral dilemmas and asymmetrical assessments of permissibility [AAP] in Doctrine of Double-Effect [DDE] cases without subscribing to the remainder of my view. Recall that making sense of the possibility of genuine moral dilemmas and AAP are central aims of this project.

Chapter 3 explained the role of an agent's action in the assessment of her moral worth. I proposed that a judgment or evaluation of moral worth is an assessment of an agent for the relationship between the reasons for which she acted and the morally significant features of the action constraining them. The relation under consideration in moral worth assessments is the responsibility relation. In that relation, the morally significant agent-independent features

of the action establish the background against which we assess agent motivations by specifying which motivations are relevant to our second-order agent-assessments. Put roughly, moral facts about the action therefore constrain which of the agent's motivations we assess in moral worth assessments. The opposite occurs in permissibility assessments.

In this chapter, I similarly argue that an agent's motivations are directly relevant to the assessment of an action's permissibility. An evaluation of an action's permissibility is an assessment of that action for the quality of its relationship to an agent's motivations. The relation under consideration is the justification relation, which I will discuss at length in Chapter 5. In that relation, the agent's motivations establish the background against which we assess the agent-independent features of the action by specifying which agent-independent features of the action are relevant to our second-order act-assessments.

I begin by rejecting what I call ordinary permissibility accounts, which typically subscribe to the following claim: (A) an action is morally permissible if and only if the action is not morally wrong. I argue that any theory of permissibility compatible with genuine moral dilemmas is incompatible with claim (A) because the duties relevant to genuine moral dilemmas are absolute duties. Any action failing to conform with an absolute duty is morally wrong because there are no exceptions to such duties. In genuine moral dilemmas, any action the agent performs is morally wrong because any action the agent performs fails to conform with one such duty. Yet, in genuine moral dilemmas, it typically seems as though there is a morally preferred solution – an action the agent permissibly performs.

The straightforward solution to genuine moral dilemmas is therefore to allow for the possibility of permissible-wrong actions [PWA] – morally wrong actions an agent nonetheless permissibly performs (the permissibility counterpart to responsibility without blame). PWA

is possible on the UMF because wrongness is a first-order act-assessment while permissibility, I argue, is a second-order act-assessment, meaning that each assessment type appeals to different assessment properties. As a result, I can plausibly argue that agents sometimes fail to conform with absolute duties, making their actions wrong, yet they do so permissibly and that the action's being permissible does mean that the duty allowed for exceptions (i.e. it need not turn out that the duty was conditional or non-absolute). I also argue that AAP in DDE cases can be explained by PWA. While both actions are morally wrong in standard pairs of DDE cases, I suggest that AAP is best explained by appeal to the agent's motivations, which establish one action as an instance of PWA.

If PWA provides a worthwhile solution to AAP and genuine moral dilemmas, then I require an account of permissibility as a second-order act-assessment because PWA is only possible if an assessment of the action's wrongness is different from the assessment of the action's permissibility, and if the agent's motivations are directly relevant to the permissibility of the agent's action. Such an account must deal with two primary concerns. First, I must explain how an agent's motivations might be directly relevant to the permissibility of an action. If permissibility assessments are act-assessments, one might wonder what role the agent's motivations serve in the assessment of her action. Second, I must explain how we can distinguish between permissibility and moral worth assessments in principle and practice despite the structural similarities when both sorts of assessments utilize the same sorts of considerations.

To address the first concern, I argue that just as an agent's action is directly relevant to assessing her moral worth, the agent's motivations are directly relevant to assessing the permissibility of the agent's action. The agent's motivations establish the background against

which we assess the agent-independent features of the action by specifying which agent-independent features of the action are relevant to our second-order act-assessments. Put very roughly, the difference between the two agents in AAP cases (e.g. the bomber cases) is that the Tactical Bomber's motivations highlight the action's dutifulness as the first-order act-assessment relevant to the action's permissibility, while the Terror Bomber's motivations highlight the action's wrongness as the first-order act-assessment relevant to the action's permissibility. The agent's motivations are relevant to the permissibility assessment because they settle the question of the description under which the action should be assessed for its permissibility.

To address the second question, I show that permissibility, as a second-order act-assessment, is the inverse of moral worth. The relation between the action and the agent, under consideration in permissibility assessments, is the justification relation, which is the inverse of the responsibility relations. Permissibility is an assessment of an action for the quality of its justification relation to the agent's motivations. On my account, not only do the two assessments differ in the target of the assessment (agent versus action), they differ in terms of the directionality of the relation under consideration (responsibility versus justification). While the responsibility relation establishes the extent to which the agent was properly motivated by the morally significant agent-independent features of the action, the justification relation establishes the extent to which the morally significant agent-independent features of the action justify the agent's motivations in acting. I conclude the chapter by showing how blameworthiness and impermissibility come apart, meaning that we can distinguish between second-order moral assessments in key cases both in theory and practice.

## 1. Permissibility

To motivate my account, I begin by rejecting what I call ordinary permissibility accounts. Ordinary permissibility reigns over actions that are typically uncontroversially permissible or impermissible. Morally right actions, supererogatory actions, suberogatory actions, and idle actions tend to fall into the category of ordinary permissible actions. Consider a few examples. When the Good-Willed Grocer treats her customers fairly, she performs an ordinary permissible right action. Using an example from Chapter 2, giving the last remaining cookie to my niece rather than my nephew, when he has had many and she none, is an ordinary permissible supererogatory action, while giving the cookie to my nephew rather than my niece is an ordinary permissible suberogatory action. And, perhaps, keeping the cookie for myself is an ordinary permissible idle action – an action with no particular moral import (as with drinking coffee, taking a short break from writing, watching Star Trek: Voyager rather than Star Trek: Next Generation, etc.).

For ordinary impermissibility, the only actions that typically fit in this category are morally wrong actions. When Chairman HARM implements the program and harms the environment, he performs an ordinary impermissible wrong action.

An ordinary permissibility account is any account that categorizes actions along these divisions, using the following bit of reasoning. Plausibly, actions qualifying as candidates for ordinary impermissibility are those that fail to conform with an operative duty, thereby making them wrong. Such actions are wrong and impermissible. Meanwhile, actions qualifying as candidates for ordinary permissibility are those that do not fail to conform with an operative duty. Such actions are non-wrong and permissible. On an account of ordinary permissibility, one might therefore submit that an action is morally permissible if and only if not wrong.

127

However, ordinary permissibility cannot help us make sense of genuine moral dilemma cases where an agent performs an ordinarily wrong action in the course of performing an ordinarily right action. If the action is permissible, then the action is non-wrong and is, instead, morally right. But then the case is not a genuine moral dilemma. In genuine moral dilemmas, such as those faced by Orestes or Medea in SIDE-TRACK, either action the agent performs is morally wrong. On the ordinary permissibility account, any wrong action is also impermissible. Nonetheless, one might think the agents permissibly perform at least one of the actions. Orestes permissibly kills his mother while Medea permissibly pulls the lever. For any proponent of ordinary permissibility accounts, this pair of claims is incompatible, and the apparent tension between simultaneous act-assessments of wrongness and permissibility leads many to deny the possibility of genuine moral dilemmas.

Even so, I have proposed to develop a framework that can be used to unify even these apparently contradictory assessments. To reconcile our intuitions, I propose a previously ignored category of moral assessment: permissible-wrong action [PWA]. PWA is a moral assessment according to which agents sometimes perform wrong actions that are nonetheless permissible.

Recall the dilemma trilemma, which arises from the following three independently plausible, yet *prima facie* jointly inconsistent, claims.

(1)     The two duties binding Orestes conflict

(2)     The two duties binding Orestes are absolute

(3)     Orestes can resolve his dilemma

In the initial characterization of the dilemma trilemma, I noted that most philosophers ask us to pick between (1) and (2). I am suggesting that we need not choose between (1) and (2) by

128

proposing the possibility of PWA. Instead, we can help ourselves to genuine moral dilemmas stemming from absolute duties and think, nonetheless, that moral dilemmas are resolvable.

*Suppose (1) – (3).*

If (1) is true, the two duties binding Orestes conflict. If (2) is true, the two duties binding Orestes are absolute. If so, then failing to conform with either duty is morally wrong. If so, then Orestes performs a morally wrong action no matter which duty he conforms with. If (3) is true, Orestes can resolve his dilemma. If so, then, even though failing to conform with either duty is morally wrong, Orestes permissibly fails to conform with at least one of his duties. If claims (1) – (3) are all true, permissibility and wrongness must, therefore, be compatible moral judgments about Orestes' action. The possibility of PWA straightforwardly resolves the *prima facie* inconsistency between (1) – (3).

I propose that PWA is only *prima facie* inconsistent, since PWA is a combination of two distinct first- and second- order act-assessments.

**Figure 1: Permissible-Wrong Action**

| | Act-Assessment |
| --- | --- |
| **First-Order** | Wrong Action (1) |
| **Second-Order** | Permissible Action (4) |

Figure 2 illustrates the conceptual niche into which I fit PWA.

While PWA may seem inherently contradictory and unfamiliar, the general category of act-assessment was first (and perhaps only) recognized by Plato. For example, in the *Crito* (47c-48b), Socrates wonders if we are sometimes allowed to act unjustly and sometimes not

129

(as "the many" say), or if, instead, we are never allowed to act unjustly as he had previously argued. Although Socrates quickly convinces Crito that we are never allowed to act unjustly, he recognizes the alternative as a view one must contend with, since a number of considerations, such as obligations to family and friends, are the sorts of considerations "the many" might use to justify their injustices.[74]

As I see it, Plato set the tone for the sort of perfectionist ethical theories present today – the ones according to which we are never permitted to perform morally wrong actions. He argues that just as a life is not worth living when the body is corrupted, a life is also not worth living when the part of us that is injured by injustice and benefitted by justice is corrupted. For, afterall, living is not our priority (at least Plato thinks as much). Instead we must live well.

However, the most pressing moral issues philosophers are concerned with involve primarily those in which the agent or her action are morally imperfect – corrupted somehow. For moral dilemmas and DDE cases, the agent chooses between two wrong actions, ultimately performing an action with significant countervailing moral considerations. The moral blight is present even if we think the agent is ultimately justified in acting. The action involves moral failure, yet it seems like we must sometimes be permitted to fail morally. In which case, we need an account of moral adequacy (rather than moral perfection) to accommodate the permissibility and necessity of living morally imperfect lives. PWA is the first steppingstone to such an account.

---

[74] For an analysis of Plato's view on the subject of moral absolutism, see L.P. Gerson's (1997) "Socrates' Absolutist Prohibition on Wrongdoing" (*Aperion*, 30.4).

A skeptical reader might worry that judgments about wrongness and impermissibility are *prima facie* the same moral assessments, which poses a problem for PWA. An action cannot be both wrong and permissible, since many philosophers define an action as impermissible just in case the action is wrong and permissible just in case the action is not wrong.[75] If so, then, as the newcomer, I bear the argumentative burden for proposing a distinction between them. Yet, my opponent might not be in the position to make any definitional suppositions of this sort (and therefore might not be in any position to foist the argumentative burden on me). When we are introducing a completely novel technical term, we are free to stipulate its meaning. However, neither 'wrong' nor 'impermissible' are completely novel technical terms – both already have a life of their own in ordinary thought and discourse – so we are not free to simply stipulate that they are synonymous, at least in our shared public language.[76] Because we disagree about the correct usage of these terms, a thoroughgoing analysis is required.

I suggest that we issue each type of moral assessment after appealing to different, though overlapping, criteria. If so, then we could reasonably expect judgments about the rightness or wrongness of an action to occasionally diverge from our judgments about the permissibility or impermissibility of the action – and they do. Permissible actions commonly include idle actions, morally right actions, supererogatory actions, and, more controversially, suberogatory actions. These actions respectively include drinking coffee, saving people from drowning, saving families from burning buildings, and driving excessively slowly during rush hour. While it may turn out that the rightness of an action makes it permissible, other sorts of

---

[75] Frederick, D. (2014). "Pro Tanto versus Absolute Rights." *Philosophical Forum*, 45(4).
[76] Thank you, David Mokriski, for thoughts on this observation.

considerations seem to bear on questions of permissibility, since many non-right actions are also permissible. Judgments about rightness and permissibility track different features of actions as predicted by my account.

If we grant that judgments about rightness and judgments about permissibility sometimes diverge, then we have *prima facie* reason to think that judgments of wrongness and judgments of impermissibility sometimes diverge (rather than thinking they always converge). The divergence is possible because they track different features of actions. In particular, I propose that rightness and wrongness are judgments about the properties of an action that determine its moral value, while permissibility is (very roughly) a judgment about an action for the quality of its relation to the motivations for which the agent performed the action. Regardless of the specifics, they are plausibly two different sorts of act-assessments derived by appeal to different features of the action.

Perhaps we, therefore, cannot take for granted that they are the same without argument, and the UMF clearly illustrates a way of conceptualizing the two types of moral assessments differently. If we can conceptualize the two types of moral assessments differently, then PWA is not ruled out. To account for PWA, I reject the ordinary permissibility claim that an action is morally permissible if and only if not wrong. Furthermore, I must develop an account of permissibility according to which these *prima facie* contradictory assessments turn out to be compatible.

## 2. Permissibility and Duty

Because ordinary permissibility accounts weave permissibility into the realm of the obligatory, I will follow suit and begin by making sense of how duty and permissibility relate in my framework. On my account, some morally wrong actions are morally permissible, resulting in assessments of PWA. PWA is only possible if we sometimes permissibly fail to conform with our operative duties, since failing to conform with one's operative duty makes one's action morally wrong. It seems that we permissibly fail to conform with at least one our operative duties in genuine moral dilemmas. If correct, then we learn a valuable lesson about the relationship between duty and permissibility. Failing to conform with one's operative duty and performing a wrong action does not on its own suffice to make one's action impermissible.

Even so, I already observed that philosophers have long thought permissible actions are such because they are non-wrong, making wrongness central to impermissibility assessments. On this point we will all agree: candidacy for impermissibility requires that the agent fails to conform with at least one operative duty. Morally wrong actions remain candidates for impermissibility because they fail from the moral standpoint in a specific way. Namely, the agent acted so as to bring about P, where P is the wrong-making feature of the action, and morality asks us to not bring about P in acting. So, we can see that unpropitious moral assessments, such as assessments of an action's wrongness or candidacy for impermissibility, have to do with certain negative features of the action – that the action failed to conform with an operative duty (among other considerations).

The above suggests that an action's wrongness does not render it impermissible while nonetheless making the action a candidate for impermissibility. In the other direction, PWA illustrates that an action's non-wrongness is not what makes an action permissible. An action

133

does not render permissible by its non-wrongness alone just as an action does not render impermissible by its wrongness alone. Instead, I propose that an action's permissibility has to do with certain positive features of the action, such as its simultaneously conforming with a countervailing conflicting duty (among other considerations), which is a key criterion for making assessments of PWA.

I have suggested that claims about an action's moral permissibility are claims about an action's moral adequacy, where permissible actions are "good enough" morally speaking. A permissible action's being good enough is one measure of moral success or the ways that the action succeeds from the moral standpoint despite any flaws. An action renders permissible when it succeeds from the moral standpoint in specific ways. An action's non-wrongness cites only the way in which the action did not fail from the moral standpoint but not the ways in which the action succeeded from the moral standpoint. So, what makes an action permissible are the other morally significant features present in the action or agent, for which the action is deemed a moral success. Of course, every non-wrong action will render permissible, meaning an action's non-wrongness entails its permissibility. This information is useful for agents deliberating about what to do, but the action's non-wrongness does not fully explain its permissibility. On my account, all of the other morally significant features present in the action or agent contribute to the overall assessment of the action's permissibility, just as they all contribute to the overall assessment of the agent's praiseworthiness or blameworthiness.[77]

---

[77] However, each feature contributes differently depending on the type and target of assessment. We can compare this way of morally assessing actions and agents to the way we might assess the quality of a soup. We can assess the quality of each of the soup's ingredient's independently as we might make first-order moral assessments independently, but, to assess the quality of a soup, we must assess the quality of all of the ingredients taken together. Yet, each of the ingredients might feature in our assessment of the soup differently depending on the type and target of the assessment. For example, the key to a good mushroom soup is the

Further, we learn that conforming with one's duty does not on its own suffice to make an action permissible, just as conforming with one's duty does not on its own suffice to make an action morally right on my account. Otherwise, we would have to think that performing the non-preferred action in a dilemma case suffices to make the action permissible and that seems unlikely. With regards to duties, this means that although conforming with at least one operative duty is permissible on the standard formulation of a dilemma, conforming with the conflicting countervailing duty is impermissible. To assess the permissibility of an action, we must consider the morally significant features of the action or the agent beyond its relationship to duty. This is consistent with the view that permissibility is a second-order act-assessment.

Meanwhile, conforming with at least one operative duty is not required for an assessment of permissibility, since supererogatory actions, suberogatory actions, and idle actions are typically morally permissible even though they are not regulated by any duties. Instead, these actions are assessed on the basis of what we have moral reasons to do or not do.

If so, one might propose that permissibility is regulated by the realm of the erogatory rather than the realm of the obligatory. Proponents of ordinary permissibility accounts are likely to endorse some version of the following claim: an action is morally permissible if and only if the action has, on balance, more and mightier moral reasons for than against. Typically, this is thought to be consistent with a second claim that an action is permissible if and only if not wrong. On such an account, we have a duty prohibiting the performance of actions with, on balance, more and mightier moral reasons against than for, which makes them wrong. I rejected the second claim, as well as the view that duties and moral reasons are tightly

---

mushrooms, but all of the other ingredients matter, whereas the same mushrooms and sundry ingredients might matter more or less when assessing a different kind of soup.

connected, because they fail to accommodate the possibility of PWA in genuine moral dilemmas. Yet, one might agree with me on these points and still propose the first claim is true – that an action is morally permissible if and only if the action has, on balance, more and mightier moral reasons for than against.

Applied to genuine moral dilemmas where nonconformance with either conflicting duty is morally wrong, we might then think Orestes' matricide, or Medea's pulling the lever in SIDE-TRACK, is permissible because pulling the lever or committing matricide is what the agent had, on balance, more and mightier moral reasons for doing than against, despite the action's wrongness. However, if this proposal is correct, then moral permissibility bears no relationship to moral duties. On such an account, nonconformance with at least one operative duty is not required for an assessment of impermissibility. As a result, some non-wrong actions could render impermissible – with suberogatory actions serving as likely candidates, since such actions are actions that we have more moral reason against doing than for.[78]

More importantly, impermissible-right actions are theoretically possible with such a view, since an action's rightness, which is regulated by the obligatory, would bear no relation to the action's permissibility. While my reader might suggest that perhaps I should swallow this tough medicine (and endorse the view according to which permissibility is regulated by the erogatory) for proposing a view allowing other apparently inconsistent moral assessments

---

[78]And, depending on what precisely makes an action impermissible, it could turn out that this move does not provide a general solution to moral dilemmas. In theory, some moral dilemmas involve conflicts between two duties prohibiting actions where all that counts in favor of conforming with one duty rather than the other is that the corresponding action is simply less bad. For example, suppose I am cruelly coerced into torturing someone to death, and I can choose between making the torture more or less painful. Presumably, I am permitted to make the torture less painful even as I am inflicting pure pain without any other benefits. Such an action might have moral reasons against and none for, yet the action is morally preferable over the alternative with greater reason against.

into my framework, such as the possibility of PWA, I would resist this move. We should want something like PWA in our moral frameworks because, as I will show later in this chapter, PWA not only helps to account for genuine moral dilemmas but also helps to make sense of AAP in DDE cases. However, I do not believe we have any reason, apart from symmetry, to want impermissible-right actions in a moral framework, and the solution I propose does not require this outcome.

I have proposed that permissibility is a second-order moral assessment the target of which is an action evaluated by appeal to interactions between at least two first-order moral assessments. This account of permissibility is consistent with the view that assessments of permissibility can be made by appeal to the agent-independent features of the action alone. For example, one might think permissibility is an assessment of the relationship between what the agent has a duty to do and what she has moral reason to do. Such a view could get the correct result for many dilemma cases. Medea acts permissibly in SIDE-TRACK, despite failing to conform with at least one operative duty thereby making her action wrong, because pulling the lever conforms with the duty to not allow the five to die and is furthermore what she had the most and mightiest moral reason to do. If Medea had failed to pull the lever, we might say that her action is impermissible because she both fails to conform with an operative duty making her action wrong, and she performs the action she had on balance the most and mightiest moral reason against doing. However, such a view cannot provide me with a complete account of moral permissibility, since actions regulated by the realm of the erogatory are not adequately accounted for. Only a hybrid account of moral permissibility, where permissibility at least sometimes involves assessing the agent and her action, can accommodate all of these desiderata.

## 3. Permissible-Wrong Action

In Section 1, I proposed the possibility of PWA to make sense of genuine moral dilemmas and demonstrated how we might make sense of this *prima facie* contradictory moral assessment using the UMF. Section 2 explored the relationship between permissibility and duty as I began to develop an account of permissibility as a second-order act-assessment capable of accounting for PWA. This section examines the remaining criteria for assessing PWA and considers additional evidence for the possibility of such an assessment.

Before considering the criteria for assessing PWA, I should first address a reasonable concern regarding my view. If morally wrong actions are sometimes permissible, one might wonder whether my view will allow bad people to freely get away with failing to conform with their duties. I contend that my reader need not fret over this concern. Ordinarily wrong actions are only ever justified in dilemma cases, where, despite failing to conform with one duty, the action nonetheless simultaneously conforms with a countervailing conflicting duty. As already mentioned, any action failing to conform with duty fails from the moral standpoint, making it a candidate for impermissibility. However, in dilemma cases, the action's failure is counterbalanced by its moral successes – the action also conforms with duty. So, although the action is a candidate for impermissibility, the action's permissibility is not ruled out. Outside of dilemma cases, an action failing to conform with an operative duty will render wrong and impermissible, which is, of course, the intuitively correct result. Meaning, my reader need not worry about the permissibility of ordinary failures to conform with one's duty. Such actions are impermissible because they fail from the moral standpoint in a number of ways. In all cases, they are wrong, nonconforming with a countervailing conflicting duty, and have moral reasons counting against them.

This brings me to the second requirement for making assessments of PWA. A wrong action must, in addition, be the morally preferred action in a dilemma scenario. Although any morally wrong action inevitably fails from the moral standpoint, the morally preferred action is the one that fails the least. Yet, PWA is not settled solely on the basis of these two conditions (conforming with an operative duty and performing the morally preferred action) because my account must also accommodate AAP in DDE cases, as described in Chapter 1.

As it turns out, PWA not only helps to resolve genuine moral dilemmas, PWA helps to explain AAP. Recall that both the Tactical Bomber and the Terror Bomber conform with duty and perform the morally preferred action (bombing the munitions factory, thereby ending the war early and saving lives on both sides), but the Tactical Bomber performs a PWA while the Terror Bomber does not. When typically discussed, the two agents perform the (structurally) same actions and differ only in their motivations.[79] Opponents of the DDE likely agree that agent motivations support asymmetrical moral assessment in some DDE cases. We correctly judge that the Terror Bomber is motivated by ill will while the Tactical Bomber is not.[80] However, they disagree about the category of moral assessment in play. Because they correctly believe that agent motivations are irrelevant to judgments about an action's rightness

---

[79] The Terror Bomber believes bombing the munitions factory will result in civilian casualties, which he hopes, in turn, will demoralize the enemy and achieve his aim, while the Tactical Bomber believes bombing the munitions factory will cripple the enemy's war machine and achieve his aim.

[80] Recall that the key difference between the Tactical Bomber and the Terror Bomber is that the Tactical Bomber instrumentally desires to perform an action and so brings about the harmful consequences, while the Terror Bomber instrumentally desires to bring about the harmful consequences and so performs the action, for both cases, in order to end the war early.

or wrongness, they conclude that the negative moral assessment that proponents of the DDE are talking about are really agent-assessments only.[81]

On this reading of the disagreement (from the standpoint of DDE opponents), the best way of framing the debate surrounding the DDE is as a disagreement about how to categorize permissibility in the FMD. The question then is whether judgments about permissibility are act-assessments or agent-assessments. We generally agree that permissibility is an act-assessment, so we might conclude that proponents of the DDE must be talking about an agent-assessment. And this seems plausible. Any wrong action for which the agent is blameworthy will not be justified, even if the action is the morally preferred one – as in the case of the Terror Bomber. The motivations for which an agent is blameworthy are motivations that fail from the moral standpoint.

However, by introducing the UMF I can propose an alternative, and perhaps more charitable, reading of the disagreement. The debate surrounding the DDE is best framed as a disagreement about whether judgments about permissibility are first-order or second-order act-assessments. On this reading, all parties to the debate successfully identify a legitimate object of moral assessment and both parties are correct in thinking that that assessment is an act-assessment. They fail to distinguish first- and second-order act-assessments.

I address DDE cases as follows. All parties to the debate should agree that there is a sense in which both actions in pairs of DDE cases should be assessed symmetrically (insofar as we would describe the actions in the two cases the same way). Opponents of the DDE propose that such actions are assessed symmetrically without regard for the agent's

---

[81] A. Markoc (2017) calls this the Objection from Confusion in "Intentions and Permissibility: A Confusion of Moral Categories" (*Journal of Value Inquiry*, 51).

motivations in acting. My account obliges this recommendation by identifying the first-order act-assessment applicable to the two cases. I propose, plausibly, that the killing of civilians is morally wrong and both act wrongly in bombing the munitions factory.[82] To judge the killing of civilians morally wrong, I need not know anything about what each agent intended. Judgments about wrongness are, once again plausibly, judgments about agent-independent features of action. So long as I have access to symmetrical assessments of wrongness in DDE cases, I am positioned to accommodate the view that two actions that are structurally the same should be judged in the same way.

Further, all parties to the debate agree that the agent motivations in pairs of DDE cases generate asymmetrical moral assessments. Surely, we morally assess the TERROR BOMBER's motivations negatively while assessing the Tactical Bomber's motivations positively. Proponents of the DDE propose the agent's motivations in acting bear on a moral assessment of the action itself. My account obliges this recommendation by identifying the second-order act-assessment applicable to the two cases. I propose, as do DDE advocates, that judgments about permissibility are judgments about actions appealing to agent motivations.

Conforming with the patterns of AAP in DDE cases, I agree that the Tactical Bomber acts permissibly while the Terror Bomber acts impermissibly. The Tactical Bomber's action is an instance of PWA and, the possibility of PWA offers the most satisfying solution to the puzzle surrounding asymmetrical assessments of permissibility in DDE cases.

---

[82] Here, I am supposing that DDE cases are genuine moral dilemmas where the two bombers have some duty to not kill the civilians but that other morally relevant considerations, such as a conflicting countervailing duty to end the war early, might outweigh this concern.

**Figure 3: Assessing DDE Cases**

|  | TACTICAL BOMBER | TERROR BOMBER |
|---|---|---|
| **First-Order Act-Assessment** | Wrong (1) | Wrong (1) |
| **Second-Order Act-Assessment** | Permissible (4) | Impermissible (4) |

Figure 3 shows how we can assess DDE cases differently using the UMF.

If so, then philosophers have already identified two distinct moral phenomena, which are legitimate targets of moral assessment, and we are in need of two distinct act-assessments in order to make distinct judgment about each of these moral phenomena. While this move is unlikely to satisfy opponents of the DDE, I am in a position to vindicate both sets of intuitions regarding the legitimate targets of moral assessment in DDE cases. If judgments about rightness and wrongness are first-order act-assessments involving only agent-independent features of actions, then we may yet use "permissibility" for second-order act-assessments involving agent motivations. This distinction in the use of the two concepts is overall consistent with current deployment in some existing theories, as with proponents of the DDE, but leaves open the possibility of novel moral assessments about PWA.

## 4. On the Direct Relevance of Motivations to Permissibility

The aim of the previous section was not to vindicate the DDE. Although some version of the DDE will be true on my account, the principle is purely descriptive and cannot ground judgments about an action's permissibility in PWA cases (non-justificatory) nor is it the sort of principle an agent can use when deliberating about what to do (non-normative). In Section

143

3, I argued that AAP in DDE cases can be explained by PWA. If PWA provides a worthwhile solution to genuine moral dilemmas as well as AAP, then I require an account of permissibility as a second-order act-assessment. PWA is only possible if an assessment of the action's wrongness is different from the assessment of the action's permissibility and if, in line with proponents of the DDE, the agent's motivations are directly relevant to the permissibility of the agent's action. Permissibility assessments are therefore structurally similar to assessments of moral worth on my account. Such an account must deal with two primary concerns, the first of which I deal with in this section. I must explain how an agent's motivations might be directly relevant to the permissibility of an action. If permissibility assessments are act-assessments, one might wonder what role the agent's motivations serve in the assessment of her action.

As Scanlon put it, permissibility assessments primarily have to do with the effect of an act on people. Although an agent's motivations determine an agent's action and therefore partially determine the effect of an action on people, an agent's motivations might seem to be relevant only insofar as they help us to predict what an agent will do and what effect an agent's action will have. This is what Scanlon calls the predictive significance of intention.

Scanlon notes that an agent's intentions often do not fully determine the effect of some action because we are often mistaken, misguided, or ignorant about things that can impact on the action and generate effects different from those we intended. As a result, it might seem that the effect of an action is more relevant to the assessment of an action's permissibility than the agent's motivations in acting because changes in the effects of actions can impact on the permissibility assessment when the agent's motivations are held constant. So, although an agent's motivations bear on the effect and, therefore, the permissibility of her action, the

144

question is whether an action's permissibility can ever depend on differences in the agent's motivations when we hold the action's effect on people constant. This is the condition on which Scanlon proposes an agent's motivations could be directly relevant to the permissibility of an action.

As I stated in the Introduction to this chapter, I propose that just as an agent's action is directly relevant to assessing her moral worth, the agent's motivations are directly relevant to assessing the permissibility of the agent's action. The agent's motivations establish the background against which we assess the agent-independent features of the action. More specifically, I am proposing that (i) the reasons for which the agent actually acted specify which of the morally significant normative features of an action are relevant to assessing its permissibility.

As with a soup, we might find it difficult to assess the overall quality without information about the kind of soup under assessment; this information helps to highlight the features of the soup most relevant to the assessment. So, to assess the quality of a soup, we need the recipe. In a mushroom soup, the type and quality of the mushroom is most important to the overall assessment of the soup, while the onion in a French onion soup will matter most. Yet, either ingredient might weigh equally in our assessment of a vegetable soup. The agent's motivations provide the recipe for the action, enabling us to determine the which features of the action justify the agent in acting.

I call the morally significant agent-independent features of the action specified in (i) the justificatory set. Only the contents of the justificatory set get to play a role in justifying or unjustifying the action. When the agent is motivated by the features of the action that make it wrong, the features privileged in our assessment of the action's permissibility are those

145

relating to the wrongness of the action. I call the relationship between the justificatory set and the agent motivations constraining them the justification relation. This section identifies key structural features of the justification relation, Chapter 5 develops an account of what makes actions justifiable and specifies what actually does the justifying.

To assess the difference between the two agents in AAP cases (e.g. the bomber cases), the motivations relevant to the assessment of the action's permissibility are only those that settled the question of whether to drop the bomb. In Chapter 2, I noted that the Tactical Bomber and Terror Bomber plausibly agree that bombing the munitions factory is the right thing to do because bombing the munitions factory, ending the war early, and killing civilians is the morally preferred solution to the dilemma. However, the Terror Bomber sees the civilian deaths as a part of what makes the action good, while the Tactical Bomber sees the killing of civilians only as something that makes the action wrong but nonetheless correctly believes these considerations are outweighed by saving the lives saved on both sides.

These motivations pick out any corresponding morally significant agent-independent features of the action, and only these features of the action can justify the agent in acting. So, this difference in motivations highlights the primary difference between the set of normative considerations playing a role in justifying the two bombers in acting. The very feature of the action that makes it wrong is specified in the justificatory set of the Tactical Bomber as a normative consideration in favor of acting while it is correctly highlighted as a reason against acting in the Tactical Bomber's justificatory set.

An action is proven permissible when the agent's motivations align in the right way with the morally significant normative features of the action. The agent's motivations align in the right way when they pick out the features of the action that make it justifiable without

146

picking out the features that make it wrong (if the action is wrong). This amounts to (ii) an assessment of the alignment between the agent's motivations and the justificatory set and (iii) an assessment of whether any inappropriate alignment is due to the inclusion of an unjustifying feature in the content of the justificatory set or the exclusion of a justifying feature in the content of the justificatory set.

Regarding (ii), if things go well, the agent aimed at performing the morally preferred action in a dilemma case and was not motivated by any aspect of the action's wrongness, as with the Tactical Bomber. The Tactical Bomber's motivations highlight the action's status as the morally preferred solution to the dilemmas and its corresponding dutifulness as central to our assessment of the action's permissibility. Because the action is justifiable on these grounds, the action is morally justified and therefore morally permissible. Regarding (iii), because considerations relating to an action's wrongness are always privileged in our moral assessments, at least on my account, the Terror Bomber's motivation to kill the civilians highlights the action's status as morally wrong as central to our assessment of the action's permissibility. Because the action is not justifiable on the basis of its wrongness, the action's wrongness therefore undermines the agent's justification in acting.

For both bombers, the action's wrongness remains relevant to the assessment of the action's permissibility. If the Tactical Bomber had been inappropriately motivated by any feature of the wrong action, then the action's wrongness similarly would have undermined his justification in acting. However, contrary to standard views on moral justification, I do not hold that the Tactical Bomber requires a justification for performing a morally wrong action. I hold that all permissible actions, even if non-wrong, require moral justification. Instead, the relevance of the action's wrongness in the permissibility assessment serves solely to

undermine the agent's justification in performing a potentially justifiable action. An agent's justification in performing a non-wrong action will simply never be undermined in this way.

However, I acknowledge that we only ever really discuss moral justification in situations where the agent performs an ordinarily wrong action in the course of performing an ordinarily right action. For ordinary wrong actions, such as those perpetrated by the Malevolent Merchant or Chairman HARM, when we know that the action is wrong, we know that the action is morally impermissible because such actions have no features for which the agent can permissibly act. The action is not morally justifiable. The action's wrongness seems to settle the question of the action's permissibility for us. The same sort of reasoning applies to right actions. Because sometimes it seems like an agent is justified in performing an ordinarily wrong action in the course of performing an ordinarily right action, and because actions in these cases are ordinarily the targets of apparently conflicting moral assessments, I understand why it might make sense to ask whether the action's ordinary rightness justifies its ordinary wrongness, at least in part, as a way of resolving the apparent contradiction in the assessments.

Instead, I propose that the agent-independent features of an action cannot settle the question of the correct description under which the action's permissibility is to be assessed in such cases, the bulk of which are moral dilemmas.[83] The action's wrongness cannot settle the

---

[83] I borrow Anscombe's phrase "under a description" as a general way of acknowledging that the very same action is sometimes appropriately characterized by different descriptions. Upon observing a man at the well, we might ask of him what he is doing. An appropriate answer to this question could be that he is moving his arm up and down, that he is pumping water from the well, that he is transporting poisoned water to the house, or that he is killing the Smiths. The action is wrong under the relevant, true description "killing the Smiths", so the action is wrong. See G. Anscombe's (1957) *Intention* (Oxford: Blackwell), also see G. Anscombe's (1979) "Under a Description" (*Noûs*, 13.2).

question of the action's permissibility in dilemma cases because the action is simultaneously wrong and dutiful. The Tactical Bomber and the Terror Bomber perform the (structurally) same morally wrong yet justifiable action, and we aim to distinguish them.

When an action's permissibility renders, we have settled the question as to which description best fits the action. The action's permissibility must be assessed "under the description" of its being wrong or dutiful but not both, since one is ordinarily considered permissible while the other is not, and the action cannot be both permissible and impermissible. In moral dilemmas generally and DDE cases in particular, further appeal to the agent-independent features of the action do not settle the question because the action consists of those features that ordinarily settle the question of the action's wrongness and those that ordinarily settle the question of the action's rightness. Because there are no additional agent-independent facts to settle the question of which description best fits the action for the sake of making the permissibility assessment, we can only appeal to the agent's motivations, which is plausible given that the agent's motivations are the action's "recipe." The agent's motivations settle the question of which description of the action is privileged in the justificatory set and helps to explain why an account like mine must privilege the action's wrongness as a defeater for the agent's justification in acting when included in the agent's justificatory set. Otherwise, we cannot solve the problem of the apparent contradiction or make sense of AAP.

A critical reader might worry I am proposing that the agent's motivations should serve as some sort of moral standard to which actions should be held. That is – we should take on a stance of radical subjectivism to see the action as good if the agent intended for it to be good and bad if the agent intended for it to be bad. I am not making any such proposal. On the UMF,

any act comprises all required agent-independent features necessary for making a permissibility assessment prior to examining the relationship between those features and the agent's motivations. So, a *post hoc* examination of the agent's motivations, cannot change which agent-independent features of the act are instantiated. Furthermore, only the agent-independent features of the action determine the appropriate first-order moral assessments. So, I am not suggesting that motivations change or modify the moral value of any agent-independent features of action. Nor do motivations contribute to an action's candidacy for permissibility or impermissibility. These questions are settled at the level of first-order act-assessments.

Instead, I propose that motivations serve only to describe the set of agent-independent features of the action central to the permissibility assessment. I am not proposing that the agent's motivations place limitations on which features of the action are morally assessible – full stop. Morally significant features of actions excluded from the justificatory set are still relevant to and morally assessible by other first- and second-order moral assessments. For example, agent motivations examined in moral worth assessments can be limited by morally significant features of an action not under consideration in the permissibility assessment. The wrongness of the TACTICAL BOMBER's action is, for example, the reason he is not a candidate for praise.

As with assessments of moral worth, some motivations matter more in our permissibility assessments than others, again, depending on the type and target of the assessment. In the next chapter, I will say more about justificatory thresholds and the way in which an agent must be motivated in cases of PWA. This discussion will include worries about how I handle cases where the agent is partially motivated by the apparent wrong-making

150

features of the action, yet the action clearly appears to be morally justified and therefore permissible.

## 5. Impermissibility & Blameworthiness

In Section 4, I provided a model addressing the first of two primary concerns for a hybrid account of permissibility. On my account, permissibility can depend on differences in the agent's motivations when we hold the action's effect on people constant, and an agent's motivations are, therefore, directly relevant to the permissibility of an action. This model also offers a solution to a second worry about a view according to which permissibility and moral worth are structurally similar. Namely, that we would have no basis for distinguishing the two types of assessments in principle because they seem to be made by appeal to the very same set of considerations.

I have provided different, though parallel, models for understanding how the assessments are made. Very generally, the object of a moral worth assessment is the agent, and the agent is assessed by appeal to the relation in which she stands to the action where facts about the action constrain our assessment of the agent. Meanwhile, the object of a permissibility assessment is the act, and the act is assessed by appeal to the relation in which it stands to the agent where facts about the agent's motivations constrain our assessment of the action. In application, the relata, the objects of the assessment, the relevant assessment properties, and the assessments themselves are different, despite any apparent similarities. With regard to the relata, for example, the inclusion of an action's wrongness in the justificatory set is contingent on the specifics of the agent's motivations in acting, while the action's wrongness is always central to a moral worth assessment, even when both are assessments appealing to the very same action. So, I have provided a principled way of distinguishing second-order moral assessment.

So far, I have dealt with two major concerns –how an agent's motivation could be directly relevant to a permissibility assessment and how to distinguish between moral worth and permissibility as second-order moral assessments. However, a critical reader might worry that linking permissibility assessments to assessments of the agent's motivations will make permissibility assessments indistinguishable from assessments of moral worth in practice if not in theory. Specifically, a critical reader might wonder whether impermissibility and blameworthiness ever come apart. Although I need not defend that these assessments ever diverge to establish the plausibility of my account thus far, I argue that the agent's motivations explain impermissibility differently from how they explain an agent's blameworthiness, and these differences sometimes result in diverging assessments. When an agent is blameworthy, the badness of the action helps to explain the badness of the agent, otherwise we would know everything we need to know to assess the agent by looking at the agent alone when making the agent-assessment. When an action is impermissible, the agent's bad motivations help to explain the badness of the action, otherwise we would know everything we need to know to assess the action by looking at the agent-independent features of the action alone.

When assessing an agent's motivations, acting, for example, from the desire to harm someone is always morally criticizeable. However, acting from a lack of moral concern is not always morally criticizeable. If I donate to charity because I desire acclaim as a philanthropist, then I act with a lack of concern for the well-being of others. Yet, my motivations themselves are not morally criticizeable. Morality does not ask of us that we always act as saints, and acting from selfishness or self-love is typically permitted as long as we do not do so at the expense of others.

Consider another familiar case, Kant's PRUDENT GROCER.

**PRUDENT GROCER:** The prudent grocer always treats his customers fairly. He always treats his customers fairly because it is prudent to do so. If this prudent grocer were to cheat his customers, he would probably lose business, so he always treats his customers fairly because it is good for profit.

The Prudent Grocer treats his customers fairly because he desires profit. While treating one's customer's fairly is what he ought to do, he nonetheless acts without concern for fairness when he acts solely on his desire for profits. While the Prudent Grocer is not praiseworthy, neither is he blameworthy. However, when Chairman HARM acts solely on his desire for profits, implementing the program and harming the environment thereby, his motivations are morally criticizeable – unlike the Prudent Grocer's motivations. Yet, he acts for the same kinds of reasons without regard for the morally relevant features of the actions. If acting without regard for the morally relevant features of one's action is not itself morally criticizeable, then some feature of the action renders Chairman HARM's motivations criticizeable. The situation is reversed for assessments of impermissibility. Some feature of the agent's motivations renders the action impermissible where we otherwise think the action is the one the agent ought to perform because the action is the morally preferred solution to the dilemma.

Before explaining how these assessments are different, I will first illustrate how we might distinguish impermissibility and blameworthiness. Consider the death of BALDUR.

**BALDUR:** Baldur, most beloved of the Norse gods, dreamt of ill befalling him. So, his mother Frigg aided him. She extracted, from each thing in the universe, a promise to not harm Baldur. She extracted a promise from everything except the mistletoe. When Loki discovered Baldur's vulnerability to mistletoe, he

constructed a spear from it and convinced the blind god Hoðr to throw the spear

at Baldur in order to honor him. Hoðr killed Baldur with the mistletoe spear.

Regarding the death of Baldur, we can ask at least two questions. Did Hoðr act impermissibly in slaying Baldur? Is Hoðr blameworthy for slaying Baldur?

Hoðr acts impermissibly in slaying Baldur. Hoðr's action, the throwing of the spear, was non-accidental. Hoðr intended to throw the spear at Baldur in order to honor Baldur's impenetrability. Despite not intending to harm Baldur, he nonetheless intentionally performed the action that killed him. Moreover, Hoðr intentionally performed a morally wrong action.[84] For plausibly, slaying Baldur is morally wrong. Afterall, Baldur's life was valuable to the other gods, and Frigg undertook the astronomical task of coercing universal benevolence towards him. Each thing in the universe was therefore bound to not harm Baldur. Even so, Hoðr slayed him.

Importantly, Hoðr was under no simultaneous obligation to throw the spear at Baldur, meaning he intentionally performed a morally wrong action for which he had no

---

[84] A critical reader might wonder whether I am making a mistake. Hoðr did not intend to kill Baldur, so while he acted intentionally he did not intend to perform a morally wrong action. Slaying Baldur was an unintended side-effect of Hoðr's action for which he perhaps cannot be held responsible. If this is a worry about Hoðr's blameworthiness, then I agree. Whether Hoðr intended the wrongness is a question for judgments about his blameworthiness and not judgments about the impermissibility of the action. Questions about permissibility, first and foremost, are questions about actions – whether they are indeed actions as such (performed intentionally) and whether they conform with some duty (or whatever the particular moral theory calls for). So, if we agree that Hoðr acted intentionally in throwing the spear, then Hoðr's action is an action properly attributable to him. Using the UMF, we assess the wrongness of the action, which is a first-order moral assessment, by appeal to the morally relevant features of the action itself. We can therefore assess the wrongness of Hoðr's action without regard for his motivations in acting. He failed to conform with the divinely-derived duty to not kill Baldur and failing to conform with a duty typically makes one's actions wrong. Thus, Hoðr acted intentionally and his action was indeed morally wrong. This is all I mean when I say that he intentionally performed a morally wrong action.

countervailing moral considerations. He threw the spear in order to honor Baldur's impenetrability, which was a nice thing to do, but something he had no duty to do. The action was therefore not a candidate for permissibility. Thus, we can say that Hoðr intentionally performed a morally wrong action, and he could not have been motivated by morally relevant considerations that could outweigh any obligation he has to not kill Baldur. Typically, any agent who intentionally performs a morally wrong action, for which he had no countervailing moral considerations, acts impermissibly. We might therefore suppose Hoðr straightforwardly acts impermissibly.

However, we may yet believe Hoðr is not blameworthy for killing Baldur. An agent is blameworthy if she performs a morally wrong action and she does so out of ill will or while indifferent to moral considerations. While Hoðr intentionally performs a morally wrong action, he is not motivated by maliciousness. Hoðr desired to honor Baldur – not to harm him. Moreover, Hoðr does not act from moral indifference. Frigg told the other gods that nothing could harm Baldur, so, when Hoðr threw the spear, he reasonably supposed nothing counted against so doing. Thus, while Hoðr intentionally performed a morally wrong action, he does not perform the action from the sorts of motivations that would make him blameworthy.

Instead we might think Hoðr acted in [non-culpable] ignorance and hence has a fully exculpating excuse. He had no reason to suppose the spear could harm Baldur. Afterall, Frigg assured him all was well. Yet, even if Frigg had revealed to Hoðr that the mistletoe made no promises for Baldur's safekeeping, not even Frigg believed mistletoe could harm anyone, which is why she never bothered to ask it to make such a promise. So, Hoðr had no reason to believe anything made from mistletoe could harm Baldur. And even if he suspected mistletoe could harm Baldur, blind Hoðr could not see that the spear was made of mistletoe. Loki tricked

him. Yet, he had no reason to suppose Loki was acting with malicious intent. Afterall, Baldur was the most beloved of gods. Moreover, we might even suppose he acted from good will, since he threw the spear out of his desire to honor Baldur. We might therefore categorize Hoðr as follows.

**Figure 4: Assessing Hoðr**

|  | Act-Assessment | Agent-Assessment |
|---|---|---|
| **First-Order** | Wrong (1) | Good Will (2) |
| **Second-Order** | Impermissible (4) | Not Blameworthy (3) |

Figure 4 shows how we might assess Hoðr using the SMF. If so, impermissibility and blameworthiness come apart in the case of Baldur.

**Conclusion**

This chapter has dealt with permissibility as a second-order act-assessment. The primary aim of this chapter has been to develop an account of permissibility consistent with the possibility of genuine moral dilemmas. One desideratum for such an account is the possibility of PWA. I argued that not only does PWA help us to make sense of moral dilemmas generally, PWA helps us to make sense of AAP in DDE. So, I argued further that we can make sense of PWA using a hybrid account of permissibility, according to which an agent's motivations are directly relevant to permissibility as a second-order act-assessment. The features of an action that make it a candidate for permissibility only ever justify the agent when those features are picked out by the agent's motivations and included in the justificatory set. The relation between the agent's motivations and the justificatory set is the justification relation, which plays a central role in my account of permissibility yet remains unaddressed. The next chapter develops what I call the adequacy account of justification – a requirement for successfully defending the possibility of PWA.

# Chapter 5: The Adequacy Account of Justification

**Introduction**

In Chapter 4, I suggested that permissible-wrong action [PWA] is the most straightforward explanation for moral dilemmas, as well as asymmetrical assessments of permissibility [AAP] in Doctrine of Double Effect [DDE] cases. PWA is incompatible with ordinary permissibility accounts, which typically subscribe to the following claim: (A) an action is morally permissible if and only if the action is not morally wrong. I believe that (A) is a corollary of another claim: (A*) an action is morally justified if and only if the action is not morally wrong. (A) follows from (A*) where an action's permissibility is dependent on its justification and an action is morally permissible if and only if the action is morally justified. So, to secure a theoretical basis for PWA, I require a non-vindicatory account of moral justification – an account of moral justification according to which the action is not wholly vindicated or shown to be non-wrong such that morally wrong actions are sometimes morally justified.

As things stand, most existing accounts of moral justification are vindicatory accounts of moral justification, the most widely held of which are objective or deeds account of moral justification [DAJ]. DAJs are committed to another familiar claim: (B) an agent's motivations in acting are irrelevant to the moral justification for the action. Claim (B) is similarly incompatible with my account of moral permissibility, since I hold that the features of the action that actually do the justifying, thereby rendering the action permissible, are specified by the agent's motivations in acting. I require this sort of hybrid view to fully account for AAP in DDE cases.

Chapter 4 demonstrated that we need not think (B) is true, so this chapter therefore takes issue with vindicatory accounts of moral justification and the commitment to (A*), situating the DAJ as paradigm of my primary opposing view in Section 1. However, justificatory vindication and not the DAJ will be the focus of the chapter. In Section 2, I argue

we have good reason to think vindication is not the function of moral justification because it rules out moral dilemmas and fails to make sense of moral residue.

So, I offer an alternative to vindicatory accounts of moral justification in Sections 3 and 4. I begin by suggesting that the purpose of justification is to overcome the problem of impermissibility and not the wrongness objection. This section develops my account of what makes actions justifiable and what ultimately does the justifying. I then provide an alternative model to justificatory vindication, which I call an adequacy account of moral justification. Adequacy accounts of moral justification employ what I call a proving method, the aim of which is to show an action is good-enough despite its flaws. In this section, I explain how my account of justification is an adequacy account. In Section 5, I deal with objections. I then conclude.

## 1. Understanding a Deeds Account of Justification

Proponents of the DAJ typically subscribe to two familiar claims: (A\*) an action is morally justified if and only if the action is not morally wrong; (B) an agent's motivations in acting are irrelevant to the moral justification for the action. Claim (B) is the defining characteristic of the DAJ, while claim (A\*) is the defining characteristic of all vindicatory accounts of moral justification – a part of the broader theoretical underpinning of the DAJ. In this section, I will focus on why proponents of the DAJ subscribe to these claims. Along the way, I will articulate the source of the justificatory puzzle for proponents of the DAJ. Taken together, I hope to explain why theories of justification in the literature have taken on the limited scope they have - with the DAJ serving as the frontrunner in our moral theorizing.

Proponents of the DAJ subscribe to moral objectivism about justification and permissibility assessments for at least three reasons. First, as Scanlon put it, permissibility assessments primarily have to do with the effect of an act on people. As a result, he thinks the effect of an action is more relevant to the assessment of an action's permissibility than the agent's motivations in acting because changes in the effects of actions can impact on the permissibility assessment when the agent's motivations are held constant. Second, we already have a category of assessment that appeals to the agent's motivations – moral worth. So, they suspect that any attempt to include the agent's motivations in a permissibility assessment, is really an assessment of the agent's moral worth. Third, they think the purpose of moral justification is to vindicate a wrong action or show it to be non-wrong  in such a way that the action's all-things-considered rightness (or something of the sort)is what justifies actions, so the agent's motivations seem irrelevant to the action's moral justification.

We can see evidence of the latter two point in how philosophers typically explain the difference between the role of justifications and excuses in moral assessments.[85] Excuses are the sorts of things that exonerate agents from blame, usually by showing that the agent was not morally responsible for the wrong-doing. To assess whether an agent is excused, we assess the role the agent's motivations played in bringing about a morally charged outcome. Justifications are the sorts of things that vindicate an agent's action, usually by showing that the action was not an instance of wrong-doing in the first place.

For example, I might attempt either to provide an excuse to exonerate me from blame or a justification to show that I did nothing wrong, if I shoot a person, thereby killing her. To provide an excuse, I might claim that I believed the gun was fake and that the shooting was, therefore, an accident. To provide a justification, I might claim that the other person was shooting at me and that the shooting was, therefore, an instance of permissible self-defense. When I offer an excuse, I do not attempt to show that shooting and killing the person was not wrong. Instead, I attempt to show that my agency was not featured in the action in a morally relevant way. However, when I offer a justification, I may very well acknowledge that my agency was featured in the action in a morally relevant way – I intended to shoot and kill my attacker. Instead, I attempt to show that this instance of intentional killing is not wrong. Thomson, for example, claims that when killing in self-defense is permissible, the attacker

---

[85]Austin, J. L. (1971). "A Plea for Excuses." *Philosophy and Linguistics* (Macmillan Education UK, 1971); Greenwalt, K. (1986). "Distinguishing Justifications from Excuses." *Law and Contemporary Problems*, 49.3; Botterell, A. (2007). "Why We Ought to be (Reasonable) Subjectivists about Justification." *Criminal Justice Ethics*, 26.1.

forfeits her right not to be killed. If this is so, then I had no duty to not kill my attacker, and I am therefore justified in doing so [86]

Suppose, for now, this account of moral justification is correct. Typically, we think failing to act in accordance with one's duty is morally impermissible. Our duties are usually only deflected or defeated when it would be extremely difficult to satisfy the duty or when the duty is in competition with another duty. So, if an agent permissibly fails to act in accordance with her duty, it must be either because the duty was too burdensome or because the duty was in competition with another duty. When a duty is too burdensome, the agent is exempted from conforming with it because the failure to conform with the duty is not a failure on the part of the agent. So, only agents bound by duties of the latter sort are puzzling for philosophers.

Now, we can put the problem a little more precisely. We are normally interested in moral principles that allow us to assess the permissibility of failing to conform to our ordinary duties in the course of conforming to other duties. Or, as I tend to put it, we are interested in moral principles that allow us to assess the permissibility of performing a *pro tanto* wrong action in the course of performing a *pro tanto* right action. And many philosophers seek to identify principles appealing only to the agent-independent features of the action – leading to claim (B) – because the purpose of justification is to vindicate the action rather than exonerate the agent who performed it – leading to claim (A*).

---

[86] To be clear – I do not endorse the distinction as articulated here, but most philosophers working in this area appear to accept it. Below, I suggest that justification is not vindicatory as specified above.

## 2. Justificatory Vindication and Moral Dilemmas

In Section 1, I presented a general overview of the motivations underlying the DAJ. I articulated an intuitive distinction between justifications and excuses, which I proposed underlies the motivation for the DAJ. Justification is vindicatory while excuses are exculpatory. Vindication places objectivist constraints on theories of justification, grounding theories like the DAJ. In this section, I argue we have good reason to think vindication is not the function of moral justification. Denying that justification is vindicatory then gives us reason additional reason to doubt accounts like the DAJ, since it undercuts the project grounding it.

Specifically, I argue we have reason to prefer a view that does not support the project of morally vindicating actions. First, I argue that vindicatory accounts of moral justification are incompatible with genuine moral dilemmas, so any philosopher invested in the possibility of genuine moral dilemmas should prefer an alternative account. Second, I argue that concerns about moral residue, when we ordinarily talk about one duty defeating another duty, suggest that moral justification does not wholly vindicate actions. Some facet of the action's moral failure shadows the agent responsible for the wrong-doing. These considerations lead me to question whether moral vindication of action is the aim of moral justification.

Showing that vindicatory accounts of moral justification are incompatible with the possibility of genuine moral dilemmas is rather straightforward. Recall that, in a genuine moral dilemma, the agent faces a conflict between two absolute duties. Any action failing to conform with an absolute duty is morally wrong because there are no exceptions to such duties. In genuine moral dilemmas, any action the agent performs is morally wrong because any action the agent performs fails to conform with one such duty. Yet, in genuine moral

165

dilemmas, it typically seems as though there is a morally preferred solution – an action the agent is justified in performing. If the agent is justified in performing the morally preferred solution to a genuine moral dilemma, then the agent is justified in performing a morally wrong action. Vindicatory accounts of moral justification subscribe to (A*), which is incompatible with the claim that agents are sometimes morally justified in performing morally wrong actions. So, we are forced to choose between justificatory vindication and genuine moral dilemmas.

While some philosophers will reject the possibility of genuine moral dilemmas, many find moral dilemmas compelling, so good moral theorizing requires that we explore the relevant philosophical space. Moreover, the best theory of moral justification will be the one that can help make sense of the broadest range of moral phenomena.

If vindication cannot help us makes sense of genuine moral dilemmas, AAP in DDE cases, and, as I will now argue, moral residue, then we should consider whether another sort of account can. Consider the problem of moral residue starting with the case of the ARSONIST.

> **ARSONIST:** An arsonist sets fire to a farmer's field because he wanted to cause
>
> the farmer to lose his livelihood.[87]

When the Arsonist sets fire to the farmer's field, he performs an action that is ordinarily morally wrong. One might think the farmer has a right against the Arsonist that he not damage the farmer's property, and the Arsonist has a corresponding duty to not damage the farmer's

---

[87] Robinson, P. H. (1975). "A Theory of Justification: Societal Harm as a Prerequisite for Criminal Liability." *UCLA Law Review*. 23.266; Robinson, P. H. "Competing Theories of Justification: Deeds v. Reasons." *Harm and Culpability*, Eds. Andrew Simester, A.T.H. Smith (Oxford Univ. Press, 1996): 46-70.

property. On a standard view, setting fire to the field is a *pro tanto* wrong action, which has not been defeated.

Yet, as we know, we are sometimes permitted to perform *pro tanto* wrong actions. Consider another example.

> **VILLAGE HERO:** A forest fire was approaching a village and the only way to save the village was to create a fire break by setting fire to the farmer's field. The village hero sets fire to the field, creating a fire break in order to save the village.[88]

When the Village Hero sets fire to the field, she performs an action that is ordinarily morally wrong as in the case of the Arsonist. In each case, the agent has a *pro tanto* duty to not set fire to the farmer's field. However, the Arsonist and the Village Hero are ordinarily assessed differently. The Arsonist acts impermissibly whilst the Village Hero acts permissibly. We might think this because the Village Hero has a further *pro tanto* duty to save the village, whilst the Arsonist does not. One might think this because the Village Hero has a further duty to save the village whilst the Arsonist does not, and saving the village is the duty the Village Hero has the most and mightiest moral reasons to perform. Thus, the Village Hero performs an ordinarily wrong action in the course of performing an ordinarily right action. For the Village Hero, the duty to save the village defeats the duty to not set fire to the farmer's field.

Although the Village Hero is morally justified in setting fire to the farmer's field, we might yet believe the Village Hero, or perhaps the village itself on behalf of which the Village

---

[88] Robinson (1975), see esp. 278.

Hero was acting, owes something to the farmer. For example, the farmer might reasonably expect compensation for the loss of the field, which was his livelihood. Similarly, the farmer might reasonably expect the Village Hero to apologize and show some remorse for inflicting considerable harm on the farmer. Such compensatory and compunctory obligations are known as moral residue, as discussed in Chapter 1. Borrowing framing from Bernard Williams, moral residue shows that there is something special about the agent's relation to performing an ordinarily wrong action, which is not eliminated by any morally justifying considerations.[89]

As a general phenomenon, moral residue poses a problem for any account of moral justification. Recall the case of VINCENT. This case is puzzling because if the ship captain was justified in tying his ship to the dock, then he did nothing wrong, since moral justification vindicates actions. If the ship captain did nothing wrong, then he did nothing for which he ought to compensate the dock owner. But he does owe compensation to the dock owner. If he owes something to the dock owner, then he did something wrong and his action was not justified. The Village Hero may very well be facing a similar situation.

Consider further the problem of compunctory obligations. Remorse is a relatively common feeling in the wake of moral failure. Recall the case of OEDIPUS. Stephen De Wijze proposes tragic-remorse as the appropriate response to moral dilemmas where the agent performs a morally justified action which is nonetheless wrong – what I call permissible-wrong action, although De Wijze has in mind Dirty Hands cases. Tragic-remorse differs from Williams' agent-regret and other relevantly similar moral sentiments in that tragic-remorse is only appropriately felt by agents, such as Orestes, the Village Hero, and the Tactical Bomber,

---

[89] Williams makes a similar point about agent-regret: "…there is something special about his relation to this happening, something which cannot merely be eliminated by the consideration that it was not his fault." See Williams, B. (1981). "Moral Luck." Moral Luck: 28.

who intentionally perform ordinarily wrong actions in the course of performing ordinarily right actions, while agent-regret and other related moral sentiments are appropriately felt by those, such as Oedipus, who accidentally engage in wrong-doing.[90]

While the specifics of the phenomenology of regret and remorse are not relevant to my discussion, something like tragic-remorse is at play in the phenomenon of moral residue relevant to considerations of moral justification for ordinarily wrong actions. Though, it seems like some form of remorse is appropriate in the wake of moral failure for any sort of morally significant situation as agent-regret shows. Alas, if justification vindicates action, then remorse renders inappropriate in cases where the action is vindicated.

In Chapter 2, I contrasted the idea of pro tanto duties with my account of absolute duties. On the former sort of view, when one *pro tanto* duty defeats another in dilemma cases, nonconformance with the defeated duty is non-wrong while conformance with the overriding duty is the right thing to do. According to most moral theories, conforming with one's singular operative duty makes one's action morally right. So, if conforming with one's singular operative duty in dilemma cases is always the right thing to do, then we should be puzzled about whether an agent ever appropriately feels remorse when failing to conform with the defeated duty.

When we feel bad about performing the morally right action, one might plausibly think that we manifest a sort of criticizeable misalignment in our character. Some kind of moral entitlement, such as righteousness, is the appropriate response to doing the right thing – not remorse. Yet, we feel remorse nonetheless. I suggest that the existence of moral residue favors

---

[90] De Wijze, S. (2004). "Tragic-Remorse – The Anguish of Dirty Hands." *Ethical Theory and Moral Practice*, 7.

a non-vindicatory account of justification. On such an account, the ship captain in VINCENT is morally justified although he performed a morally wrong action. We can make sense of why he owes compensation to the dock owner if, in fact, he did something wrong. Similarly, we can make sense of why remorse is an appropriate response to nonconformance with the defeated duty in dilemma cases if failing to conform with that duty is morally wrong even if justified.

## 3. Rejecting Justificatory Vindication

In the previous section, I argued that vindicatory accounts of moral justification are incompatible with genuine moral dilemmas and that concerns about moral residue suggest that moral justification does not wholly vindicate actions. These considerations lead me to question whether moral vindication of action is the aim of moral justification. So, suppose we reject justificatory vindication and the commitment to (A*).

Let me acknowledge that rejecting justificatory vindication seems strange. When we wonder whether an agent is morally justified for performing some action, we are not typically asking whether the agent is morally justified for performing an ordinary right action. Our intuitions on these matters appear to be settled. We tend to think that the default status for any action is permissible, such that it requires no justification. Instead, we only ever seem to ask whether an action is morally justified when we have some objection against performing it – namely that the action seems to be wrong. If the action overcomes the wrongness objection, then the action is morally justified and therefore permissible. Because most permissible actions are not objectionable in this way, we rarely scrutinize the permissibility of and the underlying justification for such actions.

As a result, questions about justification for ordinary right and wrong actions do not play much of a role in our moral theorizing. Instead, we tend to wonder whether the agent is morally justified for performing a *pro tanto* wrong action in the course of performing a *pro tanto* right action. Explaining how an agent could be morally justified for performing an action she is ordinarily unjustified for performing generates much moral theorizing.

While making headway on this question is essential to properly assessing agents performing morally suspect actions, this is not the only question we should be asking. If we

are interested in understanding how all of our moral judgments fit together, then we must investigate on what basis we make all of our moral judgments and not only on what basis we make a small subset of our moral judgments. A complete understanding of moral permissibility and moral justification thus requires that we moral theorists investigate what makes ordinarily morally permissible actions permissible.

Of course, this question is somewhat unnatural. If we do not ordinarily require a moral justification for performing ordinary permissible actions, then why should we investigate what makes permissible actions permissible? Perhaps pursuing a more complete account of moral justification of the sort I described is problematic because we have nothing to say about what justifies ordinarily permissible actions.

Perhaps, further, we have no justification in these cases. The permissibility of such actions is purely presumptive. This sort of objection misses a key distinction necessary for understanding moral justification, which further exposes the problem belying the way we have discussed justification and permissibility in the literature. Starting with the latter point, if other philosophers think no complete account of moral justification will be had, then this is something to be argued for and not assumed. At present, given that few philosophers have attempted to provide such an account, such an assumption appears premature and unwarranted. Moreover, we surely do have something to say about what justifies ordinarily permissible actions, we just never inquire about it. This brings me to the former point.

We can draw a distinction between an agent's requiring a justification for some action and an agent's having a justification for some action. When we require a justification for some action, we seek evidence that the agent has a justification for performing the action. Typically, an agent requires a justification when her action is *prima facie* on the "wrong side" of morality,

172

and an agent does not require a justification for performing, for example, ordinary right actions, which are on the "right side" of morality.

An agent has a justification for performing an action when her action has the right moral credentials or meets certain criteria. So, she may yet have a justification for performing an action even when we do not ordinarily a require a justification for performing the action.

Even if we ordinarily speak as though not requiring a justification implies that the agent has no justification, we needn't think this. For example, suppose a critical onlooker requires a justification of an agent who performs a morally right action. Perchance the onlooker mistakenly believes the agent's action is wrong rather than right. If there is no justification to be had when the agent is required to have a justification when she ought not be required to have a justification, then the agent would have no response available other than to dogmatically deny that a justification is required. Our poor agent would be unable to set things straight if her justification were purely presumptive and the moral judge failed to make the right presumptions about her action.

Perhaps instead we think the action is justified in virtue of its rightness. Maybe permissibility is built into the concept of rightness such that all right actions are also permissible ones. While I ultimately think that the former is false and the latter is trivial, the above shows that we might have something to say about what makes the action justified afterall. Our poor agent could reply that although no justification ought be required of her, she has a justification in virtue of performing a right action and not a wrong action. While, on this view, an agent has a justification on the basis of very minimal criteria in such cases, the agent nonetheless has a justification for performing her action.

If it makes sense to say that an agent has a moral justification when she performs an idle action, a right action, a supererogatory action, or a suberogatory action, then we need an account of moral justification that can explain this, and vindicatory accounts of moral justification cannot do this because none of these actions are morally wrong. So, rejecting (A*) allows for the possibility of both justified wrong-doing and PWA, as well as helps us to make sense to the tight conceptual link between justification and permissibility in general.

I propose an action is morally permissible if and only if the action is morally justified. Afterall, it certainly seems as if any unjustified action is impermissible (and vice versa). And my opponents should agree. The alternative would be to claim that some unjustified actions are permissible or that some justified actions are impermissible. In the traditional context, where questions about an action's justification are primarily applied to ordinarily wrong actions performed in the course of performing ordinarily right actions, this would be to say something like the agent permissibly performs the all-things-considered morally wrong action or that the agent impermissibly performs the all-things-considered morally right action. While some version of my sort of account could potentially accommodate such claims (as it stands, it does not), other philosophical views do not accommodate such claims.

Moral justification and moral permissibility are tightly conceptually linked. I have suggested that this is because permissibility assessments are assessments of actions that appeal to what I have called the justification relation, which is the relation in which the agent stands to the features of the action that make it justifiable. An action is justifiable when it instantiates morally significant agent-independent features ordinarily sufficient for justifying the agent in acting. I will explain which features do the justifying (or unjustifying, as the case may be) shortly, but first I examine the function of justification.

If an action is morally permissible if and only if the action is morally justified, then ordinarily permissible non-wrong actions have moral justification. Because the action is not wrong, the action does not have justification for its wrongness and, therefore, must have justification for something else. This motivates me to reject the view that the purpose of moral justification is to overcome the wrongness objection. Instead, moral justification aims at doing something different.

However, I agree with the intuition that the question of an action's justification arises when we have some objection against performing that action. I propose that every action, prior to an agent's acting, must overcome at least one objection, an existential worry from the moral standpoint – that the action risks impermissibility. I call this the problem of impermissibility. The problem of impermissibility arises for at least three reasons. First, morality is complex, onerous, and unreckonable (in its totality) and, as a result, we often fail, at least in some respects, from the moral standpoint. Second, humans, due to the countless ways in which we are fallible, always run the risk of getting things wrong and, third, sometimes want the wrong sorts of things for ourselves and those around us. Because we always run the risk of getting things wrong, we are required to keep on our toes (so to speak) with respect to moral decision making. So, before any action or inaction, as the case may be, we are challenged on the moral front to overcome the problem of impermissibility through our deliberations and actions. An action's moral justification answers the question of whether the action overcomes the problem of impermissibility, and I need a model of justification that can help us make sense of why some wrong actions overcome the problem of permissibility.

**4. The Problem of Impermissibility**

So far, I have argued that vindicatory accounts of moral justification are inadequate because they cannot make sense of several moral phenomena. Specifically, they cannot make sense of genuine moral dilemmas, AAP in DDE cases, or moral residue. Further, vindicatory accounts of moral justification cannot make sense of justification in cases of ordinary permissibility. If ordinary permissible actions are non-wrong but nonetheless justified, then we should reject the claim that the purpose of moral justification is to overcome the wrongness objection. Afterall, such actions are not subject to the wrongness objection. I suggested that the aim of justification is to overcome the problem of impermissibility instead, which leaves open the possibility of justified wrong-doing. In this section, I begin explaining how actions overcome the problem of impermissibility.

Overcoming the problem of impermissibility begins with establishing whether the action was justifiable in the first place. Only justifiable actions can be morally justified, and only the features of the action that make it justifiable can be utilized in justifying the action. However, not all of the features that make an action justifiable will always be utilized in justifying the action. As articulated in Chapter 4, the agent's motivations pick out the contents of the justificatory set, constraining which features of the action do the justifying. So, just because an action can be justified (is justifiable) does not mean an agent will be justified in acting. From this point on, I will discuss what justifies an action or what makes an action justified. I ask my reader to keep in mind that these are the very features of the action that make it justifiable.

The justification for an action is asymmetrically weighted between wrong and non-wrong actions. Non-wrong actions are justified relatively easily, while ordinarily wrong

actions are justified only with great difficulty. Afterall, idle actions are typically tacitly justified, while we frequently explicitly require an agent's justification in performing an ordinarily wrong action, even if the action was performed in the course of bringing about an ordinarily right action. Ordinarily wrong actions therefore have what I call a higher justificatory threshold than non-wrong actions.

Because only wrong actions are candidates for impermissibility, the first challenge in overcoming the problem of impermissibility lies in whether the action failed to conform with an operative duty. If the action failed to conform with an operative duty, the action remains a candidate for impermissibility and is potentially unjustified, while the action is no longer a candidate for impermissibility if the action conformed with every operative duty or if no such duty existed. Such actions inevitably meet the justificatory threshold.

Morally wrong actions remain candidates for impermissibility because they fail from the moral standpoint in a specific way. So, we can see that unpropitious moral assessments, such as assessments of an action's wrongness or candidacy for impermissibility, have to do with certain negative features of the action – that the action failed to conform with an operative duty (among other considerations). Yet, I propose that some wrong actions overcome the problem of impermissibility and render morally permissible, as in dilemma cases. Because I am interested in identifying a unified (rather than jerrymandered) explanation for what makes every kind of action morally justified, I hold that the mechanism underlying the justification for ordinarily permissible actions is the same mechanism underlying PWA. Because PWA's are wrong actions, an action's non-wrongness alone cannot be the feature of the action doing the justifying. Otherwise, PWAs would not be justifiable. I require a different explanation for what justifies an action.

177

I therefore propose that an action is not justified by its non-wrongness alone just as an action is not unjustified by its wrongness alone. Instead, I think we should look to the features of actions that we think justify actions in standard cases of performing an ordinary wrong action in the course of performing an ordinary right action. On vindicatory accounts of moral justification, such as the DAJ, an action's all-things-considered rightness (or something of the sort) does the justifying. So, I propose that an action's justification has to do with certain positive features of the action, such as its simultaneously conforming with a countervailing conflicting duty (among other considerations) in a dilemma case.

In earlier chapters, I suggested that claims about an action's moral permissibility are claims about an action's moral adequacy, where permissible actions are "good enough" morally speaking. A permissible action's being good enough is one measure of moral success or the ways that the action succeeds from the moral standpoint despite its flaws. An action renders permissible when it succeeds from the moral standpoint in specific ways. The mechanism of success for permissible actions is the moral justification. To assess whether the action is justified, we assess, first, whether the action instantiated the right sorts of positive features.

An action's non-wrongness cites only the way in which the action did not fail from the moral standpoint but not the ways in which the action succeeded from the moral standpoint. So, what justifies an action are the other normatively significant features present in the action, for which the action is deemed a moral success. I will say more about particularly troubling cases on the Objections. Of course, every non-wrong action will be justified, meaning an action's non-wrongness entails a justification in acting. This information is useful for agents deliberating about what to do, but the action's non-wrongness does not justify the action.

On my account, all of the other morally significant features present in the action contribute to the action's justification. Which features are morally significant and their relative weights will be determined by one's particular moral theory, and each feature contributes differently depending on the type and target of assessment. We can compare this way of morally assessing actions to the way we might assess the quality of a soup. We can assess the quality of each ingredient in the soup independently just as we might make first-order moral assessments independently, but, to assess the quality of a soup, we must assess the quality of all of the ingredients taken together.[91]

In dilemma cases, where the action is simultaneously wrong and dutiful, the action's dutifulness is not sufficient for justifying the action because the dutifulness is counterbalanced by the action's wrongness. Because all duties are absolute on my account and therefore all have "equal weight" when compared to each other, neither duty settles the question of its justifiability. The deciding factor in a dilemma case therefore involves the normative reasons preferring one action over the other. So, an agent is only ever justified in performing a morally wrong yet dutiful action when that action is also the action the agent has the most and mightiest moral reasons in its favor. The agent must perform the morally preferred solution in a dilemma case, as with ORESTES and Medea.

This does not mean the agent is always justified in performing the action she has the most and mightiest moral reasons to do on my account, since it should be possible for an agent to have a duty against and no duty for performing such actions. The reason the normative

---

[91] Yet, each of the ingredients might feature in our assessment of the soup differently depending on the type and target of the assessment. For example, the key to a good mushroom soup is the mushrooms, but all of the other ingredients matter, whereas the same mushrooms and sundry ingredients might matter more or less when assessing a different kind of soup.

reasons in favor of performing the action never outweigh or defeat the duty against is that the features of an action instantiating P, where P is the property of the action that makes it wrong, have a sort of lexical priority over the action's G-ness, where G is a good-making property of the action.[92] This should make sense, since I have an account according to which duties relating to wrongness are absolute, negative, and grounded in the substantive features of actions as explained in Chapter 2.

My conclusion about which features of the action do the justifying in the dilemma scenario should not be a surprise at this point. These are the conditions on PWA, initially pointed out in previous chapters. The aim of this section is to highlight that the action is justifiable because it instantiates certain positive features and that these features do not serve to justify the agent's failure to conform with the other duty in the dilemma scenario (i.e. the good does not serve to justify the bad). Instead, these are the sorts of positive features that justify any sort of permissible action, but actions in a dilemma case must meet a higher justificatory threshold.

---

[92] For a rejection of lexical priority in moral absolutism, see M. Huemer's (2010) "Lexical Priority and the Problem of Risk" (*Pacific Philosophical Quarterly*, 91).

## 5. The Adequacy Account of Moral Justification

In previous sections, I suggested that the aim of justification is to overcome the problem of impermissibility, rather than to overcome the wrongness objection, which leaves open the possibility of PWA. So far, I have given the reader a sense of how actions overcome the problem of impermissibility, especially when wrong. They must manifest certain positive features. However, I have not fully explained my model for the alternative to justificatory vindication – the adequacy account of moral justification [AAJ]. I do that in this section.

We can think of wrongness as a kind of moral marring, flaw, or impurity in the action. On vindicatory accounts of moral justification, the impurity is cut out of the action to vindicate its worth, much as with a crucible that burns the impurities out of a metal. Yet, other options for testing a metal are available, such as the proving method. A metal is proven when, despite its impurities, it meets some minimum criteria that make it good enough. For example, the metal in a sword intended for battle is proven when it withstands blows from another proven sword. Some impurities will not undermine the strength and flexibility of the metal but too many will. Perhaps, proponents of genuine moral dilemmas would reconceive of moral justification in this way.

To overcome the problem of impermissibility, the action need not render non-wrong. So, perhaps moral justification concerns the criteria for proving that one's action is morally adequate, where this means that the action can be proven or shown to be good enough despite its flaws. If so, we can count on justification to leave our intuitions about residual obligations unchallenged, since the role of justification is not to erase the way in which the action manifests moral failures.

To clarify, I do not mean to suggest that our justificatory criteria impose a literal test on the adequacy of our actions. Instead, I contrast proving an action with vindicating an action to highlight an alternative view on the function of justification. The difference is analogous to some aspects of paper grading. A student's paper can receive a passing grade even if the paper suffers from disorganization or a lack of clarity. Such a paper is proven despite clear flaws. A student's paper is vindicated when she earns higher marks by challenging the instructor's obvious misunderstanding of her argument. A proven action sustains its moral flaws even if judged adequate. A vindicated action does not. Instead, a vindicated action is shown to lack the moral flaws initially supposed.

Actions are proven on my account by showing that they stand in the right relation to the agent's motivations. Let me briefly summarise the view presented in Chapter 4. First, (i) the reasons for which the agent actually acted specify which of the morally significant normative features of an action are relevant to assessing its permissibility. I call the morally significant agent-independent features of the action specified in (i) the justificatory set. Only the contents of the justificatory set get to play a role in justifying or unjustifying the action. The justification relation is the relationship between the contents of the justificatory set and the agent motivations constraining them.

An action is proven when the agent's motivations align in the right way with the morally significant normative features of the action. The agent's motivations align in the right way when they pick out the features of the action that make it justifiable without picking out the features that make it wrong (if the action is wrong). This amounts to (ii) an assessment of the alignment between the agent's motivations and the justificatory set and (iii) an assessment of whether any inappropriate alignment is due to the inclusion of an unjustifying feature in

the content of the justificatory set or the exclusion of a justifying feature in the content of the justificatory set.

This account requires (iii) because it could turn out that the agent acted for sinister reasons but failed to instantiate any wrong-making features in acting, so the agent's motivations fail to align with the agent-independent features of the action in the right way but not because the agent's motivations picked out an unjustifying feature of the action. We might expect this sort of thing when one person attempts to murder another person by sticking a pin in a voodoo doll. Although the agent acts on objectionable motivations, those motivations fail to pick out anything objectionable in the action. Such an action would be justified just as long as the agent's motivations in acting specified any justifying feature of the action.[93] I will say more about what would count as the justifying feature of the action in the next section.

It could also turn out that the agent acted for morally irrelevant motivations while nonetheless instantiating wrong-making features in acting, such as with the case of Chairman HARM. In such a case, the agent's motivations would similarly fail to align with the agent-independent features of the action in the right way but not because the agent's motivations picked out an unjustifying feature of the action. Instead, the content of the justificatory set would simply fail to meet the justificatory threshold by failing to include the features of the action that make it justifiable. Chairman HARM's motivations pick out the action's profitableness for inclusion in the justificatory set, but the action's profitableness is not the sort of thing that can justify it when the action is wrong.

---

[93] If the justificatory set is empty because the agent's motivations were entirely dissociated from the action, then we might begin to worry about whether the agent satisfies our conditions on moral responsibility because we might not be looking at something properly called an action.

My account of moral justification is an adequacy account and not a vindicatory account because it does not aim at overcoming the wrongness objection to an action by showing the action is non-wrong. Instead, the action's wrongness is settled by the first-order act-assessment and nothing can change that, yet the wrongness plays a central role in the justificatory process. Wrongness unjustifies the action if included in the agent's justificatory set. If appropriately excluded from the justificatory set, then the action remains wrong despite the agent's justification in acting. The action is proven in the sense that, despite its moral flaws, the features of the action that make it justifiable are shown to be central to the action's justification.

.

## 6. Objections

Before concluding this chapter and this dissertation, I will deal with two objections to my account of moral justification. First, I will deal with the concern that suberogatory actions are not justified on my account. Second, I will deal with the concern that sometimes we are justified in being instrumentally motivated by an action's wrong making feature. I then conclude.

Starting with the first worry first. Suberogatory actions may appear to have difficulty with justification on my account because such actions have, on balance, more and mightier moral reason against than for. I have suggested that moral permissibility is about the way that an action succeeds from the moral standpoint and not the way in which it fails. If so, one might wonder how we can speak of the ways in which the action succeeds from the moral standpoint when the action seems wrought with moral failure.

As already discussed, moral failure is only relevant to considerations of impermissibility and failing to conform with one's duty is the only condition making an action a candidate for impermissibility. So, the moral reasons counting against a suberogatory action are irrelevant to assessing the action's justification. I proposed that any non-wrong action with normatively significant considerations in favor will count as a moral success. In many cases, the reasons in favor will be moral, as I will show with Aeneas, but we tend to think various practical and prudential reasons for action are good enough in ordinary permissible actions. This works because I do not have a view where the morally objectionable and morally desirable features of actions are weighed against each other to identify an action's permissibility, even though this does happen for other kinds of moral assessments. The

action's various charms and blemishes play different roles in different sorts of moral assessments in my framework.

So, although suberogatory actions are all-things-considered bad actions, they oftentimes have at least some moral reasons in their favor. For example, Aeneas was partially motivated to depart for Rome, leaving Dido to commit suicide, for the sake of his future children, on behalf of whom the gods reprimanded him. If Aeneas had not left for Rome, then his future children would have been deprived of many goods and opportunities as future rulers. However, his future children may well have led equally good lives as the rulers of Carthage, so these considerations most likely do not outweigh concerns about Dido's suicide. Because Aeneas performed an action with at least some moral reasons for and no duty prohibiting (making it non-wrong), the action succeeds from the moral standpoint.

Even so, some suberogatory actions have only moral considerations against and none for, as with PERRY ZOSO, who lingers at the café during lunch hour rather than giving up his seat to a hungry customer. One might worry that surely such actions fail to succeed from the moral standpoint. I propose that suberogatory actions are justifiable because they satisfy moral adequacy requirements.[94] So, AENEAS performs adequately when he performs a non-wrong action with moral reasons in favor. The action is merely adequate because the balance of moral reasons counts against performing the action.

We should think suberogatory actions like Perry Zoso's satisfy moral adequacy requirements because other ordinarily permissible actions, such as idle actions, would similarly be ruled out if they do not. Idle actions, those actions with no moral reasons against

---

[94] Minimal moral decency is the more common term, but I have mostly seen it in relation to discussions of blame where the focus is the minimally decent person, but I am interested in minimally decent action.

or for, do not appear to have anything morally positive going for them, which should lead one to wonder how they might count as moral successes. I propose that, although performing actions with moral reasons in their favor is good, performing actions with nothing for or against, morally speaking, is good enough. Merely erogatory actions are more morally desirable than idle actions, but idle actions are not morally undesirable. We think such actions are morally adequate, and I propose that adequacy contributes to an action's moral success. Afterall, even morally excellent or virtuous actions satisfy the criteria for adequacy, otherwise the actions would not be good enough morally speaking, and surely that seems like a misunderstanding of moral excellence. Such actions are good enough and more. I contend that any non-wrong action has an abundance of morally irrelevant normatively significant considerations that count in favor of performing the action, but any would suffice to make a suberogatory action adequate and therefore justifiable on my account. Just as long as PERRY ZOSO performs an action with any practical (or other) reasons counting in favor, his action renders justifiable.

On the second point. My proposal still might leave a critical reader wondering how I handle cases where the agent is partially motivated by the apparent wrong-making features of the action, yet the action clearly appears to be morally justified and therefore permissible. For example, in a genuine case of self-defense, I might shoot and kill my would-be murderer both with the intention of saving my own life and with the further intention of killing my pursuer. In this scenario, I see the death of the aggressor as a part of what makes the action good.

Killing in self-defense is typically only justified as a necessary and proportionate response to an imminent threat. Supposing these conditions apply, I take it that killing in self-defense is the morally preferred solution to a genuine moral dilemma, which, on my account,

makes such actions justifiable, but I should say more about why this might be so. Self-defense scenarios are not typical dilemma scenarios, and their assessment is complicated by my substantive commitment to moral absolutism. On a semi-standard view that is overall compatible with the UMF, killing in self-defense is permissible when the action is non-wrong. Killing in self-defense might turn out to be non-wrong because the would-be killer temporarily loses any right not to be killed when posing an imminent threat to the victim. The permissibility of the action is therefore not in question. On my account, duties do not admit of such exceptions. So, if killing in self-defense instantiates all of the properties that normally make killing wrong, then the killing fails to conform with duty, and the killing is therefore wrong. Supposing I have a duty against allowing myself to be killed, I have simultaneous occurrent conflicting duties to prevent harm to myself and to avoid harming the would-be murderer. We might therefore think of self-defense scenarios are genuine moral dilemmas.

The first complication in self-defense dilemmas is that the very feature of the action that makes it wrong is simultaneously the only (potentially justifiable) means of satisfying the duty and performing what I take to be the morally preferred action of saving myself (i.e. the killing is the only means of preventing harm to myself). The version of killing in self-defense under consideration is therefore unlike the killing of the one in SIDE-TRACK, since the killing of the one is incidental to pulling the lever and saving the five. In self-defense, the killing of the aggressor, or at least the aggressor's incapacitation through potentially lethal damage, is essential to acting in conformance with duty.

I bring up the first point because sometimes killing is a necessary means for conforming with duty, but the killing is not the sort that can be justified. This happens when the killing does not correspond to the morally preferred solution to the dilemma (e.g. killing

to keep a minor promise). Killing in self-defense is not like this because of a second complication. In self-defense cases, the aggressor is acting impermissibly and generating my dilemma thereby. My dilemma would disappear if the aggressor conformed with his own duties in the only justifiable way available and ended the attempted murder.[95] This helps to explain why killing in self-defense is the morally preferred solution to a moral dilemma involving a trade-off between one life and another life. Simply ending the attempted murder is the correct solution to the dilemma as the action both the defendee and the aggressor must take, which makes acting so as to end the attempted murder the only justifiable action I can take in resolving my dilemma (supposing, of course, that I have no means of incapacitating the aggressor through non-lethal force).

The same is not true where the killing of an innocent bystander is apparently the necessary means for conforming with duty. Killing in self-defense is sometimes justifiable because the killing of the aggressor is the only way to ensure an outcome that the aggressor and defendee are mutually required to achieve in order to put an end to my dilemma. An innocent bystander however is not generating my dilemma by acting impermissibly, so there are many actions I cannot take towards the bystander, including killing her, to end my dilemmas. I can only take an action towards the bystander that ensures something that the bystander herself was required to do. This is not to say that killing innocent bystanders is always impermissible. Again, Medea permissibly pulls the lever in SIDE-TRACK, but the conditions in SIDE-TRACK are different from those we are currently discussing.

---

[95] V. Tadros's (2016) "Permissibility in a World of Wrongdoing" (*Philosophy and Public Affairs*, 44.2) addresses similar considerations.

The puzzle is that, although killing in self-defense is the morally preferred solution to a genuine moral dilemma, it seems impossible for an agent to take the morally preferred solution in the dilemma without aiming at the aggressor's death and seeing the death as a good-making feature of the action. While a perfectly motivated person might manage to kill in self-defense solely to prevent harm to herself and not at all because she wants the other person to die, many of us will include the aggressor's death in the justificatory set in a way that seems worrisome on my account.

Here is why we can set this worry aside. Killing in self-defense can be justified even if I shoot and kill my would-be murderer both with the intention of saving my own life and with the further intention of killing the murderer. When killing the aggressor in self-defense is an essential and justifiable means for conforming with duty, then the death of the aggressor is an essential part of what makes the action conform with duty and a part of what makes the action morally preferred in the dilemma scenario. So, the death of the aggressor is a part of the description of the morally preferred solution to my dilemma, and my desire for the death helps to highlight the feature of the action that makes it dutiful and morally preferred. Because the death of the aggressor is also a wrong-making feature of the action, my desire for the would-be murder's death can help rule out the action's justification if I am only or primarily motivated to kill my pursuer and not at all motivated to prevent harm to myself. However, I am suggesting that such a desire does not, on its own, rule out the action's justification if coupled with other motivations relating to the action's conformance with duty in this sort of highly complicated dilemma. This is because the death serves both as a wrong-making feature of the action and a dutiful making feature of the action.

The example of killing in self-defense shows why an action might be justified even if the agent is partially motivated by the wrong-making features of the action. The wrong-making feature of the action must also be a part of what makes the action a necessary means of performing the morally preferred action in a dilemma. The wrong-making feature is therefore a part of the alternative action's right-making features and can be included in the agent's justificatory set when accompanying other right-making features of the action. If the wrong-making feature were included in the justificatory set on its own, then the action would be unjustified.

**Conclusion**

The real purpose of this chapter was to fill out my account of moral justification. In Chapter 4, I sketched the structural features of the view, but I did not fully explain what makes actions morally justifiable or what does the justifying on my account. Chapter 5 specifies certain positive features of actions as the justifying features and explains how those sorts of features can help us make sense of justification in cases of ordinary permissibility as well as dilemma scenarios. I am now in a position to present a summary of my model in the concluding chapter.

# Chapter 6: Conclusion

**Introduction**

I have now provided my reader with a rough outline of my Unified Moral Framework [UMF] and explained how many of our moral assessments fit into that framework. At this point, I hope my reader has been convinced of at least some aspects of my view. The basics of my framework itself, introduced in Chapter 2, can be endorsed without regard for any other aspect of the view. One might think the relationship I articulate between permissibility and moral worth holds without endorsing the basic framework. My general account of the direct relevance of motivations to permissibility assessments could play a role in the DDE debate, setting side everything else. And a reader interested in the possibility of genuine moral dilemmas might seriously consider my objections to justificatory vindication. Everything considered, this research has something to offer to philosophers with a range of different interests. Everything considered together, however, provides the reader with a novel, systematic approach to handling some of the most puzzling phenomena in contemporary ethics. In what follows, I provide the final summary of this view – the core pieces of the framework in one place for the reader's convenience.

**Overview of the Unified Moral Framework**

My project begins with a pair of distinctions. In general, moral assessments have two primary

targets – actions and agents. I call this the Fundamental Moral Distinction. Assessments are

made by appeal to certain considerations. Some assessments involve only the object of the

assessment, such as the agent-independent features of the action alone or the agent's

motivations alone. Some assessments are hybrids, appealing to interactions between both sorts

of considerations. I call the former first-order moral assessments and the latter second-order

moral assessments. This pair of distinctions establish the Simplified Moral Framework.

**Figure 1: The Simplified Moral Framework**

|  | Act-Assessment | Agent-Assessment |
|---|---|---|
| **First-Order** | Act Only (1) | Agent Only (2) |
| **Second-Order** | Act & Agent (4) | Agent & Act (3) |

*Figure 1 illustrates the SMF.*

I suggest that assessments such as right and wrong fit into Box (1), while assessments such as

an agent's good will and ill will fit into Box (2). Each sort of assessment might well appeal

only to the morally significant features of the assessment's object only. I call the morally

significant features of the assessment's object the assessment properties. I suggest that

assessments such as an agent's praiseworthiness and blameworthiness (i.e. moral worth

assessments) fit into Box (3), while permissibility assessments fit into Box (4). Each sort of

assessment might well appeal to assessment properties considered in both act and agent

assessments.

**Figure 2: The Simplified Moral Framework [Completed]**

|  | Act-Assessment | Agent-Assessment |
|---|---|---|
| **First-Order** | Right & Wrong (1) | Good Will & Ill Will (2) |
| **Second-Order** | Permissible & Impermissible (4) | Praiseworthy & Blameworthy (3) |

*Figure 2 illustrates the SMF – complete with examples of the specific moral assessments.*

Within the categories of first-order moral assessments, we find other distinctions between assessment types. Among first-order act-assessments, we can distinguish between what we have moral reasons to do and what we have duties to do. Throughout, I characterize moral reasons as the morally significant agent-independent normative features of actions. These are the substantive goods and ills of the action. Duties, on my account, are negative, prohibitions on acting so as to instantiate the wrong-making feature of the action – the action's P-ness. Among first-order agent-assessments, we can distinguish between the agent's morally significant motivating reasons *de dicto* and *de re*. An agent performs some action from the former sort of motivations when she is motivated by the action's rightness or wrongness as such, while an agent is motivated by the substantive goods and ills of the action.

**Figure 3: UMF Distinguishing First-Order Assessment Properties**

|  | Act-Assessment | Agent-Assessment |
|---|---|---|
| **First-Order** | Moral Reasons (1a) | Motivating Reasons *de re* (2a) |
|  | Duties (1b) | Motivating Reasons *de dicto* (2b) |
| **Second-Order** | Permissibility (4) | Moral Worth (3) |

*Figure 3 illustrates the core distinctions of the UMF.*

Moral reasons are the assessment properties of erogation assessments. Erogation assessments include assessments of idle, suberogatory, merely erogatory, and supererogatory actions. Duties are the assessment properties of obligation assessments. Obligation assessments include assessments of an action's rightness, wrongness, and dutifulness. Quality of will assessments utilize the agent's motivating reasons, however philosophers disagree about the category such assessments most accurately apply to. Quality of will assessment include assessments of an agent's good will, ill will, and moral indifference. For the minimal needs of my framework, I distinguish between quality of will *de dicto* and *de re*. Here is the final result.

**Figure 4: The Unified Moral Framework [Completed]**

| | Act-Assessment | Agent-Assessment |
|---|---|---|
| | Erogation (1a) | Quality of Will *de re* (2a) |
| **First-Order** | Obligation (1b) | Quality of Will *de dicto* (2b) |
| **Second-Order** | Permissibility (4) | Moral Worth (3) |

*Figure 4 illustrates the UMF – complete with examples of the specific moral assessments I propose for each box. My reader can agree to the structure of Figure 4 without agreeing to the content.*

Moral worth and permissibility are my primary examples of second-order moral assessments. As second-order moral assessments, moral worth and permissibility are both assessments the objects of which are evaluated by appeal to interactions between the assessment properties considered in first-order moral assessments. Moral worth is most naturally considered a second-order moral assessment about agents, since moral worth is often characterized as a

hybrid assessment appealing to the morally significant features of the action and the motivations of the agent under assessment.

Specifically, on my account, moral worth is an assessment the subject of which is an agent evaluated by appeal to the quality of the relationship between the reasons for which she acted and the morally significant features of the action constraining these reasons. The morally significant agent-independent features of the action provide the background against which the agent's motivations are assessed by specifying which of the agent's motivations are relevant to a particular moral worth assessment. In brief summary, (i) the morally significant normative features of the action specify which of the agent's motivations are relevant to assessing the agent's moral worth. I call the motivating reasons specified in (i) the oneratory set. The contents of the oneratory set establish the way in which an agent is morally responsible for the moral quality of some action. I call the relationship between the oneratory set and the features of the action constraining them the responsibility relation. So, the assessment properties of the action pick out any of the agent's corresponding morally significant motivating reasons, and only these motivating reasons can onerate or exonerate (i.e. burden or unburden with responsibility) the agent.

The moral worth of an agent is determined by the way in which the action's assessment properties align with the agent's motivating reasons. The action's assessment properties align in the right way when they pick out motivating reasons indicating good will without picking out motivating reasons indicating ill will. This amounts to (ii) an assessment of the alignment between the action's assessment properties and the oneratory set and (iii) an assessment of whether any misalignment is due to the inclusion of an exonerating motivation in the content of the oneratory set or the exclusion of an onerating motivation in the content of the oneratory

set. This account explains the direct relevance of the action in moral worth assessments of agents.

Regarding (ii), if things go well, the agent's good will alone is a member of the oneratory set and the agent performed a morally right or supererogatory action that specifies those motivations as relevant to the moral worth assessment. Regarding (iii), exonerating and onerating motivations for praiseworthiness and blameworthiness are inversed. The good will serves as an onerating motivation for praise and an exonerating motivation for blame, while ill will serves as an onerating motivation for blame and an exonerating motivation for praise.

Praiseworthiness is an assessment of the agent for performing a morally right action from the motive of good will. When an agent is assessed for performing a morally right action, the agent is typically only praiseworthy she performs the action for its rightness *de dicto* and not for its rightness *de re*. I think this because rightness, on my account, does not correspond to any substantive moral goods.

Supererogation assessments however do correspond to substantive moral goods. A supererogatory action is an action the agent has on balance more and mightier moral reasons in favor of performing than against and no duty for or against performing it. When an agent is assessed for performing a supererogatory action, the agent is typically only praiseworthy when she performs the action for its supererogatoriness *de re* and not for its supererogatoriness *de dicto*. I think this because supererogation, like all erogation, is an assessment of the action's normative moral reasons. Moral worth, as a second-order moral assessment, appeals only to the assessment properties considered in the corresponding first-order moral assessment, so, in the case of supererogation, moral worth only considers the morally relevant substantive goods. Because supererogation is an assessment of such goods and is not itself a good, the motive to

perform an action because the action is supererogatory as such is not a potential praiseworthy motivation.

Blameworthiness is an assessment of the agent for performing a morally wrong action with a lack of good will in acting. An agent's lack of good will can manifest as ill will, which is the motive to perform an action for its wrongness *de dicto* or *de re*. An agent's lack of good will can also manifest as moral indifference, which is the motive to perform an action for morally irrelevant reasons or without sufficient regard for the morally relevant reasons *de dicto* or *de re*. Acting with any of these motivations while performing a morally wrong action suffices for making an agent blameworthy because wrongness is related to the substantive features of action in a way that rightness is not.

Permissibility is a less natural candidate as a second-order moral assessment. As Scanlon put it, permissibility assessments primarily have to do with the effect of an act on people. Although an agent's motivations determine an agent's action and therefore partially determine the effect of an action on people, external factors and the agent's own fallibility can lead to unintended effects which oftentimes directly impacts our permissibility assessments. As a result, one might reasonably think that an agent's motivations are never directly relevant to permissibility assessments. However, the asymmetrical pattern of permissibility assessments in Doctrine of Double Effect [DDE] cases suggest the opposite.

I propose that permissibility is a second-order act-assessment, which is structurally similar to moral worth assessments, appealing to the same sorts of assessment properties. Specifically, on my account, permissibility is a moral assessment the object of which is an action evaluated by appeal to relationship between the reasons for which the agent acted and the morally significant features of the action constraining these reasons. The agent's morally

significant motivating reasons provide the background against which the action is assessed by specifying which agent-independent features of the action are relevant to a particular permissibility assessment.

In brief summary, (i) the reasons for which the agent actually acted specify which of the morally significant normative features of an action are relevant to assessing its permissibility. I call the morally significant agent-independent features of the action specified in (i) the justificatory set. Only the contents of the justificatory set get to play a role in justifying or unjustifying the action. The justification relation is the relationship between the contents of the justificatory set and the agent motivations constraining them.

An action is proven when the agent's motivations align in the right way with the morally significant normative features of the action. The agent's motivations align in the right way when they pick out the features of the action that make it justifiable without picking out the features that make it wrong (if the action is wrong). This amounts to (ii) an assessment of the alignment between the agent's motivations and the justificatory set and (iii) an assessment of whether any inappropriate alignment is due to the inclusion of an unjustifying feature in the content of the justificatory set or the exclusion of a justifying feature in the content of the justificatory set.

On this account of permissibility and justification, the action's justifiability is established by morally significant agent-independent features of the action, but the agent's motivations determine whether the features of the action that make it justifiable will serve to justify the agent in acting. My account of moral justification is what I call an adequacy account of moral justification [AAJ] because this account of moral justification allows for the possibility of justifying actions manifesting certain kinds of morally significant failures as

long as those failures are not among the contents of the justificatory set. Specifically, my account makes room for justified-wrong actions, which makes room for the possibility of *prima facie* inconsistent assessments of permissible-wrong actions. Permissibility and wrongness are only prima facie inconsistent on my account because wrongness if a first-order act-assessment, while permissibility is a second-order act-assessment that considers a more diverse set of assessment properties.

My account is therefore in opposition to what I call vindicatory accounts of moral justification. According to vindicatory accounts of moral justification, an action is morally justified if and only if the action is not wrong, and the function of justification is to demonstrate that the action under consideration was indeed not wrong. Justification on this account therefore vindicates the action. I reject vindication as the function of justification and instead endorse an account of justification according to which its function is to prove or test an action for its adequacy, as described above.

**Solving Some Contradictory Puzzles**

First, consider my solution to the problem of suberogation. Erogation assessments, including assessments of the suberogatory, follow directly from my account of first-order act-assessments. If we can distinguish between what we have duties to do and what we have moral reasons to do, then we can make assessments of actions while considering only the morally relevant reasons that count for and against performing the actions. Suberogatory actions are actions we have more and mightier moral reason against performing than for and no duty prohibiting or requiring the action.

Second, consider my solution to genuine moral dilemmas. Recall the dilemma trilemma, which arises from the following three independently plausible, yet *prima facie* jointly inconsistent, claims.

(1)     The two duties binding Orestes conflict

(2)     The two duties binding Orestes are absolute

(3)     Orestes can resolve his dilemma

In the initial characterization of the dilemma trilemma, I noted that most philosophers ask us to pick between (1) and (2). I am suggesting that we need not choose between (1) and (2) by proposing the possibility of permissible-wrong action [PWA]. Instead, we can help ourselves to genuine moral dilemmas stemming from absolute duties and think, nonetheless, that moral dilemmas are resolvable.

Suppose (1) – (3).

If (1) is true, the two duties binding Orestes conflict. If (2) is true, the two duties binding

Orestes are absolute. If so, then failing to conform with either duty is morally wrong. If so,

then Orestes performs a morally wrong action no matter which duty he conforms with. If (3)

is true, Orestes can resolve his dilemma. If so, then, even though failing to conform with either

duty is morally wrong, Orestes permissibly fails to conform with at least one of his duties. If

claims (1) – (3) are all true, permissibility and wrongness must, therefore, be compatible moral

judgments about Orestes' action. The possibility of PWA straightforwardly resolves the *prima

facie* inconsistency between (1) – (3).

I propose that PWA is only *prima facie* inconsistent, since PWA is a combination of

two distinct first- and second- order act-assessments.

**Figure 5: Permissible-Wrong Action**

|  | Act-Assessment |
|---|---|
| **First-Order** | Wrong Action (1) |
| **Second-Order** | Permissible Action (4) |

*Figure 5 illustrates the conceptual niche into which I fit PWA.*

PWA is possible on my account because I reject vindication as the purpose of moral

justification and opt instead for the AAJ. On the AAJ, an action is sometimes justified despite

significant moral flaws. The question, on my account of moral justification, is whether the

action's wrongness is featured in the action's justificatory set, which would be sufficient for

unjustifying the action.

Third, consider my solution to asymmetrical assessments of permissibility [AAP] in DDE cases. To assess the difference between the two agents in AAP cases (e.g. the bomber cases), the motivations relevant to the assessment of the action's permissibility are the assessment properties considered in the appropriate first-order agent-assessment for the agent's motivations in acting. For the two bombers, only those motivations that settled the question of whether to drop the bomb serve as the relevant assessment properties. The Tactical Bomber and Terror Bomber plausibly agree that bombing the munitions factory is the right thing to do because bombing the munitions factory, ending the war early, and killing civilians is the morally preferred solution to the dilemma. However, the Terror Bomber sees the civilian deaths as a part of what makes the action good, while the Tactical Bomber sees the killing of civilians only as something that makes the action wrong but nonetheless correctly believes these considerations are outweighed by saving the lives saved on both sides.

These motivations pick out any corresponding morally significant agent-independent features of the action, and only these features of the action can justify the agent in acting. So, this difference in motivations highlights the primary difference between the set of normative considerations playing a role in justifying the two bombers in acting. The very feature of the action that makes it wrong is specified in the justificatory set of the Tactical Bomber as a normative consideration in favor of acting while it is correctly highlighted as a reason against acting in the Tactical Bomber's justificatory set.

If things go well, the agent aimed at performing the morally preferred action in a dilemma case and was not motivated by any aspect of the action's wrongness, as with the Tactical Bomber. The Tactical Bomber's motivations highlight the action's status as the morally preferred solution to the dilemmas and its corresponding dutifulness as central to our

assessment of the action's permissibility. Because the action is justifiable on these grounds, the action is morally justified and therefore morally permissible. Because considerations relating to an action's wrongness are always privileged in our moral assessments, at least on my account, the Terror Bomber's motivation to kill the civilians highlights the action's status as morally wrong as central to our assessment of the action's permissibility. Because the action is not justifiable on the basis of its wrongness, the action's wrongness therefore undermines the agent's justification in acting.

Fourth, consider my solution to praise-blame asymmetry [PBA] in Side-Effect Effect cases. To assess the difference between the two agents in PBA cases (e.g. the chairmen cases), the agent-independent features of the action relevant to the assessment of the agent's moral worth are the assessment properties considered in the appropriate first-order act-assessments of the action for which the agent is being assessed. To assess Chairman HELP, we look to the right-making features of the action, and to assess Chairman HARM we look to the wrong-making features of the action.

The assessment properties of the action pick out any of the agent's corresponding morally significant motivating reasons, and only these motivating reasons can onerate or exonerate (i.e. burden or unburden with responsibility) the agent. So, the difference between the two chairmen's actions highlight the primary difference between how we assess the set of motivating reasons playing a role in onerating them. Because both chairmen are motivated by profit, nothing relating to the morally significant features of the action are present in the oneratory set, except, perhaps, their genuine indifference to the relevant moral considerations and their intentions to perform the action despite these considerations.

206

Although the two chairmen both act with moral indifference, their motivations play different roles in the assessment. Chairman HELP is not praiseworthy because her right action fails to pick out anything indicating that she acted with a good will in the oneratory set. Meaning, the onerating motivation, the motivation that burdens Chairman HELP with responsibility for the rightness of her action and makes her praiseworthy, is excluded from the oneratory set. Chairman HARM is blameworthy because his wrong action fails to pick out anything indicating that he acted with a good will in the oneratory set. Meaning, the exonerating motivation, the motivation that unburdens Chairman HARM with responsibility for the wrongness of his action and makes him not blameworthy, is excluded from the oneratory set. The good will is relevant to the action's wrongness insofar as it deals with the agent's appropriate responsiveness to the wrong-making features of the action. Instead, we find only Chairman HARM's indifference to the action's harmfulness and his intention to perform the action despite its wrongness in the content of the oneratory set, which clearly saddles him with the burden of responsibility for the action's wrongness.

Fifth, finally consider my solution to the problem of moral residue. Moral residue shows that there is something special about the agent's relation to performing an ordinarily wrong action, which is not eliminated by any morally justifying considerations. Moral residue includes residual compensatory and compunctory obligations for performing a morally wrong action, even if apparently justified. Vindicatory accounts of moral justification have difficulty making sense of moral residue. If the action is not morally wrong, then the action should not generate compensatory or compunctory obligations. So, if we sometimes think something is owed in the aftermath of acting wrongly, then we are either mistaken or the action was not vindicated

– the action sustained criticizeable moral flaws. Since moral residue is a well-established moral phenomenon, with roots as far back as Greek literature, I propose that we are probably not mistaken. Instead, I hold that we can simultaneously judge that an action is wrong and justified in dilemma cases. If the action is fully morally wrong, then we can make sense of the source of our residual obligations. I take this to be a consideration in favor of the AAJ.

**Conclusion**

As moral judges, we often find it difficult to accurately morally evaluate other agents and their actions. We might think of the very same agent that she is simultaneously vicious yet praiseworthy or think of the very same action that it is wrong yet permissible. On the face of it, such judgments are contradictory, but we cannot help but feel the pull of these incongruent evaluations when faced with the complex moral dimensions of human lives. This dissertation has provided a new framework for understanding the plausibility of such complex and contradictory assessments.

# Works Cited

Anscombe, G.E.M. (1957). *Intention* (Oxford: Blackwell).

Anscombe, G.E.M. (1958). "Modern Moral Philosophy." *Philosophy*, 33(124).

Anscombe, G.E.M. (1979). "Under a Description". *Noûs*, 13(2).

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford University Press).

Austin, J.L. (1971). "A Plea for Excuses." *Philosophy and Linguistics* Ed. Colin Lyas (Macmillan Education UK).

Bennett, J. (1974). "The Conscience of Huckleberry Finn." *Philosophy*, 49(188).

Bennett, J. (1980). "Killing and Letting Die" The Tanner Lectures in Human Values, Delivered at Brasenose College, Oxford University (May 9,16, and 23).

Bennett, J. (1995). *The Act Itself* (Oxford University Press).

Botterell, A. (2007). "Why We Ought to be (Reasonable) Subjectivists about Justification." *Criminal Justice Ethics*, 26(1).

Burgess, A. & Plunkett, D. (2013). "Conceptual Ethics I." *Philosophy Compass*, 8(12).

Burgess, A. & Plunkett, D. (2013). "Conceptual Ethics II." *Philosophy Compass*, 8(12).

Chisholm, R. (1963). "Supererogation and Offence: A Conceptual Scheme for Ethics." *Ratio*, 5.

Chisholm, R. & Sosa, E. (1966). "Intrinsic Preferability and the Problem of Supererogation." *Synthese*, 16.

De Haan, J. (2001). "The Definition of Moral Dilemmas: A Logical Problem." *Ethical Theory and Moral Practice*, 4(3).

De Wijze, S. (2004). "Tragic-Remorse – The Anguish of Dirty Hands." *Ethical Theory and Moral Practice*, 7.

Driver, J. (1992). "The Suberogatory." *Australasian Journal of Philosophy*, 70.

Foot, Phillippa. (1977). "The Problem of Abortion and the Doctrine of Double Effect." *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford University Press).

Frederick, D. (2014). "Pro Tanto versus Absolute Rights." *Philosophical Forum*, 45(4).

Gardner, J. (2004). "The Wrong-Doing That Gets Results." *Philosophical Perspectives*, 18, Ethics.

Gardner, J. (2005). "Wrongs and Faults." *The Review of Metaphysics*, 59(1).

Gerson, L.P. (1997). "Socrates' Absolutist Prohibition on Wrongdoing." *Aperion*, 30(4).

Gowans C.W. (Ed) (1987). *Moral Dilemmas* (Oxford University Press).

Green, J. (2009). "The Cognitive Neuroscience of Moral Judgment." *The Cognitive Neurosciences IV* Ed. M.S. Gazzaniga (MIT Press).

Greenspan, P.S. (1983). "Moral Dilemmas and Guilt." *Philosophical Studies*, 43.

Greenspan, P.S. (1995). *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms* (Oxford University Press).

Greenwalt, K. (1986). "Distinguishing Justifications from Excuses." *Law and Contemporary Problems*, 49(3).

Harman, E. (2016). "Morally Permissible Moral Mistakes." *Ethics*, 126.

Hawley, P. (2008). "Moral Absolutism Defended." *Journal of Philosophy*, 105(5).

Heyd, D. (2002). "Supererogation." *The Stanford Encyclopedia of Philosophy*. www.plato.stanford.edu

Hieronymi, P. (2009). "The Will as Reason." *Philosophical Perspectives*, 23, Ethics.

Hindriks, F. (2008). "Intentional Action and Praise-Blame Asymmetry," *The Philosophical Quarterly*, 58(233).

Huemer, M. (2010). "Lexical Priority and the Problem of Risk." *Pacific Philosophical Quarterly*, 91.

Hursthouse, R. (1995). "Fallacies and Moral Dilemmas." *Argumentation*, 9.

Hursthouse, R. (2002). *On Virtue Ethics* (Oxford University Press).

Jackson, F. & Smith, M. (2006). "Absolutist Moral Theories and Uncertainty." *Journal of Philosophy*, 103(6).

Johnson King, Z. (2019). "Praiseworthy Motivations." *Nous*, 54(2).

Kant, I. (2002). *Groundwork for the Metaphysics of Morals*. Ed. Allen Wood (Yale University Press, first published 1785).

Kant, I. (1994). "On a Supposed Right to Lie on Altruistic Motives." *Ethical Philosophy*, (2nd Ed.) trans. Ellington, J. W. (Hackett Publishing, first published 1799).

Knobe, J. (2003). "Intentional Action and Side Effects in Ordinary Language." *Analysis*, 64.

Markoc, A. (2017). "Intentions and Permissibility: A Confusion of Moral Categories." *Journal of Value Inquiry*, 51.

Mason, H.E. (Ed) (1996). *Moral Dilemmas and Moral Theory* (Oxford University Press).

Mason, H.E. (2013). "Objectivism and Prospectivism About Rightness." *Journal of Ethics and Social Philosophy*, 7(2).

Massoud, A. (2016). "Moral Worth and Supererogation." *Ethics*, 126.

McMahan, J. (1994). "Revising the Doctrine of Double Effect." *Journal of Applied Philosophy*, 11(2).

McConnell, T. (1978). "Moral Dilemmas and Consistency in Ethics." *Canadian Journal of Philosophy*, 8(2).

Mellema, G. (1987). "Quasi-Supererogation." *Philosophical Studies*, 52(1).

Nelkin, D.K. & Rickless, S.C. (2014). "Three Cheers for Double Effect." *Philosophy and Phenomenological Research*, 89(1).

Otsuka, M. (1994). "Killing the Innocent in Self-Defense." *Philosophy & Public Affairs*, 23(1).

Pickard, H. (2011). "Responsibility Without Blame: Empathy and the Effective Treatment of Personality Disorder." *Philosophy, Psychiatry, and Psychology*, 18(3).

Pickard, H. (2014). "Responsibility without Blame: Therapy, Philosophy, Law." *Prison Service Journal*, 213.

Quinn, W.S. (1989). "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing." *The Philosophical Review*, 98(3).

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

Rachels, J. (1970). "On Moral Absolutism." *Australasian Journal of Philosophy*, 48(3).

Ramirez, A. J. (2014). "Righteous Villainy: A Defense of Offence" (PhD thesis, PDF retrieved from personal correspondence: June 01, 2015).

Raz, J. (1975). *Practical Reason and Norms* (Oxford University Press).

Robinson, P.H. (1975). "A Theory of Justification: Societal Harm as a Prerequisite for Criminal Liability." *UCLA Law Review*, 23(266).

Robinson, P.H. (1996). "Competing Theories of Justification: Deeds v. Reasons." *Harm and Culpability*, Eds.Andrew Simester & A.T.H. Smith (Oxford Univ. Press).

Saul, J. (2011). "Just Go Ahead and Lie." *Analysis*, 72(1).

Scanlon, T.M. (2008). *Moral Dimension: Permissibility, Meaning, and Blame* (Harvard University Press).

Shakespeare, W. (2005). *Titus Andronicus*. Eds. B. Mowat, P. Werstine. (The Folger Shakespeare Library, first published 1594).

Shelley, M. (1994). *Frankenstein*. S. Applebaum, C. Ward (Eds.) (Dover thrift Editions, first published 1818).

Shiffrin, S. (1999). "Wrongful Life, Procreative Responsibility, and the Significance of Harm." *Legal Theory*. 5(2).

Tadros, V. (2016). "Permissibility in a World of Wrongdoing." *Philosophy and Public Affairs*, 44(2).

Thomson, J.J. (1985). "The Trolley Problem." *The Yale Law Journal*, 94(6).

Thomson, J.J. (1991). "Self Defense." *Philosophy and Public Affairs*, 20(4).

Varden, H. (2010). "Kant and Lying to the Murderer at the Door... One More Time: Kant's Legal Philosophy and Lies to Murderers and Nazis." *Journal of Social Philosophy*, 41(4).

Vincent v. Lake Erie Transp. Co., 109 Minn. 456, 124 N.W. 221, 1910 Minn. LEXIS 588 (Minn. 1910)

Watson, J.B. & Rayner, R. (1920). "Conditioned Emotional Reactions." *Journal of Experimental Psychology*, 3(1).

Williams, B. (1965). "Ethical Consistency." *Proceedings of the Aristotelian Society*, 39.

Williams, B. & Nagel, T. (1976). "Moral Luck" *Proceedings of the Aristotelian Society* (*Supplementary Volumes*), 50.

Zimmerman, M. (2014). *Ignorance and Moral Obligation* (Oxford University Press).