Sources and Reasons: Moral Responsibility and the Desert of Praise and Blame

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of The Ohio State University

By
Audrey L. Anton, M.A.
Graduate Program in Philosophy

The Ohio State University
2011

Dissertation Committee:
Timothy Schroeder, Advisor
Tamar Rudavsky
Allan Silverman
Copyright by

Audrey L. Anton

2011
Abstract

This dissertation is an inquiry into the natures of moral responsibility and an agent’s worthiness of praise and blame and the relation between these two phenomena. The project commences with a reflection on how contemporary views of moral responsibility and praise-and blameworthiness owe a significant heritage to a division that can be seen in Aristotle’s view—that of sourcehood and reasons-responsiveness. Aristotle was one of the first to argue that an agent ought only to be praised or blamed for voluntary behavior. For Aristotle, the voluntary involves the agent’s particular knowledge of her situation as well as her control over her behavior as its source. Since Aristotle’s presentation of these issues, scholars tend to highlight one of these aspects over the other, indicating which is more central to establishing moral responsibility and the worthiness of praise and blame.

Then, I show that contemporary views of each type hold that moral responsibility is prior to (i.e., a precondition for) an agent’s worthiness of praise and blame. In my dissertation, I challenge this presumption. First, I show that moral responsibility is a type of causal responsibility. Then I argue that there are instances of agents being praise- and blameworthy that lack a causal element on the part of the agent, thus showing that one need not be morally responsible in order to be worthy of praise or blame. From here, I consider what grounds praise- and blameworthiness, offering my own account, the moral
attitude account, which is in part inspired by my interpretation of how Aristotle deemed emotions to be deserving of praise and blame. Finally, I consider the possibility that moral responsibility requires agents to be praise- or blameworthy and I ultimately endorse this claim. I suggest that sourcehood accounts are applicable to moral responsibility only and that reasons-responsiveness accounts are better equipped to incorporate my conclusions into their views.
Dedication

To Mary Ellen Anton…as promised.

I dedicate this dissertation to my late mother, Mary Ellen Anton. While my mother did not live to see the completion of this dissertation or my defense of this work, she believed in my abilities as a teacher and scholar each and every day that she lived. An accomplished and beloved teacher herself, my mother has inspired me to find innovative and intuitive ways to incorporate moral education and personal development into all of my lectures and assignments, regardless of the subject matter. A paragon of virtue, she taught by example not only how to behave well and treat others fairly, but also how to love and forgive unconditionally. She demonstrated remarkable courage and selflessness in the face of certain demise; on her deathbed, she continuously instructed my brother and I to never give up and to always finish what we start. I dedicate the completion of this project, a treatise on moral responsibility and praise- and blameworthiness, to my mother, who never gave up, who always fulfilled her responsibilities, and who will remain eternally deserving of praise.
Acknowledgments

Several people have aided and guided me throughout the process of writing this thesis. My dissertation advisor, Timothy Schroeder, has given his assistance on multiple drafts as well as in many conversations. His guidance and encouragement has been invaluable to me as I wrote this thesis. My committee members, Tamar Rudavsky and Allan Silverman, have also provided support throughout the writing of this work, and for that I am grateful. Anthony J. Lisska, my informal mentor and friend, offered much sage advice throughout the writing process as well. I would also like to thank Jonathan Jacobs for taking the time to review Chapter 2 and providing me with many helpful suggestions. Finally, I would like to thank my loving husband, Stephen Kershner, for his daily encouragement and for performing the minor miracle of helping me maintain my sanity throughout a very difficult year.
Vita

1998……………………………………….Palmer High School

2002……………………………………….B.A. Philosophy and Social and Rehabilitation Services, Assumption College

2006……………………………………….M.A. Philosophy, State University of New York at Buffalo

2009……………………………………….M.A. Philosophy, The Ohio State University

Publications

-“BREAKING THE HABIT: Aristotle on recidivism and how a thoroughly vicious person might begin to improve.” Dialogue with Ancient Philosophy, a special issue of Philosophy in the Contemporary World, 13.2.

-“DUTY AND INCLINATION: Defending Kant against the “moral automaton” problem.” Southwest Philosophy Review, 22.1.

Table of Contents

Abstract .................................................................................................................................................. ii
Dedication .............................................................................................................................................. iv
Acknowledgments ................................................................................................................................... v
Vita ........................................................................................................................................................ vi
List of Figures ......................................................................................................................................... viii
Chapter 1: Introduction ......................................................................................................................... 1
Chapter 2: Aristotle: the dual causation of human action ................................................................. 22
Chapter 3: Contemporary Sourcehood Theories ................................................................................. 86
Chapter 4: From Sources to Reasons-Responsiveness .................................................................. 122
Chapter 5: The Voluntary Nature of Aristotelian Emotions ........................................................... 148
Chapter 6: The Moral Attitude Account ............................................................................................ 176
Bibliography .......................................................................................................................................... 215
List of Figures

Figure 1: Common Taxonomy: moral responsibility dependent upon free will.........10

Figure 2: Sourcehood and Reasons-responsiveness Taxonomy..........................11

Figure 3: Intentional Responsibility: a neo-Fischerian view.............................186

Figure 4: Types of Responsibility According to Entailment Views..................191

Figure 5: The Moral Attitude View.................................................................208
Chapter 1

Introduction

Despite relatively consistent outcomes, the defense lawyers of persons such as Robert Alton Harris, Patty Hearst, and the Menendez brothers frequently launch a defense that claims essentially the same thing—that other adult persons were responsible for shaping their motives, and in particular, those motives that led to the commission of serious criminal and moral offenses. While such a defense is rarely successful in the courtroom, recent literature in the fields of ethics and moral psychology suggests that this defense does indeed elicit some sympathy in many philosophers considering such scenarios. For example, free will theorist Robert Kane suggests that we consider the case of a young man on trial for a vicious assault in which his victim was beaten to death. Kane comments, “His crime was heinous. But as we listen daily to how he came to have the mean character and perverse motives he did have—a story of parental neglect, child abuse, sexual abuse, bad role models—some of our resentment against the young man is shifted over to the parents and others who abused and mistreated him.”¹ While Kane

¹ Fischer et al. (2007): 6-7.
admits that such blame shifting does not necessarily or automatically entail that all blame is lifted from the defendant, he does suggest that much of it is lifted.

It seems to me that this tendency to redistribute blame is closely linked to a common folk judgment regarding the relationship between moral responsibility and praise- and blameworthiness. That view of the relationship is as follows. One is only blameworthy or praiseworthy insofar as one is responsible for that which constitutes or grounds one’s blameworthiness or praiseworthiness. That is to say, if one is blameworthy in virtue of beating another to death, one is only as blameworthy as that person is responsible for causing the death. However, if we can point to other partial or sufficient causes, we find the need to divide up responsibility as well as blame.

In this thesis, I aim to show that this common view is flawed, and that this flawed view fuels much of the conflict in contemporary debates concerning the relationship between moral responsibility and praise- and blameworthiness as they relate to an agent’s ability to respond to reasons as well as the agent’s control over the sources of her motives. First, the common view rests on the assumption that the kind of moral responsibility tied up with praise- and blameworthiness is prior to praise- and blameworthiness. That is, the assumption holds that this moral responsibility is a necessary condition for one’s being worthy of praise or blame. Second, since the kind of responsibility being considered is viewed as a species of causal responsibility, its distribution (and the justified distribution of anything else dependent on it) follows our assumptions of how causes work. For example, we often consider how something can causally contribute to an event even if the cause is not alone sufficient for the occurrence
of that event. For instance, it might be said of a forest fire that it was partially caused by
the prolonged drought that had occurred prior to the event. However, what sparked the
fire was some other agent or event (e.g., lightning, a careless person disposing of a lit
cigarette, etc.). Since each aspect was necessary for the event’s occurring, each aspect is
considered when contemplating the whole cause. If neither part is a sufficient cause
alone, then neither part could be deemed fully causally responsible. Furthermore, even if
both were sufficient causes for such an event, whether one in fact contributed more than
the other to the actual event is something of interest to us. In such cases where agents are
concerned, moral responsibility can be doled out to whomever is the partial cause, but we
tend to think that some responsibility must be reserved for the other factors. This is
especially the case if other factors include other agents and, in particular, other agents
over whom the original subject has little or no control.

Such a view suggests that since there is a limited amount of causal responsibility
for an event that can be attributed to various causal factors, so too is there a limited
amount of blame appropriate and it can be distributed to a single agent or across multiple
agents. Therefore, when blame is shared, it gets diffused; the more people that are
deemed blameworthy, the less blame each person deserves. This consequence is
problematic—why should one be less blameworthy simply because others are also
blameworthy? The problem rests on a common assumption that we are only to blame for
what we cause.

In general, philosophers addressing moral responsibility or praise- and
blameworthiness seem to hold that moral praise- and blameworthiness requires moral
responsibility, which in turn requires causal responsibility. This assumption seems to me to be false and, therefore, I shall challenge it in this thesis. I shall argue that blameworthiness and praiseworthiness are not solely related to causal responsibility. Indeed, the degree or amount of praise- or blameworthiness might have nothing to do with the degree or amount of causal (and thereby, moral) responsibility. This is to say, there is and ought to be plenty of praise and blame to go around.

**Ways of Being Responsible**

Before we can begin to see how this is the case, it is important to distinguish the different senses of moral responsibility that are being discussed in the literature and isolate the sense of the term that I shall be considering throughout this thesis. Garrath Williams identifies at least four main types of moral responsibility attributable to individuals: moral agency, prospective responsibility, responsibility as a virtue, and retrospective responsibility.\(^2\) *Moral agency* involves whether an individual has satisfied the minimum requirements to be considered a moral agent and a legitimate member of the moral community. Most people hold that individuals such as very young children, squirrels, and rocks do not meet these minimum requirements. Such requirements might include: a certain set of cognitive capacities, the capacity to perceive one’s surroundings, acknowledgement of and appreciation for moral norms, etc. Responsibility in the sense of moral agency is a baseline or *eligibility requirement* for all other types of moral

\(^2\) Williams, Garrath (2006).
responsibility. That is to say, if one does not qualify as a moral agent, one cannot be a responsible person, have responsibilities, or be legitimately held accountable for past behavior.

Prospective responsibility is the sense in which a person can be said to have a responsibility. That is, a person has prospective responsibilities if she has duties or obligations in virtue of some role that she holds or some contract that she has entered. For example, a mother has the prospective responsibility to provide food for her children. She has this responsibility in a way that others do not. Her role and position as a parent grounds her duty and responsibility to perform this task well.

Responsibility in the sense of being virtuous is a way to assess a person’s moral character in general. This is the sense in which we might consider a person to be a particularly responsible person. This is not to say that the person is necessarily responsible for much (i.e., in the sense that she caused very much). Instead, the responsible person is one who is concerned with discharging her prospective responsibilities and does so reliably. If we say someone is responsible in this sense, we mean that the person can be counted on to do the right thing as well as to hold herself accountable on the rare occasion that she fails to do the right thing.

The fourth and final type of moral responsibility identified by Williams is retrospective responsibility. Retrospective responsibility is the type of responsibility often considered when discussing agents’ praise- and blameworthiness for or in virtue of things that have already happened. It is called “retrospective” in that one is assessing the degree of blame or praise warranted by “looking back” to the past. When assessing this
type of responsibility, we consider the agent’s motives, intentions, control over what she was doing, etc., all of which, in a very qualified way, determines the agent’s desert of praise or blame.

It should be noted that all four of these types are distinct from more legal notions such as liability. While there are certainly cases in which one is morally responsible and liable for the same consequences, these notions can come apart. For instance, cases of liability are distinct from moral responsibility sometimes when an agent cannot be legitimately held responsible for things that she did or did not do as a result of lack of control, but nevertheless must make restitution. For example, one who accidentally breaks china in a china shop is liable to pay for it, even if the person did not do it on purpose and could not have avoided doing so. Similarly, a company might be strictly liable for cleaning up a toxic spill even if it has been proven that a rogue employee caused the spill and the company had taken every reasonable step to avoid such spills and to hire good people. The company is not at fault, yet the company is liable.

The type of responsibility that I wish to address in this thesis is the fourth type of moral responsibility mentioned above—retrospective responsibility. It is this type of responsibility that is most often associated with praise- and blameworthiness. While the association is frequently noted, precisely how praise- and blameworthiness are connected to retrospective responsibility remains unclear in the contemporary literature. The task of this thesis is to consider this connection and offer a positive theory of how praise- and blameworthiness are connected to moral responsibility along with a positive theory of praise- and blameworthiness themselves. In order to offer such an account, I first
consider what contemporary theories have said about praise- and blameworthiness or moral responsibility. In doing so, I shall focus entirely on theories of moral responsibility and praise- and blameworthiness that are found in the literature on free will. While there are important contemporary views of moral responsibility that have not engaged the free will literature directly (e.g., the work of Barbara Herman), I have found the free will literature’s treatment of these issues to be particularly interesting, and particularly worth focusing on. I hope to turn to other theories of moral responsibility and praise- and blameworthiness (derived purely from concerns in ethics, for instance) in another work.

**Realism**

Before mentioning the specific theories we are about to review, a preliminary remark concerning the type of theories we shall engage is in order. This thesis addresses what I shall call *realist views* of moral responsibility, praiseworthiness, and blameworthiness. Realist views hold that moral responsibility and praise- and blameworthiness are real features of persons (i.e., properties belonging to those persons). In other words, such views consider our practices of holding people morally responsible and judging them to be blameworthy or praiseworthy to be practices that deal with features of the world that are separate from such practices and judgments. These views allow that such practices can be mistaken (or unjustified) and such judgments can be accurate or inaccurate (and
propositions expressing such judgments can be true or false). Antirealist accounts, in contrast, reduce moral responsibility and praise- and blameworthiness to other factors.\textsuperscript{3,4}

I am going to argue that debates over moral responsibility can often be understood as being derived from one of two underlying approaches to the topic, which I will call the sourcehood approach and the reasons-responsiveness approach. Sourcehood approaches focus on how an agent can be said to be the source of her behavior. For example, sourcehood views might focus on the way that an agent causes her behavior in contrast to possible alternative influences. Reasons-responsiveness views, on the other hand, focus on the motives for behavior. For example, reasons-responsiveness views might focus on the way an agent uses or fails to implement her capacities for judgment when acting.

The views that I have chosen to showcase and criticize (in chapters 3 and 4) are representative of a recent trend to champion either sourcehood or reasons-responsiveness as integral to moral responsibility and praise- and blameworthiness. The selection is by

\textsuperscript{3} For example, “Strawsonian” accounts (named after P.F. Strawson’s account presented in “Freedom and Resentment”) tend to reduce the properties of moral responsibility, praise and blame to social constructions. For Strawson, moral responsibility is justified by the social practices and human attitudes that are associated with assessing moral responsibility and that no external justification is required; that is, the social system justifies itself as a basic fabric of humanity. Since Strawson’s paper, many philosophers have adopted similar views but with realist bents (e.g., Arpaly (2006), Wolf (1990)). Therefore, while some of these accounts consider themselves “Strawsonian” due to their heritage rooted in the ideas of P.F. Strawson, I consider their disagreements regarding the realism question to be significant and distinguishing ones. An even clearer case of antirealism about moral responsibility is advanced by Daniel Dennett (1984). According to Dennett, we are justified in holding people morally responsible just in case the consequences of not doing so are worse than those of praising/rewarding/blaming/punishing them. Therefore, being morally responsible just is being in a position where being held morally responsible would yield favorable consequences for society at large (\textit{mutatis mutandis} for praise- and blameworthiness).

\textsuperscript{4} It is worth noting that the above distinction allows one to be a realist even if one does not believe in moral responsibility. For instance, Derk Pereboom argues for hard incompatibilism—the view that moral responsibility is neither compatible with indeterminism nor determinism (2001). While he denies that human beings are capable of moral responsibility, Pereboom illustrates what moral responsibility would be like. Still, I count Pereboom among the realists since, like other realists, he holds that if there is moral responsibility, it is due to some fact concerning, or feature of, agents.
no means exhaustive, and the proposed trend is not inclusive. That is to say that some views are not mentioned for the sake of brevity and the focus of each chapter. Also, some contemporary views do try to place equal value on both sourcehood and reasons-responsiveness (one of which will be addressed) while other contemporary views abandon both elements in the place of a third priority (e.g., sentiments).\(^5\) Therefore, I do not mean to suggest that all contemporary views will fit nicely in one of these two categories. What I do mean to suggest is that many if not most views qualify as either sourcehood views or reasons-responsive views and that it will be intellectually productive to look at the literature through this lens.

**Taxonomies of Theories**

Before we can closely examine each trend, we must first consider how my division of theories differs from the dominant way to conceptually distinguish contemporary theory types. Therefore, let us now consider how the division of contemporary theories in the way that I am suggesting is distinct from the traditional way to divide them up. The traditional typology begins with a division between *compatibilists* and *incompatibilists*. While these terms will remain convenient for our purposes, it is important to see how my viewpoint of these theories includes, but does not focus on, this division.

The established taxonomy of contemporary theories of moral responsibility involves segregating them into two main groups: those that are incompatibilist and those

---

\(^5\) For example, P.F. Strawson’s view depends on our reactive attitudes whereas R. Jay Wallace’s view focuses on moral sentiments.
that are compatibilist. The incompatibilist type is further subdivided into libertarians and hard determinists, whereas the compatibilist type is essentially a free-for-all of ways of viewing determinism, free will, and moral responsibility so that they are not mutually exclusive. The common taxonomy can be illustrated thusly:

![Diagram of taxonomy]

**Figure 1: Common Taxonomy: moral responsibility dependent upon free will**

I have no interest in challenging this conception of views of free will; however, in this thesis I shall offer an alternate taxonomy of contemporary views based, not on free will specifically, but rather on moral responsibility and praise-and blameworthiness. My taxonomy can be illustrated as follows:
While these two pictures differ in organization, the theories represented are, for the most part, the same. The first figure illustrates common ways to view answers to the question:
are free will and determinism compatible? The second figure illustrates how we might see the relationship between views that answer the question: what are the conditions for moral responsibility (or praise- and blameworthiness)?

The second figure does not include commitments about free will because not every moral responsibility theorist is committed to the view that free will is a precondition for moral responsibility. People who hold this belief might consider a diagram of moral responsibility to mirror precisely that of the first figure, since the assumption of dependency is often taken to mean that what is true about the moral responsibility of an agent is also true concerning the freedom of the agent’s will. However, recently proposed views do not fit nicely into such a pattern. For example, John Martin Fischer espouses a view he calls semicompatibilism. According to semicompatibilism, determinism alone does not threaten moral responsibility, but it does rule out free will. Therefore, since we might still have moral responsibility in a determined world but not free will, free will is not a precondition for moral responsibility. Views such as semicompatibilism suggest that it is worth asking the two questions mentioned above separately. This thesis concerns only the second question; however, most of our discussion will include theories of free will since free will, moral responsibility, and praise- and blameworthiness are intertwined in many theories.
Sourcehood Views

Sourcehood views of moral responsibility and praise- and blameworthiness are theories that are focused on the question of whether the potentially responsible (or praise- or blameworthy) agent is the distinguished causal origin of the action for which she is responsible (or praise- or blameworthy). Such views might be concerned with whether the agent is the only sufficient cause of her behavior, or whether the agent had a certain amount of control over her behavior. What all sourcehood views share in common is that they identify the agent’s status as the source of her behavior as the fundamental underlying condition that, if satisfied, can establish the agent as the appropriate object of various assessments (e.g., that she was free to do otherwise, that she is responsible for what she caused, etc.). All sourcehood theorists agree that what is important to an agent’s freedom and moral responsibility is that she be the origin or source of her behavior in some important sense. Oftentimes this “important sense” involves defining sourcehood negatively in order to bar certain alternative influences. For example, Robert Kane writes, “Free Will also seems to require that the sources or origins of our actions lie “in us” rather than in something else (such as the decrees of fate, the foreordaining acts of God, or antecedent causes and laws of nature) outside us and beyond our control.”6 Which sources a theorist is concerned about excluding can determine whether the theorist is a source compatibilist or a source incompatibilist. For our purposes, we shall consider both types of scholars to be sourcehood theorists.

---

Views of free will and moral responsibility that focus on an agent’s being the source of her behavior are usually incompatibilist views. Kevin Timpe defines *source incompatibilists* as those who “think that free will is primarily a matter of being the ultimate source of one’s actions.”7 While I agree with Timpe that this thought is common among all source incompatibilists, I also contend that this thought can be found as the grounding for several compatibilist views as well. For example, compatibilist A. J. Ayer argues that one is free if one causes herself to act in such a way that is distinct from coercion and compulsion—meaning that no external or internal force is the source of her behavior.8 While Ayer maintains that determinism does not threaten one’s being this kind of source of her choice, Ayer does acknowledge that forces such as external threat or internal compulsion (such as that involved with kleptomania) preclude the agent as being the appropriate source, and this suggests that Ayer’s focus is on whether the agent plays a distinguished role in the causation of her action.

While many sourcehood theorists are incompatibilists, they are incompatibilists because of what they take to be necessary conditions for sourcehood to be achieved. That is to say, such theorists believe that robust alternative possibilities that only an indeterministic world could provide are required for sourcehood. As sourcehood theorist and hard incompatibilist Derk Pereboom puts it, “Moral responsibility requires actions to have indeterministic actual causal histories, or more fundamentally, to have causal histories that make agents the ultimate sources of their actions.”9 It is this adjective

---

7 Timpe (2008): 77.
8 Ayer (1982).
“ultimate” that marks the difference between source compatibilists and incompatibilists. Some might hold that one must be the source of one’s behavior, thereby excluding other potential sources aside from antecedent causes and laws of nature. It would be open to such a theorist to maintain compatibilist principles. On the other hand, other theories maintain that the agent must be the only sufficient cause or origin of her behavior, in which case, incompatibilism seems a natural route to take.

Regardless of whether one is a source incompatibilist or a source compatibilist, all sourcehood theorists maintain that sourcehood is the most important condition for freedom and moral responsibility. Other issues such as control, motives, and awareness of reasons are also important; yet these are secondary to the primary requirement that one be the source of one’s behavior.

Reasons-Responsiveness Views

In contrast to sourcehood theorists, reasons-responsiveness theorists are first concerned with the way in which agents can and do respond to moral reasons. While each reasons-responsive theory is unique, I follow Fischer and Ravizza’s description of what reasons-responsiveness in general involves. According to Fischer and Ravizza, the phenomenon of an agent responding to a reason can be illustrated by a model of the process as consisting of two steps.

The first step or aspect of reasons-responsiveness is recognition. Agents first come to see certain features of the world or possibilities in a given situation in a certain
light. Such features are viewed as either desirable or good, or as undesirable or bad. The agent considers the desirability or goodness of the feature to constitute or be indicative of a reason to promote or avoid that feature. For example, in the case of a good feature, Juan might see a piece of cake in a display case at a coffee shop. He might view that cake as desirable or good. In so viewing it, Juan considers possible actions that might lead to his acquisition of the cake to also be good. For instance, he might consider the act of taking out his wallet and informing the attendant that he would like to buy the cake to be good acts to do since they appear to promote his acquisition of the desirable object. In this way, we seem to first recognize the value the desirable object appears to have and such recognition then generates other mental states, such as a further recognition of what potential actions could promote the end in question as well as viewing such actions to be worthwhile. This type of scenario can happen in a negative repulsive way as well. For instance, if Juan were to look more closely at the cake and catch a glimpse of the mold on the far side of the pastry, he might come to view the object as undesirable. In so doing, he might then consider avoiding the cake a worthwhile pursuit. He might even judge that nothing behind the display case should be trusted, thereby deeming no such items to be valuable. He would presumably then judge that he ought to put his wallet away and sit down. In each of these cases, that of attraction and that of repulsion, the agent recognizes something as good or bad and worth pursuing or avoiding and considers possible actions to have similar values according to their likelihood of contributing to the end in question.

While reasons-responsiveness involves recognition of reasons to act or refrain from acting, recognition of the reason is only the first step. For one could potentially
recognize a reason to act and fail to be motivated by that reason to act. Fischer and Ravizza remind us that a good account of reasons-responsiveness acknowledges “the required connection between reason and action,” which entails that agents act intentionally and for that reason for the action to be “appropriately connected” to an agent’s reason. So a typical reasons-responsive event would include both the recognition of a reason to perform an action as well as such recognition being followed by an action done for that same reason.

Finally, it is important that the agent recognizes and reacts to reasons that make sense in order to be considered sufficiently reasons-responsive to warrant ascriptions of moral responsibility and praise- and blameworthiness. What an agent recognizes as a reason must fit what Fischer and Ravizza consider a coherent pattern of reason-recognition. This means that a third party must be able to comprehend why that which the agent recognized as a reason really would be a reason for the action in question. To borrow and modify one Fischer and Ravizza’s own examples, imagine that there is a man armed with a large and dangerous saber on a ship. He is about to murder the passengers with his saber. Just before he swings his saber high above his head, he stops, puts the saber down, and announces his decision to not kill anyone. If, when asked what changed

---

11 Fischer and Ravizza also add that acting for a reason is required to exclude wayward causal chains. It is unclear as to whether Aristotle worried about such cases. The closest discussion that comes to mind is his treatment of chance and luck (Physics II. 4-6). He describes chance as something that happens randomly; luck is reserved for those instances of chance that do not have final causes, but could have (i.e., there exists a potential for a rational agent to choose that action for some end or purpose). Aristotle does mention that character, virtue, and happiness are matters too important to leave up to chance—perhaps suggesting that chance ideally has no role to legitimately play in such matters.
his mind, the would-be mass murderer replied that he would have killed everyone only if he had not seen someone smoking a Gambier, listeners would be rightfully perplexed.\textsuperscript{13} Why is that the reason to refrain from murder? Fischer and Ravizza seem to suggest that such a reason is not the kind of reason good receptivity latches on to. It is as if the would-be murderer has not recognized a real reason (or not the right kind anyway) to not murder. In addition, we can imagine that there were probably better reasons to refrain from murdering that he did not recognize (i.e., that it is immoral, senseless, sure to land him in jail for life, etc.).

I take this description to be quite basic and acceptable to most if not all reasons-responsive theories. If this is what it is to respond to reasons, then a reasons-responsive view of moral responsibility, or of praise- and blameworthiness, will be one in which an agent’s responsibility (or praise- or blameworthiness) is established primarily by reference to the agent’s responsiveness to her reasons. Responsiveness to reasons is, on its face, quite different from playing a particularly distinguished causal role; hence reason-response theorists take quite a different approach from that of sourcehood theorists.

**Outline of Chapters to Come**

The next chapter is devoted to a careful examination of the ideas of sourcehood and of reasons-responsiveness as found in Aristotle’s work on the voluntary and on praise- and

\textsuperscript{13} Fischer and Ravizza (1998): 55-6.
blameworthiness. In this chapter, I shall first explain how Aristotle distinguished internal sources of movement from external sources. Then, I shall analyze the requirement that an agent have sufficient knowledge of the particulars of her situation in light of how this might ground an ability to recognize (and thereby react to) reasons. Finally, I shall present an interpretation of Aristotle as being a hybrid of a sourcehood theorist and a reasons-responsive theorist.

Once we have seen how sourcehood and reasons-responsiveness can be maintained on an account as equal partners, I shall begin outlining contemporary theories that do not aim to value each component equally. Chapter 3 has several goals. The first goal is to establish the selected views as sourcehood views. The second goal is to illustrate how all such views involve a commitment to alternate possibilities. Like the reasons-responsiveness theorists, I do not think that alternate possibilities are required for moral responsibility and praise- and blameworthiness; therefore, this universal characteristic binds these views together as equally vulnerable to complaints that alternate possibilities do not matter.

In Chapter 4 I survey contemporary reasons-responsiveness theories. Here we first consider how Harry Frankfurt’s article, “Alternate Possibilities and Moral Responsibility,” created new hope for compatibilist views that were not concerned with sourcehood but rather motives for acting. We then consider some such views, for instance, those offered by Frankfurt himself, Gary Watson, Susan Wolf, and Nomy Arpaly. In each case, I first aim to demonstrate why each view qualifies as a reasons-responsiveness view. Then, I draw attention to the commitment a view has to the idea
that moral responsibility and praise- and blameworthiness are only about actions—a claim that I shall later dispute.

In chapter 5 we return to Aristotle’s theory briefly to consider the importance of action in a theory of moral responsibility and a theory of praise- and blameworthiness. Here I argue that, for Aristotle, one can deserve praise or blame for actions and emotions. While one is not directly responsible for one’s emotions (and, in some cases, might not be responsible at all), this does not prevent Aristotle from finding emotions to be blameworthy. I end this chapter by motivating my own account and why I think that Aristotle is right to allow for the possibility that we might be praise- and blameworthy in virtue of mental states as well as actions that result from our mental states.

Chapter 6 consists of my alternative view—the moral attitude account. First I argue that moral responsibility is not primary to praise- and blameworthiness contrary to views presented earlier in chapters 3 and 4. I show that retrospective responsibility is a species of causal responsibility. Then, I argue that there are cases where an agent is praise- or blameworthy without being equally causally responsible. Some examples illustrate an asymmetry in the degree of responsibility and the degree of blameworthiness, while others illustrate a degree of blameworthiness absent any responsibility whatsoever. From here I conclude that praise- and blameworthiness do not depend on moral responsibility and, therefore, are not exclusive to things we cause (e.g., actions). In the place of actions, I offer an account that holds that we are praise- and blameworthy in virtue of our moral attitudes. I argue that moral attitudes are prior to both intentions and actions; while they often lead to the formation of intentions and their
resultant actions, such consequences are not required for an agent to deserve praise or blame. However, the consequences of intentions and actions are required to warrant moral responsibility. For this reason moral responsibility depends on a prior state of either praiseworthiness or blameworthiness, but praise- and blameworthiness do not require moral responsibility.
In this chapter, I shall show how the roots of sourcehood views and reasons-responsiveness views can be found in Aristotle’s claims. I shall then argue that Aristotle ostensibly held each in equal esteem regarding their relation to moral responsibility and praise- and blameworthiness because sourcehood and reasons-responsiveness are intimately connected. In essence, I hold that Aristotle’s theory is one that addresses sourcehood, reasons-responsiveness, moral responsibility, and praise- and blameworthiness. While I contend that Aristotle likely considered these four to be importantly connected (and that, if he did, I would agree with him), I do not think that Aristotle has clearly shown how these four concepts are related. It is as if four ingredients one intuitively knows go into a dish have been diligently collected and placed on the counter, but the recipe for how they are to go together is missing. In subsequent chapters, I shall attempt to pick up where Aristotle left off. But for now, let us consider how each element can be found at Aristotle’s metaphorical table.
The assumption that moral responsibility is a type of causal responsibility is a widespread one. However, while contemporary philosophers tend to have a very narrow view of what constitutes a cause, Aristotle held that there were four main types of causes of an object or state of affairs. One type of cause that is still recognized today is called the efficient cause. The efficient cause is what sets something in motion or initiates its coming to be. In the example of one domino falling, the efficient cause might be its neighbor domino that fell on it. In the case of an object’s coming to be, my parents can be said to be my efficient cause. They are causally responsible for setting in motion the cells that later developed into my body. In each of these instances, the efficient cause is prior to the effect as well as external to the effect. The former claim must always be true on an Aristotelian program. However, the latter might not be required. For instance, while I am not my own efficient cause, I am the efficient cause of much of my behavior to date.

Certainly, in these cases, my decision to act is prior to my acting. However, it is not necessarily accurate to say that my decision to act is external to my acting.

14 The proceeding discussion can be found in *Metaphysics* (hereafter *Met.* ) I. 3 (esp. 983a 24-32) and *Physics* (hereafter *Phys.* ) II.1-3. It is true that contemporary philosophers recognize the explanatory power of the three non-efficient causes. For instance, many philosophers follow Donald Davidson in holding that reasons are causes because they are helpful in the explanation of a purpose of a voluntary action. Indeed, one significant difference between contemporary philosophy and Aristotle’s view is that, for Aristotle, everything had a final cause whereas, for contemporary philosophers, we only consider final causation in cases involving rational or voluntary action. Even when contemporary scholars appear to agree with Aristotle’s characterization of causation, the common practice is to isolate efficient causation as the cause of an event or object and treat other aspects of causation as merely explanatorily helpful, somehow dependent upon the efficient cause, or, at least, of secondary importance to efficient causes.
A second type of cause is the *material* cause. As the name suggests, this is the part of an object or state of affairs that constitutes the matter. For example, the material cause of a human being would be the human body.

A third type of cause is called the *formal* cause. The formal cause determines the type of thing an object is. For instance, the form of human being is one of the causes of a particular human being; neither you nor I could exist as we do, says Aristotle, if not for the fact that we are constituted by the form of human being, namely the soul. The formal cause is intimately tied up with the function of a being and its essential nature. That is to say, the form of *human being* grounds the possibility of the function of the human being; to do what human beings do by nature (reason) and to do it well. This aspect of function ties the formal cause to the final cause.

The fourth type of cause is the final cause. This cause is of particular interest to Aristotle. The final cause is the end or purpose of a thing. It is a goal. The final cause of an acorn is to be an oak tree; the final cause of an airplane is to fly. Actions and events have final causes as well. For example, the final cause of my going to the grocery store could be that I acquire a loaf of bread. I say that this cause is of particular interest to Aristotle because he holds that there are final causes in nature; that is, nature is teleologically oriented.

While Aristotle holds that all four causes are relevant to understanding anything as it is, I would like to focus on the relationship between two causes in particular—

---

15 According to the doctrine of hylomorphism, all physical objects have both form and matter (see Aristotle’s example of a bronze statue at *Met.* VII. 10 1035a 5-10). However, Aristotle distinguishes the form of artifacts from the form and natures of natural objects in *Phys.* II (see esp. 192b 10-20). For an interesting investigation into the ontological status of Aristotelian artifacts, see Katayama (1999).
efficient and final. In this chapter, I shall argue that the reason why Aristotle does not seem to value either sourcehood or reasons-responsiveness over the other is because each exemplifies one of these types of causes and, when it comes to voluntary actions that are praise- and blameworthy, these causes work in consort. I suggest that the efficient cause exemplifies the element of source in morally responsible behavior whereas the final cause is illustrated in the element of reasons-responsiveness. In short, there is a dual causal relationship between the recognition of reasons and motivation in light of them that requires both aspects to be intact for one to deserve praise. On the other hand, if an agent recognizes good moral reasons to do something above all other options, yet fails to be so motivated, blame is appropriate.

**The Voluntary (hekousion) as Sourcehood**

The most notable discussion of praise- and blameworthiness is found in Aristotle’s *Nicomachean Ethics (EN)* in the first half of the third book. While the subject of this book is the voluntary, its nature, and how it compares to the nonvoluntary, we are told that all of this discussion is taking place since the student of ethics ought to know what to praise and what to blame (i.e., what is worthy of praise and what is worthy of blame). That is not to say that all instances of the voluntary ought to be responded to in one way or another; instead, it is the case that no instance of something that is either praiseworthy
or blameworthy could also be nonvoluntary.\textsuperscript{16} Therefore, it is an open question as to whether there is anything that is voluntary yet neither praiseworthy nor blameworthy. Indeed, common sense would suggest that there are several such instances (e.g., I may voluntarily stretch without being worthy of some feedback for it). However, the suggestion here is that there are no instances of praise- and blameworthiness that are not also instances of the voluntary.

Aristotle does not come right out and give a theory of the voluntary. Instead, he carefully discusses what the voluntary is not in order to clear the way for a more precise understanding. First, Aristotle gives us two types of actions that are generally held to be \textit{involuntary} (\textit{akousion}): those done under compulsion and those done through ignorance.\textsuperscript{17} Aristotle then clarifies the true nature of \textit{compulsion} and \textit{acting from ignorance}. It is confusion about these two phenomena that might cause one to misdiagnose a voluntary act as an involuntary one.

Compulsion is a phenomenon not well-understood in our own times, and things seem to have been no different in Aristotle’s day. For Aristotle it is clear that compulsion (\textit{bia}) involves the origin or source of the act being located externally to the agent: “What

\textsuperscript{16} Here, I shall speak generally of the nonvoluntary. Aristotle does distinguish between the \textit{merely nonvoluntary}, or behavior performed accidentally but that does not elicit pain or regret, and the \textit{involuntary} behavior, which is nonvoluntary behavior that also elicits pain and regret. In later sections and chapters, this distinction will become more important. For now, I only need to distinguish the broad genera of \textit{voluntary} and \textit{nonvoluntary}.

\textsuperscript{17} “Now it seems that things coming about by force or because of ignorance are involuntary” (\textit{EN} 1110a 1-2). Henceforth, quotations and references to Aristotle’s works that point to specific lines will use the Bekker number convention.
is forced has an external principle, the sort of principle in which the agent, or [rather] the victim, contributes nothing." 18 That is, force is always external. 19

Aristotle does, however, address one type of apparent compulsion that the common opinion (endoxa) considers internal: that pertaining to fear of greater evils. In such cases of “mixed acts,” the act itself is something that, in general, would not be voluntarily done. Still, under the circumstances of the particular instance of such an act, it might be one worth doing. Aristotle gives examples of acting under the coercion of a tyrant who is threatening one’s family, or deciding to throw cargo overboard a ship in order to prevent it from foundering in a severe storm. In such cases, what one would normally want to do is either not available or is truly unattractive. Indeed, Aristotle tells us that, “the end or motive of an act varies with the occasion, so that the terms ‘voluntary’ and ‘involuntary’ should be used with reference to the time of action.” 20 He declares that such actions are voluntary, since “at the actual time when they are done…the origin of the movement of the parts of the body instrumental to the act lies in the agent.” 21 In other words, the property of being voluntary hinges on the token instance of an act and whether that token act was initiated by movement internal to the agent.

18 EN 1110a 3.

19 It would seem as though Aristotle was not sympathetic to what modern thinkers might consider internal compulsions, such as addictive desires, fears, etc. For all such urges are internal to the agent. On this account, the only way to conceptualize the agent such that these urges render behavior involuntary is to consider the agent a subsection of the soul. Some later views identify the agent in ways that allow for a distinction between internal and external urges. For example, some identify the agent with a particular subsection of the soul. For instance, one might identify the agent by noting which desires are of the highest reflective order (Frankfurt) or which involve values as opposed to mere desires (Watson).

20 EN 1110a 12.

21 EN 1110a 16-17.
In *Nicomachean Ethics*, Aristotle suggests that the external factor that renders an action involuntary is such that the agent is entirely passive and contributes nothing to the act: “What is forced, then, would seem to be what has its principle outside the person forced, who contributes nothing.”\(^{22}\) Aristotle’s own example involves a man being swept up by a strong gust of wind or carried off by a large group that has physically overcome him.\(^{23}\) Such an image calls to mind physical forces that redirect the course of one’s body, or perhaps constrain the body from moving. In any case, *external* here suggests movement initiated in such a way so as to be detached from and foreign to the soul. And Aristotle is clear that it is because such an external force is the *source* (*arché*) of the behavior (and not the man) that the act is rendered involuntary. Therefore, we may infer from this discussion that the voluntary requires that the initiating movement (*kinesis*) of that which is voluntary be located within the agent:

Criterion 1: *sourcehood*: the origin of the motion of that which is voluntary must be internal to the agent that is being considered for praise or blame.

Immediately from this discussion we can see how *sourcehood* relates to one’s worthiness of praise or blame. One cannot be apt for praise or blame unless one’s behavior is voluntary. Since one’s behavior cannot be voluntary without being initiated from within (i.e., the agent is the source), then *a fortiori* one cannot be eligible for praise or blame unless one is the source of her own behavior. Therefore, it would seem on

---

\(^{22}\) *EN* 1110b 17.

\(^{23}\) “…if, for instance, a wind or people who have him in their control were to carry him off” (*EN* 1110a 4).
Aristotle’s account that *sourcehood* is relevant (and, indeed, required) for praise- and blameworthiness.

The location of the origin or source is at issue here not simply because we look to which object to praise or blame. That is, if we were to discover that the source of a given token behavior was, in fact, the wind (i.e., something external to the human being in question), we would not apply praise and blame to that which *is* the source. Therefore, either being the source alone is insufficient for an object to be the proper target of praise or blame, or Aristotle has a narrow sense of *source* in mind. Indeed, I think that both are true.

Let us first get clear on Aristotle’s notion of *source*. Aristotle clarifies his claim that one’s behavior is voluntary if one is the source of that behavior: “if the principle of the actions is in him, it is also up to him to do them or not to do them.”

24 In other words, being the source of an action entails that one has an ability to do otherwise. Now, there is a broad notion of the term *source* for which this claim might be false. If we take *source* to mean the physical body from which the movement comes we can quickly imagine several counterexamples to this claim. For example, one might experience a nervous twitch that was not triggered by any external stimuli. The person would be the source of the twitch. However, it seems odd to hold that the person could have controlled the twitch and elected to not twitch. But there is a narrower sense of *source* that might exclude such cases: a *source* could be that which efficient-causes the effect in question. In the example of the nervous twitch, we need not say that the person *qua agent* caused themselves to

24 *EN* 1110a 18.
twitch; on the contrary, it is more accurate to say that some non-agential part of the person’s body (blood sugar levels, individual neurons, etc.) caused the agent to twitch. On the other hand, if an agent causes herself to pick up a book, it would also be true that she had the power to cause herself to refrain from picking up the book.

This clarification of what is entailed by sourcehood for Aristotle reveals a further salient question: how is it that one causes oneself to act? I shall now argue that an Aristotelian account of causing oneself to act necessarily involves a type of reasons-responsiveness. Following this discussion, I shall return to the discussion in *Nicomachean Ethics* where a second criterion is suggested and illustrate how this requirement, knowledge of the particulars, implicitly suggests reasons-responsiveness as well.

**Aristotelian Reasons-Responsiveness**

That reason will be related to moral responsibility and praise- and blameworthiness is obvious in some ways. For instance, for Aristotle, man is a *rational* animal, and this aspect of man’s soul and capacity for deliberative choice and intellectual thought is what distinguishes human beings from other similar creatures.\(^\text{25}\) So, broadly speaking, reason in general is essential to everything human. In addition, Aristotle never attributes praise- or blameworthiness to any non-human animal, and that suggests that Aristotle himself holds the view that animals are not morally responsible for their behavior, even though

\(^{25}\) *De Anima* (hereafter *DA*); c.f. *EN* 1097b 20-1098a 9 in which a discussion of reason being the distinguishing mark and function of man can be found.
their behavior can be voluntary.\textsuperscript{26} If so, moral responsibility would also be something uniquely human on Aristotle’s view. If something is unique to human beings, it is very likely related to that which is unique in human beings—reason. Finally, Aristotle pays close attention to our capacity for \textit{practical reason}, which is a capacity distinct from theoretical reason. Practical reason determines not what things are or are like (as does theoretical reason), but rather, what the agent is to do. Therefore, action is governed by reason for Aristotle. So, these are a few pieces of evidence that reason in general, and that practical reason in particular, are tied up with the voluntary and how agents deserve praise or blame. I do not suppose that any scholars deny these basic elements of interpreting Aristotle. So why ought we to address them? The answer is simple. For Aristotle, reason is important to moral responsibility and praise- and blameworthiness insofar as it is a \textit{faculty} or aspect of the soul intimately wrapped up in all that constitutes voluntary and praise- and blameworthy behavior. In addition, it seems that for Aristotle, \textit{reason qua faculty} can recognize and respond to considerations that modern and contemporary philosophers term \textit{reasons}. In other words, I shall argue that, for Aristotle, an element of moral responsibility depends on the human faculty and its ability to respond to moral considerations.

\textsuperscript{26} At \textit{EN} 1111a 25, Aristotle suggests that animals behave voluntarily. Here, he addresses why actions directly resulting from emotions ought to still count as voluntary since, if the opposite were the case, no animal or child could be said to behave voluntarily. He echoes this sentiment again later in his discussion of why the voluntary in general is distinct from decision or choice. At \textit{EN} 1111b 8-9, Aristotle tells us that animals and children do not have a share in decision; and the tone of the prose suggests that animals and children serve as counterexamples to any claim that all that is voluntary is decided. The subtle implication here is that we should want to have an account that preserves some intuition that animals \textit{do} have the capacity for voluntary behavior. This is not sufficient to suggest though, that animals deserve praise and/or blame according to Aristotle. While Aristotle says that we only apply praise and blame to the voluntary (if praised or blamed, then voluntary), it is not obvious that the converse holds (if voluntary, then deserving of praise and blame).
While considering the evidence that is about to be presented, let us recall from chapter 1 the basic structure of the process that is called *reasons-responsiveness*. First, an agent *recognizes* considerations, facts, etc. as constituting reasons for or against performing a certain course of action. Second, the agent *reacts* to this recognition by moving towards behaving in the way that the agent thinks the consideration justifies. Therefore, in our exploration of Aristotle’s theory of action, we shall consider both how he might have viewed the awareness of moral considerations as well as motivation towards acting in light of them.

**Actions Have Ends**

It seems as though Aristotle accounted for something like the recognition of reasons in his system. For example, Aristotle tells us that all actions have ends or goals. A *telos*, or end, seems to be an ancient version of our notion of a reason to act. When one’s action has an end, there is a purpose for that action; and an action done for a purpose is done for a reason. This is true both in the sense of explanation as well as that of intention. For example, if one were to ask why Juan gave the cashier a $10 bill, a reasonable explanation might be that he wanted to purchase the cake. So, in the sense of an explanation, citing the *telos* can make sense of an action. But this kind of sense is also *intentional*.

---

27 EN 1094a 1-2.
There are all sorts of reasons for behavior, some of which are the kinds that involve *telos* and some of which do not. For example, we can imagine that Juan is neurologically odd, and any time that he starts to desire a pastry of any sort, his hand automatically reaches into his pocket, pulling out whatever contents it finds, and extends them outwards. In fact, this neurological tic is rather embarrassing for Juan, since he does not mean for this behavior to occur and he seems powerless insofar as his ability to prevent such behavior from being exhibited. In this sense, we might cite the same reason when asked why Juan reached into his pocket, grabbed a $10 bill, and extended it towards the cashier. That is, he exhibits such behavior because he wants cake. However, in this case, he had no purpose or intention in behaving as such.

On Aristotle’s system, we can distinguish between each example concerning Juan based on the type of cause to which we point. That is, in the first case, the cause of Juan’s action that we cite as a reason is a *final cause*. In the second case, the cause of Juan’s action that we cite as a reason is an *efficient cause* and not a final cause. Citing a reason in each case is explanatory and helpful to onlookers. However, only in the case of a reason that is a final cause can there be something that speaks to the voluntary nature of Juan’s behavior. In each case there is a sense in which Juan is the efficient cause (i.e., the source) of the behavior; but only in the first instance can Juan act voluntarily, as the final cause is in some important way related to Juan’s motive. And, as Aristotle tells us, we can only be praised and blamed for that which is voluntary.

Now that we have seen how the voluntary involves some kind of motive, let us examine further the epistemic aspect of such motives (i.e., how they are recognized by
the agent) and how correct epistemic grasp of salient moral reasons can render an agent praiseworthy.

The End as that which Appears Good

While there is one common end or goal that all human beings naturally pursue—eudaimonia—28—we aim to achieve this end in very different ways. Aristotle tells us that, while all men aim to be happy, we tend to disagree widely about what constitutes and what leads to happiness.29 And while we are all naturally inclined towards happiness, and while happiness is in fact the highest most self-sufficient and complete good, we are not motivated by what is in fact good. On the contrary, we are moved by what Aristotle calls the apparent good.30

Aristotle’s notion of the apparent good and how it factors in to being motivated to act is similar to one theory of moral psychology prevalent today. If one were to analyze Aristotle’s view of moral psychology and human motivation and anachronistically apply our terms to his thought, an accurate description of Aristotle’s view might be what philosophers of mind today call motivational internalism. According to the most common way to express the thesis of motivational internalism, believing or holding something to be good is sufficient for being motivated towards acquiring, pursuing, or preserving that object. As Miller explains, this view holds that: “Necessarily, for any agent S, if S judges

28 EN I.4.
29 Aristotle discusses the many different notions of “the good life” at EN I.5.
30 See EN III.4.
that some available action is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), then S is motivated at least to some extent to perform (refrain from performing) that action.”

Few philosophers have advocated the strong version of this thesis: that not only does considering something to be good motivate one towards it, but in every case one is sufficiently motivated to act by what one considers best:

Necessarily, for any agent S, if S judges that some available action is morally right (or good, or obligatory, or . . .) for S to perform (refrain from performing), then S is motivated at least to some extent to perform (refrain from performing) that action and the added motivation that the agent has as a result of having formed this judgment, has its source solely in the judgment itself, and not in the individual or joint contribution of any of S’s other mental states.

On such a view, *akrasia* or weakness of will is not possible. We know that Aristotle indeed wanted to preserve the appearance and support the intuition that akrasia is a fact of human nature, since he makes it a point to reject Socrates’ strong motivational internalism. However, it is clear that Aristotle does advocate a weaker type of

---

32 ibid.
33 Since *akrasia* is the phenomenon of knowing the better and doing the worse, strong motivational internalism is incompatible with the possibility of akrasia. If one believes or knows something to be a better course of action than others, according to strong motivational internalism, that knowledge would sufficiently motivate the person to act and preclude her voluntarily choosing an inferior option. C.f. Miller (2008).
34 *EN* VII.2 1145b 23-29.
motivational internalism, as he holds that all actions aim towards some good and that human beings act in order to achieve what appears good to them as individuals.

Our ability to retroactively name Aristotle a motivational internalist supports my claim that he had an element of reasons-responsiveness in his theory of the voluntary and praise- and blameworthiness. While it is not necessarily the case that motivational externalists are precluded from being reasons-responsive theorists, the fact that Aristotle was an internalist about motivation invites us to consider what distinguishes movement towards the apparent good from reasons-responsiveness. Indeed, it seems to me that, especially in Aristotle’s case and concerning the terms as we have identified them, there is no difference. For what it is to recognize something as a reason to act is to view an object as good in the relevant way that affects motivation to act. The fact that Aristotle holds that all actions are done for ends and that ends are adopted because they appear to be good and worth pursuing suggests that Aristotle’s theory is a reasons-responsiveness view.

As a reasons-responsive view, Aristotle’s theory is concerned primarily with our rational nature. We saw earlier how a wide sense of the term sourcehood might include parts of the agent that are external to her reason. For example, in the wide sense of the term, an agent is the source of her nervous twitching—the twitching originates in her (i.e., her body). For this reason, I prefer to use a more narrow sense of sourcehood since sourcehood theorists aim to capture some notion of control. It is important to see how reasons-responsiveness, like our modified view of sourcehood, involves a limited part of the human psyche.
Parts of the Soul

Not only are praise- and blameworthiness characteristic of the voluntary, but they are also characteristic of behavior emanating from a limited part of the human soul. Aristotle tells us that there are three main parts or aspects of the soul. The first is the plantlike or nutritive part of the soul. This aspect takes care of bodily tasks such as digestion, growth, and reproduction. Human beings have this aspect of the soul in common with all living things. A second aspect of the soul is animalistic, or appetitive. The appetitive part of the soul is responsible for desires, emotions, and mobility. Aristotle explains, “What is called desire is the sort of faculty in the soul which initiates movement.”35 Human beings have this aspect of the soul in common with all other animals. The third part of the soul is unique to human beings—reason. It is this aspect of the soul that performs both theoretical and practical inquiries and deliberations.36

Aristotle is clear that we are only going to be concerned with the rational and appetitive aspects of the soul when we discuss subjects such as virtue and the voluntary. He writes, “let us leave aside the nutritive part, since by nature it has no share in human virtue.” There must be something about the other two aspects that renders them relevant to such discussions.

35 DA III.10, 433a31-b1. For animals, appetite alone moves them, since they have no rational capacity (DA III.9, esp. 432b 14-33a 5). However, humans move with appetite and sometimes against it, as is the case with the continent person (DA III.10, 433a 17-19; EN I.13 1102b 26).
36 C.f. EN VI.8, esp. 1143a 35-1143b 10.
The relationship between reason and appetite is revealed in a further distinction that Aristotle makes. In *Nicomachean Ethics*, Aristotle divides these three aspects of the soul further into two main parts: the *rational* and the *nonrational*: “We have said, for instance, that one [part] of the soul is nonrational, while one has reason.”

The rational aspect of the soul consists solely of reason. The nonrational part of the soul consists of the appetitive and the nutritive parts. At first glance, this would seem to put appetite into the class that should be disregarded. But Aristotle tells us that there is another sense in which appetite is rational: “this [part] as well [as the rational part] appears, as we said, to share in reason. At any rate, in the continent person it obeys reason; and in the temperate and the brave person it presumably listens still better to reason, since there it agrees with reason in everything.”

Therefore, the appetitive part of reason is related to the rational part, since it is able to *listen to* and *obey* reason:

For while the plantlike [part] shares in reason not at all, the [part] with appetites and in general desires shares in reason in a way, insofar as it both listens to reason and obeys it….The nonrational part also [obeys and] is persuaded in some way by reason, as is shown by correction, and by every sort of reproof and exhortation….If, then, we ought to say that this [part] also has reason, then the [part] that has reason, as well [as the nonrational part], will have two parts. One will have reason fully, by having it within itself; the other will have reason by listening to reason as to a father.

---

37 *EN* 1102a 29-30.
38 *EN* 1102b 26-29.
39 *EN* 1102b 30—1103a 4.
So virtue and, *a fortiori*, the voluntary will involve the appetitive and rational parts of the soul. Indeed, Aristotle tells us that reason alone moves nothing, and that appetite is responsible for mobility.\(^4^0\) If this is correct, then it is safe to say that *causing oneself to act* involves both reason and appetite. Still, while appetite would seem to be required for mobility, and is, therefore, an efficient cause of action, appetite is *subordinate* to reason. Ideally, appetite *listens* to reason. But what does reason tell appetite? It would seem that reason *recognizes* some action as worthwhile and commends such an end to appetite, which in turn often leads appetite to so act. When appetite receives the command, it initiates movement, which is the start of action. In other words, Aristotle’s comments lead one to believe that, if the only two aspects of the soul involved in virtue are reason and appetite, and if appetite alone has the power to move, *but* appetite is supposed to listen to reason, then reason must provide some end (final cause) and, in response, appetite initiates motion (efficient cause).

### Acting On Purpose

So far, we have considered how it is that Aristotle’s moral psychology allows for agents to recognize reasons to act, and we have stipulated that such recognition triggers reaction. Since Aristotle seems particularly concerned with the reaction to reasons, it is worth our

\(^4^0\) Aristotle claims that “Thought alone moves nothing,” (*EN* VI. 1, 1139a 35-36; *DA* III.9 432b26–7) and “what moves is the single faculty of desire” (*DA* III.10 433a22). However, he also claims that there are three types of desires: that of appetite (*epithumia*), spirited emotion (*thumos*), and rational desire (*boulésis*). Therefore, since Aristotle appears to have allowed for there to be rational desires, it might be too crude to consider Aristotle’s view to be one in which agents are helpless against the motivation of their appetites. Still, under every standard interpretation, it is accurate to understand Aristotle as claiming that reason does not move us unaided by desire.
while to examine more closely this aspect of reasons-responsiveness on Aristotle’s account.\textsuperscript{41} In this section, let us consider how recognition can lead to reaction in a normal situation where an agent consciously considers her action before deciding to perform it.

We can learn more about how Aristotle might have had a view of reasons-responsiveness if we carefully scrutinize his discussion of deliberation and choice. Whenever we deliberate, according to Aristotle, we first somehow adopt an end and then deliberate. But, as we noted above, it is not as though we adopt any old end. Aristotle makes it clear that we adopt ends that we think we can achieve: “[we deliberate about what is up to us, i.e., about the actions we can do.”\textsuperscript{42} For Aristotle, it is clear that we do not deliberate about beliefs, as if we had control over their truth or falsity; nor do we deliberate about particulars, as if we could make something other than it is by deliberating alone.\textsuperscript{43} Also, we do not deliberate about what we deem to be impossible, for that is what distinguishes deliberation from a general wish; we can wish for the impossible, but “[we do not decide on impossible things—anyone claiming to decide on them would seem a fool.”\textsuperscript{44} In other words, we only deliberate about what we think is possible for us to achieve because the goal of deliberation is choice. Also, Aristotle adds:

“The principle of an action—the source of motion, not the goal—is decision; the

\textsuperscript{41} This is evidenced by the fact that Aristotle rejects the idea that one could remain virtuous a lifetime if asleep or completely idle (\textit{EN} 1095b 35-1096a 4). Also, virtue is at the very least necessary for happiness and, quite possibly constitutive of happiness (\textit{EN} I). Happiness, for Aristotle, is not a state, but an activity (\textit{EN} 1098a 2-21). Also, as we shall see, human beings become virtuous by \emph{doing} the types of actions that the virtuous person does \emph{repeatedly} (i.e., \emph{not simply in recognizing the good}) (\textit{EN} 1103b 14-21).

\textsuperscript{42} \textit{EN} 1112a 32.

\textsuperscript{43} \textit{EN} 1113a 1-5: Aristotle’s own example concerns why it would be absurd to deliberate about whether a loaf of bread is or ought to be a loaf of bread.

\textsuperscript{44} \textit{EN} 1111b 20-25: Aristotle adds here that we might wish for something impossible, for example, immortality, as well as for things that are possible.
principle of decision is desire and goal-directed reason. That is why decision requires understanding and thought, and also a state of character; for acting well or badly requires both thought and character. As Al Mele interprets Aristotle to understand choice, “[choice] is the proximate efficient cause of action and that it, in turn, is caused by desire and reasoning with a view to an end.” In other words, choice is often both the result of deliberation and the source of motion of actions. We only deliberate about what we believe we can possibly affect since the point of deliberation is to find and initiate a choiceworthy action. Therefore, Aristotelian deliberation is a kind of reasons-responsiveness: like reasons-responsiveness, deliberation begins with the recognition of a telos and, all things considered, results in some kind of movement or reaction aiming toward the achievement of the telos.

The fact that one must believe an action to be possible to bring oneself to deliberate about its end is an important one. It may be intimately related to the idea that reasons-responsiveness and deliberation are about purposeful action. However, this reasons responsiveness is amenable to multiple deliberations with sub-goals culminating in multiple choices and actions over time. Aristotle’s notion of actuality and potentiality in voluntary human action can help explain how it is possible for deliberation to ultimately result in reactions to reasons even if such reactions are not immediate.

---

45 EN 1139a 31-36.
47 I cautiously use the word ‘often’ here because I remain unconvinced that Aristotle held all voluntary actions to be chosen actions. To my knowledge, scholars agree that choice or decision is the product of deliberation (see EN 1112a 14-16). However, some scholars, such as Al Mele, consider deliberation to be present in all actions in some sense. I do not follow Mele in this regard, because Aristotle suggests that children and animals have a share in the voluntary, but not in decision (EN 1111b 9).
According to Aristotle, there are multiple ways that one can have the potential to do or be something, as well as multiple ways that one can actually do or be something. Aristotle’s own example is how one might say that a young boy has the dispositions required such that he could be a general someday. At the moment of this utterance, the boy is not yet a general. And to say that he has the potential to become one is not to suggest that the boy could simply choose right then and there and become one. What is meant by the claim is that, if the boy chooses to do the many acts that could lead to his becoming a general, and if nothing impedes his activities, he will eventually become a general. When he does grow up and finally become a general, this is what Aristotelian scholars call a first actuality. The general is, in fact, a general. However, he is a general even when asleep. If he performs the duties of a general, he is accessing yet another level of actuality; in performing the acts that he may perform because he is a general, he is demonstrating a second actuality. Similarly, we can imagine these levels with abilities to speak a language. I do not speak French, but there is some sense in which I can speak French (should I perform the actions that make one learn French). Again, once I have done that, there is still a different sense in which it can be said that I can speak French even though I am asleep, or speaking English, or otherwise occupied. Finally, there is a

48 While there are many ways to be potential and actual, Aristotle tells us that there may not be specific terms to distinguish such types (c.f. Met. 1048a 37).

49 DA II.5.

50 Following scholars such as Anthony Kenny in his recent volume on Ancient Philosophy, Volume 1 of A New History of Western Philosophy, Oxford University Press (2007).
third way in which I can be said to be able to speak French, and that is when I actually utter words and phrases in French.\textsuperscript{51}

Layers of potentiality and actuality can be helpful to understanding how perceived possibility factors into free action because the type of possibility perceived might influence the course of deliberation. For example, if I am considering an end such as drinking my soda, I have the first potentiality to drink it. I happen to have a soda, it is in front of me, and I am currently physically capable of grabbing it and pouring its contents down my throat. In this case, deliberation will be brief and will result in a choice rather directly. I might think, I want soda, there is my soda, (therefore) I choose to grab the soda (and I do). On the other hand, I might be considering an end such as speaking French. I do not deliberate and directly choose to speak French, thereby simply doing it. On the contrary, upon recognizing that I want to learn French, I would start a series of subsets of deliberation—a network of practical syllogisms. I would deliberate about which actions that are in my immediate control might contribute to my achieving my goal to speak French at some later time. It is in this way that the type of potentiality involved and detected by the agent might affect the course of deliberations and choices. Also, depending on my perception of the type of potentiality I presently have to speak French, my larger goal (speaking French) can generate smaller goals (e.g., buy a French textbook).

While Aristotelian deliberation is not about ends, this tenet of Aristotelian motivation is compatible with deliberation revealing multiple reasons that lead to other

\textsuperscript{51} For a discussion of the difference between possessing and exercising knowledge, see DA II.1.
actions. In other words, it is possible that such deliberation would not take place without the agent first having a larger goal in mind. However, the goal is not the only thing judged to be good. Anything that is conducive to that goal can also be, in virtue of this fact, good. As Aristotle tells us, “We wish for the end more [than for what promotes it], but we decide to do what promotes the end.”\(^{52}\) While the end is what is desirable and, perhaps, what is required for deliberation to take place, it is not as though we can simply decide to achieve the end. For example, if Sam has a goal of being an owner of a large company, Sam does not simply decide to be an owner of a large company. He wishes to be one; but he decides to do tasks in his immediate control that he has discovered through deliberation are conducive to achieving his ultimate goal (e.g., he decides to enroll in a local business school). Therefore, while it is, in a sense, possible that Sam become an owner of a large company, it is not possible at this moment for Sam to do so. It is not in his immediate control. He cannot will himself into this role. There are steps he must take to put himself in that position. So, in such a case, deliberation reveals other reasons for other actions one might take in efforts to achieve a certain goal. In Sam’s case, the goal that he did not deliberate over is the goal of becoming the owner of a large company. This end was not revealed in deliberation; on the contrary, recognition of this end triggered deliberation. While further deliberation reveals other goals (e.g., attending business school, applying for a small business loan, etc.) these other goals are means to a greater goal. Their value depends upon its worth.

\(^{52}\) EN 1111b 29, Bracketed phrase Terrence Irwin’s.
In normal cases in which agents act rationally, the judgment that an end is choiceworthy, even if only indirectly due to a relationship to another end, is enough to make the agent move towards achieving that end. Therefore, not only is Aristotle’s theory one where agents recognize reasons to act, in normal healthy cases, this recognition can smoothly transition to reaction to the reason (i.e., acting for that reason). But the reaction to the reason is the initiation of the physical action. Therefore, in such normal cases, an agent is the efficient cause of a voluntary action as a result of recognizing a reason to act.

The above suggests that, for Aristotle, responding to reasons is causally relevant to sourcehood, since the second stage of reasons-responsiveness, reaction, initiates movement in the sense of an efficient cause, and sourcehood involves efficient causation. In this sense, sourcehood depends on reasons-responsiveness because, without the first aspect of reasons-responsiveness, recognition, there would not be a voluntary efficient causal initiation of an action—a reaction, which grounds the agent’s sourcehood. Therefore, it would seem that Aristotle’s view is neither a sourcehood view nor a reasons-responsive view since both aspects are extremely important to his notion of the voluntary.  

---

53 Indeed, there seems to be a puzzle regarding how reasons-responsiveness is primary to voluntary behavior since Aristotle holds that animals share in the voluntary, yet they do not have a rational aspect of their souls the way that human beings do. For an interesting discussion of this worry, see Christopher Shield’s article, “Aristotle’s Psychology,” Stanford Encyclopedia of Philosophy. Shields considers Aristotle’s comment at DA III.10, 433a17–19 suggestive of ambivalence as to whether animals had something like practical reason. Aristotle does claim that imagination is like a kind of thought and animals have imagination (though they do not have calculation): “These two at all events appear to be sources of movement: appetite and thought (if one may venture to regard imagination as a kind of thinking; for many men follow their imaginations contrary to knowledge, and in all animals other than man there is no thinking or calculation but only imagination). Both of these then are capable of originating local movement…It follows that there is a justification for regarding these two as the sources of movement, i.e., appetite and practical thought; for the object of appetite starts a movement and as a result of that thought gives rise to movement, the object of appetite being to it a source of stimulation.” (DA 433a 9-20).
The Second Criterion for Voluntary Agency

We have just seen how, in order for Aristotelian agents to meet the first criterion properly, sourcehood, such agents also behave in ways that contemporary philosophers call reasons-responsive. We are now in a position to consider Aristotle’s second criterion for voluntary behavior—knowledge of the particulars. I contend that recognizing the significance of the second criterion will further support my claim that all that is voluntary is intimately related to proper reasons-responsiveness.

When Aristotle discusses how voluntary behavior is inconsistent with external force, he does not conclude his investigation there. Instead, he adds to it a discussion of a certain type of knowledge that is necessary for actions to qualify as voluntary. This kind of knowledge, knowledge of the particulars, suggests that agents must have a certain kind of epistemic access to the facts in order to truly behave voluntarily.

In the first chapter of the third book of Nicomachean Ethics, Aristotle suggests that one’s ignorance of particulars precludes voluntary behavior: “Everything caused by ignorance is nonvoluntary.”54 For an action to be caused by ignorance means “the cause is ignorance of the particulars which the action consists in and is concerned with.”55 In other words, the agent might be ignorant of certain facts relevant to her actual and immediate circumstances (not general laws or principles). Aristotle tells us that a man

54 EN 1110b 19.
55 EN 1111a 2.
might be ignorant of who is doing the act. A contemporary example might be John Perry’s description of how he followed a trail of sugar on the floor in a supermarket with the intention of stopping the shopper from letting his/her cart spill sticky crystals all over the establishment, only to realize that he had walked in a circle because it was he who was walking around with a leaky sack!\textsuperscript{56} An agent might also be ignorant of what she is doing. For example, a person might give another a refreshing drink without realizing that the liquid is poison. She is not only giving her companion a drink, but she is also killing him. She may not be aware that her action is also a killing. A third type of ignorance of the particulars is “about what or to what he is doing it.”\textsuperscript{57} For instance, Oedipus knew he was killing a man, but did not know that the man he was killing was his father. In a sense, it can be said that he voluntarily committed homicide, but his ignorance of the particulars might mean that he nonvoluntarily committed patricide.\textsuperscript{58} To these kinds of examples, Aristotle adds “sometimes also what he is doing it with—what instrument, for example; for what result, for example, safety; in what way, for example, gently or hard,” etc.\textsuperscript{59}

\textsuperscript{56} Perry (1979): 3.
\textsuperscript{57} EN 1111a 5.
\textsuperscript{58} Though Aristotle does not discuss Oedipus here in this context, scholars of ancient philosophy use this example to illustrate how something that is, in a sense, voluntary is under one important interpretation nonvoluntary. For instance, see Irwin (1990): 599. Irwin considers Aristotle’s notion of “negative causes” (\textit{Phys.} 195a 11-14) to be explanatory here. Essentially, Oedipus is responsible for the event that is identical to patricide, but since patricide was not the goal of Oedipus’s action, the source of what makes it patricide is, says Irwin, \textit{external} to Oedipus.
\textsuperscript{59} EN 1111a 6.
By negation, we can infer that, if the absence of such knowledge or awareness renders behavior non-voluntary, then its presence must be required for voluntary behavior. Therefore, we may add to the first criterion the following condition:

**Criterion 2: knowledge**; voluntary behavior must involve the agent having a certain kind of epistemic access to the morally relevant facts (particulars) of her action.

Just what counts as an agent having the right kind of epistemic access to morally relevant facts of her situation will occupy the majority of our attention for the remainder of this chapter. For now, let us be satisfied with the claim that the agent must be aware of these facts at the time of action for her behavior to be voluntary. This claim shall be revised as we examine certain problems more carefully; for example, it will soon be clear that ignorance of particulars caused by the agent herself might be an exception to this claim.

The second criterion for voluntary behavior also supports my claim that an agent’s ability to respond to reasons is extremely important to Aristotle. Knowledge of the particulars in no way affects whether one is the efficient cause of one’s behavior. On the contrary, it seems as though the majority if not all of the examples of nonvoluntary behavior issuing from ignorance of the particulars involves activity in which the agent was the source of her own movement. For example, imagine that I am ignorant of the fact that someone on the other side of a door has knelt down on the floor to inspect the doorknob. When I open the door I inadvertently hit her in the face, breaking her nose. Now, Aristotle would say that such behavior, the breaking of her nose, is nonvoluntary because it was done due to ignorance of the particulars. However, my action can also be
re-described as *the opening of the door*. In this sense, I was the source of the action.
Nothing external to me caused me to open the door. I even thought that I was voluntarily
opening the door; because I saw no reason not to open the door, and I had some reason to
open the door that I recognized and to which I reacted. So, there is some sense in which
my behavior is under my control. I controlled my bodily movements to do precisely what
they did do.

Be that as it may, no one would blame me for breaking the woman’s nose, and
Aristotle seems to concur with this judgment; for he considers my knowledge of the
woman’s location to be requisite for my desert of blame. I do not think that Aristotle’s
acknowledgment of this requirement is merely an *ad hoc* attempt to preserve our
intuitions—I think it speaks to his underlying concern that appropriate reasons-
responsiveness is required for praise- or blameworthy behavior. My ignorance of her
location and vulnerability precluded me from recognizing these facts as reasons to open
the door and responding to such reasons. If I *had* known of her position and vulnerability
and if I *did* open the door with these reasons in mind (presumably because she is an
enemy and I intend to harm her), I would be quite blameworthy. In other words, I was
epistemically prevented from responding to reasons to do what I, in fact, did. This is to
say, I did not break her nose for any reason. Also, my ignorance of the woman’s location
and vulnerability epistemically prevented me from recognizing the reasons *not* to open
the door. Since these facts were not epistemically available to me, my action *qua* nose-
breaking is not voluntary. While I was the source of the act that was identical to a nose-
breaking, I was not responding to any reason to break the women’s nose. I was the source
of my behavior in that I was the efficient cause of my movement. I was also the efficient cause in an intentional sense; in moving to open the door, I was reacting to some reason to open the door. For example, I did not open the door as a result of some involuntary spasm. Therefore, there is one clear sense in which I was the source of my behavior.

Aristotle certainly would not be alone in absolving me of blame in the above scenario. Contemporary sourcehood theorists would likely concur. What makes this aspect of Aristotle’s theory intriguing for our purposes is that such behavior for Aristotle is not even voluntary. Sourcehood theorists might consider the behavior voluntary, but consider the ignorance of the particulars to be what exonerates the agent from blame. Perhaps a sourcehood theorist might say that since I was not the source of my ignorance, I am not at fault for what led me to break the woman’s nose. But, for Aristotle, such ignorance destroys the behavior’s status as voluntary altogether. I surmise that, for Aristotle, a lack of knowledge of the particulars entails an impairment of reasons-responsiveness. Even if one’s behavior is responsive to some reasons, Aristotle is sensitive to the fact that the agent must be able to know of the morally salient reasons, which includes correct judgment of the facts. Still, Aristotle is not alone in this matter. Many sourcehood theorists are likely to agree with him. However, since Aristotle seems to hold the knowledge requirement to be as important as the source requirement, he is not a sourcehood theorist in our sense of the term.
Implications Concerning How to View Incontinence and Vice

So far I have argued that, for Aristotle, reasons-responsiveness is as important as sourcehood to qualifying behavior as voluntary and worthy of praise and blame. However, there are two types of action that, on the surface, appear to contradict this claim: akratic or incontinent behavior and vicious behavior. I say that these types of actions seem difficult to reconcile with the proposed view for the following reasons. In neither of these types of actions is the agent capable, in some sense, of fully responding to reasons. While the akratic may recognize good reasons to behave well, she does not react to them. The vicious person, on the other hand, is so entrenched in her wickedness that she no longer can even recognize, let alone react to, reasons to behave well. However, while neither type of agent successfully responds to reasons, in both types of actions the agent is the source and efficient cause of her own movement. And, most importantly, Aristotle considers these types of actions to be the standard types of actions that are to be blamed. Therefore, we must examine and explain how these types of behaviors are blameworthy when, on the surface, it appears that they only qualify as behaviors for which an agent is a source.

In the following sections, I shall argue for several points. The first is that failure to respond to reasons might not result from an inability to respond to reasons. This is to say that, while the akratic person systematically fails to follow through with reasons-responsiveness, this poor track record in no way entails that the akratic cannot do what she deems best. This brings us to our second point. An ability to be the efficient cause of
an omitted course of action would seem to be related to sourcehood. Once again, sourcehood remains a necessary condition even if it is not the only or most critical condition for praise- and blameworthiness. Finally, I shall point out that even Aristotle made ambiguous moves regarding the relationship and the primacy of sourcehood and reasons-responsiveness. While I hold that sourcehood seems to be, on such an account, embedded in reasons-responsiveness, it is still likely that Aristotle relies on the nature of sourcehood in the account to handle potential objections.

**Akrasia: When Recognition and Deliberation are Not Enough**

While Aristotle’s moral psychology explains how, ideally, voluntary praise- or blameworthy acts result from the recognition of and reaction to reasons, it is not the case that, empirically, all instances of human choice and action work this way. In fact, many do not. The main example of how this process can go awry is that of the *akратic*—one who knows the better, but does the worse. For example, Emma might consider how much she values her health. She might consider this value and recognize that it gives her reasons to exercise at the gym. While she reflects and concludes that she should go to the gym, she instead seats herself in front of the television and proceeds to watch a movie. In such instances, the agent is *weak willed* since she does not bring herself to do what she has recognized that she should do.

In examples such as these, something goes terribly wrong somewhere either in the moment that one considers an action, chooses an action, or executes that choice.
Aristotle’s precise account as to how reasoning goes awry is famously difficult to discern. Despite the fact that almost an entire book of *Nicomachean Ethics* is devoted to the subject, Aristotle makes apparently contradictory claims. For example, Price points out that Aristotle sometimes suggests that akrasia is caused by the agent’s sentiment (*pathos*): “The acratic person, knowing that what he does is bad, does it because of sentiment.”  

For example, Aristotle writes, “the incontinent fails to abide by reason because of too much [enjoyment].” However, at other times Aristotle suggests that the akratic acts from some specific type of ignorance or confusion. For example, the agent might be confused about whether a premise in a practical syllogism is universal or particular, or whether it is a universal statement concerning human beings, or concerning the agent *herself* as a human being, or concerning some type of object. I am unsure of what to make of Aristotle’s disparate comments regarding akrasia. I am not opposed to the idea that Aristotle had many classes or types of akratic action. However, for our purposes, we need not decide on an interpretation to see the relevance to Aristotle’s view of akratic action and how his treatment of incontinence suggests that Aristotle had an agenda centered on reasons- responsiveness.

Regardless of which account was really Aristotle’s view, all accounts have something in common—Aristotle blames the akratic for her behavior. It is easy to see

---

60 Price, 236.

61 *EN* 1151b 25-6.

62 See *EN* VII.2 esp. 1145b 23-9.

63 For a discussion of the importance of the two types of premises, see *EN* VII.3 1146b 35-1147a 4; for a discussion of how to apply a universal premise to an agent, to humanity, or to the object of the considered action, see *EN* VII. 3 esp. 1147a4-10.
why Aristotle would blame the akratic if she fully understood why her behavior was inferior to alternatives she could have done. She had the knowledge of what it was best to do, and did not apply that knowledge. In such a case, we might say that she recognized a sufficient reason to \( \varphi \) and chose to \( \sim \varphi \). While she recognizes such a reason, recognition is only part of the reasons-responsive behavior. She does not react to that recognition.

It seems that Aristotle is justified in finding the akratic to blame given the interpretation of akrasia as caused by some confusion or ignorance since Aristotle notes that there are many ways that one can be ignorant, and not all ways are excusable. In fact, in *Nicomachean Ethics* III.1, Aristotle explicitly describes the type of ignorance that can constitute an excuse:

> Action caused by ignorance would seem to be different from action done in ignorance. For if the agent is drunk or angry, his action seems to be caused by drunkenness or anger, not by ignorance, though it is done in ignorance, not in knowledge. Certainly every vicious person is ignorant of the actions he must do or avoid [and viciousness is no excuse]...[When ignorance constitutes an excuse] the cause is ignorance of the particulars which the action consists in and is concerned with...For an agent acts involuntarily if he is ignorant of one of these particulars...[such as] who is

---

64 It is worth noting that Aristotle does not seem to be claiming that the akratic *cannot* bring herself to do what she judges best. Instead, his account seems more descriptive; the akratic *does not* do what she judges to be best. Also, it is not entirely clear that all akratics act akratically on all occasions, or even on each occasion of a certain type. Indeed, Aristotle notes at *EN* 1146b 4-5, “For no one has all the types of incontinence, but we say that some people are simply incontinent.” I have taken this claim to be suggestive that our language is hyperbolic at worst and dangerously convenient at best. In other words, either we are exaggerating when we say one’s weakness of will is incorrigible, or we mean rather that the person very often behaves akratically. In any case, the power or ability to act as one judges fit does not seem to be absent in the case of the akratic.
doing it; what he is doing; about what or to what he is doing it; …what he
is doing it with…for what result…\(^{65}\)

What it is to be ignorant of the particulars is a bit easier to discern than to act in
ignorance; however, I think that we can infer the latter from an explication of the former.
To be ignorant of a particular is to be unaware of some relevant detail or feature
concerning the actual particular situation in which the agent acts. For example, Aristotle
tells us that one might injure another with a spear accidentally if she believes it to have a
button on the tip when it does not, or kill someone by giving him a drink in efforts to save
his life (for instance, if he is allergic to the plant that made the juice).\(^{66}\) In such cases, the
harm done is not intended and it is reasonable that the agent not have known about the
details of which she was ignorant.

Acting in ignorance, on the other hand, seems not to be about particulars. When
one is drunk, or full of rage, or vicious, that person seems to be in what I shall call a
global state of ignorance. They are either ignorant of some universal (e.g., that torturing
is bad), or they are in a frame of mind that inhibits full comprehension even if facts are
revealed (e.g., a drunkard might not appreciate the seriousness of having lost his fortune
in a game of cards until the next day, when sober). In any case, acting in ignorance is no
excuse and does not render an action nonvoluntary.\(^{67}\)

---

\(^{65}\) EN 1110b 25-1111a 6.

\(^{66}\) EN 1111a 13-4.

\(^{67}\) Incidentally, it seems to me that Aristotle would most likely hold that akrasia happens as a result of strong
appetitive desires overpowering the rational desires that accompany, at least, partial knowledge of what one
is doing. I not only say this because Aristotle declares that with “the continent and incontinent person we
praise their reason, that is to say, the [part] of the soul that has reason, because it exhorts them correctly and
toward what is best,” which suggests that the incontinent person is not in fact ignorant. I also think that the
The above distinction suggests that, in *Nicomachean Ethics*, Aristotle was careful to distinguish the kind of ignorance that constitutes an *excuse* from the type of ignorance that very well might constitute the *offense*. In the case of akrasia, whatever ignorance is involved, it is an inexcusable variety. This suggests that Aristotle might have been aware of the potential for such confusion and made sure to explain how akratics can be blameworthy for bad behavior.

Indeed, some have speculated that differences between the arguments of *Eudemian Ethics* (*EE*) and those in the (presumably) later *Nicomachean Ethics* were intentionally made with such a worry in mind. In her article, “Aristotle on the Voluntary”, Susan Sauvé Meyer argues that one of the reasons why Aristotle might have revised the account found in the *Eudemian Ethics* is that it makes little sense of akratic action as voluntary. In the *Eudemian Ethics*, Aristotle describes the voluntary as that which accords with impulse and contrasts it with the involuntary, which is that which is contrary to impulse since “what is done according to appetite is voluntary.”

Presumably, when an akratic person fails to do what she deems best to do, she acts on some other inferior motive. But, if Aristotle’s theory is one infused with motivational internalism, the akratic also must have had some desire (albeit a weak one) to do what the agent deemed best. So, in every case of akratic action, the agent would do something that above interpretation is the only way to preserve a real distinction between the incontinent person and the vicious person. For the vicious person does not recognize the good ever. If the incontinent or akratic behaves poorly because of ignorance, and it is the kind of ignorance for which she must be blamed, then it would seem to be that she acts *in ignorance*; and this is the defect characteristic of the vicious person. Admittedly, it is possible that there is no real difference between the akratic and the vicious except for degree. Still, I believe Aristotle’s practice of distinguishing the two as different types suggests that there was something truly different about how the akratic acts weak-willed where the vicious person wholeheartedly does wrong.

---

68 *EE* 1223b 1-2.
accords with one impulse and, at the same time, contrasts with another impulse. Both impulses would properly belong to the agent. Presumably, this would also be true for the continent agent, who recognizes something as good, brings herself to actively pursue it, but experiences pain or regret in doing what she thought to be best. She too would seem to have opposing impulses. As Aristotle also holds that nothing can be both voluntary and involuntary, this result would not do: “for to act incontinently is to act through appetite contrary to what the man thinks best; whence it results that the same man acts at the same time both voluntarily and involuntarily; but this is impossible.”69 This is one reason that, according to Meyer, Aristotle replaced his notion of contrary to impulse with the notion later discussed in Nicomachean Ethics—that of external force.

Let us imagine such a case of incontinent behavior in which an agent acts contrary to impulse but as a result of some other internal impulse. Such examples can also help to illustrate why reasons-responsiveness is as important as sourcehood for Aristotle. For instance, imagine that Clyde the coward is in battle, recognizes a reason to stand firm, is overcome by fear, and chooses to flee. Given an assumption of motivational internalism, Clyde’s recognition of a reason to stand firm would include an impulse to stand firm. Alas, another impulse motivates him to flee. Given this assumption, we can presume that akratic behavior is voluntary and, therefore, a fair candidate for praise or blame.70 In this scenario, we can see how Clyde is blameworthy for fleeing. He has reason to stand firm, yet he does not. He is blameworthy for a failure to fully respond to this reason. He is the

---

69 EE 1223b 9-10.
70 Indeed, we may do more than assume that this is Aristotle’s view since he so indicates in several places, most notably in EE BK II Ch. 8 and EN BK VII.
source (and, therefore, efficient cause) of this failure, and the fact that he almost responded to reason to stand firm is insufficient to exonerate him of this failing. Additionally, he did respond to a reason to flee, which is not the correct reason to follow. Therefore, in this instance, Clyde is both reasons-responsive and the source of his behavior.

On the other hand, we can imagine a slightly different case, in which Clyde the coward does stand firm, yet he is not praiseworthy as a result of a failure to respond to reasons in the right way. Imagine that in the second scenario, Clyde recognizes a reason to stand firm, is simultaneously motivated by fear to flee, yet his fear is so paralyzing that he cannot even move to flee! In a sense, he stood firm in the face of battle. However, his reason for doing so was his paralyzing fear, and not the call to demonstrate courage. Aristotle surely would not praise Clyde for this action. While the source or efficient cause of Clyde’s stance was within him, it was in response to his fear, not his recognition of his martial duty, which moved him to stand firm.

What the example of the two Clydes can show us is precisely how a praiseworthy act embodies both a final and an efficient cause. The final cause is the telos recognized by the agent as good. The efficient cause is the movement of the body (initiated from within) as a response to the recognition of the telos. Only when the final cause is properly identified and the efficient cause follows naturally and non-deviantly can an act be a

---

71 This example is similar to those offered by Donald Davidson and others who undertake the problem of wayward or deviant causal chains. However, in this case, it is clear that the right action was achieved not by being caused by a good reason but rather by its opposing desire. See page 79 of Davidson (1973) in Davidson (2001). There Davidson provides a similar example of a rock climber doing something he recognizes he has reason to do as a result of a wayward causal chain.
candidate for praise. However, this model is a bit complicated, and when things are complicated, there is much opportunity for minor flaws to derail an entire operation. It is for this reason that Aristotle warns us that there is but one way to go right, but many ways to go wrong:

Moreover, there are many ways to be in error—for badness is proper to the indeterminate, as the Pythagoreans pictured it, and good to the determinate. But there is only one way to be correct. That is why error is easy and correctness is difficult, since it is easy to miss the target and difficult to hit it… ‘for we are noble in only one way, but bad in all sorts of ways’.

Similarly, there is but one way to act such that the action is praiseworthy, yet many different ways, to act such that the action is blameworthy. For example, one might act incontinently by knowing the better and failing to bring oneself to do it. One could have known the better and lost sight of it, thus doing the worse. In some places, Aristotle seems to suggest that one could become confused about what is better and act incontinently due to miscalculation. It seems that, in all such cases, if the act is voluntary, the agent is blameworthy. To me, this suggests that any failure of reasons-responsiveness deserves blame for Aristotle; and that all aspects of the reasons-responsive process are equally important. In short, recognizing the good (i.e., having it as your *telos*) is not enough to deserve praise or even pardon for the failure that follows. Also, a failure to react to such recognition is not an excuse. Indeed, there seems to be an underlying

---

72 EN 1106b 30-35.
normative claim in theories such as Aristotle’s that one should (qua rational being) recognize and react to the reason to do what is best (or judged best, as the case may be). Therefore, failure to react to the recognition of a sufficient reason to act cannot be an excuse—it is, in fact, the offense.

If my interpretation is correct, Aristotle considers the akratic’s failure to react to the best reasons blameworthy. However, there is another aspect of the akratic’s behavior that could ground blame—the fact that the akratic indeed responds to inferior (i.e., bad) reasons. When akrats behave incontinently, Aristotle does not suggest that they recognize a good reason to do X and, for no reason whatsoever, do Y instead. On the contrary, Aristotle claims that all actions are done for some reason. Therefore, even akratic actions are done for reasons. When the akratic does Y instead, she does Y for reasons. However, she also knows that these reasons are not as good as the reasons to do X. So, in the case of akratic, there are two aspects of her poor behavior that jointly make her behavior blameworthy: she fails to respond to the good reasons that she recognizes and she responds to lesser reasons that she recognizes to be such.

This second flaw in the akratic’s behavior is shared in common with the vicious person. However, the vicious person does not share with the akratic the first flaw. The vicious person cannot recognize the good. Her viciousness precludes such moral perception. The vicious person also behaves for reasons. She responds to the apparent good. Still, Aristotle finds the vicious person the most deplorable.
Viciousness: When Recognition is Not Possible

We have just seen how the akratic agent’s failure to fully respond to reasons is blameworthy. While it is true that the akratic is blameworthy for a failure of follow-through, it is not entirely clear that such instances of blameworthiness depend on the initial successful recognition of reasons to act other than the akratic does act. In fact, the vicious agent is incapable of recognizing moral reasons, and this agent is deemed the most blameworthy of all character types according to Aristotle. Some scholars have been puzzled by this fact. Some might think that the akratic is a worse person because she, at least, knows what she ought to do and still does not do it. As the saying goes, she knows better. The vicious person does not. Therefore, if the vicious person cannot recognize good moral reasons, how can we be justified in blaming the vicious person when she fails to react to what she cannot detect? In this section, I shall attempt to show that all such concerns rest on a mistaken interpretation of Aristotle’s commitment to reasons-responsiveness. Indeed, a clearer picture of how reasons-responsiveness and sourcehood are related on Aristotle’s moral psychology will illustrate why a present inability to respond to reasons is not an excuse to behave poorly. While reasons-responsiveness might be at issue in the case of the vicious person, it was not always at issue.

73 For example, see Bondeson (1974) and Brickhouse (1976) for a discussion of the fairness of blaming vicious people if they cannot willfully change. See Fairbrother (1897) for a discussion of the difference between doing wrong and being a wrong-doer—a defense of Aristotle’s assessment. See Duff (2002) for an argument for why our system of legal liability should not take into account Aristotle’s distinction between incontinence and vice.
These objections rest on an idea that a vicious agent is somehow unable to behave precisely how she should. This worry has been captured in contemporary philosophy by the principle of possible prevention, a principle named and described by Peter van Inwagen.\textsuperscript{74} According to this principle, one can be blamed only if one could have prevented oneself from doing the bad act (i.e., either by doing nothing at all or by doing the act that ought to be done). Scholars who worry about the legitimacy of blaming the vicious on an Aristotelian program do have some cause for this worry since, ostensibly, it would seem that Aristotle holds this view. If he does, his characterization of the blameworthy vicious person could become problematic. While I do not believe that Aristotle has this problem, I do think that seeing how he avoids the apparent tension can illuminate something about character development and reasons-responsiveness that is important for our discussion.

First, let us consider why one might claim that Aristotle’s view is congenial to the spirit of the principle of possible prevention. First, Aristotle makes it clear that we are responsible for what we do because what we do is up to us (\textit{eph’heemin}).\textsuperscript{75} He also declares that if it is up to us to $\varphi$, it is also up to us to $\sim \varphi$: “For when acting is up to us, so is not acting, and when no is up to us, so is yes. And so if acting, when it is fine, is up to us, not acting, when it is shameful, is also up to us; and if not acting, when it is fine, is up to us, then acting, when it is shameful, is also up to us.”\textsuperscript{76} These two claims (that what we do is up to us \textit{and} that what is up to us to do is also up to us to not do) in conjunction

\textsuperscript{74} Van Inwagen (1986).
\textsuperscript{75} EN 1113b 6-13.
\textsuperscript{76} EN 1113b 6-11.
suggest the following: our responsibility requires that we are not only in control of our behavior, but that we could control ourselves in such a way to act differently from the way that we do. This result might seem to some people to be in tension with Aristotle’s characterization of the vicious agent.

For Aristotle, the vicious agent does not recognize the true good. Like everyone else, she pursues the apparent good. However, her character colors her perception of what is good, as is the case with all character types, since “his character controls how the end appears to him.” That is, whenever the vicious agent considers something to be good, it is quite likely to be, in fact, not good. Her viciousness leads her to take pleasure in things that are bad. By taking pleasure in a bad end, she is likely to view that end as good and pursue it. However, the more that she pursues such an end, the more pleasure she will get from achieving it and the better it will look! It seems to be the case that the vicious person cannot do other than she does in a controlled way (i.e., on purpose) since her psychology is wired in such a way that she will never recognize a reason to pursue the actual good.

Several philosophers have addressed this worry in Aristotle’s corpus and I am unsatisfied with their solution. Most scholars attempt to resolve the conflict by appealing to a notion of conditional freedom. Conditional freedom is the ability one has to φ based on the following condition:

\[ EN \text{ 1114a 30-1114b 3.} \]

\[ 78 \text{ For example, Everson (1990) considers a nuanced version of conditional freedom. He begins his account taking from A.J. Ayer’s seminal paper “Freedom and Necessity” (1963), citing Ayer’s conditions of freedom as an ability to act otherwise such that “I should have acted otherwise if I had so chosen” (Ayer, 282). Everson adds to this naïve conception a notion of acting \textit{qua} individual and acting \textit{qua} member of the} \]
**Conditional freedom:** If S had chosen to φ, S would have φ-ed.

In this sense of freedom, even the vicious person is free. For instance, a vicious person who is unaware of a good reason to do other than she does is still conditionally free if, *had she chosen* to do otherwise, she would have done so. This is the case even if there is no possible world in which she does choose to do otherwise. I do not think that this is the only sense of our behavior being “up to us” that Aristotle had in mind. In fact, I think that a more complete picture of character development can illuminate a richer notion that is directly linked to reasons-responsiveness.

First, let us consider why this cannot be the only notion Aristotle had in mind. Aristotle’s notion of our behavior being “up to us” must be richer than contemporary notions of conditional freedom for the following reason. Aristotle does hold that ignorance of the particulars makes an agent’s action *nonvoluntary*. As it is nonvoluntary, the behavior is *not* up to the agent. However, in the conditional sense of something’s being up to us, such a behavior *is* up to the agent. For example, imagine that Brad is trying to do a good deed by helping two strangers load heavy boxes into their van. Unbeknownst to Brad, the heavy boxes contain stolen electronics, and his assistance human race. While Everson claims that he holds that “The difference between a voluntary and a compelled action does not rest upon the fact that in the case of the first, but not the second, the agent could have acted other than he did” (97), he proceeds to describe an account that distinguishes an agent acting on her own beliefs and desires and not on her brute human nature. While this sounds like a step towards an interpretation like mine concerning conditional reasons-responsiveness (see upcoming discussion), it is distinct from my account. According to Everson, an agent acts from her beliefs and desires if it is not the case that her action is “not done because of pressure which overstretches human nature and which no one could endure” (*EN* 1110a 23-6/p. 88, his emphasis). To my mind, there is an underlying reason to be preoccupied with this passage: even if one were to deliberate and choose to do otherwise, one would not actually do otherwise; human nature would take over and necessitate the behavior. Everson’s account is still a very nice view that tries to reconcile conditional freedom with a more sophisticated Aristotelian view. However, he does not seem to consider the fact that, for Aristotle, it matters that the agent’s could have had something as a reason. Instead, he is more concerned with what belongs to the agent as an individual and what belongs to the agent as a human being.
facilitates the criminals’ safe getaway. On Aristotle’s account, Brad involuntarily facilitated a crime. However, according to conditional freedom, it was up to Brad to facilitate the crime because, had he chosen not to help the men, he would not have done so. For instance, if Brad had decided to keep walking to spare his back, he would not have facilitated the crime. Still, this sense of Brad’s behavior being up to him cannot be what Aristotle has in mind; for Aristotle would not blame Brad and he would not blame him because he would not consider his behavior to be up to him in the relevant sense. In short, it was up to Brad whether or not he performed the act, but that the act was up to the agent is not a sufficient condition for voluntariness.

So, what is the relevant sense? I suggest that the relevant sense involves a kind of conditional reasons-responsiveness that grounds the agent as the source of her present abilities (or lack of ability) to respond to certain reasons. That is to say, if it was ever reasonably in Brad’s control (i.e., “up to him”) to recognize the good reasons to do the right thing, then a failure to recognize such a reason is blameworthy. However, if no one (even a virtuous person) would have seen the truth, then Brad’s behavior is not up to him, in the relevant sense. Therefore, let us define conditional reasons-responsiveness as such:

**Conditional Reasons-responsiveness**: an agent is conditionally responsive to reasons when \( \phi \)-ing at time \( t_1 \) if she \( \phi \)’s as a response to some reason and she had the ability at some prior point in her past \( (t_1-n) \) to be the efficient cause of her own recognition of some reason(s) at \( t_1 \) to \( \sim \phi \).

I propose this model as a way to use contemporary concepts to understand and accurately interpret Aristotle’s thinking. I think that Aristotle’s view of the relationship between
final and efficient causality in human action (what I have identified as also a relationship between reasons-responsiveness and sourcehood) allows Aristotle to make claims that contemporary theorists might consider forced hybrids of competing views. However, on Aristotle’s view, these claims seem perfectly consistent and natural. In order to describe my proposed account of conditional reasons-responsiveness, let us first review the apparent problem.

**The Problem—vicious blindness to good reasons**

The account that I want to propose should make sense of how the following Aristotelian claims are consistent:

1) We are worthy of praise and blame because what is to be praised or blamed is *up to us*.

2) If something is *up to us*, it is voluntary; therefore, we are the source of the behavior as the efficient cause as well as agents that respond to final causes.

3) When something is *up to us* to do, it is also *up to us* to not do it (omit it).

4) The vicious deserve blame for their wicked behavior.

5) The vicious act wickedly in response to wicked reasons.
6) The vicious are not receptive to alternative reasons to act well.

The apparent problem seems to derive from the following observations. According to 1, 2 and 3, if we deserve praise or blame for our behaviors, we were able to be efficient causes that respond to final causes to do otherwise. However, according to 5 and 6, the vicious agent is the efficient cause that responds to final causes of the behavior she does, in fact, perform, but could not have recognized (let alone, react to) alternative reasons. This seems to contradict 4, which holds that the vicious deserves blame for their wicked behavior, which, in conjunction with 1-3, would suggest that the vicious could have responded to reasons to do otherwise.79

Character development

According to Aristotle, there are four main human character types that we may develop over time: virtuous, continent, incontinent, and vicious.80 The virtuous person knows the good, does what is good, and feels appropriately about it. The continent person recognizes the good, does what is good, but does not feel appropriately about it. The

79 This issue is also applicable to the virtuous agent, who, it seems, cannot psychologically fail to do what is good or right voluntarily. I elect to focus on the vicious person here since, as an agent who is blamed, unfair treatment seems most salient. However, our arguments and explanation for the vicious agent’s moral standing and blameworthiness can be analogously applied to the virtuous agent’s inability to act poorly (yet, she remains praiseworthy).

80 I specifically call these “human character types” because Aristotle mentions other states such as “brutishness” that human beings can fall into, but are more like the states of beasts. Such states do not appear to involve much application of the rational faculty, if any. Similarly, the superhuman or “divine” character does not apply to human beings.
incontinent person recognizes the good, fails to do it, and has inappropriate feelings towards the good. The vicious person does not recognize the good, as a result, fails to do the good, and feels inappropriately towards the good.\textsuperscript{81} It is worth noting that the only character type that feels appropriately is the virtuous, and the only character type fully ignorant of the good is the vicious.

While these are the main character types one might acquire, none of these is bestowed upon a person at birth:

None of the moral excellences arises in us by nature; for nothing that exists by nature can form a habit contrary to its nature. For instance the stone which by nature moves downwards cannot be habituated to move upwards, not even if one tries to train it by throwing it up ten thousand times.\textsuperscript{82}

Here Aristotle seems to be using the apparent obvious fact that characters do change as a counterargument for the idea that excellence or virtue is natural. It is true that we are naturally capable of becoming excellent, as Aristotle writes, “we are adapted by nature to receive them [the virtues], and are made perfect by habit.”\textsuperscript{83} At least one type of character, the virtuous type, is one that is not inborn, yet is, in a sense, available to everyone.

Indeed, it would seem that all four types of character are acquired. For Aristotle also writes, “It is from the same causes and by the same means that every excellence is

\textsuperscript{81} See \textit{EN} VII (esp. Chapters 1, 3, 5, 7 and 8).

\textsuperscript{82} \textit{EN} 1103a 19-23.

\textsuperscript{83} \textit{EN} 1103a 24-28.
both produced and destroyed.” If virtue is acquired by habitually doing the types of acts that virtuous men do, then the destruction of virtue would entail habitually doing the types of acts that non-virtuous men do. Therefore, while we are naturally able to have a character, the type of character that we acquire is dependent on the types of actions habituated.

For Aristotle, characters are acquired by doing certain types of acts repeatedly until the habit of doing such acts is formed. In addition, it seems to be a common phenomenon of human nature that we gradually become more inclined to do actions that are similar in nature. In other words, while we might experiment with different types of behavior often in youth, we tend to homogenize our behavior as we age. Aristotle has a very simple explanation for this phenomenon. Oftentimes, we do certain types of acts repeatedly because we experience pleasure in doing them; others we avoid because of previous experiences of painful consequences. In time, we become so accustomed to certain types of acts that it is like second nature for us to do them. As a disposition to act in a particular way develops, the individual acquiring the character begins to take pleasure in doing the acts that accord with this habit: “for no one would call a person just, for instance, if he did not enjoy doing just actions, or generous if he did not enjoy generous actions.” In other words, we are inclined to behave in the ways that we

---

84 EN 1103b 6.
85 Aristotle tells us in the NE that the just man becomes just by doing just acts (EN 1103a 33 (learn by doing), EN 1105b 9-10 (the just man)).
86 EN 1099a 19-20.
associate with pleasure, and the more that we exhibit such a behavior, the more we enjoy that activity.

From this observation we can see how, once acquired, a character state perpetuates itself. That is, the agent’s moral perception is informed by the character that she has; vicious people view bad ends as good ones, while virtuous people have keen moral perception.\(^{87}\) Since, ultimately, every human being acts for some good,\(^{88}\) what is viewed as the good by a given individual is what she pursues as well as that in which she delights when she achieves it. Therefore, a character state moves an agent towards doing the kinds of acts that led to the formation of that character.\(^{89}\) Since character is formed and solidified by repetition, repeating like acts only strengthens a character. That is why when someone acquires either a vicious or a virtuous character, her character is a relatively fixed state that determines much of her actions and feelings.\(^{90}\)

\(^{87}\) EN 1113a 31-34, EN 1114b 2, EN 1114b 21-24.

\(^{88}\) EN 1110b 11.

\(^{89}\) This is, in part, due to the fact that, for Aristotle, reason alone does not act (EN 1112b 13). Therefore, an incontinent person might know what is better, she might not be able to do what is better based on this knowledge. Her character is of the type that it is attracted to the worse, and for this reason it selects worse ends for action. Also, see Tuozzo (1991) for a convincing discussion of why, for Aristotle, the types of ends of acts are not chosen (via deliberation).

\(^{90}\) There is a wealth of literature on this consequence of Aristotelian character concerning the possibility of character change. T. Brickhouse (1976) suggests that one cannot alter one’s character because once sufficiently fixed, it would be impossible to act contrary to that character (which is what is required to develop a contrary state). Bondeson (1974) suggests that Aristotle did believe in character change, but that it was difficult given the proclivities of the agent and the types of ends towards which she acts. Ott (2000) discusses the argument Aristotle develops for our being responsible for our character, arguing that Aristotle jumps to his conclusion. While we might be responsible for individual acts, it does not follow, says Ott, that we are responsible for our character. Di Muzio (2000) argues that character can be altered if a vicious person were to happen to imitate a virtuous person’s actions for some other goal. In another paper (Anton 2006), I argue that fixed characters can be altered, but only if the change is initiated from without (i.e., external). If a person is influenced by others to act contrary to character (that is, others select her ends for her and manipulate her into pursuing them), the fixed habit can be weakened.
So it is easy to see how, on Aristotle’s account of character acquisition, one could become vicious in such a way that the person does not recognize good ends. Our observation accords with Aristotle’s own claims concerning the vicious person’s ability to recognize good reasons, since Aristotle does say that the vicious person is not only blind to good reasons to act,\textsuperscript{91} but that he is also one who cannot be reasoned with by means of argument alone.\textsuperscript{92} However, it is still unclear as to why the vicious person is to blame for her behavior when she no longer is capable of voluntarily acting differently.

Aristotle does have an explanation for how human beings are responsible for their characters even if a character perpetuates itself and hinders the agent’s ability to act differently. At one point, Aristotle makes the following clarification:

Yet it does not follow that if he wishes he will cease to be unjust and will be just. For neither does the man who is ill become well on those terms—although he may, perhaps, be ill voluntarily, through living incontinently and disobeying his doctors. In that case it was \textit{then} open to him not to be ill, but not now, when he has thrown away his chance, just as when you have let a stone go it is too late to recover it; but yet it was in your power to throw it, since the moving principle was in you. So, too, to the unjust and to the self-indulgent man it was open at the beginning not to become men of this kind, and so they are such voluntarily; but now that they have become so it is not possible for them not to be so.\textsuperscript{93}

\textsuperscript{91} For example, at \textit{EN} 1150b 36, Aristotle writes “Vice is unconscious of itself.”
\textsuperscript{92} \textit{EN} 1095b 3-8; \textit{EN} 1179b 4-31.
\textsuperscript{93} \textit{EN} 1114a 13-22.
In this passage Aristotle makes it clear that being *ultimately* responsible for one’s character is sufficient for being responsible for the consequences of that character—including behavior that issues from it. The fact that, at the moment of such behavior, the agent could not have acted differently, is irrelevant since there was *some moment* in which it was open to the agent to act differently than she did act.

This result would seem to have a strange consequence when we return to the idea of how final and efficient causation works in the case of voluntary behavior. Earlier, we had decided that identifying a goal was a final cause, which after a mental process such as deliberation occurs, initiates the efficient causation of the agent *qua source* in the execution of the behavior deemed to be worthwhile. Now, we recognize that the efficient causal aspect of such actions is responsible for *which* final causes are in play in future scenarios. Therefore, it would seem that the relationship between final and efficient causation in voluntary and free action on Aristotle’s account is a reciprocal symbiotic one, requiring both aspects equally. In the case of the vicious person, her response to bad reasons grounds her blameworthiness. However, she is also responsible for becoming blameworthy. It appears that, had the agent not been the source of her inability to recognize good moral reasons, Aristotle might have struggled with how to justify blaming her. So, in this regard, sourcehood is important. However, in the next section we shall see that the sourcehood required is minimal. The agent must have had some opportunity to have turned out differently, but Aristotle maintains that we all have such an opportunity. Aristotle is not sympathetic to excuses of bad experiences or poor
upbringing, which I consider to be evidence that reasons-responsiveness maintains an important status on Aristotle’s view.

**Conditional Reasons-Responsiveness**

Some might worry that this ability to act differently is not distinct from the ability targeted in the concept of *conditional freedom* mentioned (and rejected) earlier. That is, this is an ability such that, had the agent chosen to act differently, she would have done so and she would have developed a different kind of character with different proclivities that selected different apparent goods as ends or reasons to act. However, there is something distinct from the contemporary view of conditional freedom in Aristotle’s notion of ultimate responsibility. Not only was it possible, in the past, for a vicious person to choose differently in the conditional sense, it was possible in a reasons-responsive sense too. I believe that there are two reasons why this is the case. First, Aristotle specifically says that we are all, in some way, aware of doing bad acts prior to developing a vicious character. Second, I believe that this claim is evidence that no one goes directly from having no character state straight to viciousness—that is, all vicious people must have first been incontinent prior to becoming blind to reasons. If these two claims are correct, then there is a very important sense in which we are all able to recognize and react to reasons to do good acts.

Let us first consider Aristotle’s claim that we are all initially capable of recognizing reasons to act well rather than poorly. In the fifth chapter of the third book of
Nicomachean Ethics, Aristotle says that only a thoroughly senseless person, one he calls the *komide anaistheton*, is not responsible for his character. The utterly senseless person is never capable of making choices. However, everyone else is, at some point, able to choose from among multiple actions. William Bondeson\(^9^4\) suggests that Aristotle might have thought knowledge of habituation a first principle given this remark: “Now not to know that it is from the exercise of activities on particular object[s] that states of character are produced is the mark of a thoroughly senseless person.”\(^9^5\) Aristotle claims that such knowledge is evident, “for it is activities exercised on particular objects that make the corresponding character. This is plain from the case of people training for any contest or action; they practise the activity the whole time.”\(^9^6\) In other words, everyone should know that if you do actions of a certain type regularly, you will become very good at doing those sorts of actions. Perhaps also, the success of doing such actions well will only encourage one to continue such training, as athletes rarely enter but one contest.

Now, this is not to say that all people are aware of all of the consequences of developing certain habits. For example, one who practices gymnastic activities such as tumbling habitually in the gym might know that such exercise will help her to do the same type of activity naturally and with ease without realizing that, in ten years, it will also produce arthritis in her overused joints. Likewise, a youth who imitates the crimes of a local drug-dealer might recognize that doing similar acts will gain her immediate fear and respect among the members of her community without realizing that, once caught,

\(^9^4\) Bondeson (1974).
\(^9^5\) 1114a 8-10.
\(^9^6\) 1114a 6-8.
she could spend the rest of her life in jail. One might think that this interpretation favors the importance of sourcehood over reasons-responsiveness since a vicious person is the source of actions that caused her to become vicious and when she chose to perform those actions she was not fully aware of the consequences of her choices. However, it does not seem to be the case that Aristotle is concerned with whether a vicious person was aware of all of the consequences of her earlier behavior. He seems only to be concerned with showing that 1. the vicious person voluntarily did bad acts prior to becoming vicious in response to some reason, 2. the agent could have chosen and become the efficient cause of alternative behaviors and 3. she was aware that being the efficient cause of such acts could only make her more accustomed to doing them. Indeed, each of these descriptions is still compatible with Aristotle’s overall program.

So far, we have seen how Aristotle would legitimate the claim that an agent should know that choosing to do similar acts repeatedly involves a risk or expectation that a habit will form. Still, we have not seen how, prior to becoming vicious, one would be able to recognize reasons to perform good acts. In this section, let us examine how the vicious person might have a history of being capable of reasons-responsiveness concerning good reasons to act.

Aristotle declares that one might discover the right ends to pursue acts in one of two ways: either that person is led to the right ends by another or she happens upon them.

---

97 Indeed, it is often held that only the phronimos (the practically wise person) fully recognizes, comprehends, and appreciates the good for what it is. This is to say that lesser persons might know that something is good or the right thing to do, but very few agents have full knowledge of the reasons for acting and all potential consequences of so acting.
by chance.\textsuperscript{98} Let us examine each of these possibilities and see if they are common enough to justify Aristotle’s claim that we all know enough prior to developing a vicious character to be responsible for doing so.

Let us consider first the idea that someone might lead us to the right ends. This seems to be obviously the case if one is fortunate enough to have good parents. But what if one does not? I think that Aristotle believed enough in the social systems of city-states and the social nature of human beings to rule out a person’s “slipping through the cracks.”\textsuperscript{99}

First of all, Aristotle is clear that man is a social and political animal by nature.\textsuperscript{100} He also suggests that leaving all of society behind would destroy a person’s humanity.\textsuperscript{101} So a youth is unlikely to miss chances to receive some amount of guidance from others in a community. Also, it is likely that the truly vicious comprise a very small proportion of any community.\textsuperscript{102} Therefore, a young member of a community is more likely to

\textsuperscript{98} C.f. EN 1105a 20-25. Aristotle discusses the possibility of someone producing something grammatical either by chance or through instruction. For Aristotle, if we do what virtuous people do either by chance or instruction, we are not behaving virtuously (we must have a virtuous character state). However, it is, in principle, possible that we do the type of acts that are done by the virtuous either by chance or by instruction.

\textsuperscript{99} I do not mean to overlook the apparent elitism that rings throughout Aristotle’s corpus. It is possible that Aristotle was unconcerned about certain people being at a disadvantage. Still, his political theory suggests that a good society would be one in which everyone has sufficient exposure to examples and rules to follow in order to have a chance at virtue and happiness.

\textsuperscript{100} Politics (hereafter Pol.) 1253a 2.

\textsuperscript{101} For example, in Pol. I.2, Aristotle tells us that one who lives away from society is either a beast or a god.

\textsuperscript{102} For example, “…the bestial person is also rare among human beings…We also use ‘bestial’ as a term of reproach for people whose vice exceeds the human level” (EN 1145a 30-35). Bestiality is a character type that is sub-human. Technically, it is beneath even vice. Here, Aristotle suggests that we often conflate thorough viciousness (i.e., in the sense of many or all extreme vices) with a bestial state, and this is possibly due to the fact that the two are so terrible and so rare. Admittedly, virtue too might be rare. But since virtuous, continent, and incontinent persons have some knowledge of the good, and since continence
encounter elders who are virtuous, continent, or incontinent than she is to encounter persons who are vicious. Since the virtuous, continent, and incontinent all have some grasp of the good, their advice is likely to be good advice. Finally, and most importantly, Aristotle stresses the importance of politics as a science.\textsuperscript{103} He suggests that the role of political agents is to construct laws that deter citizens from behaving poorly and guide them towards doing the types of acts that make them virtuous so that they may become happy.\textsuperscript{104} All of these influences, those of the guidance of one’s parents, the advice and examples provided by fellow citizens, and the political machine would seem to provide reasons for youngsters to perform better actions rather than worse ones. It would seem that, if this picture is accurate, the vicious person would have had to consider and reject reasons to behave differently than she did when developing her present character.

In addition to the above system producing reasons to behave well, it is possible that one might behave well by chance. Aristotle was aware of the possibility of good fortune causing one to do what happens to be the right thing, and he was cautious not to count such behavior as truly virtuous. By way of analogy, he explains why:

But surely actions are not enough, even in the case of crafts; for it is possible to produce a grammatical result by chance, or by following someone else’s instructions. To be grammarians, then, we must both

\textsuperscript{103} EN I.4.
\textsuperscript{104} EN I.4, I.9 (1099b 32) and X.9 (esp. 1180a 17-1181b 14).
produce a grammatical result and produce it grammatically—that is to say, produce it in accord with grammatical knowledge in us.\(^{105}\)

We can imagine that a youth deciding which type of person to become is similar to a tourist deciding between two equally unfamiliar dishes to eat for dinner. For example, when I visited India, I found myself at a restaurant that had menus written entirely in Hindi. The waiter spoke no English, and I, no Hindi. I had to choose what to eat by randomly pointing to an item on the menu. It is possible that, on some rare occasions, deciding which actions and persons to imitate is a similar experience.

Some might object that this outlook puts a person’s character in the hands of randomness and chance, not in the person’s own hands. However, it is possible to interpret Aristotle as being satisfied with the control an agent has if she responds to some reason to act rather than to refrain from acting, even in those instances when full comprehension of consequences is not possible. For instance, when I was in the predicament described above, it is still true that I had a reason to choose an item from the menu, which I could not read. I was hungry, and choosing \textit{anything} would lead to my receiving nourishment. I did not know what I was choosing, but I knew \textit{that} I was choosing from among options and I knew \textit{that} one important consequence of my choice was that I would be fed. Likewise, one deciding which kind of person to be, as a human being, possesses the desire for happiness. As doing nothing at all will not change one’s situation (and thus, not bring one to happiness), adopting some way of life is more likely to trigger development of excellence than adopting none.

\(^{105}\) \textit{EN} 1105a 21-26.
Still, objectors might protest, my knowledge of nourishment is more certain than a youth’s knowledge of which lifestyle might be conducive to happiness. On the other hand, Aristotle seems to be aware of this worry when he declares that matters as important as the highest good ought not to be left up to chance. In this sense, I propose that the two options, guidance and chance, are not really independent options. That is to say, Aristotle does seem unsatisfied with complete randomness. However, he seems confident that there are reasons available to be recognized to do the right kinds of actions. I say that there is some element of chance involved too, though, since, without prior understanding of why this reason to $\phi$ is better than that reason to $\sim\phi$, it is difficult to see how an agent can, in full knowledge, choose to do the right thing. This element of Aristotle’s theory has always seemed a bit like bootstrapping. However, to his credit, it does seem like the way that society is set up with good advice and deterrents to acting badly could be enough to help young human beings quickly develop enough common sense to follow the right kinds of advice over any alternative temptations.

**Prior Knowledge**

If the above discussion is correct, we can see how Aristotle might have thought that just about everyone knows enough about what types of behaviors are those that the virtuous person does, and which types of behaviors lead to viciousness. If so, every person would have sufficient prior knowledge to be held responsible for continually making bad

---

106 *EN* 1099b 24.
choices that lead to a vicious character. Indeed, Aristotle may have thought it reasonable to expect youths to be aware of certain basic laws (e.g., prohibition on murder, theft, treason, etc.), and that such awareness is enough to constitute an availability of reasons to act differently from vicious behaviors even if the reasons available do not illuminate all of the truth about why such behavior is detrimental. Much like when we are children, we are given rules and prescriptions by our parents that we only later come to fully understand and appreciate long after practicing them.

This general common knowledge had by youths could be the key to understanding how the vicious person, in her global state of ignorance of the good, is still to blame for her ignorance and poor behavior. Aristotle seems to suggest that, not only are vicious people not born that way, they hardly develop into vicious persons overnight. Such a disposition certainly requires a significant amount of repetition of similar types of acts. Presumably, if one were breaking the law frequently, one would be reprimanded frequently. Logistically speaking, it might take a person a significant portion of her lifetime to develop into a thoroughly vicious person. She would not only have to have had the occasion to voluntarily do many bad acts, but she would have to have chosen as such so many times that her habit outweigh all of the countervailing reasons being provided by outside sources.

If our analysis is correct, all vicious agents would have had a history of incontinence prior to becoming full-blown vicious agents. I think it is this fact that leads Aristotle to say that the vicious is to blame for her character and all of the behavior that issues from it, despite the fact that the vicious person is in no position to recognize and
respond to reasons to do other than she does. The vicious person is the source of her own blindness. She created her blindness over time and during a time when she saw reasons to do otherwise clearly. The fact that her choices constituted an efficient cause of rendering her insensitive to such reasons (thereby making her the source of her insensitivity) does not diminish the fact that the vicious agent was, at one time, open to those reasons and able to respond to them. She had to disregard alternative reasons knowingly in choosing to behave poorly; and in choosing to behave poorly she became more accustomed to poor behavior. Eventually, behaving poorly was the apparent lone choice. But the agent knowingly created this appearance.

Ignorance of the Particulars

Aristotle’s exoneration of those who act due to ignorance of the particulars can be seen as further evidence that praise- and blameworthiness require some prior epistemic availability of reasons for alternate actions. As we noted in previous sections, in the cases of acting due to ignorance of the particulars, it would appear that the agent does something voluntarily. That is, the agent moves her body in such and such a way with some end or reason for so doing. So, unlike the exempting factor of external force, this excuse does not entail that an agent’s behavior is, in every way, not up to her. Indeed, the agent is often the source of such behavior.

Aristotle seems to suggest that ignorance of the particulars is excusing because, while the movement of the agent’s body is up to her, the movement of her rational
deliberation is inhibited in ways that are not her fault. For example, had Oedipus known that his would-be victim were his father, this knowledge might have factored in significantly in his deliberations. That is, while he is considering one syllogism concerning fighting a man, he could consider another syllogism concerning honoring and protecting your kin. But without this knowledge that the man before him is his kin, he would not be able to deliberate on any ends or reasons concerning kinship.

It is possible that Aristotle’s motivation for including these types of exceptions has to do with his belief that, had such persons been aware of the particulars (and, therefore, had access to such reasons), such persons would have chosen to act differently. Aristotle distinguishes between his earlier claim that all actions caused by ignorance of the particulars are nonvoluntary and a claim concerning necessary conditions for an action to be involuntary: “…what is involuntary also involves pain and regret. For if someone’s action was caused by ignorance, but he now has no objection to the action, he has done it neither voluntarily, since he did not know what it was, nor involuntarily, since he now feels no pain.”

It appears to me that one must respond to a reason to do an action to have done it voluntarily; as in the case of Oedipus, he did not, in fact, respond to a reason to kill his father.

On the other hand, Aristotle suggests that a natural and necessary consequence of behaving in a way that is against one’s will is pain and regret. This last part is interesting and deserving of more attention than we can allot it here; however, I shall take a guess at its meaning. It seems to me that in order to act involuntarily, the agent must be intending

---

107 EN 1110b 20-22.
to do something that is incompatible with what, in fact, ensues. For example, if I buy a lottery ticket because I would like to win $10,000, and I win, but the prize is $100,000, or $9,000, I will not feel pain or regret for having bought the ticket. It does not go against my will to win a different amount of money. I simply wanted to win some money. On the other hand, in the case of one friend offering another a drink that turns out to be poison, the friend feels pain and regret. She feels pain because her intention was to help and not to harm her friend’s health and physical well-being.

In short, the absence of awareness of particulars constitutes an absence of internal reasons to do otherwise. Oedipus’ killing of his father was not “up to him” because he was not able to access a reason to not kill his father. His failure to access such a reason stems from ignorance of the particulars, and not a global ignorance of universal principles that he, himself, precipitated by prior behavior, as is the case with the agent who acts viciously.

The above analysis supports my claim that Aristotle’s theory of freedom and responsibility is centered on conditional reasons-responsiveness and not a wider notion of conditional freedom to act. We are responsible for our behavior and deserving of praise or blame for it if it is, in fact, issued by a response to reasons. But such a response to reasons must be “up to us” such that we were able, at some point prior to acting, to respond to alternative reasons to act differently than we in fact do act.
Conclusion

In this chapter, I have argued several points. The overarching position has been that Aristotle’s account anticipates contemporary views of sourcehood and reasons-responsiveness. In demonstrating this fact, I have argued that, for Aristotle, both sourcehood and reasons-responsiveness are required for voluntary behavior and, therefore, for the aptness of praise and blame. In attempts to explain how both are necessary, we saw that one, sourcehood, depends upon the other, reasons-responsiveness. Not only would proper sourcehood not be possible without reasons-responsiveness, but a complete execution of reasons-responsiveness has sourcehood among the constituent elements. The way that sourcehood and reasons-responsiveness seem to operate on an Aristotelian picture of praise- and blameworthy behavior involves what I have called a dual causal activity. Reason supplies the end as a final cause (telos) and, in this sense, has recognized a reason. The recognition of this reason enables appetite to serve as an efficient cause (kinesis) in the right way by giving it a command to which it listens or obeys. We then discussed how this mechanism can go awry, and which types of interference constitute mitigating excuses and which do not. We determined that, so long as agents have conditional reasons-responsiveness, the agent can be praised or blamed for her behavior. I argued that, for Aristotle, it is likely that every individual has conditional reasons-responsiveness regarding universal propositions concerning the good. Therefore, only in cases of ignorance of the particulars could ignorance compromise the conditional reasons-responsiveness of an agent. In the case of the incontinent person,
ignorance of the particulars is not at play. The fact that a reason to do otherwise is recognized is sufficient for the akratic’s omission of right action to qualify as blameworthy. In the case of the vicious person, only global ignorance of the universal is preventing the agent from deliberating about good ends. However, the fact that this agent has a causal history of sabotaging her own ability to recognize alternative reasons to act is sufficient grounds for blaming her when she fails to recognize (and, therefore, react to) reasons to behave well. Therefore, agents are worthy of praise and blame in virtue of their ability to correctly implement a dual causal relation between recognizing the ends to pursue and reacting to such recognition as the proper source of the behavior.

---

108 All things considered—that is, this also supposes that the agent is not paralyzed, or tied down, etc.
Chapter 3

Contemporary Sourcehood Theories of Moral Responsibility and Praise- and Blameworthiness

As Figure 2 from chapter 1 suggests, contemporary views of moral responsibility and praise- and blameworthiness can be divided into those that consider sourcehood a primary condition for moral responsibility and those that consider reasons-responsiveness the primary condition. As I suggested in the previous chapter, sourcehood resembles Aristotle’s first criterion for the voluntary (that the action not be forced but initiated by internal motion), which involves efficient causality. In this chapter let us consider how prominent contemporary theories address the criterion of sourcehood.

Sourcehood views focus on how agents cause their behavior. More specifically, sourcehood views hold that an agent must be a privileged cause of her behavior in that *she* be the causal source for the behavior over and above other persons and influences. Therefore, the sourcehood theorist is not only concerned with the fact that the agent causes her behavior, but also that she does so in a sufficiently controlled manner.

In this chapter, I shall present various sourcehood views in light of how such views detect a sufficient degree of control on the part of an agent. The views that I shall
present all focus on *alternate possibilities*. Alternate possibilities are options from which an agent can choose. These sourcehood theorists have conceptualized a sort of test to determine whether an agent has proper control over her behavior or whether something else is ultimately the source of such control: consider whether the agent had *alternate possibilities*. The idea behind this criterion is that, if the agent was the source of her behavior and nothing external to her was that source, then the agent could have controlled her behavior in a multitude of ways. If she could only do as she in fact does, presumably, this would be due to the fact that *something* external to her was *forcing* her to behave in this restricted manner.

The earliest sourcehood views have been called *Leeway views.* Leeway views hold that an agent must be able to do otherwise at the time of an action in order to act freely and to be morally responsible for such behavior. In other words, if an agent acts freely and responsibly, then it must be true that, just prior to her action, she could have caused herself to act in a way other than how she does act. For instance, if I am said to freely and responsibly choose to give a friend a gift, then it must be true that I could have freely and responsibly chosen not to give the friend a gift. If it was not possible that I refrain from the gift-giving, according to leeway views, I did not give the gift freely or

---

109 I include leeway views as sourcehood views since, by having the ability to do otherwise immediately prior to an action, an agent is also the source of that behavior. Admittedly, later sourcehood views distinguish themselves from leeway views in two ways: 1) they put priority on sourcehood over alternate possibilities (which, I shall argue is merely an appearance) and 2) they do not require alternate possibilities immediately prior to action. Therefore, some consider leeway views ancestors of sourcehood views, yet distinct from them. I shall simply lump them together for purposes of this discussion. Also, it is worth noting that there are leeway compatibilists as well as leeway incompatibilists. For a discussion of the latter, see Timpe (2008): 71-73.
responsibly. On such views, the ability to do otherwise makes one the source of one’s behavior since it is up to one which of the alternative behaviors she enacts.

Some such sourcehood views are compatibilist; this is to say that such views have an understanding of an agent being the appropriate causal source of her action at some relevant time in which the agent could have done otherwise in a conditional sense. For example, in his paper “Freedom and Necessity,” A. J. Ayer first posits that freedom is contrary to constraint, but it is not contrary to cause.\textsuperscript{110} If I am tied to a chair, I am constrained and cannot act other than I do. If I were to choose to get up, I would not be able to do so under such circumstances. However, if I am determined to sit in a chair, I am not so constrained. If I were to choose to get out of the chair, presumably, I would. The fact that I will not choose to do so is irrelevant. In the first instance, I am constrained to sit in the chair. In the second instance, I cause myself to sit in the chair. Only the first instance, says Ayer, is an example of constraint that precludes freedom and responsibility.

As Ayer says, “to say that I could have acted otherwise is to say, first, that I should have acted otherwise if I had so chosen; secondly, that my action was voluntary…and thirdly, that nobody compelled me to choose as I did.”\textsuperscript{111} In other words, the agent is the source of the act, which results from a choice initiated within, and had a different choice been issued, the act would have been different. Presumably, the caveat ruling out the compulsion by another is added to exclude cases where sourcehood can be

\textsuperscript{111} Ayer (1982): 117.
traced beyond the agent in question to another agent. On such an account alternative possibilities and sourcehood are intimately connected. If one is the source, one will also have had conditional freedom.

In addition to requiring conditional freedom, Ayer considers it essential to acting freely that such acts are caused by the agent’s choice in a particular way. The choice itself must be caused by a person’s tendencies, character, background, etc. For Ayer, if one’s choices are not caused by such things, then the choice is arbitrary, unpredictable, and we look at the person not as a moral agent, but rather as a lunatic.\(^{112}\) Being caused by one’s nature is what makes a choice belong to an agent in the relevant sense. In other words, the agent can only be the source of her behavior if her behavior is intimately connected to who the agent is.\(^{113}\)

Some leeway theorists do not find conditional freedom very substantial since it is compatible with the truth of determinism. Such a *leeway incompatibilist* might argue that one must have an ability to do otherwise prior to acting in a robust sense (i.e., stronger than the conditional sense). That is, the agent must have been able to choose otherwise without changing the past, the circumstances, or the laws of nature. Peter van Inwagen holds such a position. He first illustrates that determinism rules out this ability to do

---

\(^{112}\) Ayer (1982): 112.

\(^{113}\) It is arguable that the reference to avoiding lunacy suggests that Ayer was primarily concerned with rationality, and thus was more of a reasons-responsive theorist. It is certainly likely that Ayer had such concerns in mind; however, as his theory centers on how an act is caused and how such causes admit of abilities to do otherwise, it seems more likely that his focus was on sourcehood. Be that as it may, this does not mean that rationality played no role in his theory. It is for this reason that I distinguish views based on what they hold to be primarily related to moral responsibility.
otherwise. From his argument, it can be seen that this ability is ruled out because factors outside of the agent can equally be described as the source of the agent’s action.

Peter van Inwagen advances an argument called *The Consequence Argument*. Van Inwagen argues that if determinism is true, then every event including people’s actions are the necessary consequences of the past and the laws of nature. But the past is not up to us (e.g., it is not up to us what went on before we were born). Likewise, it is not up to us what the laws of nature are. Since our actions are the necessary consequences of these two, it is not up to us what we do. Something is “up to us” if and only if it is up to us to do it or to not do it. Since it is not up to us what we do, we cannot do other than what we do.

Van Inwagen does not explicitly state that his concern is with ensuring that a free agent have a distinguished causal role in generating her actions.\(^{114}\) Still, it is easy to see how sourcehood is embedded in the worries van Inwagen does express. The argument is ostensibly about how alternate possibilities afford us control over our actions. What the argument suggests, but does not mention, is that there is a sort of competition between the agent and the pair constituted by the past and the laws of nature to be the cause of the agent’s action. That is, upon assuming the truth of determinism, the argument shows that an agent’s actions are consequences of *both* the agent’s previous states (presumably including intentions, choices, etc.) and the past and the laws of nature. But this is not to say that an agent’s behavior is overdetermined; for the agent’s intentions and choices are also consequences of the past and the laws of nature. In short, this argument can be

\(^{114}\) Incidentally, van Inwagen is not only concerned with alternate possibilities supported by physical indeterminism; he also requires that agents not be psychologically determined as well.
compelling because it suggests that the true source of an agent’s behavior (if the thesis of determinism is true) lies outside of and prior to the agent herself.

I think that this suggestion motivates a lot of libertarians to object to source compatibilist views. On source compatibilist views such as Ayer’s, all that is required for an agent to be the source of her action is that the action be immediately caused by the agent’s choice in such a way that, had the agent chosen differently, she would have successfully caused a different outcome. Since the agent’s ability to choose differently lies in influences beyond her control (her constitution and nature, the past, the laws of nature, etc.), libertarians “trace” the locus of causation beyond the agent to such influences. For example, imagine that Joshua commits a brutal murder and performs this act as a result of his choice. Now imagine that determinism is true. It follows that some $P_0$ at some time $t_0$ (prior to Joshua’s birth) in conjunction with the laws of nature is sufficient for both Joshua’s choice and the murder. Therefore, while Joshua’s actual choice is a sufficient cause of the murder, so is some other state of affairs that did, in fact, obtain prior to Joshua’s birth. Furthermore, that other state of affairs is not only a sufficient cause of the murder, but also a sufficient cause of Joshua’s deliberation and choice to commit the murder. Since each is a sufficient (and actual) cause of the murder, but only one is the sufficient cause of the other, it seems reasonable to the incompatibilist to consider that cause the true cause (or, at least, the more fundamental one). Therefore, while Joshua is the immediate cause of the murder, it would seem as though his past and the past prior to his existence in conjunction with the laws of nature are just as much causes as (and, perhaps more to blame than) Joshua is.
The idea of tracing back the causal chain of an agent’s behavior on a sourcehood compatibilist account has raised an additional concern. Not only might it be problematic that early factors are causally responsible for an agent’s behaviors such as determinism, the past, etc., but also some of those factors might be paradigms of the kinds of causal influence that mitigates all responsibility on any account. For example, some of those influences might be manipulative influences, and we intuitively do not feel justified in blaming the behavior of one agent that was manipulated to so act by another agent. For example, imagine that Jeffrey robs a bank today because 10 years ago neuroscientists implanted microchips into Jeffrey’s brain that generated neurological signals that ultimately resulted in Jeffrey’s desires for malfeasance (including his present desire to rob banks). Now, imagine that we were unaware of this fact when we first learn of Jeffrey’s crime. Initially, we feel justified in blaming him. However, once we learn of the neuroscientists’ manipulative tricks, we begin to question the validity of our judgment of Jeffrey’s blameworthiness. Indeed, many might prefer to shift blame to the neuroscientists. In any case, the fact that such views allow the possibility that the past (including past manipulations) could be as responsible as the agent when she acts has caused some compatibilists to abandon the sourcehood approach as well as inspiring some sourcehood theorists to embrace incompatibilism.
Proximal and Distal Causes—the Historical Requirement

I suspect that worries such as these have motivated compatibilists to refocus on what is the primary condition for moral responsibility. Indeed, the majority of compatibilist views espoused today are reasons-responsive oriented. However, a small minority of source compatibilists remains, and source compatibilists seem to be implicitly responding to worries just like the above. One standard type of response might be that an agent is responsible for that for which she is the proximal cause.\footnote{For example, see Mele (2006) Chapter 7, where Mele advances a compatibilist approach focused on an historical requirement.} For example, in the case of Joshua the murderer, compatibilists might argue that the above worry is mistaken. The fact that Joshua is the proximal cause, or the cause that is immediately prior to the event of his victim’s death, is what makes him responsible. In order to avoid manipulation arguments, theorists might augment such a view with a stipulation concerning the history of the agent’s desires and motives. Therefore, the fact that Joshua is the proximal cause of the murder coupled with the fact that no one manipulated Joshua to act in such a way are jointly sufficient for holding Joshua accountable.

Al Mele’s compatibilist account is an example of a view based on proximal causation that is immune to manipulation arguments. Mele holds that in order for one to be a free and autonomous agent, one must be free of compelled or coercively produced attitudes.\footnote{Mele (1995): 187; (2006): 164.} In order to illustrate this requirement, Mele asks us to consider two
colleagues, Ann and Beth.\textsuperscript{117} Ann is naturally hard working and reliable. Beth is naturally lazy. A team of neuroscientists examine Ann to see how she operates and then secretly manipulate Beth’s brain patterns such that she feel and act similarly. As Mele puts it, “Beth is now, in the relevant respect, a “psychological twin” of Ann.”\textsuperscript{118} The manipulation example is supposed to elicit the following worry in readers. After manipulation, Beth appears to deliberate rationally based on her beliefs and desires and she makes praiseworthy choices as a result. These factors, on traditional compatibilist accounts, are sufficient for an agent to qualify as praise- or blameworthy. Yet, in this example, we think Beth only meets these requirements accidentally and as a result of interference and manipulation of others. For these reasons, we are reluctant to praise her as we would Ann.

Mele agrees that Ann would seem to deserve more praise than would Beth (that is, if Beth deserves any). However, contrary to many incompatibilist complaints, Mele does not hold this view because of some concern with determinism. We are not reluctant to praise Beth because she was \textit{determined} to act in praiseworthy ways. We are, however, reluctant to praise Beth because of \textit{how} she became so determined. Mele comments, “the difference in their current status regarding freedom would seem to lie in how they \textit{came} to have certain of their psychological features, hence in something \textit{external} to their present psychological constitutions. That is, the crucial difference is \textit{historical}: free agency is in some way history-bound.”\textsuperscript{119}

\textsuperscript{117} The example is first proposed in Mele (1995): 145 and is reiterated in Mele (2006): 164-166.
\textsuperscript{118} Mele (2006): 165.
\textsuperscript{119} Mele (2006): 166.
Mele’s reasoning for an historical requirement amounts to a concern for sourcehood. Like Aristotle, Mele is opposed to considering Beth responsible for her behavior if such behavior was caused by something external to Beth. However, this is not the only factor that qualifies Mele’s view as a sourcehood view.

In addition to his concern with the history of desires and motives, Mele is equally worried about the resistibility of desires, which amounts to a concern with alternate possibilities. In his paper “Irresistible Desires,” Mele distinguishes between incontinent agents who “succumb to desires that they could have resisted”\textsuperscript{120} and agents who succumb to desires that they could not have resisted. For Mele, “a desire is irresistible if and only if it is both unconquerable and uncircumventable.”\textsuperscript{121} A desire is unconquerable when an agent cannot intentionally combat that desire, neither by sheer effort of will nor indirectly by trying to, “manipulate his motivational condition, his environment, or both in such a way as to bring it about that he does not act on the desire…”\textsuperscript{122} That is to say, the agent cannot create incentives for herself, paper her wall with motivational post-it notes, or get her friends to tease her enough to alter her motivation. For Mele, such attempts to induce the resistance to the desire are not even possible when the desire is unconquerable. However, Mele contends that it is possible for a desire to be unconquerable but still circumventable. In order for a desire to be circumventable, an agent must have another desire that she could intentionally fulfill; the fulfillment of that

\textsuperscript{120} Mele (1990): 456.
\textsuperscript{121} Mele (1990): 467.
\textsuperscript{122} Mele (1990): 459.
desire would result in her not fulfilling the desire she is circumventing. Mele considers such desires linked to *W*-alternatives:

First, let us say that intentional actions A and B are *W*-alternatives at (during) t for an agent if and only if the agent has a t-desire to do A and a t-desire to do B and he will not be moved by both desires to perform their targeted intentional actions. Then we can say that a desire, D, to A at t is circumventable at t for an agent, S, if and only if, independently of any attempt or intention at t to resist D at t, S is able at t to perform at t an intentional action that is a *W*-alternative to a D-motivated intentional A-ing.123

Consider the following case of an unconquerable desire that is circumventable. Imagine that Angela is an alcoholic and she has an unconquerable desire to drink at any time t such that drinking is made possible. No matter how many times she has tested her will by walking into a pub with the intention of just ordering a soda, writing affirmations in her journal, or receiving encouragement from her friends, she always opts to consume alcohol when it is available. However, she frequently desires to spend long hours at Chuck-E-Cheese’s, where alcohol is not available. If she acts on a desire to go to Chuck-E-Cheese’s, she has circumvented her desire to drink. However, if the option were available, Angela could not willfully resist the opportunity to drink.

Mele seems to exclude cases like Angela’s from the category of *irresistible desires* on the basis that it is possible for Angela to circumvent this desire, which is

---

another way of saying that it is possible for Angela to intentionally do other than she does. However, since Mele thinks that the irresistibility of desires is relative to specific instances,\textsuperscript{124} he would find Angela’s desire to be irresistible \textit{at the moment} that she is in the pub. While in the pub, Angela cannot conquer her desire to drink nor can she circumvent it, since all such opportunity was lost the second she entered the door.

Therefore, Mele’s compatibilist view is a sourcehood view. It is a sourcehood view in the sense that the historical requirement is an example of barring particular instances of external causation to rule out past manipulation \textit{and} it is a sourcehood view since a lack of alternate possibilities threatens the agent’s control in such a way that she is no longer considered free and responsible. These two aims of his seem focused on ensuring that the agent plays a special causal role in producing her actions; therefore, Mele’s is a sourcehood theory.\textsuperscript{125}

\textsuperscript{124} Mele (1990): 455.

\textsuperscript{125} While Mele’s view does address the concerns of distal causation and manipulation, it has not won many converts from other sourcehood theories. I imagine that the way in which Mele addresses these worries is what leaves sourcehood theorists unsatisfied. Rather than showing how a sourcehood compatibilist could hold a theory immune to such objections, Mele simply claims that such cases are exceptions. That is to say, Mele claims that proximal causation is sufficient. He also holds that if an agent is manipulated or in the grips of an irresistible desire, then she is not free or responsible. However, he has not shown how moral responsibility can be grounded in a compatibilist sourcehood view; instead he has declared that the objections raised above justify excluding such problem cases. These declarations might strike philosophers as ad hoc.
Another compatibilist sourcehood account is John Martin Fischer and Mark Ravizza’s *semicompatibilism*. While Fischer and Ravizza offer extensive discussions of what reasons-responsiveness is and how it factors in to moral responsibility, they also maintain that a certain type of control is both necessary and sufficient for moral responsibility. Fischer and Ravizza consider such control to be of the utmost importance because behaving in a reasons-responsive manner at any given moment is not sufficient for an agent to be responsible. An agent’s reasons-responsive mechanism must be her own, which means that she cannot be manipulated, she must satisfy a basic historical requirement, and she must take responsibility for her actions as her own. If all such requirements are met, the agent can be said to have the kind of control required for moral responsibility. In holding these requirements, Fischer and Ravizza seem to be primarily concerned with sourcehood.

Fischer and Ravizza consider their account a version of *semicompatibilism*. Semicompatibilist views hold that moral responsibility is compatible with determinism *even if* free will is incompatible with determinism. According to their view of semicompatibilism, moral responsibility requires a certain kind of control that Fischer and Ravizza call *guidance control*. As Fischer and Ravizza tell us, “guidance control of

---

126 While *semocompatibilism* is endorsed their co-authored work, Fischer has expounded upon his commitments to the view in independently authored pieces. For this reason, I shall sometimes speak of him exclusively. When I refer to both Fischer and Ravizza, I am paraphrasing remarks concerning their theory as it is presented Fischer and Ravizza (1998).

127 Fischer and Ravizza (1998): 51-4. Fischer believes that free will in the sense that requires alternate possibilities is incompatible with determinism, as he professes to be convinced by Van Inwagen’s Consequence Argument and finds accounts of conditional freedom unhelpful to resolving the problem posed by Van Inwagen’s argument (Fischer et al. (2007): 49-56).
an action involves an agent’s freely performing that action,” where the term *freely* is to be understood as involving intentional behavior with control of the movement of the agent’s muscle’s when executing a choice. Guidance control is best described as the type of control the agent has in any Frankfurt-style case where the mechanism involved in the action is the agent’s own and the mechanism that blocks alternatives remains inactive.

Guidance control can be distinguished from a different type of control called *regulative control*, since:

Regulative control involves a *dual* power: for example, the power freely to do some act \( A \), and the power freely to do something else instead (where “doing something else” may be simply refraining from acting at all, or “doing nothing”). Alternatively, one could say that regulative control involves the dual power to exercise guidance control: the power to

---


129 Fischer and Ravizza’s own Frankfurt-style case used to illustrate guidance control concerns a driver named Sally: “Let us suppose that Sally is driving her car. It is functioning well, and Sally wishes to make a right turn. As a result of her intention to turn right, she signals, turns the steering wheel, and carefully guides the car to the right” ((1998): 30). We can see that Sally has guidance control in turning the car to the right because Sally seems to have a reason to turn right (her wish or desire), she forms an intention to turn right, and she turns the car right without any interference from anything external to her (i.e., the car is functioning well, no other cars collide with hers before the turn, etc.). To illustrate how Sally’s control of turning the car to the right need not involve genuine alternate possibilities (and is therefore compatible with things that might rule out alternate possibilities, such as determinism), Fischer and Ravizza alter the example by adding to it that Sally is in the process of receiving a driving lesson from a driver’s ed. instructor in a car that has dual controls, the dominant set being on the side of the car where the instructor sits: “We can further imagine that the instructor is quite happy to allow Sally to steer the car to the right, but that if Sally had shown any inclination to cause the car to go in some other direction, the instructor would have intervened and caused the car to go to the right (just as it actually goes)” (1998, 32). For example, let us suppose that Sally has already signaled right and moved the car into a right-turn-only lane. The instructor is perfectly glad to let Sally continue to make the right turn. However, if Sally begins to do anything else the instructor will take over the controls of the car and make it go to the right (presumably to avoid accidents, ensure their safety, follow traffic rules, etc.). Fischer and Ravizza use two distinct prepositions when describing the kind of control that Sally has and the kind of control that her driving instructor has. When Sally controls the car to go to the right, says Fischer and Ravizza, she has control of the car. She has guidance control of turning the car to the right. But since the instructor was there ready to take over control of the car in the event that Sally neglect to turn the car to the right as she had indicated that she would, Sally does not have control over the car. She can only have control over the car if she had guidance control plus something extra, making that type of control a bit more complex.
exercise guidance control of $A$, and the power to exercise guidance control of something else (instead of A).\textsuperscript{130}

In other words, regulative control is the type of control one has when she exhibits guidance control and at the same time could have exhibited guidance control in an entirely different way.\textsuperscript{131} For Fischer and Ravizza, though, regulative control is sufficient but not necessary for moral responsibility. Ostensibly, not requiring regulative control amounts to being unconcerned with alternate possibilities. While Fischer and Ravizza are not concerned with the kind of alternate possibilities that only indeterminism can ground, they are concerned with alternative possibilities of another sort.

The guidance control necessary for moral responsibility involves the operation of a certain kind of mechanism that is reasons-responsive in a particular way.\textsuperscript{132} As we saw in Chapter 1, reasons-responsiveness requires receptivity to reasons and then a reaction to them. In order for one’s behavior to qualify as having guidance control (and, therefore, moral responsibility), Fischer and Ravizza hold that one’s actions must stem from a mechanism that is at least moderately reasons-responsive.

For Fischer and Ravizza, Moderately Reasons-Responsive (MRR) mechanisms involve a certain asymmetry between their tendencies towards recognition and reaction:

We contend that the reactivity to reasons and receptivity to reasons that constitute the responsiveness relevant to moral responsibility are crucially asymmetric. Whereas a very weak sort of reactivity is all that is required, a

\textsuperscript{130} Fischer and Ravizza (1998): 31.

\textsuperscript{131} Fischer and Ravizza stress that “regulative control requires that the agent exercise guidance control (or, alternatively, act freely) in both the actual sequence and the alternative sequence” ((1998): 31).

\textsuperscript{132} A mechanism is loosely defined as the process that leads to the relevant behavior or action ((1998): 38).
*stronger* sort of receptivity to reasons is necessary for this kind of responsiveness.\(^{133}\)

In order for an agent to be operating with a MRR mechanism, she must be regularly receptive to reasons (some of which are moral reasons) and there must be at least one possible world such that there is a sufficient reason to do otherwise, the agent recognizes that reason and does otherwise for that reason. The first condition is included because “we want to know if (when acting on the actual mechanism) he recognizes how reasons fit together, sees why one reason is stronger than another, and understands how the acceptance of one reason as sufficient implies that a stronger reason must also be sufficient.”\(^{134}\) When all of this is true of an agent, he can be said to have an appropriate pattern of reasons-recognition since it is an “understandable pattern of (actual and hypothetical) reasons-receptivity.”\(^{135}\) For an agent’s pattern of reason to be understandable, it must involve recognition of somewhat objective reasons (“relative to a given set of preferences, values, and beliefs”\(^{136}\)) as well as an accurate “grading” of strengths of reasons (e.g., if the fact that a dress costs $1,000 is sufficient reason to refrain from buying it, so too would it costing $2,000). In other words, being regularly receptive to reasons is a sign that one is sufficiently rational to count as a rational, moral agent.

The second requirement for a MRR mechanism is that it must be *weakly reactive* to reasons. That is, there must be at least one possible world in which there is a sufficient


\(^{134}\) Fischer and Ravizza (1998): 71.

\(^{135}\) Ibid.

reason to do otherwise, the agent recognizes that reason and *translates* it into a choice and resulting behavior *because* she grasps that reason as a sufficient reason for action. This aspect of their theory also suggests some affinity for sourcehood concerns; for requiring certain behaviors in possible worlds sounds like a concern for alternate possibilities. If this is the case, Fischer and Ravizza seem to be implicitly committing to alternative possibilities in a conditional sense. What makes the mechanism viable for issuing acts with guidance control is that it is capable, even if only on one possible occasion, of issuing an *alternate* behavior. As Fischer and Ravizza state, “if an agent’s mechanism reacts to some incentive to (say) do other than he actually does, this shows that the mechanism can react to any incentive to do otherwise.” As I recognize sourcehood compatibilism as a theory concerned with conditional freedom, this explanation of MRR suggests to me some substantial sourcehood thinking on the part of Fischer and Ravizza.138

Fischer and Ravizza’s second requirement for a mechanism to be viable is that the mechanism must be an agent’s own. In order for a mechanism to be an agent’s own, the agent must satisfy an appropriate historical connection to the mechanism (and, therefore, must not be manipulated by anyone) and the agent must take ownership of her own

---

137 Fischer and Ravizza (1998): 73.

138 On the other hand, Fischer and Ravizza claim that they add this condition to the qualifications for MRR mechanisms because it rules out the possibility of *irresistible desires*. They offer an example of a weak-willed addict named Brown. Usually, when faced with sufficient reasons to refrain from taking his drug, Plezu, Brown takes the drug anyway. However, if he were told on one occasion that taking the drug would cause an extremely undesirable consequence (e.g., death), he would, under those circumstances, recognize that as a sufficient reason and refrain from taking Plezu (Fischer and Ravizza (1998): 69-70). Since it is possible for there to be a scenario presenting a sufficient reason to do otherwise that the agent not only recognizes but also reacts to, the desire is not irresistible. We can infer from this discussion that, according to Fischer and Ravizza, an essential feature of irresistible desires is that experiencing such a desire entails that there is no possible world in which one has the desire and is not motivated to action by it.
mechanism. Both the historical requirement and the requirement of taking ownership are suggestive of the importance Fischer and Ravizza place on sourcehood.

Fischer and Ravizza contend that while “The properties in virtue of which a mechanism (or sequence) is reasons-responsive are modal or dispositional properties of an agent at (roughly) the time of the action, omission, or production of the consequence,”¹³⁹ there remain, however, responsibility-undermining conditions that could preclude one from being morally responsible even if the agent has and acts from said properties. These responsibility-undermining conditions include deviant causal chains in the history of the mechanism in question.

In order to illustrate their reasons for holding this view, Fischer and Ravizza distinguish between views of moral responsibility that consider only a “current time-slice” and those that consider historical facts relevant to ontology of the present object or scenario. One crude example of the difference might involve a perfect imitation of a Picasso painting. Even if the painting had precisely the right colors made from the same materials as those used by Picasso and placed precisely in the right amount on the canvas (also identical to the material used by Picasso) in exact relations to other colors on the canvas such that the copy is an exact replica, only the original is “a Picasso.” Fischer and Ravizza contend that what makes the Picasso a Picasso and the replica merely a replica is that the original has an historical causal link to Picasso himself and can, therefore, be said to be his work.¹⁴⁰

The Picasso example is meant to be analogous to how we evaluate instances of moral responsibility. Fischer and Ravizza offer the following example:

It is often thought that an individual’s background can have a crucial impact on his subsequent moral responsibility. If an individual has been subject to significant sorts of mental and physical abuse as a child or young adult, this may well imply that the agent is not subsequently morally responsible for at least some of his behavior.\textsuperscript{141}

Essentially, what Fischer and Ravizza might ask us to imagine is something like the following. Imagine that two individuals have been caught committing exactly the same vicious criminal offense. We initially consider each equally morally responsible. But if we were to learn, as the comment above suggests, that one of the two were abused in the past in ways that can, perhaps, explain his present behavior, we no longer consider the person morally responsible. If this is true, then it must be because history matters to the kind of guidance control one exhibits; if it did not matter, there is no justification for different judgments since both offenders have the same present behavior.

Fischer and Ravizza also address more immediate examples of influence and manipulation. In the above case, it might be argued that the abused person had significant time and alternate influences to have become a different person. To such objections, Fischer and Ravizza consider the more science-fiction examples of a neuroscientist manipulating an agent’s first-order desires as well as her higher-order desires or judgments such that an agent who was previously unlikely to behave immorally now responds to reasons to do evil acts regularly. In such an instance, we are to imagine that the change has happened over night. Fischer and Ravizza are not satisfied with the apparent rationality of the victim the next day when she recognizes and responds to

\textsuperscript{141} Fischer and Ravizza (1998): 187.
reasons to harm others. They write, “When the harmony (or mesh) in selected mental ingredients—whatever they are—is produced in these ways, the mechanism that issues in the relevant behavior is not, in an important intuitive sense, the agent’s own.”

Here, “the mesh” is taken to be the agent’s present rational mechanism operating with the agent’s present mental states.

In addition to the claim that an agent must have an appropriate causal connection to her reasons-responsive mechanism, Fischer and Ravizza add that, “Taking responsibility, we believe, is a necessary feature of moral responsibility. It is part of the process by which a mechanism leading (say) to an action, becomes one’s own.”

Taking responsibility involves an agent first seeing “himself as an agent; he must see that his choices and actions are efficacious in the world.” In addition to this requirement, “the individual must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts,” where the agent considers himself an apt candidate in certain, non-arbitrary contexts (i.e., it is not necessary that these contexts be 100% in line with which ones morality, in fact, is involved). Even if the agent has not reflected on the nature of morality, it is sufficient if he forms “a judgment about our social practices…the individual must see that in certain contexts it is “fair”…for others to subject him to the reactive attitudes in certain circumstances.”

Fischer and Ravizza admit that there are cases in which an agent might seem to act via a moderately reasons-

---

146 ibid.
responsive mechanism, yet fail to satisfy these requirements. Regarding such instances, Fischer and Ravizza provide the following commentary:

The fact that we might further blame someone who fails to react appropriately toward himself indicates that we believe the person is an appropriate candidate for the reactive attitudes. If, however, a person resolutely shows no moral response or appreciation of the moral force of the attitudes we take toward him, then eventually we must concede that he is not an appropriate partner in the conversation.\(^\text{147}\)

Therefore, the requirement that an agent affirm himself as an agent who is the source of his behavior is a necessary one, the absence of which undermines any other apparent reasons-responsiveness.

We can now see why Fischer and Ravizza’s view ought to be classified as a sourcehood view. First, the notion of history and ownership of a mechanism seem characteristic of a sourcehood account. An agent must have the appropriate causal connection to the past influences that shaped his reasons-responsiveness. In addition to the fact that the agent must have a certain causal status concerning his present state, the agent must also now be the source of his legitimacy for being judged as responsible. In both instances, it is “up to” the agent whether he qualifies as a candidate for ascriptions of moral responsibility.\(^\text{148}\)


\(^{148}\) While I admire Fischer and Ravizza’s view, I cannot endorse several of their main claims. First, as I shall later show, Fischer and Ravizza consider responsibility an eligibility requirement for praise- and blameworthiness. Therefore, anything that is required for moral responsibility is also required for praise- and blameworthiness. I shall dispute this view in subsequent chapters. Also, since Fischer and Ravizza offer an account of moral responsibility based solely on control of bodily movements, their account of moral responsibility and praise- and blameworthiness is entirely about actions. Therefore, I cannot endorse this view, as I contend that praise- and blameworthiness can be about more types of behavior than simply actions.
As we recall, van Inwagen’s original worry involved the truth of universal determinism ruling out free action. However, later he argues that psychological determinism creates the same problem.\textsuperscript{149} According to van Inwagen, we are caused to act by the reasons we have—a claim that I shall refer to as \textit{Davidson’s Thesis}.\textsuperscript{150} Like many, van Inwagen holds that it is not under our control which reasons we have for acting. Unlike many libertarians, van Inwagen concedes that, if given a strongest reason, we shall be caused to act on it. Therefore, van Inwagen concludes, we are not free when we act on the strongest reason. He remains a libertarian, however, because he holds that there are occasions when there is “a tie” between reasons and an agent then has the power to choose between them.

Since none of us starve to death when confronted with equally appealing choices at a restaurant, van Inwagen concludes that we have a capacity to choose even when no decisive reason is present.\textsuperscript{151} That is, we are free in the sense that we enjoy a \textit{liberty of indifference}. Only when reasons run out, or are incommensurable, or are of equal strength

\textsuperscript{149} Van Inwagen (1995): esp. 223-225. See also Wolf (1997): esp. 199-201. Wolf argues that, when an agent does something morally wrong, it must be because she did not have a reason to do something better (due to psychological determinism). Therefore, she cannot be blamed for her behavior. Agents can only be praised for doing good acts (since they responded to good reasons for acting), but cannot be blamed for doing bad acts. Hers is a hybrid of incompatibilism and compatibilism.

\textsuperscript{150} I call the claim that reasons for action must be causes of that action “Davidson’s Thesis” after Donald Davidson’s 1963 influential paper, “Actions, Reasons, and Causes.” I have also seen this view referred to as “the causal theory of action.”

\textsuperscript{151} Van Inwagen seems to be convinced that we are not in danger of succumbing to \textit{Buridan’s Ass Problems}. Buridan’s Ass examples illustrate the potential problem of rational agents being incapable of choosing when their options are equally good. John Buridan, a medieval philosopher, hypothesized that, if we can choose only on the basis of decisive reasons, an ass seated an equal distance from several equally appealing bales of hay would not be able to choose any bale to feast upon. The ass would starve to death. But such a scenario seems absurd; we presume the ass, like we ourselves, would simply pick a bale of hay at random. See van Inwagen’s discussion of these and similar cases at (1995): 233-234.
do we have free will. Van Inwagen surmises that the number of occasions upon which we are faced with such choices is few and far between. Thus, he concludes that “being able to do otherwise is a comparatively rare condition, even a very rare condition”\textsuperscript{152}—a claim that has led others to declare that we have precious little free will.\textsuperscript{153}

This consequence of van Inwagen’s view results from the fact that he is a leeway incompatibilist. That is, since van Inwagen holds that one must be able to do otherwise at the moment of acting and since we are only able to do otherwise in a handful of moments over the course of a lifetime, we are very rarely free. It seems that van Inwagen is more comfortable with this consequence (that we are almost never free and responsible) than he would be with including determined acts that have an indeterministic history. In other words, for van Inwagen, the agent must be the direct source of her behavior for it to count as free and responsible. For this reason, I consider van Inwagen’s requirements for free and responsible action to be indicative of the importance of sourcehood in his account; if one is not consciously exerting effort to choose between equally good reasons to act, something else is responsible for the agent’s behavior.

Some theorists feel that van Inwagen’s requirements are too strict. Surely there is some sense in which my free behavior today justifies praising me tomorrow for related but unfree behavior. While I cannot act other than I do tomorrow, I could have acted otherwise if I had acted otherwise today (which I was able to do). This reasoning depends on a strategy of tracing, or holding an agent responsible for behaviors and effects that result from some free action performed by the agent, even if proximate causes of the behavior or effect were not up to the agent. We saw in Chapter 2 that Aristotle’s view

\textsuperscript{152} Van Inwagen (1995): 223.  
\textsuperscript{153} Van Inwagen (1989): 405.
incorporates such assessment of agents based on earlier control. Robert Kane holds such a view—one that he calls *Ultimate Responsibility*.\(^{154}\)

**Robert Kane’s Theory of Ultimate Responsibility**

Robert Kane is an event-causal libertarian. He believes that the thesis of determinism is incompatible with free will and moral responsibility; however, he also believes that determinism is false and the kind of indeterminism that we have in our world allows us to have free will and be ultimately responsible for our actions. Kane sets out to give a positive account of free will in which free and responsible actions are event-cause in a probabilistic manner that secures an agent’s *ultimate responsibility* for subsequent behavior resulting from said free acts.

Kane argues that ultimate responsibility is what really concerns us when we worry about free will and moral responsibility.\(^{155}\) He claims that our intuitions about being free reflect a sense of control over our actions such that we could have done other than we did, as well as a sense that we are the sources of our actions. It is the second that he thinks is primary and the first is entailed by the second. To be ultimately responsible is to be the sole cause of any sufficient motive for a particular action or effect.\(^{156}\) Kane’s focus on our being the sole sufficient cause for our behavior qualifies him as a sourcehood theorist. As it turns out, this condition of ultimate responsibility can be satisfied even if


many of our actions for which we are responsible lack alternate possibilities at the time of action.

Kane argues that if we can trace an agent’s action back to a moment when the agent does determine some sufficient cause for the act in question, the agent is ultimately responsible for the present act. For example, if Gary hits a child while driving intoxicated and kills her, it might be argued that Gary had no alternative action available to him seconds before the accident since he was already in the car, it was moving in that direction, and his motor responses were compromised. However, we typically do not exonerate people like Gary simply because they could not do otherwise in such cases. Instead, we say that Gary should not have gotten so intoxicated in the first place, or he should not have driven and he should have secured a safe ride instead, or he should have surrendered his keys prior to drinking. We find him responsible for his present actions and their consequences because we consider these to be consequences of previous actions. And we presume that Gary had alternate possibilities open to him immediately prior to those past acts and at that time did not choose the correct route.

Such actions Kane terms *Self-forming Actions (SFAs).* 157 When one performs a SFA, one is torn between equally compelling motives. Kane’s famous example is of a woman (following Pereboom, let’s call her Ann) who witnesses a mugging while on the way to an important business meeting at work. The woman is torn between a desire and will to help the victim and a desire and will to impress her boss and be punctual. 158 According to Kane, if she desires and wills each equally, she must choose which kind of person to be. Depending on which one she does, her future motives and future self are

---


sure to be affected. If she assists the victim, she is more likely to have stronger charitable motives in the future. If she continues walking, she is likely to be less sensitive to the needs of others and more focused on personal gain. For this reason, these actions are considered *self-forming* since they contribute to the kind of person the agent later becomes and the degree of probability that the agent will act in certain ways in the future.

Kane’s view is very similar to van Inwagen’s view; however, he distinguishes his view from van Inwagen’s in two important ways. Like van Inwagen, Kane believes that psychological determinism holds when we are not torn. Therefore, the only contexts in which free will is possible are the same for both van Inwagen and Kane. However, unlike van Inwagen, Kane holds that we can be responsible for future unfree behavior if our SFA’s are sufficient causes for such behavior. Since SFA’s are free actions, and since they result in the agent becoming a certain kind of person, the agent is ultimately responsible for being the kind of person who, in the future, may act unfreely in certain morally salient ways.

Kane also distinguishes himself from van Inwagen in that he illustrates how the brain sorts out such decisions and defends this picture against potential objections of randomness. Kane explains how an agent is able to move forward from the position of being torn in such a way that, whichever action is chosen and performed by the agent is also an action the agent was trying to do.\(^{159}\) For instance, imagine that our driver, Gary, was in the following situation. While out for a night of drinking, Gary’s friend offers to

---

\(^{159}\) The reason why SFA’s qualify the agent as being the source of her evolving into such and such a type of person is that SFA’s result from what Kane calls, *will-setting actions* (WSAs) (Kane (1998): 113-115, 128; Fischer et al. (2007): 20). Kane’s discussion of will-setting focuses more on the neurological network and probabilistic causation, whereas his discussion of self-formation is meant for a higher order of things, such as a person’s aspirations of what type of person to become. SFW’s occur only under certain conditions. Like in the case of a SFA, for a willing to be a SFW the agent must be “torn” between alternative courses of action, and it is not determined which one she will do.
pay for a cab for Gary whenever Gary has had enough of drinking and decides to go home. When faced with this opportunity, Gary is actually torn. On the one hand, he thinks it is the moral and responsible thing to do to take his friend’s offer. On the other hand, he hates the idea of the trouble and inconvenience he will have the next day when he figures out how to get his car back (imagine that Gary is very busy the next day, must be somewhere early and would prefer to roll out of bed hung over the moment before the meeting rather than wake up any earlier than he would have to, etc.). This is a case of conflict between moral and prudential considerations. This is when Gary’s brain begins “parallel processing.”¹⁶¹ There is a neural network sorting out motives to do the one thing and another sorting out motives to do the other thing and as these two networks process, eventually one “wins out” and sets the will. Kane insists that cases of such will-setting are indeterminate—the agent’s will could be set either way.¹⁶²

Kane thinks that his account answers the worries of voluntary control and randomness (which he calls “luck”) since, on his view, agents have plural voluntary control. That is, whichever way the agent wills, when she wills she wills voluntarily (i.e., controlled), intentionally (on purpose) and rationally (for a reason).¹⁶³ This is true regardless of which motive “wins out” since in each case, Kane says, the agent was trying to do what the motives prescribed. It is the plural voluntary control that grounds an agent’s sourcehood, and having plural voluntary control entails having alternate

¹⁶⁰ Kane suggests that there are various types of situations that might cause an agent to be torn. This is one type. Other types include being faced with incommensurable values, conflicts between short term and long term interests, and “Buridan’s ass” cases where all options seem equal (though you cannot choose both), etc.


possibilities. Kane admits that while the agent does not control which of the two options “wins out”, when one of them does, the agent controls that action. That is, while there is an element of luck, it is in the outcome and not in the willing.

An agent is responsible for the outcome of this will-setting action because whichever option is willed is one she was trying to do. Kane argues that we find people responsible for things that they tried to do that involve an element of chance external to the trying all of the time. One of his examples is that of an angry husband who slams his fist on his wife’s beloved glass tabletop in a fit of rage.164 Kane asks us to imagine that the potential breaking of the glass is an indeterministic event (the glass’s disposition to break involves probabilities that under certain fist-pounding conditions it is likely (though not guaranteed) to break). Kane says that if the glass breaks and the husband excuses himself, “It’s not my fault that the table broke; chance broke the table!” we rightly consider this response “lame” (Kane’s term) and unsatisfactory. The fact that he was trying to break the table is enough for him to be responsible for its breaking since the chance is in the outcome of what he wills, not in his willing it. Therefore, Kane argues, his account meets the control objection (the agent is controlling what she does for reasons and could also control for other reasons doing something else) and the luck or randomness objection (since the “luck” that is involved is in an irrelevant part of the event-causal chain).165

164 The example is used in Fischer et al. (2007): 27-28. For a similar example and discussion, see Kane (1998): 55-56.

165 I think that Kane’s view is a noble attempt to meet the objections to incompatibilist event-causalists. But I think it fails. Kane admits that, in such instances, the agent is trying to do multiple incompatible acts, and it is the trying that makes us responsible. If this is true, then isn’t the agent responsible for A-ing and also for not-A-ing? If the answer is “yes”, that is absurd. How can one be responsible for an action and its contrary/omission simultaneously? If we’re responsible for both, then everyone has a moral status that is exactly balanced between positive and negative alternatives. Everyone is simultaneously praiseworthy for every good thing she might have willed as a WSA and SFA and blameworthy for every bad thing she might
Carl Ginet has developed a noncausal account of free action that involves a conscious denial of Davidson’s thesis and an awareness of possible challenges such a move might bring.\(^{166}\) On Ginet’s account, an agent’s act is free if it is undetermined, has an *actish phenomenal quality*, and is done for a reason.\(^{167}\) The fact that an act is undetermined (i.e., completely spontaneous with no antecedent cause) secures sourcehood for the agent. However, if the act is not caused, it is also not caused by a reason. Ginet, then, must explain how the uncaused act can be done *for* a reason without being caused.

On Ginet’s view, when an agent acts freely, a volition is noncausally formed. Such a volition brings with it an *actish phenomenal quality*, which is the seeming to the agent that she is directly causing her act on purpose.\(^{168}\) According to Ginet, the actish phenomenal quality affords the agent both ownership and control of the action. An agent owns her actions for two reasons: 1) they originate in her brain and 2) the actish...
phenomenal quality includes a sense of ownership. In addition, it yields a sensation of control. That is, it seems as if I am *intentionally* raising my hand.

In addition, this view rules out the possibility of reasons causing an agent to act freely. But, Ginet does think that free actions are done for reasons. Therefore, he argues, there is a way to act for a reason that does not involve being caused by it. A reason can be part of a reasons-explanation if the agent:

1) Has the reason

2) Recalls it when forming an intention to \( \phi \)

3) Intends that her action satisfy that desire/reason

In other words, Ginet holds that one acts for a reason when one has a desire to do the act and recalls this desire (i.e., it comes to mind) *simultaneously* with the formation of the uncaused volition.\(^{169}\)

Views such as Ginet’s definitely establish sourcehood as a condition for free action and, therefore, moral responsibility. If there is some point in the causal chain of an action in which *no* cause was required *and* this point in the causal chain exists within the agent’s mental makeup, then the agent is definitely the *starting point* of all sufficient causes of the agent’s behavior. However, Ginet’s account is, in turn, susceptible to concerns about control. It would seem that what establishes free actions is spontaneous and, therefore, random. If this is the case, it is difficult to see how such mental events can be said to have been controlled by the agent. They might seem merely *related* to the agent, or, as we worried earlier about other views, a matter of locating the spontaneous

\(^{169}\) Ginet (1990): especially chapters 1, 4, and 6.
event within the agent. Worries about control in this sense have inspired agent-causal accounts of free and responsible action.

Agent Causal Accounts—Chisholm and O’Connor

Objections to noncausal accounts of free action make apparent our intuitive desire for an account that incorporates a certain amount of causal control on the part of the agent. As I mentioned when discussing my concerns with Kane’s account, sourcehood theorists typically seek a description of the agent as a source that is much more intentional and controlled than that of a mere location of spontaneous events. In the following sections, we shall examine two accounts that attempt to describe such intentional control.

Roderick Chisholm’s agent-causalist view is a good example of the standard basic view of agent causation. In order for Chisholm to defend a libertarian view that maintains control in the face of indeterminism, he suggests that we are free, yet we are “unmoved movers.” First, he claims that, if a man is responsible for something he does, it must be because it was up to him. But, if it is up to someone to φ, it must also be possible (i.e., in her power) to not φ. Finally, Chisholm concludes that these two claims entail that, in order for one to φ freely, φ-ing must not have been caused by anything that was not also in the agent’s control: “it could not have been caused or determined by any event that was not itself within his power either to bring about or not to bring about.”170 Therefore, if reasons are, as I suggested above, something over which we have little or no control, then they cannot be causes of free actions on Chisholm’s account.

Chisholm holds that free action stems from *agent causation*, which is the causation of an event or state of affairs directly by an agent. Chisholm (1997) calls event causation *transuent causation* and agent causation *immanent causation* (148).

That is to say, no external or antecedent factors caused the agent to act as she does. The agent is considered a metaphysical substance with the power to spontaneously cause intentions that lead to actions. Agent causation is distinct from Ginet’s view of uncaused intention formation because the intention, according to an agent causal account, is an effect of a prior cause—the agent. However, that cause does not trace backwards in time beyond the agent, as is the case with event causation. Essentially, the buck stops with the agent.

The fact that Chisholm’s view includes a requirement that an agent be able to control anything that causes her behavior is indicative of the importance such views place on sourcehood. For this agent causalist, the agent is the source of the behavior as constituting a substance that is both necessary and sufficient in moving itself. Finally, like all of the views we have considered in this chapter, it is essential to Chisholm’s view that the agent cause have alternative possibilities. The agent cause has more robust possibilities than mere conditional freedom, for according to Chisholm, it is not enough that the agent would have done otherwise had she chosen to do otherwise; Chisholm insists that freedom requires that *she could choose* to do otherwise even if we do not alter antecedent conditions. And Chisholm establishes this ability to do otherwise regardless of the past and present influences by linking it to this unique type of causation—*agent causation*.

---

171 Chisholm (1997) calls event causation *transuent causation* and agent causation *immanent causation* (148).


More recently, Timothy O’Connor advances an agent causal account that attempts to describe the process of intention formation as well as how agents might freely act for reasons. Unlike Chisholm, O’Connor does not hold that agents are unmoved movers; instead, he claims that they are “not whollyMOVED movers.”¹⁷⁴ It is the weakening of this idea that allows him to develop a sophisticated account of acting on reasons as an agent cause. On O’Connor’s account, when agents act freely they directly cause immediately executive intentions¹⁷⁵ and these intentions are of a sort to act here and now for this reason.¹⁷⁶ The very idea of acting for a reason is then built into the idea of intending. The immediately executive decision is the beginning of an event-causal chain, the end of which is the observable physical action of the agent.

While O’Connor considers reasons-responsiveness in his view, it remains a sourcehood view since, when an agent acts freely, the buck stops with her. That is to say, it is an important part of O’Connor’s view that there are no free acts that have sufficient causes external to the agent performing the act. Regardless of the presence of influences, a free agent causes herself to act.¹⁷⁷ Therefore, sourcehood is the primary concern for O’Connor. Like other source incompatibilists, it is essential that the agent be the source of her intention and that the production of her intention be free in that she could have produced alternative intentions and that the agent be able to do this without a requirement that any of her motives or desires be different from what they are in the actual scenario.

¹⁷⁶ O’Connor (2000): 86.
While O’Connor does insist that free and responsible action is such that an agent causes her immediate executive intentions, he does not consider such a genesis to happen independently of motive. He writes, “an agent causally initiates his action and … agent causation is conceptually tied to the agent’s having a reason for acting. (We might say that agent causation is a triadic relation. Agents cause intentions in virtue of reasons; there can be no such thing as agents causing intentions without any motivation at all.”\(^{178}\)

According to O’Connor, when an agent acts freely:

1) She had a desire that \(\psi\) prior to acting.

2) Her action was initiated in part by her own self-determining causal activity (which had an event component of a coming-to-be-of-an-action-triggering-intention-to-so-act-here-and-now-to-satisfy-\(\psi\))

3) She continues to desire that \(\psi\) and intends of this action that it satisfy (or contribute to satisfying) that desire.

4) Her intention concurrent with the action is directly caused by (or is a continuation of) the action-triggering intention and is sustained throughout the completion of the action.\(^{179}\)

On O’Connor’s view, while agents are not caused by reasons (i.e., reasons are not a sufficient cause of an agent’s action), agents act for reasons. Reasons are necessary but not sufficient conditions for action.\(^{180}\) When an agent acts for a reason, the reason


\(^{179}\) O’Connor (2000): 86.

\(^{180}\) Since, for O’Connor, reasons are not sufficient, there is no contrastive explanation (i.e., no way to explain why the agent A-ed rather than B-ed). O’Connor denies that there can be “negative events” and therefore feels that there is no need to provide such an explanation. If reasons are so involved in the fruition of free action, they can serve in noncontrastive explanations of the action.
structures the action. This is why reasons can be explanatory. So reasons can explain action even if they do not directly produce it; the agent cause directly produces the action, and while this agent cause requires a reason, the reason does not produce the act. According to O’Connor, agent causation is “conceptually tied to” the agent’s having a reason for acting since “agents cause intentions in virtue of reasons” and “there can be no such thing as agents causing intentions without any motivation at all.”\textsuperscript{181}

O’Connor has constructed a sophisticated account that seems to focus on sourcehood without overlooking reasons-responsiveness. O’Connor writes of his account that “It allows the reason to influence the agent’s producing the outcome, while not (directly and independently) causing it.”\textsuperscript{182}

\textbf{Rejecting Sourcehood}

While I find O’Connor’s account the most appealing of sourcehood incompatibilist views, I do think that it places the cart before the horse regarding moral responsibility and praise- and blameworthiness. That is to say, I do not think that any privileged type of causation or any type of alternative possibilities is necessary to ground praise- and blameworthiness. In essence, O’Connor (like all sourcehood theorists) has explained far more than is necessary to ground praise- and blameworthiness and moral responsibility and, in so doing, has left these issues behind.

In order to demonstrate this point, let us now turn to reasons-responsiveness views. In the next chapter, we shall see how many of these views seem to offer a

\textsuperscript{181} O’Connor (2000): 88.
\textsuperscript{182} O’Connor (2000): 89.
satisfactory explanation of desert of praise or blame while keeping all conversation focused on reasons and rationality. In so doing, we shall see why I think that alternate possibilities are not necessary for praise- and blameworthiness. While I do think that these views are a step in the right direction, they too have a central flaw that renders them imperfect: contemporary reasons-responsiveness theories consider actions or intentions the only appropriate targets of ascriptions of moral responsibility, praise and blame, whereas I do not think that this is the case.
In the last chapter, we saw how sourcehood theorists aim to develop an account of the right kind of source required for praise- and blameworthy behavior. We also saw that sourcehood theories have been subject to objections concerning their compatibility with basic claims about the proximity of causation, the possibility for manipulation to qualify as free, and worries about rationality, where rationality means an agent’s ability to act for reasons in such a way that citing those reasons explains the behavior. Most inclusively though, we saw how every sourcehood account, regardless of what objections pertained to it, considered some notion of alternate possibilities essential to sourcehood. Reasons-responsiveness theorists, on the other hand, do not consider alternate possibilities to necessary for free and responsible action. Therefore, our first task is to consider why reasons-responsiveness theorists differ in this regard.

Alternate Possibilities Reconsidered

Reasons-responsive theories have become more popular since the publication of Harry Frankfurt’s 1969 paper, “Alternate Possibilities and Moral Responsibility.” In this paper,
Frankfurt challenges what he calls the *principle of alternate possibilities* (PAP). This principle holds that an agent can be morally responsible only if that agent could do other than she, in fact, does. The principle seems intuitive for a multitude of reasons. For example, if a person being blamed defends herself by declaring, “I could not help it” or “there was nothing else that I could do,” we often accept such a defense as mitigating or justifying. Indeed, Frankfurt too finds such appeals compelling, but warns us that the principle, as it is generally understood, is too broad. He shows this by offering counterexamples to the general claim.

Frankfurt constructs scenarios where an agent cannot do other than she does, and we have the intuition that she is still morally responsible. For example, imagine that Mark is considering stealing a loaf of bread. Imagine that the owner of the bakery, Bernie, has a rival, Wally the wrap-maker, who wants Mark to steal the loaf of bread from his competitor. Also imagine that, in his spare time, Wally has learned how to manipulate a person’s mind via radio waves and is able to do so with precision. Finally, imagine that Wally is an excellent judge of other people’s thoughts and can tell the second that Mark starts to decide not to steal the bread. If this happens, Wally will manipulate Mark’s brain so that he in fact steals the bread from Bernie’s Bakery. But Wally only wishes to use his talent if he has to. So if he detects no sign that Mark will refrain from the theft, Wally will leave Mark be. If Mark is about to change his mind, he will do something that Wally...

---

183 This example is my own. Frankfurt uses an example where Jones is about to perform an act that Black wants him to perform, and Black has the means and is willing to force Jones to do just that if he hesitates. But Black prefers Jones do the act on his own, if possible. This article has generated an enormous amount of literature—so much, in fact, that the number of similar examples in articles and book chapters in response to the 1969 paper that a class of philosophy examples has been named, *Frankfurt Style Cases*. In the wake of modern neuroscience, many of the examples involve a nefarious doctor who has *somehow* implanted a mechanism inside the agent’s brain that can remain inactive, but if activated, would override the agent’s capacity for choice and/or action.
is able to detect and, as a result, Wally will interfere with Mark’s brain. In such a case, Mark cannot do otherwise. If he were to consider doing otherwise, he would be manipulated by Wally to steal Bernie’s bread. Now imagine that without any hesitation, Mark steals the bread. He had no beginnings of choosing to do otherwise, so Wally does nothing. It seems as though Mark is responsible and blameworthy for the theft. Yet Mark could not have done otherwise. It seems as though the ability to do otherwise is not that important after all.

In this same article, Frankfurt admits that had a person like Mark tried to refrain and had Wally interfered, Mark would not be morally responsible. However, Frankfurt maintains that:

…If someone had no alternative to performing a certain action but did not perform it because he was unable to do otherwise, then he would have performed exactly the same action even if he could have done otherwise. The circumstances that made it impossible for him to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way.184

Frankfurt adds:

The fact that he could not have done otherwise clearly provides no basis for supposing that he might have done otherwise if he had been able to do so…this, then, is why the principle of alternate possibilities is mistaken. It asserts that a person bears no moral responsibility—that is, he is to be

excused—for having performed an action if there were circumstances that made it impossible for him to avoid performing it.\textsuperscript{185}

While he does reject the traditional principle of alternate possibilities, Frankfurt does agree with many of the verdicts handed down by the principle. For example, he writes, “We often do, to be sure, excuse people for what they have done when they tell us (and we believe them) that they could not have done otherwise. But this is because we assume that what they tell us serves to explain why they did what they did.”\textsuperscript{186} Frankfurt agrees that it is possible that an inability to do otherwise excuses poor behavior. However, he does not think that an inability to do otherwise always excuses poor behavior. For Frankfurt, if the fact that an agent could not do otherwise is the reason that agent did what he did, this reason might shed light on their reasoning for the behavior in such a way that an excuse is appropriate. For example, Ron might tickle his friend, Tony, until Tony is in great pain. Normally, we would find this behavior blameworthy. But if Ron tickled Tony because a third party, Dr. Nefarious, threatened to shoot them both if Ron did not tickle Tony, we are inclined to forgive Ron. Because of examples like this one, Frankfurt offers a revised PAP where we hold that one is not responsible if one could not do otherwise \textit{and} the reason that she does what she does \textit{is} the fact that he cannot do otherwise.

Frankfurt’s article supports reasons-responsive theories because, in efforts to refute PAP, Frankfurt has also shown that the reasons one behaves a certain way are crucial to one’s praise- or blameworthiness for one’s behavior. That is to say, Frankfurt cases seem to positively single out an agent’s reason for action as morally most relevant.

\textsuperscript{185} ibid.

\textsuperscript{186} Pereboom (1997): 164-5.
while apparently rendering irrelevant the question of whether the agent’s causal role is privileged in any particular way.

The controversy over alternate possibilities remains. Source incompatibilists continue to argue against Frankfurt-style examples. For instance, some philosophers believe that something suspicious lurks behind the thought experiment, a so-called “flicker of freedom,” that accounts for our intuitions. However, reasons-responsive theorists find Frankfurt’s argument compelling. Although this controversy exists, I will simply presuppose, for the purposes of this dissertation, that the defenders of Frankfurt’s position are correct. The purpose of this dissertation is not to defend Frankfurt, but to take inspiration from Aristotle in articulating both a new way of thinking about theories of responsibility, praise- and blameworthiness themselves. Because the theory that I wish to develop is in the reasons-responsiveness camp, I wish to articulate my reasons for rejecting the views of the sourcehood theorists, as I have just done. But in this dissertation I will only be defending my Aristotelian-inspired theory against its closest competitors, and those are other reasons-responsiveness theories. Accordingly, I now turn to describing these views.

187 Pereboom (1997): 165. This revision of PAP opens the door for libertarian retorts. For all a libertarian has to argue is that determinism is just like Wally using radio waves consistently throughout everyone’s life. While Frankfurt tries to remain agnostic about such claims, his contemporaries take his findings in many different ways. For example, libertarians like Robert Kane and Timothy O’Connor argue that the only reason we blame Mark is because we think he had the power to try to do otherwise. Explanations like these have prompted more elaborate Frankfurt-style cases preventing even the attempt to do other than the agent does. Be that as it may, libertarians maintain that Mark had it in his power to be the one who chooses his action or to do something that would cause the mechanism to force him to do the action. This alone, they argue, implies an ability to do otherwise. Even though Mark would not ostensibly do otherwise since he would still steal the bread, in one scenario he determines himself to steal the bread and in another, he does something that prompts the mechanism to take over (and cause him to steal the bread). Controversies such as these have generated a debate about the possible flicker of freedom lingering in such examples. Some, like John Martin Fischer, complain that any such flicker is not robust enough to hang freedom and responsibility on. The debate is far from settled.
I have chosen what I take to be representative theories of major recent movements. First, we shall look at Frankfurt’s own positive explanation for free action as a representative of *hierarchical models*. Then, we shall consider a critic of Frankfurt’s, Gary Watson, who advances what I shall call a *value-oriented theory*. From here, we shall move to Susan Wolf’s asymmetrical view as an example of *sane self theories*. Then we shall see how the division proposed does not encompass all theories, and how some theories, such as that of Fischer and Ravizza, seem to be trying to cobble both source and reasons-responsive aspects together. Finally, we shall consider Nomy Arpaly’s theory of praise- and blameworthiness based on an agent’s quality of will. While some objections shall be discussed that are specific to individual theories, I shall pay close attention to the fact that every theory focuses on an agent’s praise- or blameworthiness for actions or willings. Ultimately, I shall explain why I prefer Arpaly’s views to all others for its characterization of how agents are sensitive to moral reasons. Still, I shall not endorse any of these views since, as I shall argue in subsequent chapters, I do not hold that all praise- and blameworthy behavior is active behavior.

**Frankfurt’s Hierarchical Model**

Harry Frankfurt constructs a theory of free action based on his view of what it is to be a person. Presumably, this strategy follows because, for Frankfurt, one must first be a person to be able to act freely. In order for one to be a person, says Frankfurt, one must have what he calls *second-order volitions*. As it turns out, we act freely when our highest-order volition is satisfied.
Before this can be made clear, we must first consider how desires are ordered according to Frankfurt. On Frankfurt’s view, all humans and many animals have desires. But not all desires are of the same type. For example, all desires that are desires to perform an action are what Frankfurt calls “‘first-order desires’ or ‘desires of the first order,’ which are simply desires to do or not to do one thing or another”.188 These desires are the most basic type in the hierarchy relevant for action.189 All creatures that move on their own do so on the basis of such desires. However, more sophisticated creatures also have what Frankfurt calls desires of the second-order. A desire of the second order has a first-order desire as its object. For example, since I know that vegetables are good for my health, I might desire that I desire to eat vegetables. Another example might be a desire to know what experiencing some desire is like. For instance, I might listen to my neighbor talk about how much he wants to be an Olympic athlete, and I might have a desire to know what it feels like to want to do something so much that I would train 18 hours a day to achieve it. The first example of a second-order desire (that of a desire to desire to eat vegetables) is an example of what Frankfurt calls a second-order volition.190 A second-order volition is a second-order desire that a first-order desire be acted upon. That is, not only do I want to want more vegetables, but I also want the desire for vegetables to be effective in motivating me to actually eat vegetables. That is, I want my desire to eat vegetables to become my will. As Frankfurt tells us, “Whichever of these first-order desires

189 Frankfurt (1988c): 13-14. To my knowledge, Frankfurt does not entertain the possibility that there are desires simply for objects. For example, he says nothing of a desire for ice cream. I suppose that if I were to tell him that I had a desire for ice cream, he would say it amounted to a first-order desire to eat ice cream. That is to say, perhaps all desires for objects are truly desires for actions that involve those objects, and can be re-described as such. However, I imagine that one could desire that someone love them the way that they are, which does not seem translatable on this schema. For our purposes, let us simply note that Frankfurt is considering the desires to act or omit action to the exclusion of any other possible desire types.
desires actually leads to action is, by virtue of that effectiveness, designated the will of the individual whose desire it is.”

Therefore, a second-order volition is a second-order desire for the success of a first order desire.

Persons have the potential to act freely, which is different from an ability to exercise freedom of the will. Frankfurt tells us “freedom of action is (roughly, at least) the freedom to do what one wants to do.” However, “a person enjoys freedom of the will… [If] he is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or to have the will he wants.” In other words, a person exercises freedom of the will by “securing the conformity of his will to his second-order volitions.” Therefore, it is possible that an agent act freely but, at the same time, not enjoy freedom of the will.

Acting with freedom of the will is not required for moral responsibility. For example, Frankfurt states, “It is not true that a person is morally responsible for what he has done only if his will was free when he did it. He may be morally responsible for having done it even though his will was not free at all.”

---

191 Frankfurt (1988g): 164. See Frankfurt (1988c): 14 for a distinction between will and intention; a will is distinct from an intention since a will is what one actually tries to do whereas an intention is what one identifies as what one most wants even though the strength of a competing desire might be greater, thus overpowering the intention.


193 Ibid.

194 Ibid.

195 It is worth noting that Frankfurt seems to require alternate possibilities for one’s will to be free. However, Frankfurt is not a sourcehood theorist, since a free will is not required for moral responsibility.


197 In his discussion of how a willing addict acts freely but without a free will, Frankfurt states, “Given that it is therefore not only because of his addiction that his desire for the drug is effective, he may be morally responsible for taking the drug.” (1988c, 25) In other words, the fact that he acts freely might be enough to ground his moral responsibility.
Harry Frankfurt is not a sourcehood theorist. First, as we have just seen, he does
not consider alternate possibilities relevant to moral responsibility; yet sourcehood
theorists consider alternate possibilities essential to moral responsibility. Without
alternate possibilities, there is no way to say that an agent is the distinguished causal
origin of an action, and this is what the sourcehood theorist wishes to say. In addition,
Frankfurt has made comments ruling out sourcehood as a relevant criterion for moral
responsibility:

Suppose first that a man comes to a fork in the road, that someone on the
hillside adjoining the left-hand fork threatens to start an avalanche which
will crush him if he goes that way, and that the man takes the fork to the
right in order to satisfy a commanding desire to preserve his own life.
Next, suppose that when the man comes to the fork he finds no one issuing
threats but instead notices that on account of the natural condition of
things he will be crushed by an avalanche if he takes the left-hand fork,
and that he is moved irresistibly by his desire to live to proceed by the fork
to the right…Whether he is morally responsible for his decision or action
in each case depends not on the source of the injury he is motivated to
avoid but on the way in which his desire to avoid it operates within him.\textsuperscript{198}

It is clear that the source of our motives is not at issue for Frankfurt. Yet, it is also clear
that the nature and operation of our desires is relevant. Reasons-responsive theorists
typically consider the operation of desires when contemplating how one is able to
respond to reasons.

\textsuperscript{198} Frankfurt (1988h): 45.
Frankfurt is also unconcerned with an agent’s history. His original view says nothing about history, nor does he make any attempts to accommodate history in his later refinements of his view. These facts suggest that Frankfurt is not concerned that we be the appropriate source of our own beliefs and desires. So long as we are agents acting on our attitudes, we are in the realm of potential praise- and blameworthiness, according to Frankfurt.

In addition to a lack of concern for traditional sourcehood requirements, Frankfurt also makes positive claims that suggest that his view is a reasons-responsiveness account. For example, he writes, “There is naturally an intimate connection between what a person cares about and what he will, generally or under certain conditions, think it best for him to do.” What a person cares about is deeply rooted in the person’s character and, oftentimes, what is wholeheartedly wanted. Thinking something is best for one to do is to hold that there is a reason for one to do the action. If we act with responsibility when we have the will that we want, as Frankfurt has suggested, then we act responsibly when we are doing what we have reason to do. In short, Frankfurt’s dismissal of the need for

---


200 In his discussion of “the unthinkable,” Frankfurt mentions that a person might be incapable of willing herself to do a certain action if it conflicts with what she cares about (e.g., she cannot will to protect her finger if doing so would entail the destruction of the world, and she cares about many things related to the world remaining intact); but this is not a lack of freedom or an exoneration of responsibility. On the contrary, it might be a manifestation of her deeply rooted rationality ((1988i): 190).

201 It seems that caring is related to wholeheartedness and identification since Frankfurt writes, “A person who cares about something is, as it were, invested in it. He identifies himself with what he cares about.” (83). However, it is not the case that what we care about is identical with wanting something. Frankfurt points out on page 83 that, “A person might stop caring about something because he knew he could not have it. But he might nonetheless continue to like it and to want it.” His reasoning behind this claim is that we care about things that are present and future-oriented. If it is not possible for something to happen, we cannot care about it. However, we can desire things that are impossible (e.g., the desire to wish one had never been born). Therefore, it seems as though if we care about something, then we have desires related to it and reasons for actions, but it does not follow that if we desire something we care about it.
alternate possibilities, along with his focus on wholeheartedness, caring, and taking something as worth doing lead me to consider him a reasons-responsiveness theorist.

Finally, like other reasons-responsiveness theorists, Frankfurt is concerned with agents being responsible for actions and their outcomes and not other types of behavior. For example, in “The Importance of What We Care About,” Frankfurt states, “The outlook of a person who cares about something is inherently prospective; that is, he necessarily considers himself as having a future.” He also adds, “A person might stop caring about something because he knew he could not have it.” If caring is future-oriented and it dissipates when an agent does not consider what he cares about achievable, this suggests a connection to action. Much like how we saw Aristotle claim that we only deliberate about what we think is possible, if we only care about what we think is possible, it is likely because it is inherent in the nature of deliberation and the nature of caring that we eventually do something. Finally, Frankfurt suggests that:

Since the making of a decision requires only a moment, the fact that a person decides to care about something cannot be tantamount to his caring about it. Nor is it a guarantee that he will care about it. By making such a decision, the person forms an intention concerning what to care about. But whether that intention is truly fulfilled is quite another matter. A decision to care no more entails caring that a decision to give up smoking entails giving it up. In neither case does making the decision amount even to initiating the state of affairs decided upon unless that state of affairs actually ensues.

203 ibid.
This would hardly be worth pointing out except that an exaggerated significance is sometimes ascribed to decisions, as well as to choices and to other similar “acts of will.” If we consider that a person’s will is that by which he moves himself, then what he cares about is far more germane to the character of his will than the decisions or choices he makes. The latter may pertain to what he intends to be his will, but not necessarily to what his will truly is.204

This passage makes it clear that Frankfurt thinks that overt behavior is a natural consequence of caring and responding to the reasons that we have in virtue of what we care about. More interestingly for our purposes, Frankfurt, like the other reasons-responsive theorists presented in this chapter, ascribes praise and blame to actions. That is to say, while he ascribes an importance to what we care about, his view of our cares and concerns as solely related to those choices and decisions that we do act out renders actions (and, perhaps, attempts to act) the only appropriate objects of praise and blame for Frankfurt.

Valuing—Watson’s View

Gary Watson has objected to Frankfurt’s view on the grounds that it allows for behavior related to desires and reasons of any sort to qualify as responsible behavior. Watson contends that only desires reflective of values should qualify, as Watson considers values to be the key to free agency.

204 Frankfurt (1988e): 84.
All types of desires are distinct from values, which Watson describes as “more or less long-term aims and normative principles that we are willing to defend.”\textsuperscript{205} While it is true that valuing always entails wanting, what makes values distinct is that they come from our reason and judgment, whereas other desires might be purely appetitive.\textsuperscript{206}

According to Watson, the \textit{valuational system} “of an agent is that set of considerations which, when combined with his factual beliefs (and probability estimates), yields judgments of the form: the thing for me to do in these circumstances, all things considered, is \textit{a},”\textsuperscript{207} and beings that make judgments of this sort have \textit{free agency}. Such beings “assign values to alternative states of affairs, that is, rank them in terms of worth.”\textsuperscript{208} \textit{The motivational system}, in contrast, “is that set of considerations which move him to action,” and “the possibility of unfree action consists in the fact that an agent’s valuational system and motivational system may not completely coincide.”\textsuperscript{209}

Watson worries that sometimes an agent is unable to get what she values “and this inability is due to his own ‘motivational system’.”\textsuperscript{210} For example, agents who suffer from compulsive disorders might do what their \textit{strongest} desire dictates, and this is a matter of their motivational systems, but they might have no rational control whatsoever over their behavior. In such instances, “the obstruction to the action that he most wants to do is his own will.”\textsuperscript{211} These struggles happen, says Watson, because what one \textit{values} might not be what one is most moved to do, since the extent to which one values

\textsuperscript{205} Watson (1975): 25.
\textsuperscript{207} Watson (1975): 25.
\textsuperscript{208} Ibid.
\textsuperscript{209} Ibid.
\textsuperscript{210} Ibid.
\textsuperscript{211} Ibid.
something is not determined by the strength of the person’s desires. That is to say, the rank of something insofar as it is valued might not map on precisely to that object’s rank according to desire strength. In fact, we might have desires so foreign to our values that there is no correspondence whatsoever. Watson’s own example is of a frustrated parent who, in an instant, feels an urge to drown her child.212 The parent might have a desire to do such a horrendous act without having any corresponding value of baby drowning.

Watson’s view of the difference in kind between desires and values and how only values can ground reasons to act makes his view clearly an example of a reasons-responsiveness account. His notion of free action depends on the connection of an agent’s values and behavior insofar as those values provide reasons for such behavior. In other words, if the agent acts in response to reasons connected with her values, she acts freely. While Watson does not explicitly declare what he thinks about responsibility, he seems to implicitly accept the idea that free action is required for moral responsibility. If this is correct, then, for Watson, moral responsibility is substantially related to an agent’s ability to respond to reasons.213 Furthermore, Watson is not concerned with the alternate possibilities or the history behind the agent’s attitudes—two standard characteristics of sourcehood views. These aspects of his view distinguish him from sourcehood theorists in a fundamental way.

213 It might be objected that Watson’s discussion of the source of our desires makes him a sourcehood theorist. For Watson, our reason generates and identifies values, whereas our appetite is the source of our desires. That is, the distinction between values and desires is not “a distinction among desires or wants according to their content,” but rather the distinction “has…to do with the source of the want or with its role in the total “system” of the agent’s desires and ends. It has to do with why the agent wants what he does,” ((1975): 20-21) and oftentimes the reason an agent wants something can be found in the source of the desire. Be that as it may, I do not think this feature of Watson’s view qualifies him as a sourcehood theorist. Watson is concerned with whether the agent’s reason or appetite is the source of her behavior and not whether the agent as a whole has a privileged causal status. He is not concerned with alternate possibilities. And, finally, his concern with the source of the desires that move an agent is subordinate to the fact that rational desires make one free because only they can provide reasons.
In her book, *Freedom Within Reason*, Susan Wolf addresses Frankfurt and Gary Watson’s views and other views like them. Wolf considers Watson’s view to be an example of what she terms *Real Self Views*. On such views, actions are free and responsible if they are governed by the desires of the true or real self. Watson’s view is a paradigmatic example.

While Wolf appreciates the spirit of Watson’s view, she finds much at fault with it. Wolf wonders why actions being reflective of the real self would be sufficient for moral responsibility and an agent’s worthiness of praise or blame. For instance, we can imagine an agent, Edgar, who suffers from paranoid schizophrenia. It is of the utmost importance to Edgar that he keep all of his information private. One day, while in a coffee shop, Edgar “senses” that a nearby patron of the shop is reading his mind. In efforts to deter this evil intruder of Edgar’s thoughts, and as a seemingly justified defense of himself, Edgar purchases steaming hot coffee and throws it into the patron’s face, screaming “Stop reading my mind!” Surely, all of Edgar’s behavior is reflective of his true values. However, Wolf is reluctant to find Edgar to blame for this behavior. The fact that his behavior reflects his values means nothing if his values are so warped and deranged that Edgar could not have behaved well. In essence, Wolf contends that the

---

215 Wolf also claims that perverse individuals are not subject to moral censure if their perverse real selves are caused by horrific childhoods (1990): 37-38). This is another interesting point of contention between Wolf and Watson, as Watson famously discusses the case of a terrible murderer, Robert Alton Harris, concluding that being truly evil is blameworthy, regardless of what made a person that way (see Watson (1988)).

136
problem with views like Watson’s is that they focus on a person’s actual values rather than the person’s relationship to what is, in fact, objectively valuable. After all, we do not praise and blame people based on their personal psychological valuation system; that is, we do not praise serial killers when they kill efficiently because it is the standard their personal systems suggest. On the contrary, we compare them to the general common sense set of values we all try to live by, and insist that others also live by. Indeed, Wolf thinks that the entire practice of holding others morally responsible depends on the existence of a universal standard: “We cannot hold persons morally responsible if we reject the objectivity of value.”

Wolf’s likely assessment of Edgar would not depend entirely on the observation that Edgar is insane. In fact, the view that Wolf espouses has higher standards for an agent to qualify as blameworthy than it does for an agent to qualify as praiseworthy. This feature of her view is called the asymmetry of freedom. For Wolf, the majority of responsibility theorists make the mistake of considering the condition of freedom, the requirement that agents be free and have their actions under their control, ontologically prior to and necessary for the condition of value, or, the requirement that one be a moral agent for praise and blame of her actions to be legitimate. In fact, Wolf holds that the reverse is the case: one cannot be a free agent if one is not capable of being a moral agent. What is salient about cases like Edgar’s is that he is not able to see the true and good reasons to refrain from his behavior. If he is unable to see such reasons, he cannot react to them. This lack of ability exempts Edgar from deserving blame.

In order to illustrate why she holds this claim, Wolf first discusses the importance of *psychological determinism*. According to Wolf, an agent’s acts are psychologically determined if they are determined by her interests, and if those are determined by her genes and her environment.²¹⁸ Wolf is sympathetic to a strong interpretation of Davidson’s thesis, arguing, like Ayer did before her, that if our actions are not deterministically caused by our interests, then either they are not caused by us, or they are caused by us and are superficial or crazy, since they amount to mere whims.²¹⁹

According to Wolf, the likely fact of our psychological determinism does not threaten our agency, as incompatibilists might claim, nor does it guarantee it, like Ayer supposed. Wolf finds a compromise between these two kinds of views. She claims that, when we consider someone to be praiseworthy, we do not require that the person could have done otherwise. She explains:

> For if phrases like ‘I couldn’t help it,’ ‘he had no choice,’ ‘she couldn’t resist’ typically count as excuses intended to exempt the agent from blame for an action, they do not ordinarily serve as grounds for withholding praise. ‘I cannot tell a lie,’ ‘he couldn’t hurt a fly’ are not exemptions from praiseworthiness but testimonies to it. If a friend presents you with a gift and says she “couldn’t resist,” this suggests the strength of her friendship and not the weakness of her will.²²⁰

---

²¹⁸ Ibid.
Indeed, Wolf declares that “an action is praiseworthy only if it is done for the right reasons,” and, as we have seen already, on her account, acting for reasons in a sane way requires psychological determinism. So, in the case of praiseworthy action, an agent need not be able to behave differently and, most likely, needs to be psychologically determined. Indeed, this finding is foundational to the main claim of Wolf’s own position, the Reason View. According to the reason view, “responsibility depends on the ability to act in accordance with the True and the Good. If one is psychologically determined to do the right thing for the right reasons, this is compatible with having the requisite ability.”

On the other hand, the same result is not true of blameworthy actions. Wolf argues that we do tend to withhold blame when we find that an agent could not do other than she does. This is not to say that we are concerned with metaphysical or physical possibilities, per se. It is not as though we require that the agent be able to do any old alternative action. On the reason view, an agent is free insofar as she is able to do what her reason judges to be good and how accurate her reason is in detecting the true and the good. We want some assurance that the agent knew better and could have done a better act than the one for which the agent is about to be blamed. But one who is psychologically determined to behave badly seems impaired on this view:

For if one has to do the wrong thing, then one cannot do the right, and so one lacks the ability to act in accordance with the True and the Good. The Reason View is thus committed to the curious claim that being psychologically determined to perform good actions is compatible with

\footnote{221 Wolf (1997): 204.}
\footnote{222 Wolf (1990): 79.}
deserving praise for them, but that being psychologically determined to perform bad actions is not compatible with deserving blame.\(^{223}\)

At first glance, this passage seems to suggest that Wolf is concerned with alternate possibilities in the same way that sourcehood theorists are. However, a more careful examination shows that her primary concern is with the agent’s ability to respond to the True and the Good, which is a matter of reasons-responsiveness. This can be seen when we focus on Wolf’s asymmetrical claim. Of this asymmetry, she writes:

> There is an asymmetry in our intuitions about freedom which has generally been overlooked. As a result, it has seemed that the answer to the problem of free will can lie in only one of two alternatives: Either the fact that an agent’s action was determined is always compatible with his being responsible for it, or the fact that the agent’s action was determined will always rule his responsibility out. I shall suggest that the solution lies elsewhere…what we need in order to be responsible beings, I shall argue, is a suitable combination of determination and indetermination.\(^{224}\)

In essence, Wolf claims that we need determination in acting rightly, and some level of indetermination when acting wrongly in such a way that the agent can be blamed. Therefore, some might consider her a sourcehood theorist for blameworthiness, but not for praiseworthiness. I prefer to consider her a genuine reasons-responsiveness theorist, though, for one very important reason—the primary requirement for Wolf is an ability to respond to the True and the Good. An ability to do otherwise is not required, on Wolf’s view, for praiseworthiness since this primary requirement is satisfied without it.

\(^{223}\) ibid.

Moreover, the ability to do otherwise that is required for blameworthiness is *only so required* because the primary requirement (ability to respond to the True and the Good) is not satisfied. Perhaps Wolf considers indeterminism helpful here because she thinks it justifies a claim that this primary requirement *could have* been met if the world were not determined. If this is correct, her interest in indeterminism to create alternate possibilities is dependent on her main concern not being maintained. Her main concern is that we are able to respond to the True and the Good. It seems to me that another concern creates the tension; Wolf seems to be committed to the *ought implies can* doctrine. As the above passage suggests, Wolf thinks that, if a person who acts badly cannot act well, such an inability undermines that person’s moral agency and, *a fortiori* her moral responsibility for her bad behavior. While one might claim that commitment to the *ought implies can* doctrine amounts to a commitment to a requirement of alternate possibilities, this does not seem obvious to me.\textsuperscript{225}

Wolf’s view is definitely a reasons-responsiveness view. Her view focuses on how reasons are psychologically available to agents and how, when an agent acts, the action is appropriately connected to her true self. Finally, like other views that are reasons-responsiveness theories, Wolf’s addresses only how an agent can be praise- and blameworthy for actions, but no other forms of behavior.

\textsuperscript{225} For an interesting discussion of whether the *ought implies can* doctrine is distinct from alternate possibilities, see David Copp (2006). Copp argues that we can derive PAP from the *ought implies can* doctrine (so long as the *ought implies can* doctrine is augmented by a few key premises). If this is true, then Wolf’s apparent commitment to *ought implies can* might amount to a commitment to PAP. However, I am skeptical as to the truth of the augmented premises needed for the argument to work.
Moral Qualities of Will—Nomy Arpaly

According to Nomy Arpaly, it is our ability to respond to reasons that grounds the kind of freedom necessary for praise- and blameworthiness as well as that which distinguishes us from lower life-forms.\(^\text{226}\) In her work, *Merit, Meaning, and Human Bondage*, Arpaly sets out to explain this fundamental difference as well as the conditions for our warranting moral praise or blame for exercising or failing to exercise this ability.

Arpaly is concerned with determining what *warrants* praise and blame. For Arpaly, warrant of either praise or blame depends on the relationship between the reasons an agent has for acting and the features of the act that make it right or wrong. For example, when discussing Kant’s grocer, who gives a child patron the correct change not because it is the right thing to do, but rather because he is concerned about his reputation and future business, Arpaly writes, “His *reasons for action* do not correspond to the action’s *right-making features*.”\(^\text{227}\) Therefore, doing the right thing is not sufficient for an agent’s behavior to warrant praising the agent. In an earlier paper, Arpaly explains:

For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, in response to the features that make it right (the right reasons clause); and an agent is more praiseworthy, other things being equal, the stronger the moral concern that has led to her action (the concern clause). Moral

\(^{226}\) See Arpaly’s discussion of how individual persons are distinct from a rattlesnake ((2006): 3-5).

concern is to be understood as concern for what is in fact morally relevant and not as concern for what the agent takes to be morality.\textsuperscript{228}

Arpaly acknowledges that such standards for praiseworthiness entail that there can be multiple ways to fail acting in praiseworthy ways. For this reason, she distinguishes between \textit{good will, lack of good will, and ill will}. The praiseworthy agent who “does the right thing out of responsiveness to moral reasons does it out of \textit{good will}.”\textsuperscript{229} However, one might either fail to do the right thing or do something that is, in fact, wrong as a result of failing to be moved by morally pertinent reasons. Such a person demonstrates what Arpaly calls \textit{lack of good will} or, alternatively, \textit{moral indifference}. This type of blameworthiness is distinct from that which involves either performing the wrong action or failing to perform a right action “for the very reasons that make his course of action wrong.”\textsuperscript{230} A person who behaves badly for the reasons that make the action a bad act acts out of \textit{ill will}.

For Arpaly, praise- and blameworthiness depend on the quality of will one has when acting. Therefore, one is \textit{praiseworthy} on Arpaly’s account if and only if the person takes the morally right course of action out of good will. One is \textit{blameworthy} if one omits the morally right action and/or takes a morally wrong action \textit{either} out of moral indifference or as a result of ill will. Of course, there is a qualitative difference between blameworthy behavior that results from moral indifference and blameworthy behavior that results from ill will, and perhaps there are distinctions within these categories. For

\textsuperscript{228} Arpaly (2002): 233.
\textsuperscript{229} Ibid.
\textsuperscript{230} Arpaly (2006): 15.
this reason, Arpaly claims that the degree of praise- and blameworthiness corresponds to the degree of the quality of the will.\textsuperscript{231}

It is clear that Arpaly’s view is a reasons-responsive account. Not only does she declare it to be such, but she also focuses on response to reasons and dismisses many worries put forth by sourcehood theorists. First, what grounds praise- and blameworthiness is an agent’s quality of will \textit{in relation to} moral reasons dictating certain behavior in that agent’s situation. Therefore, Arpaly is not concerned with subjective views of morality, woefully misguided “good” intentions, or other such aspects of voluntary behavior. She homes in on the agent’s quality of will towards \textit{moral reasons}.

Second, Arpaly denies that even manipulation arguments should convince us that “the history through which we have reached our desires, values, and beliefs”\textsuperscript{232} matters to moral responsibility. She first argues that there is no morally significant difference between “being shaped by the gloomy weather in your town [and]…being shaped by the manipulations of a fellow human being.”\textsuperscript{233} That is to say, if we consider the agent who lives in a cloudy town to be just as responsible as one who lives in a sunny one, then we ought to consider the agent who has been manipulated to behave similarly morally responsible as well. Bad weather and Dr. Nefarious are simply two different sources of influence. And for Arpaly, the source of our influences is irrelevant.

Arpaly maintains the position even in the face of classic paradigm examples of manipulation. For example, Arpaly considers even a person who, like Patty Hearst, was

\begin{itemize}
\item \textsuperscript{231} Arpaly writes, “Furthermore, (1) other things being equal, a person is \textit{more} blameworthy for a given course of action if she acts out of ill will than if she merely acts out of lack of good will, and (2) other things being equal, a person is \textit{more} praiseworthy for a given good course of action the more good will she demonstrates in taking that course of action…” ((2006): 15).
\item \textsuperscript{232} Arpaly (2006): 112.
\item \textsuperscript{233} ibid.
\end{itemize}
“born again” and experienced a significant overhaul of character as a result of victimization to be blameworthy for later acts. Prior to being kidnapped, Patty Hearst was a normal teenager. After being kidnapped, she sided with her captors and became, like them, a terrorist. Arpaly contends that Hearst is still responsible for her bad acts after being held hostage and becoming a member of the group. While, Arpaly admits, such a story might be one of “bad constitutive moral luck,” this fact does not mitigate Hearst’s blame.234 When she does act, she acts in response to reasons. Therefore, Hearst qualifies as blameworthy. Her history as a somewhat nice person and victim is irrelevant to whether she deserves blame now. The person she is now is quite blameworthy.235

Furthermore, while Hearst’s case strikes us as unusual, instances of abrupt changes in our beliefs and motives are not unusual. Arpaly points out that we often change our beliefs or desires simply because we get older. Puberty, Arpaly notes, is a very abrupt change in such mental states. At one moment, little Jimmy could not stand little girls. At the next moment, he finds himself strangely attracted to their presence.236 In short, Arpaly states, “if we treat nonrational personality changes as excluding subsequent autonomy, too few people are autonomous;”237 in fact, “even in the most

235 Arpaly explains that part of our reluctance to blame Hearst might have to do with the fact that she is not the same person she was prior to the kidnapping. Arpaly’s response to this worry is simple—she’s not, and that is precisely the point. If Hearst’s identity has changed, then the Hearst we are blaming is distinct from the Hearst towards whom we felt sympathy. While Arpaly does not get into what constitutes such a change in personal identity, she seems to suggest that, on a rare occasion, the identity of a person is so significantly overhauled that connections to the past person ought to be discounted or otherwise completely overlooked. Arpaly holds this for both blaming a person and exonerating a person. For example, she suggests that if a person were a criminal as a youth, but lead an exemplary life for a significant time between her criminal activity and being caught by the police, it might be best not to blame the person since the person in handcuffs is not the person who committed the crimes anymore ((2006): 113-116).
normal lives or the lives of the most reasoning and reflective people, this history is checkered.”

So strong influences that change a person’s behavior do not exonerate a person from moral responsibility for Arpaly. Such a commitment would seem to preclude her from the class of sourcehood theorists. However, there does seem to be one aspect of Arpaly’s view left to examine before making this determination. Arpaly admits that if the manipulation involves the acquisition of a mental disorder, say, one that convinces the agent that everyone she sees is trying to kill her and other innocent people, this agent deserves little, if any, blame.

I think that Arpaly’s reasons for this claim have nothing to do with the fact that the agent is not the source of her delusions. Had the agent acquired such delusions when, reaching the age of 25, her genetic predisposition to develop schizophrenia kicks in, Arpaly would likely excuse that person as well. The reason for her excuse is not that she has acquired false beliefs from an unnatural or a natural source. The reason for her excuse, I believe, is that this extent of delusional beliefs interferes with the agent’s ability to rationally connect with the real world. If this is correct, than the excuse is still rooted in reasons-responsiveness. However, this does cause some tension between Arpaly’s claims. Earlier, we saw how Arpaly was unmov ed by the fact that some people might do evil things and believe that they are acting with good will. She thinks that such persons deserve blame simply because they are not responding to the good reasons to do otherwise. Their cognitive confusion was irrelevant. Here, we see that some cognitive confusion is relevant. It could be that sourcehood is a concern for Arpaly after all.

---

However, it could be that some reliable access to reality is a basic requirement for qualifying as an agent in the minimal sense. Since she so clearly rejects alternate possibilities and any history requirement, I shall conclude that a charitable reading would consider the second explanation the likely one.

While Arpaly’s view is my preferred view of those surveyed in this and the last chapter, I do have concerns about her view as well. These concerns are related to how Arpaly’s view, like all other in this chapter, consider actions (or, in her case, *wills*, which are intended to result in actions) the exclusive candidates in virtue of which agents can be praise- and blameworthy. While I do agree that agents are responsible for such things, I disagree that actions and willings exhaust the list of possibilities for what can be praise- or blameworthy. While there exist reasons for actions, there are reasons for many other things such as feelings, attitudes, beliefs, etc. On my account, since we are capable of responding to such reasons without such a response being an action or an act of will, these instances too are apt for praise and blame. This concern will be raised in the subsequent two chapters, to which we now turn.
Chapter 5

The Voluntary Nature of Aristotelian Emotions

As I noted at the end of the last chapter, I believe that there is a viable interpretation of Aristotle under which emotions and feelings are praise- and blameworthy. While I do not agree with Aristotle that the true object is simply our emotional reactions, I do think that much of what he claims can massage our intuitions and motivate the account that I ultimately wish to endorse.

For Aristotle, actions are not the only appropriate targets of praise and blame. He as much as says so in his opening line of the account he presents in the first chapter of the third book of *Nicomachean Ethics*. Aristotle tells us that we must look into the nature of the voluntary, “since excellence is concerned with passions and actions, and on voluntary passions and actions praise and blame are bestowed.”"239 Not only is Aristotle stating that, if we want to understand virtue, we must investigate the nature of the voluntary, he is also making implicit claims about the nature of emotions that are not universally accepted. Subtly embedded in the claim just quoted are two additional claims: 1) that emotions can be worthy of praise and blame and 2) that emotions are voluntary. It seems that the majority of contemporary and historical thinkers have taken for granted that the opposite

\[239\] *EN* 1109b 30-33.
is the case; emotions are not praise- or blameworthy, and their involuntary nature is what grounds this fact. I find it interesting that Aristotle makes these claims and, as I shall argue in the next chapter, I agree with him in these regards. However, the fact that Aristotle holds a minority view coupled with the fact that he neglects to explicitly defend these claims places a burden on those of us who are inspired by him to take up the task of providing an explanation and defense within Aristotle’s own framework. Therefore, in this chapter, I shall do just that. I shall try to provide evidence, when it is available, for Aristotle’s commitments that suggest his sincerity in subscribing to the above claims. In addition to motivating the view that Aristotle meant what he said when he wrote this claim, I shall provide argumentation in defense of the consistency of these claims with the rest of his view.

The claim of interest here, upon which a clearer picture of Aristotle’s account of praise-and blameworthiness rests, is that there are voluntary emotions. We can infer this claim from the above comment by imagining a simple valid argument expressing the relations between Aristotle’s thoughts:

1) If we are justified in praising or blaming something, then that object is voluntary.

2) There are occasions on which we are justified in praising and blaming emotions (pathos).

3) Therefore, there are occasions on which emotions are voluntary.

While the argument is valid, we must consider whether it is sound. The first premise is stipulated by Aristotle. Since we are considering whether emotions can be voluntary on
Aristotle’s account, we may take this premise as given. The second premise, as it is in existential form, will require both a negative argument as well as arguments based on intuitions to prove.

The goal of the negative argument is to establish that it is *not* the case that all emotions are *non-voluntary*. If it were the case that all emotions were non-voluntary, then premise 2 would be false. Aristotle held that no one object was both voluntary and non-voluntary. If it were the case that all emotions were non-voluntary, then there could not be an emotion that is voluntary. And if there could not be an emotion that is voluntary, there could not be an emotion that is deserving of praise or blame (per premise 1).

It is an empirical question as to whether or not there are instances when we are justified in praising and blaming emotions. I contend that there are intuitive cases that support this premise and that Aristotle may have had similar cases in mind, given his commitments regarding related matters. Therefore, after arguing for the possibility that emotions count as voluntary on Aristotle’s account, we shall consider cases where it is intuitive to blame or praise an agent for an emotion or emotional response.

In order to show that it is possible for emotions to count as voluntary and, therefore, be appropriately subject to praise and blame, let us consider the general nature of emotions on Aristotle’s view and compare that with typical excusing conditions that render behavior non-voluntary. If something about the nature of emotions is proven to be an example of an excusing condition, then emotions are non-voluntary by nature. If this is not the case, then it remains open that emotions could be voluntary and deserving of praise or blame.

---

240 c.f. *EE* 1223b 9-10.
Many scholars agree that Aristotle does not provide us with a properly theoretical account of what emotions are. That is to say, there is no extant text where Aristotle describes the general nature of emotions. The most thorough account that provides a glimpse into Aristotle’s view of emotions is found in the second book of *Rhetoric*. Aristotle begins the second book of *Rhetoric* with the following description of emotions: “The emotions are all those feelings that so change men as to affect their judgments, and that are also attended by pain and pleasure.” Far from a definition, this description is the only comment Aristotle makes on the general nature of all emotions. From here Aristotle provides a long list of specific emotions. The explanations of these emotions mirror those of particular virtues in *Nicomachean Ethics*. However, unlike *EN*, there is no earlier book on what all emotions share in common. This we must intuit for ourselves.

John Cooper has argued that Aristotle believed that there are at least three central elements that all emotions share in common. The first is that all emotions are agitated affected states of mind. They are mental states that have sensory aspects. The second element is that this agitated quality often depends on and arises from “the ways events or

---

241 For example, John M. Cooper (1996) makes this claim at pp. 238-239 and Gisela Striker (1996) seconds this sentiment at pp. 286-287 and Martha Nussbaum (1996) notes at p. 305 that the account in *Rhetoric* is distinct from Aristotle’s ethical views and is more of an inquiry into the nature of individual emotions rather than an inquiry into the nature of emotions in general.


243 I am thinking of how in the second book of *EN* we are told that virtue is the mean between two vices—one of excess and the other of deficiency. Interestingly, the mean and its excesses are measured by emotion (e.g., the excess of courage is a lack of fear and the deficiency of courage involves too much fear).

244 Cooper (1996): 250-252.
conditions strike the one affected.” That is to say, appearance and perception will inform an emotion. Finally, emotions are desires “for a specific range of reactive behaviors or other changes in the situation as it appears to...be.” For example, if I am afraid of a spider, this fear is also a desire to get away from the spider.

Martha Nussbaum has also offered an account of what is essential to emotions for Aristotle. She writes, “In Aristotle’s view, emotions are not blind animal forces, but intelligent and discriminating parts of the personality, closely related to beliefs of a certain sort, and therefore responsive to cognitive modification.” From here, Nussbaum adds three characteristics that she feels are standard to the common view of Aristotle’s time period. The first condition is that emotions are what we now call intentional states as they involve intentional awareness: “They are forms of awareness directed at or about an object, in which the object figures as it is seen from the creature’s point of view.”

The second characteristic is that emotions are intimately related to beliefs in such a way that they can modify beliefs as well as be modified by beliefs. For an example of the first modification, imagine how one’s affection for one’s child can modify one’s belief that the mischief caused by that child on a particular occasion is less offensive and less in need of correction than it probably is. For an example of the second modification, imagine that you have been told unflattering facts (or perhaps even lies) about a new colleague before meeting him. These beliefs might inspire less hospitality, concern for and interest in him than you might have had absent such beliefs. Finally, Nussbaum adds,

246 Ibid.
248 Ibid.
“All this being so, emotions may appropriately be assessed as rational or irrational, and also (independently) as true or false, depending on the character of the beliefs that are their basis or ground.”250 In other words, the legitimacy of the emotions that we have depends on the legitimacy of their corresponding beliefs; if an emotion is based on an irrational belief or a false belief, the emotion is not appropriate and is not reflective of the truth. For example, if Wanda is afraid of the teddy bear to her left, this fear is illegitimate. We can imagine that Wanda’s fear is either linked to a false belief that teddy bears are dangerous, or perhaps the fear stems from an irrational belief that all things are dangerous whenever they are positioned to her left.

From Aristotle’s own text, we learn several characteristics of specific emotions that would seem to apply universally. For example, Aristotle writes, “if a man desires anything and has good hopes of getting it, if what is to come is pleasant, it seems to him that it is sure to come to pass and will be good; but if a man is unemotional or in a bad humor, it is quite the reverse.”251 This comment supports the view that emotions affect beliefs and judgments. In addition, it seems that emotions typically motivate actions and provide an agent with goals. For instance, Aristotle describes anger as something that seeks revenge.252 This last point suggests that Aristotle believed that emotions resided primarily in the appetitive part of the soul, for it is this aspect of the soul that is responsible for motivation and movement. Gisela Striker seems to agree when she points out that Aristotle’s discussion in EN at book I chapter 13 suggests that “The nonrational part of the soul whose virtues are the virtues of character can be regarded as primarily the

250 Ibid.
251 Rhet. 1378a 2-4.
252 Rhet. 1378a 31-34.
In this discussion, Aristotle informs us that the nonrational part of the soul is comprised of two parts: the nutritive part and the appetitive part. The nutritive part is responsible for nourishment, growth, and other biological functions that all living things share in common. But the nutritive part “by nature…has no share in human virtue.” Its virtues are not, as Striker says, “the virtues of character,” since “the virtue of this capacity is apparently shared, not human. For this part and this capacity more than others seem to be active in sleep, and here the good and the bad person are least distinct.” The appetitive part, on the other hand, shares in reason. But judgment also affects emotion. Aristotle writes, “The [part] with appetites and in general desires shares in reason in a way, insofar as it both listens to reason and obeys it…[it] also is persuaded in some way by reason, as is shown by correction, and every sort of proof and exortation.” This last comment suggests that, with time and rehabilitation, emotions might change as a result of reason. This too suggests that appetite is the seat of emotions, since appetite is most notably changed by repetition and training involving pleasure and pain.

In short, a plausible picture of Aristotelian emotions would include the following. Emotions are mental states that aim at or are about beliefs and their objects. Emotions have causal influence over beliefs, but they are also dependent upon and can be shaped by beliefs. Emotions have an affective quality that is felt as well as some motivating aspect, which makes them related to the appetitive part of the soul. While they are

---

254 *EN* 1102a 34-1102b 4.
255 *EN* 1102b 13.
256 *EN* 1102b 4-8.
257 *EN* 1102b 14-30.
mental, intentional states intimately connected with the nonrational part of the soul, they have normative conditions that can be met or missed. That is to say, an emotion, because of its relationship to beliefs and actions, can be justified or not, appropriate or not, and, as I shall now try to demonstrate, often worthy of praise or blame.

**Excusing Conditions**

Recall that in Chapter 2, we saw that Aristotle provides us with two criteria for the voluntary. However, Aristotle defines the voluntary largely in contrast to the involuntary: “Those things, then, are thought involuntary, which take place under compulsion or owing to ignorance.” Therefore, let us establish two excusing conditions that ground the non-voluntary nature of that which is to be excused:

1) If something takes place under compulsion, then the agent is excused from responsibility.

2) If something takes place owing to ignorance, then the agent is excused from responsibility.

I take it for granted that, for Aristotle, an agent excused from responsibility is thereby rendered unworthy of either praise or blame. Now, let us consider what it means for something to take place under compulsion and what it means for something to take place owing to ignorance. If we can determine that some aspect of emotions on Aristotle’s view counts as either taking place under compulsion or owing to ignorance universally, then emotions will always be non-voluntary. However, I shall now argue that there is nothing

---

258 *EN* 1110a 1-2, emphasize mine.
inherent to emotions that renders them compulsory or products of ignorance given Aristotle’s notion of compulsion and ignorance of the particulars.

The Negative Argument—External Force as an Excuse

Let us consider first the description of that which is done out of compulsion as non-voluntary. As we saw in Chapter 2, Aristotle elaborates that something qualifies as “compulsory of which the moving principle is outside, being a principle in which nothing is contributed by the person who acts or is acted upon, e.g., if he were to be carried somewhere by a wind, or by men who had him in their power.”\(^{259}\) Such an external principle entails that the agent or victim contributes nothing. This would seem to suggest that if the agent contributes anything at all to the behavior in question, such behavior cannot qualify as non-voluntary.

When considered by itself, the exclusion of external force cannot render the emotions non-voluntary. Indeed, emotions are paradigms of *internal* force. In other words, since emotions are inherently internal movements of agents, they can never qualify as external force. Since they can never qualify as external force, emotions cannot be non-voluntary in virtue of the excusing condition of compulsion. Therefore, since there is nothing essential to the nature of the emotions that could qualify emotions as compelled, it remains open that emotions could be appropriate objects of praise and blame.

\(^{259}\) *EN* 1110a 2-4.
Aristotle acknowledges that it might be argued that, while emotions reside within the agent’s soul, and do not appear to be literally stimulated by physical contact with an external source, there are other ways to causally influence one’s psychological states from without. That is, it is possible that our emotions are non-voluntary in nature because of their relationship and susceptibility to factors external to us and beyond our control. While such an objection might be compatible with other theories of moral psychology, Aristotle’s theory blocks this worry. Aristotle presents this worry as such: “But what if someone says that pleasant things and fine things force us, on the ground that they are outside us and compel us?” Here, Aristotle seems to be considering the common belief (endoxa) that we are compelled by what we believe to be good since much of what appears to be good is not under our control and the appearance of something as good motivates us to pursue it. Aristotle dismisses this worry in saying that “For him, then, everything must be forced, since everyone in every action aims at something fine or pleasant.” In other words, the person who holds this objection must believe that nothing he does is voluntary—a conclusion that Aristotle seems to consider patently false. Aristotle adds that when we are forced to act, such force makes the act painful. But acting for pleasure is not painful; therefore acting for pleasure is not an example of force. Finally, Aristotle suggests that this objection rests on a desire to escape blame for our depravity: “It is ridiculous, then, for him...to hold himself responsible for his fine actions, but pleasant things responsible for his shameful actions.” Since it is evident to

---

260 EN 1110b 10-11.
261 EN 1110b 11-12.
262 EN 1110b 14-16.
Aristotle that we are deserving of praise for our fine actions, our intuitions for symmetry dictate that we must accept responsibility for our base actions as well.\footnote{While contemporary blame incompatibilists like Susan Wolf challenge this claim of symmetry, it seems to have been an accepted premise, given how Aristotle relies upon it to make his case.}

Aristotle’s discussion of so-called “mixed actions” also clarifies how external force is distinct from compelling reasons to feel or act a certain way. We have just seen how “the good” cannot be rightly said to compel an agent. In his discussion of mixed actions, Aristotle shows that circumstances cannot be said to compel us either.

As we saw earlier in Chapter 2, in his discussion of mixed actions, Aristotle notes that the common opinion (endoxa) would have it that there are some actions that are involuntary because the agent is “forced” to make a choice to do something undesirable. Paradigm cases that Aristotle points to include things that are done from fear of greater evils or for some noble object.\footnote{EN 1110a 5.} I believe that the latter case, that of a noble object, can be dismissed based on the discussion above against compulsion of the good. However, the idea that we are compelled by fear of greater evils is more difficult to explain away, since there is a sense in which agents acting under such conditions are reluctant to do what they do and behave, in a sense, unwillingly. For instance, take Aristotle’s example: no one would voluntarily throw his cargo overboard a ship; but if weather conditions made such an act necessary for the survival of the ship and its passengers, a captain is likely to reluctantly surrender his cargo.\footnote{Aristotle suggests something similar at EN 1110a 9-12.}

When considering cases like this one, Aristotle distinguishes the situation from the act in order to determine if it is voluntary, and then reconsiders the act in light of the situation to determine praise or blame. First, he notes that if we ignore the situation and
the type of act that is done (i.e., that it is a discarding of property), it is undeniable that the agent is the source of motion and knows the particulars. Therefore, his discarding property is voluntary. Reconsidered in light of the situation, we can see that discarding property was the best action available, and therefore desirable at the time of acting. Therefore, the agent ought to be praised for doing the right thing and saving lives. We act voluntarily under sub-optimal conditions often. The fact that we may not like our options is not enough to make our choosing one involuntary.

We have just seen how neither the allure of external goods nor the repulsiveness of unfavorable circumstances qualifies as external force for Aristotle. If such things do not qualify as excusing conditions for actions, they cannot qualify as excusing conditions for emotions either. Even if our emotions have external sources, this does not mean that they are caused by external force. For example, a sunny day might put me in a good mood. However, I am not forced from without to be in a good mood. I might be influenced by factors outside of me to be in a good mood. But this kind of influence cannot be considered external force. For one, it is not painful to be in a good mood. Since Aristotle contends that what is involuntary by means of external force is always painful, this condition excludes the possibility of my good mood being involuntary.

While it is tempting to continue pushing this example concerning negative influences, Aristotle has another principle that precludes such an argument. For example, we might consider whether my sadness or anger caused by a cloudy day could count as involuntary, since sadness and anger are painful. But given Aristotle’s claim of symmetry, that if we are responsible for our good behaviors we must also be responsible

\[266\] EN 1110a 14-18.
for negative behaviors, this argument cannot work. In other words, while pain is a necessary condition for something to be involuntary, it is hardly a sufficient condition. Similarly, while external causation is necessary for the excuse of external force to apply, it is hardly sufficient.

It appears that the excusing condition of external force does not render emotions, by nature, involuntary.\textsuperscript{267} Let us now consider whether the excusing condition of \textit{owing to ignorance} could apply universally to emotions.

\textbf{Case Studies—Feeling Badly with and without Ignorance of the Particulars}

In addition to actions issued by external principles, there may be involuntary as well as non-voluntary actions done due to ignorance. Both instances of actions must be first distinguished from a type of action caused by ignorance, which is not excused. In \textit{Nicomachean Ethics}, Aristotle is careful to distinguish two types of ignorant actions: those done \textit{because of ignorance} and those \textit{done in ignorance}. Actions done because of ignorance are done from ignorance of the particulars.\textsuperscript{268} Recall the case of Oedipus. Because he did not know that the man whom he killed was his father, Oedipus did not commit patricide voluntarily (though he did commit homicide voluntarily). Actions done \textit{in ignorance}, however, are those done because the agent was in a state conducive to a more general, global ignorance. Examples of these are drunkenness and viciousness.\textsuperscript{269}

\begin{itemize}
\item \textsuperscript{267} I only add the qualification “by nature” here since contemporary theorists might disagree with cases of artificial manipulation (for example, were a neuroscientist to somehow implant different emotions in a patient without consent). I doubt that Aristotle was concerned with such matters; however, to be clear, let us set such possibilities aside.
\item \textsuperscript{268} EN 1111a 1-20.
\item \textsuperscript{269} EN 1110b 25-30.
\end{itemize}
The drunkard is disoriented in several ways. The vicious is blind to reason and virtue—*he is ignorant of the good.*\(^{270}\) It is because of their *condition* that these people perform the bad actions that they do, not because of ignorance of particular facts about their current situation.

Actions done in an ignorance to which the agent voluntarily succumbed are also not excusable. The man who elects to get drunk is both responsible for being drunk and for what he does while drunk. The vicious person is responsible for becoming vicious, and is therefore responsible for all he does as a vicious person. Aristotle claims that actions done in ignorance where the agent is responsible for her ignorance are voluntary (i.e., they are subject to praise and blame).\(^{271}\) Actions done because of ignorance of the particulars, however, may be involuntary.\(^{272}\)

### A Case that Does Not Satisfy this Excusing Condition

We have just seen that actions done in ignorance where the agent is responsible for her ignorance are voluntary and fair candidates for blame. However, we are interested in learning whether it is plausible, on Aristotle’s view, that *emotions* that result from voluntary ignorance are praise- or blameworthy. Aristotle makes several comments that

\(^{270}\) C.f. *EN* 1150b 36-1151a 20.

\(^{271}\) *EN* 1113b 29-33.

\(^{272}\) Only some such actions qualify as involuntary. Aristotle distinguishes between actions caused by ignorance of the particulars that are later accompanied by pain and regret and those that are not. If the action is not accompanied by pain and regret, Aristotle considers it *merely* non-voluntary. If pain and regret follow, the action is not only non-voluntary, but also involuntary: “Everything caused by ignorance is nonvoluntary, but what is involuntary also involves pain and regret. For if someone’s action was caused by ignorance, be he now has no objection to the action, he has done it neither willingly, since he did not know what it was, nor unwillingly, sine he now feels no pain” (*EN* 1110b 19-22). C.f. Meyer (1989).
suggest this might be the case. For example, in book I, chapter 9 of *Rhetoric*, Aristotle writes:

> Praise is the expression in words of the eminence of a man’s good qualities, and therefore we must display his actions as the product of such qualities. Encomium refers to what he has actually done…Hence it is only when a man has already done something that we bestow *encomiums* upon him. Yet the actual deeds are evidence of the doer’s character; even if a man has not actually done a given good thing, we shall bestow praise on him, if we are sure that he is the sort of man who *would* do it.\(^{273}\)

Here Aristotle suggests that we praise a man’s *good qualities*. We can praise character even if the character is not actively doing a praiseworthy act. Since character is about both actions and feelings, it would seem that we are justified in praising emotions in such cases.

Certain discussions of the virtues also suggest that emotion is an important factor that is to be praised or blamed. In regards to courage, Aristotle writes, “Those things, also, are noble for which men strive anxiously, without feeling fear; for they feel thus about the good things.”\(^{274}\) Here it seems that what makes such behavior noble is the absence of fear. Could it then be the case that feeling fear on such instances would be *ignoble*, and therefore blameworthy? Of the coward, he writes, “The person who is excessively afraid is the coward, since he fears the wrong things, and in the wrong way, and so on.”\(^{275}\) Cowardice is a blameworthy state. Here, Aristotle defines cowardice in

\(^{273}\) *Rhet.* 1376b 27-34.

\(^{274}\) *Rhet.* 1376a 15-16.

\(^{275}\) *EN* 1115b 34.
terms of an emotion. If excessive fear constitutes cowardice, and if cowardice is blameworthy, then it would appear that excessive fear is blameworthy. Finally, Aristotle states that the “attitude to frightening things that the brave person really has is the attitude that the rash person wants to appear to have.”  

Rashness is a vice. However, the rash man appears to be just like the brave man when he acts, since he imitates the brave man. Aristotle makes the striking diagnostic claim that “most of them are rash cowards; for, rash though they are on these [occasions for imitation], they do not stand firm against anything frightening.” Unfortunately, Aristotle does not make clear how they do not stand firm against anything frightening. Is it that they imitate brave persons when there is no danger? Or is it that they imitate brave persons when there is danger, but they do not stand firm against anything frightening because it is not frightening for them, for they do not feel the appropriate fear?

If I am taking these suggestions fairly, it seems that Aristotle was comfortable praising and blaming people in virtue of the emotions that they have. However, Aristotle does not provide us an example of a praise- or blameworthy emotion independent of any discussion of actions or character. Still, we might consider how Aristotle’s view would analyze such an example of our own making, should my above suggestions be correct.

Imagine that Eva loathes her co-worker, Betty. Betty regularly irritates Eva by attempting to make small talk, smiling, and asking Eva how she is doing—all behaviors Eva cannot tolerate. One day, Betty is killed in a horrific automobile accident. Upon hearing the news, Eva cheers for joy. She is delighted by the news. She has epistemic

---

276 EN 1115b 31.
277 EN 1115b 33.
278 EN 1115b 35.
access to the relevant facts; Betty is actually dead and she did die in an automobile accident. Eva is also not experiencing any external force in having this emotional release, let us stipulate. She satisfies no excusing conditions, yet she strikes me as deserving of blame; given Aristotle’s criticism of those who feel fear and other emotions inappropriately, I believe that Aristotle would agree. Since, as we illustrated at the beginning of this chapter, something deserving of praise or blame is necessarily voluntary, then Eva’s emotion is voluntary on Aristotle’s account.

As our original argument suggests, if Eva’s emotional response is blameworthy then, on Aristotle’s account, it is voluntary. Therefore, if Aristotle would agree with the above assessment of Eva’s desert of blame, then it would appear that some emotional responses on Aristotle’s account are voluntary and deserving of praise and blame. If my assessment is correct, we need not look further for our negative argument. However, I want to propose another hypothesis to test out related to this excusing condition: perhaps some emotions that are caused by ignorance of the particulars remain, nonetheless, praise- or blameworthy and, in virtue of our original argument, voluntary. I think that this hypothesis is true, both for Aristotle and in general. I also think that understanding the truth of this hypothesis can motivate my attitudinal account that will be advanced in the next chapter. In order to prepare our intuitions for my theory, let us briefly consider how the excusing condition of something coming about due to ignorance of the particulars might not apply universally to all emotions caused by such ignorance.
Cases that Satisfy the Excusing Condition, Yet Remain Inexcusable

Like actions caused by ignorance of the particulars, it is easy to imagine that certain emotions and, in particular, emotional reactions to a situation might be caused by a misperception of that situation (i.e., ignorance of some facts regarding the situation). For example, if a just and virtuous *phronimos* were out for a walk and spotted what appeared to be two hooligans assaulting a little old lady, his deep sense of justice and righteous indignation would cause him to be outraged. This “just walker” would intervene and be very angry with the perpetrators. However, once he recognizes that he had happened upon a movie set and that the scene was a fake, two results would follow: he would no longer be angry and no sensible person on the set would blame him for having gotten angry. They would understand that it was due to his ignorance that he acted inappropriately and had the inappropriate feeling.

So, it seems as though we can imagine cases of having an emotion that satisfy the second excusing condition—being caused by ignorance of the particulars. Indeed, we can imagine that the actors forgive the just walker for any harsh words or actions that he took as a result of his ignorance. Onlookers might pity the just walker for his embarrassment. Therefore, there is a sense in which we can imagine pardoning and pitying someone who has emotional responses caused by ignorance. Since apt pardon and pity are characteristic of the involuntary, we can imagine such cases might qualify as involuntary emotions.

However, while pardon and pity are appropriate when one acts (and perhaps feels) involuntarily, it is not the case that these reactions are exclusive to involuntary behavior. That is, Aristotle acknowledges that we might pardon someone for voluntarily offending
us if the person repents. We might pity someone who has foolishly though voluntarily gambled away her life savings if we feel that her present suffering is excessive in proportion to her offense. The fact that we might pardon or pity someone who experiences an emotion as a result of ignorance of the particulars is not sufficient evidence to show that some emotions are involuntary.

On the other hand, the fact that some emotions might be involuntary does not entail that all emotions are involuntary. And for our case, we only need that some emotions are voluntary. I contend that either all Aristotelian emotions are voluntary or at least some are voluntary. As to the first option, we can imagine that the above scenario pans out differently such that it makes sense to call the emotion voluntary and the actions that follow from it involuntary.

Do we really want to say that such emotions are involuntary and therefore not subject to praise and blame? It seems more plausible that such emotions could still be apt candidates for praise and blame; however, it might be the case that we can justifiably praise emotions that accompany actions that might have been blameworthy had they been voluntary (i.e., if the just walker had known that the people were actors yet been enraged, we would not praise his emotional response). In other words, I think it is not that such emotions are not subject to praise and blame, but rather that whether it is praise or blame might depend on the knowledge of the particulars. For example, recall the just walker. Not only would we pardon his intrusion, but also we might praise him for being the type of person who would be outraged and intervene in such a situation. I can imagine the

---

279 Rhet. II.3.

280 See Rhet. II. 8, esp. 1385b 12-15 where Aristotle claims that we pity those we feel suffer in ways that are undeserved.
little old lady saying something like “Well, at least I’m glad to know that you would save me if I really were being mugged! Thank you so much for your concern!” While the action of intruding was due to ignorance and therefore involuntary, the just man’s anger seems slightly different. While his anger is an effect of his ignorance in that he would not have become angry had he known of the particulars, it is also an effect of the type of character that he has. It is for this reason that we ought still to thank the just man for offering to help (and not merely forgive him for trying to help where it was not actually needed). We should praise his getting angry because it is what the virtuous person should have felt had he actually witnessed a mugging.

We just saw how we might praise the just walker for his actions and the sentiments that motivated his actions when he feels and acts as a result of ignorance of the particulars. However, such examples might be confusing in that we might find the behavior excusable and the emotion laudable. Let us consider an example with more symmetry to test our intuitions about the appropriate response.

Consider Carmine, who is homophobic, quick to anger and who frequently instigates physical altercations. One day, while in Carmine’s presence, a bug flies into Buford’s eye and Buford blinks desperately to flush it out. However, Carmine mistakes Buford’s blinking for a seductive wink. Carmine is quickly enraged by this apparent advance and pins Buford up against a wall by his neck while shouting obscenities in a threatening tone.

It is true that Carmine acted on an emotion (anger at Buford’s apparent advance). It is also true that this emotion is one that was caused by ignorance of the actual facts of Carmine’s situation. However, I think it makes little sense to forgive Carmine for
anything. We should not forgive him for his anger since it is morally deplorable. We should not forgive him for his actions because what Carmine took to be justification of such behavior (i.e., being flirted with by a man) is not, in fact, justification for such behavior. Anger at mild same-sex flirtation is never justified so no action motivated by such anger can be justified. Also, since Aristotle does not excuse ignorance of the good, the fact that Carmine thinks that he is justified is no reason to pardon him. At least, in Carmine’s case, we are justified in blaming him without any confusion of a sensible alternative to pardon him.

There is no doubt that people have emotional responses due to ignorance of particulars. But the question that Aristotle did not answer for us is this: are emotions excusable in the same way that actions due to ignorance are? I have answered that there are at least some cases in which they are not. This consequence, if true, would be very interesting indeed; for it would suggest that most, if not all emotions are voluntary. However, I admit that it remains open to interpretation as to whether there are any cases that constitute involuntary emotions due to ignorance. As interesting a consequence as this would be, what is most important for our purposes is that we now see which aspects of Aristotle’s view support his claim at the beginning of Nicomachean Ethics, III.1.

An Alternative Interpretation of These Results

We have seen that some emotions are apt candidates for praise and blame on Aristotle’s account. I contend that the emotions themselves are a proper and direct target of praise and blame for Aristotle. However, an alternative explanation for why emotions are
praise- and blameworthy has been proposed: perhaps Aristotle was willing to praise and blame emotions because emotions are often effects of our characters. Since we are responsible for our characters, we are also responsible for their effects, which include actions and emotions. Therefore, while we are praise- and blameworthy for our emotions, we are only praise- and blameworthy for them derivatively.

This alternative interpretation that we are now considering conflicts with my interpretation, since it justifies praise and blame solely on prior voluntary action—making a connection to action necessary for an emotion to warrant praise and blame. However, I do not think Aristotle says that this is the case and I do find this interpretation problematic. Since this interpretation is a prominent one, I shall try to motivate why I think my interpretation is preferable.

I do not believe that I can dispel this interpretation entirely, as Aristotle’s texts are both incomplete and ambiguous about such matters, rendering multiple interpretations plausible. I do contend that our previous discussion has successfully established that Aristotle had the resources, given his original view of the voluntary, to deem emotions praise- and blameworthy without an indirect approach. Now, I would like to briefly address the attractiveness of the alternative view and attempt to show that my view is more attractive still.

Kosman’s View

One of the clearest arguments for the view that we are only derivatively responsible for our emotions can be found in L.A. Kosman’s 1980 article, “Being Properly Affected:
Virtues and Feelings in Aristotle’s Ethics.” Here, Kosman explains the apparent problem in detail.

Kosman begins by pointing out that “virtues are dispositions toward feeling as well as acting,”\(^\text{281}\) which raises the question as to whether we are responsible for both. Kosman thinks that if we are responsible for virtue, we are probably responsible for its aspects and components as well. Kosman then adds that the state of being virtuous is constituted by both *praxis* and *pathos*. It is the second aspect that perplexes us, because *pathos* is passive; that is to say that we can “describe the moral virtues as dispositions toward *action* and *passion*—toward characteristic modes of conduct, in other words, in which the virtuous person acts and is acted upon—is the moral subject, as it were, of active and passive verbs.”\(^\text{282}\) Kosman adds that since “Aristotle sees fear, anger, desire, pleasure, and pain as *pathe*, as passions, he views what we would call feelings or emotions as modes of a subject being acted upon.”\(^\text{283}\) The passive nature of emotions inspires in us skepticism of their voluntary and/or praise- and blameworthy nature. Much like our earlier discussion of *external force*, the fact that emotions involve an agent being *acted upon* suggests that they are acted upon by something else that is outside of the agent. But, Kosman contends, “‘Activity’ here must take on a sense broad enough to include both acting and being acted upon—must, that is, include modes of active human being in which the human individual is both subject and object of the action, both agent and patient.”\(^\text{284}\)

\(^{281}\) Kosman (1980): 104.

\(^{282}\) Kosman (1980): 104.

\(^{283}\) ibid.

A second, though related, worry stems from the fact that, “It appears to be a distinction between our actions and our passions that actions are within our control, whereas passions are not; we are the initiating principle of what we do, but not of what is done to us.” Kosman thinks that this is a worry for two reasons. First, he is concerned about the asymmetry of characteristics between two objects that share the property of being praise- and blameworthy. Second, he feels that choice is important for our deserving praise- and blame for our emotions. Let us consider each in turn.

I remain unconvinced that the asymmetry noted by Kosman is a cause for any concern, as Aristotle merely states that praise and blame are bestowed upon both actions and feelings; he does not claim that actions and feelings share praise- and blameworthiness in common due to their similar natures. Actions and emotions are different kinds of things. Yet praise- and blameworthiness seem to be properties that can appropriately apply to different kinds of things. However, we see such differences regularly without finding such examples confusing. For example, a dog and a kangaroo are of different kinds. However, it is possible that the dog and the kangaroo share in common a color. For instance, they could both be tan. It would be absurd to ask why the dog is tan since, we know all kangaroos are tan and kangaroos have entirely different natures and origins from those of dogs! So why, I ask, does it matter that actions are active and emotions are passive? This alone does not mean that the two cannot share properties in common. Just because two objects have distinct natures does not mean that all properties must be similar. In Aristotelian-speak, it is as though Kosman is contending

---

that everything that is praise- and blameworthy must have the same substance in order to have any of the other nine categories in common. However, this claim is obviously false. The second worry seems to suggest that the problematic aspect of the asymmetry is the unchosen nature of our emotions. However, I feel that Kosman is making an error concerning the necessary conditions that Aristotle laid out for something to qualify as praise- or blameworthy. Everything that is praise- or blameworthy is also voluntary. However, contrary to our modern intuitions, not everything that is voluntary is chosen.

Kosman addresses passages that say things such as the following: “…we feel anger and fear without choice, but the excellences are choices or involve choice.” However, it is important to note the context of the discussion. In this part of the text, Aristotle is telling the reader how virtues and vices are not identical to emotions. Similarly to what we have just argued, the fact that virtues and vices are not members of a species or genus known as emotions by no means indicates that characters and emotions have a different status regarding the voluntary.

Most importantly, Aristotle tells us elsewhere that while choice is a sufficient condition for the voluntary, it is hardly a necessary one:

Choice, then, seems to be voluntary, but not the same thing as the voluntary; the latter extends more widely. For both children and other animals have a share in voluntary action, but not in choice, and acts done on the spur of the moment we describe as voluntary, but not as chosen.287

The fact that our emotions are not chosen does not mean that they are not voluntary. There is another class of the voluntary that does not involve choice. Also, choice or

286 EN 1106a 3-4.
287 EN 1111b 6-10.
decision has been defined by Aristotle as the conclusion of a practical syllogism (i.e., the act of practical deliberation).\textsuperscript{288} Aristotle states that decision “is voluntary, but not everything that is voluntary is decided. Then perhaps what is decided is what has been previously deliberated. For decision involves reason and thought.”\textsuperscript{289} However, it is not evident that every action (let alone any emotion) results from deliberation. Indeed, since deliberation requires reason, and since animals do not have a rational aspect of their souls and children have yet to develop and utilize their rationality, then the only way that animals and children could have a share in the voluntary is if their voluntary behavior is due to some voluntary aspect of appetite. Therefore, since choice requires reason and the voluntary does not require reason, choice cannot be required for the voluntary. Furthermore, if choice is not required for the voluntary, and the voluntary is required for emotions to be praise- and blameworthy, then it remains open that emotions could be praise- and blameworthy regardless of the fact that emotions are not chosen. Indeed, I contend that what grounds the voluntary (and therefore praise- and blameworthy) nature of emotions stems from that class of the voluntary that is unrelated to choice.

Finally, I think that there are reasons to prefer an interpretation of Aristotle’s view as an account that justifies praise and blame for individual actions and emotions rather than simply all actions and only those emotions related to character. Interpreting Aristotle Kosman’s way subjects Aristotle’s view to several flaws that my interpretation does not.\textsuperscript{290} First, it would seem that one has to act a certain way repeatedly to acquire a character. Once acquired, the person is responsible for having that character given her

\textsuperscript{288} C.f. \textit{EN} III. ii.

\textsuperscript{289} \textit{EN} 1112a 14-17.

\textsuperscript{290} Admittedly, it could be the case that Aristotle’s view has these flaws. However, mine is the more charitable interpretation.
prior acts. That is to say, since her character is a result of such prior acts, she is responsible for having that character. But we think of being responsible for consequences or states of affairs that issue from behavior for which we are also responsible. Therefore, if she is responsible for her character, it is because she was responsible for the prior acts that caused her character to come about. So far, this seems consistent with Kosman’s view.

What Kosman’s view cannot allow is that anyone would be praise- or blameworthy for an emotion prior to acquiring a character. This, to my mind, seems odd. It makes sense that actions of any type—those performed prior to acquiring a character and after character acquisition—would be behaviors for which we are responsible. It also makes sense that we could only be responsible for the character that we have after we acquire it, for no one is born with a character. If something (e.g., having a character) is not true of a person, then it cannot be something for which the person is responsible. What makes little sense to me is that an agent might be responsible for her emotions only if they are consequences of her prior voluntary behavior. If a youth has a hateful and angry emotional response towards another who does not deserve it, I find such a response blameworthy even if the youth has yet to develop a character of any particular sort.

**Conclusion**

In summary, I have just shown that tracing the source of what makes us responsible for our emotions to prior voluntary acts is not necessary to sketch a picture of how Aristotle might have considered us praise- and blameworthy for emotions. Aristotle’s technical use
of terms such as voluntary, force, ignorance, and choice leave it open that emotions are praise- and blameworthy on his account. Furthermore, Aristotle’s attention to the attitudes of various virtues and vice as that which constitutes their praise- and blameworthiness suggests that emotions were appropriate targets for praise and blame according to Aristotle. I have also suggested that it is preferable to view Aristotle’s justification of praising and blaming in the way that I have proposed over the standard interpretation, since the standard line of explanation creates additional explanatory problems.

On the other hand, I have not shown (nor do I think it possible to show) that Kosman’s interpretation is inconsistent with Aristotle’s writings. For example, Aristotle does seem to say, at times, that the vicious person is responsible for her character because it was up to her to become vicious or not.\textsuperscript{291} Indeed, my discussion of dual causation in Chapter 2 suggests that it is both that we are the source of our characters and that we respond to reasons the way that we do. Nonetheless, I do think that Aristotle was making, at the very least, ambiguous moves towards a theory of praise and blame of emotions that does not require controlled causation. In the coming chapter, I shall take up this task myself.

\textsuperscript{291} C.f. \textit{EN} III.v.
Chapter 6

The Moral Attitude Account

David Hume famously suggested that actions are essentially symptoms of passions.\(^{292}\) Therefore, when we blame a person for certain behavior, we really ought to be blaming that person’s character. My view is somewhat similar, with the exception of what is the appropriate target of praise and blame. The states of being that I shall argue are the appropriate targets of praise and blame will be referred to as *moral attitudes*. I take a moral attitude to be an agent’s attitude towards moral considerations (i.e., moral reasons), which include attitudes of indifference or even willful disregard.\(^{293}\)

In order to comprehend the type of mental state that is the appropriate bearer of ascriptions of praise and blame, let us consider the following case. Imagine that Jesse used to be a good person, but has recently become amused by the suffering of others. He frequently laughs at others when they are hurt or have lost something valuable to them. Even when he does not show outward signs of such amusement, he is still overcome by joy at the sight of the suffering of others. In fact, he constantly wishes and hopes to


\(^{293}\) One reason why I wish to distinguish my view from views that praise and blame on the basis of character is that I hold that an instance of a moral attitude can be apt for praise or blame *even if* that attitude is out of character for the person, and *even if* the person has yet to solidify a character of any particular type. In other words, we are justified to praise and blame persons for specific mental events; that is, we are not justified to praise or blame during a given event because of who the person involved in it is.
encounter future displays of pain and suffering, and he is completely unapologetic for his feelings.

I contend that Jesse is worthy of blame for having these attitudes towards his fellow human beings. Unlike Hume, I do not hold that these attitudes must constitute part of Jesse’s character. Indeed, they are blameworthy even if they are surprisingly out of character for Jesse. As I stipulated above, Jesse was not always such a person. Therefore, the first instance he held such an attitude was “out of character” for him. Be that as it may, it does not make the attitude any less blameworthy.

In addition, Jesse does not need to act on these attitudes to be worthy of blame. He is blameworthy just in case he has malicious attitudes. Indeed, when he does act on these attitudes, we take notice and blame him. However, we are not blaming him solely for what he does; we are blaming him for the attitudes that he has as well. If Jesse cackles at the sight of a little old man being run over by a truck, we do not reprimand him because we think he should keep such feelings to himself, or because we worry that the old man’s feelings might be hurt if he hears Jesse’s laughter (that is, provided that he survives). Certainly, both of these judgments are sound. But they are not the main issues in this scenario. We blame Jesse for having that attitude regardless of whether or not he acts on it. Keeping his attitude to himself would be preferable to his present behavior, but what is truly morally required is that he not have such an attitude towards the suffering of another in the first place. The fact that outward behavior provides epistemic reason to believe that someone is worthy of blame ought not to be confused with the state of affairs that warrants blame regardless of our awareness of it. In other words, while it is true that
we are usually made aware of appropriate instances of blame by outward behavior, it is not the case that all such instances involve outward behavior.

In this chapter, I shall articulate and defend a new theory of moral praise- and blameworthiness. I shall argue that praise- and blameworthiness do not require causal responsibility, and so a fortiori do not require moral responsibility. I shall present my positive account of praise- and blameworthiness in light of this conclusion. My positive account is inspired by Nomy Arpaly’s view. My view, however, has a much broader scope than that of Arpaly.

There is much about Arpaly’s view with which I agree. However, there is a fundamental point of departure; on Arpaly’s view, agents are praise- and blameworthy as a result of the quality of their wills. Since one’s will is understood in the traditional sense of that which actually motivates one to action in response to a reason to act, Arpaly is committed to the restriction that we are only praise- and blameworthy for actions (which includes willful omissions and tryings that were thwarted). While I do agree that we can be praise- and blameworthy in virtue of attitudes that result in overt behavior, I do not wish to restrict the category of what we are praise- and blameworthy for to solely attitudes that result in actions. That is, I believe that there are moral attitudes for which we are praise- and blameworthy that do not factor in to the issuing of overt behavior.

For example, Arpaly dedicates a discussion to what she terms The Insignificance of the Moral Emotions.294 Here, Arpaly argues that, while moral emotions are sometimes indicators of qualities of will, they are not always accurate. She explains that multiple factors can affect whether one feels something like fear. For example, the presence of a

larger menace, the effects of a childhood trauma, and (I suppose) even aspects such as our mood or blood-sugar levels can affect whether we feel fear and how much. I do not deny these psychological and physiological facts. However, I do think that it is easy to consider such facts as part of the unique situation that renders an emotion or behavior appropriate. In other words, I think that there are times when the mere having of an emotion or other attitude is, in itself, praise- or blameworthy. This aspect of my view is partly inspired by my interpretation of Aristotle’s assessment of emotions. However, I also want to add other moral attitudes to emotions that might be deserving of praise or blame.

This is not to say that I wish to show that every mental state is praise- or blameworthy. For example, Arpaly notes that we have aesthetic emotions and these seem like paradigm examples of emotions that do not qualify as praise- or blameworthy. Arpaly rightfully worries that our aesthetic and moral emotions might compete and we are only psychologically capable of appreciating one at a time. For instance, she comments that it is not blameworthy to find Nabakov’s Lolita beautiful or intriguing even though one also deplores pedophilia and other forms of child abuse.\textsuperscript{295} I agree with Arpaly that this is the case. However, I do not think that the category of emotions towards objects that are, under certain descriptions, morally relevant exhausts the economy of moral attitudes that we possess towards morally relevant features we encounter. In fact, it is not obvious that we must be tending to such attitudes to have them. For example, a racist attitude is blameworthy even if a racist person is not consciously paying attention to her racist attitude. Therefore, I agree that emotions in particular can be misleading—

especially if we take their absence to indicate an absence of a moral attitude. Still, I intend to show that some emotions and other moral attitudes are apt candidates for praise and blame.

In order to present my positive account, it is first necessary to demonstrate how it differs from standard views and why those views ought to be abandoned. In general, I find that primacy is granted to moral responsibility. For example, in order to determine whether Robert Alton Harris\textsuperscript{296} is to blame for his behavior, theorists first argue over whether he can be legitimately said to have been morally responsible for his motives. In contrast to this assumption, I advance an argument for reversing the roles of moral responsibility and the worthiness of praise and blame.

**Necessary but not Sufficient Condition**

What is the relation between moral responsibility and praise- or blameworthiness? First, one might say that being morally responsible is necessary but not sufficient for one to be worthy of praise or blame. A clear articulation of this position is found in the work of John Martin Fischer and Mark Ravizza.

On John Martin Fischer and Mark Ravizza’s account of moral responsibility, one can be morally responsible without being praiseworthy or blameworthy. However, one is not praiseworthy or blameworthy without being morally responsible.\textsuperscript{297} First, let us

---

\textsuperscript{296} Robert Alton Harris was a criminal who was executed in 1992 by the State of California for two felony murders that he committed in 1978 at the age of 25. His case is often considered in discussions of moral responsibility because he allegedly suffered an extremely abusive childhood. In fact, an appeal was made to the Governor of California on the grounds that his abusive childhood mitigated his responsibility for his acts. The appeal was denied.

\textsuperscript{297} See Fischer et al. (2007): 185-186, for a discussion of the eligibility requirement.
consider their view in general. Then, we shall see how this view allows for moral responsibility and the worthiness of praise and blame to come apart.

On Fischer and Ravizza’s account of moral responsibility, one is morally responsible for what one does if one has guidance control.\textsuperscript{298} Recall that one behaves with guidance control if and only if one satisfies two conditions. First, the agent must have the kind of control afforded agents in Frankfurt-style cases.\textsuperscript{299} The other requirement Fischer and Ravizza have for guidance control is that this control be issued from the agent’s own moderately reasons-responsive mechanism.\textsuperscript{300} In other words, the agent regularly recognizes reasons to do things (including reasons to do differently from how the agent, in fact, does) and there exists at least one possible world in which the agent also responds to a sufficient reason to do otherwise.

It is easy to see how, on this view, agents can be morally responsible without being praiseworthy or blameworthy. There are countless actions in which agents engage each and every day that involve guidance control. When I tie my shoes in the morning, I exhibit guidance control over my shoe-tying and I am receptive and reactive to the very good reasons of wanting to walk outside without tripping, getting my feet wet, etc. Yet, if I had a very good reason not to tie my shoes, I would be able to go out with them untied. Since acting with guidance control is sufficient for moral responsibility, I am morally responsible for tying my shoes in the morning. But Fischer and Ravizza think that,

\begin{footnotes}
\item[299] Recall that a Frankfurt-style case is one in which an agent is prevented from doing otherwise, but the agent appears to act voluntarily and, in turn, be morally responsible for what she does (Frankfurt 1969/1988b).
\item[300] Recall that a mechanism is moderately reasons-responsive when the mechanism is regularly receptive to reasons (some of which are moral) and at least weakly reactive to them. Fischer and Ravizza (1998): Chapters 2-3.
\end{footnotes}
because it is a matter of moral indifference whether I tie my shoes, I am neither praiseworthy for doing so nor blameworthy. Hence I am morally responsible, but neither praiseworthy nor blameworthy, for my act. The fact that the same rational capacity would be in play when I make moral choices, such as whether to tell the truth or whether to keep a promise, is all that is required for me to be morally responsible for my actions. The fact that I am not *presently* responding to moral reasons, on this account, is irrelevant. So long as the mechanism that issues in my behavior is my own and one and the same mechanism that can respond to moral reasons is sufficient for my shoe-tying behavior to be behavior for which I am morally responsible.

In addition to such morally insignificant behaviors as shoe-tying, Fischer and Ravizza also hold that moral responsibility can exist *without* praise- or blameworthiness with morally significant behaviors as well. In an exchange with Derk Pereboom, Fischer explicitly takes this position.\(^{301}\) Here, Fischer explains that moral responsibility is an “eligibility requirement” for praise and blame.

On my view, judgments of moral responsibility and (say) blameworthiness are two separate, but obviously related moments in our evaluation of behavior. To establish moral responsibility is to show *eligibility* for certain responses; but such eligibility does not in itself imply that the responses ought to be made in a particular context.\(^{302}\)

At first blush, Fischer’s claim seems confused. What, exactly, does it mean to be eligible for certain responses? Is it not the same thing as being *worthy* of such responses? And, if one is not worthy of a certain response, does that not also mean that one is not *eligible* for

\(^{301}\) The exchange is published in Fischer et al. (2007): 184-203.

\(^{302}\) Fischer et al. (2007): 186.
such a response? I suspect that Fischer is not adequately distinguishing an agent’s desert of a certain kind of judgment, such as blameworthiness and praiseworthiness, from the appropriateness of certain kinds of treatment, such as expressing such judgments or justifying and implementing punishments based on such judgments.

My suspicions seem correct when we consider Fischer’s own illustration of the alleged difference between being eligible for blame and being deserving of blame. Fischer offers an example of a battered woman who kills her husband. The woman is morally responsible for killing her husband; but given the abuse that she suffered, she is not blameworthy, according to Fischer, for the murder. Fischer argues that she is not blameworthy because the abuse that she has suffered is sufficiently mitigating. However, since she had guidance control over her behavior when killing the man, she is morally responsible for the death. Therefore, for Fischer, moral responsibility is the minimum requirement for praiseworthiness and blameworthiness and other factors determine whether the agent is merely morally responsible or worthy of praise or blame as well.

This reasoning, to my mind, is flawed. Fischer seems to want to say that we should not imprison or publically humiliate the battered woman, and that might be right. It might also be the case that we should refrain from such actions because the woman does not deserve blame; for if she does not deserve blame, it is likely that she does not deserve punishment. On the other hand, it is not clear that these claims are true and that the woman is morally responsible for the death of her husband. What does it mean for a

---

303 Fischer et al. (2007): 185-186.

304 Just what these other factors might be is not made explicit for us by Fischer. Since his is a theory of moral responsibility only (i.e., he does not purport to be offering a theory of praise and blame), these factors are left for other works and discussions. However, as I suggested at the start of this paper, Fischer and Ravizza periodically make mention of praise and blame because, I suspect, our notions of them are intimately related to our intuitions concerning moral responsibility.
person to be morally responsible for the death of another if not that the responsible party
deserves at least blame and perhaps also punishment?

Like others, I find it counterintuitive to consider activities such as shoe-tying and nose-scratching to be morally responsible actions. Actions for which one is morally responsible ought to be actions that have some moral significance. While I agree with Fischer that certain factors can count as mitigating when assessing one’s blame, I find it strange to say that one could be morally responsible without deserving any blame for a morally wrong act. Perhaps the woman is less blameworthy than she would have been had her husband treated her well all these years. But if she is morally responsible for a bad act, I am hard-pressed to clear her of all blame. On the other hand, if these factors are sufficiently mitigating and she is not blameworthy in the least (i.e., it was a justifiable homicide), to continue to consider her morally responsible seems odd as well. She might be causally responsible for the death of her husband, but there is a big difference between being causally responsible for something and being morally responsible for it. And this world of difference consists of moral significance.

Fischer and Ravizza might reply that it is not to be denied that this type of responsibility does have moral import; it is the import of having guidance control. But such a response would be unsatisfactory. Merely having control over something does not entail that to exercise such control is to behave in a morally relevant way. To say otherwise is to water down the true meaning of the terms.

Still, even if Fischer and Ravizza were to refrain from considering average shoe-tying acts morally responsible acts, they might insist that the case of the killing of the

305 In particular, Derk Pereboom. For a brief excerpt on how Pereboom views what he calls “bona fide notion of moral responsibility” see Fischer et al. (2007): 197-198.
abuser has moral import since, of all behaviors, killing is certainly morally relevant. If this is true, they might argue that they still have one example of how a person could be morally responsible without being blameworthy or praiseworthy. However, the moral import here belongs to the nature of the act and not the person performing the act. Much as contemporary consequentialist theories conceptually separate agents from actions, I think we naturally separate responsibility from action-types. That is, responsibility is a property of a person whereas the goodness or rightness of an act can be said to belong to the act itself. In general, killing a man is morally wrong. But if a woman, on one particular occasion, kills a man and she is not blameworthy due to her mental state and particular history with the victim (or since it was justifiable as an act of self-defense), it would seem that the same would go for her moral responsibility. For when we find one to be entirely void of blame, we consider that person a fortiori to be void of moral responsibility as well (even if we continue to judge the person’s act to have been a good or a bad thing to do overall).

On the other hand, like Fischer, I agree that such activities are importantly distinct from nervous twitches, involuntary behaviors, etc. Even though there is no moral significance to my shoe-tying, and it seems odd to consider it something for which I am morally responsible, it does seem to be a rational behavior that I exhibit voluntarily. At this impasse, I would like to offer a compromise: what Fischer calls moral responsibility I suggest we call intentional responsibility. The subset of behaviors for which one is intentionally responsible and involve moral reasons (in virtue of which agents are

---

306 This is not to say that agents too cannot be assessed for goodness or badness. On the contrary, the character and intentions of agents can still be evaluated. The suggestion here is that sometimes we evaluate actions and agents separately.
praiseworthy or blameworthy) might be the set of morally responsible behaviors. So far, our picture looks like this:

![Diagram](image)

**Figure 3: Intentional Responsibility: a neo-Fischerian View**

This suggestion has the consequence of strengthening the relationship between moral responsibility and praise- and blameworthiness. It follows from this suggestion that if one is morally responsible for some behavior, she is also intentionally responsible and either praiseworthy or blameworthy. If one were neither praiseworthy nor blameworthy, one might still be *intentionally responsible* (if she had something akin to Fischer and Ravizza’s guidance control). Therefore, it could still be said that what one did was *on purpose* even though no moral evaluation of the agent is called for. However, this proposed alteration of Fischer and Ravizza’s view holds that it is not possible to be
praiseworthy or blameworthy without also being morally responsible. Ultimately, I shall argue that this claim is false.

But first, let us consider an established view that holds both that one cannot be morally responsible without being blameworthy or praiseworthy and that one cannot be either praiseworthy or blameworthy without being morally responsible. Derk Pereboom’s hard incompatibilism is such a theory. Let us see how his view endorses the account of moral responsibility proposed above. Then we can critically assess Pereboom’s theory as a theory of praise and blame and show whether the relationship among moral responsibility and praise- and blameworthiness underlying his theory is perhaps too strong.

Entailment

Derk Pereboom seems to have such a view. He writes, “an agent’s being blameworthy for an action is in fact entailed by his being morally responsible for it in the sense at issue in the debate, together with his understanding that the action was in fact morally wrong.” I take the final qualification (“together with his understanding that…”) to be a qualification concerning which of the two, praise or blame, is called for, and not that the agent could be morally responsible and not to blame if the act were done in ignorance of the moral value of it or its consequences. That is, Pereboom’s present disagreement with

\[\text{Entailment}\]

\[\text{Derk Pereboom seems to have such a view.}^{307}\] He writes, “an agent’s being blameworthy for an action is in fact entailed by his being morally responsible for it in the sense at issue in the debate, together with his understanding that the action was in fact morally wrong.”^{308} I take the final qualification (“together with his understanding that…”) to be a qualification concerning which of the two, praise or blame, is called for, and not that the agent could be morally responsible and not to blame if the act were done in ignorance of the moral value of it or its consequences. That is, Pereboom’s present disagreement with

\[\text{307 Actually, Derk Pereboom’s view is what I call a “hypothetical theory of responsibility, praise and blame.” Pereboom does not actually believe in moral responsibility or the justification of praise and blame. However, he offers a hypothetical account of what these would be, were humans equipped to be bearers of such properties.}\]

\[\text{308 Fischer et al. (2007): 99.}\]
Fischer coupled with his conclusions in his work, *Living Without Free Will*, suggest that on his view, moral responsibility and an agent’s worthiness of praise or blame stand or fall together. Therefore, from this claim it follows that, had the person not understood that the action was morally wrong, he would not have been morally responsible.

While I agree with Pereboom that “an agent’s being blameworthy for an action is in fact entailed by his being morally responsible for it…,” I do not agree that an agent is blameworthy if and only if she is responsible for an act that she knew to be wrong. That is, I do not agree that one’s being blameworthy or praiseworthy entails that one is morally responsible in the sense at issue.\(^{309}\) In order to illustrate my reasons for diverging from both Fischer and Ravizza and Pereboom’s assessment, we must consider the role of causation in moral responsibility.

**Moral Responsibility and Causal Responsibility**

One of the reasons that I suggest that we recognize a higher class of behaviors as intentionally responsible is that we seem to implicitly recognize an even broader class of actions as responsible actions *simpliciter*. I am referring to the apparent assumption that moral responsibility is a species of responsibility in general. And this general sort of responsibility seems to be causal responsibility.

\(^{309}\) This is a consequence of his view, *hard incompatibilism*. Pereboom argues that we lack the metaphysical makeup to be responsible for our behaviors and, for this reason, praise and blame are unjustified (Pereboom (2001); Fischer et al. (2007): 110-114). Therefore, while Pereboom does not explicitly declare this to be part of his view, it is a natural consequence of it. If praise and blame are unjustified because agents are not morally responsible, then, according to Pereboom, moral responsibility is required for praiseworthiness and blameworthiness.
When we speak of one being morally responsible, we consider that person to be morally responsible for some event or outcome and it must be the case that that person be the one who caused said outcome. If it is learned that a person is not causally responsible for an outcome, the person is cleared of moral responsibility. For example, imagine that Christine’s bicycle was stolen, and she believes it was stolen by Eric. Naturally, she would blame Eric and hold him responsible. However, if she were to learn that it was, in fact, stolen by Tom, she would no longer blame Eric or hold him responsible. The knowledge that it was not he who stole the bike alters such a judgment because it matters that the person being judged caused the harm.

When an agent has caused something of moral significance, we tend to include both the actions and their consequences among the class of things for which the agent might be responsible. That is to say, when we treat people as morally responsible for P (a bodily movement) we also typically treat them for many of the things caused by P; so it seems that we treat causal involvement as pro tanto reasons supporting responsibility claims. For example, if George voluntarily and maliciously shoots Martha, we do not say that George is responsible only for pulling the trigger; nor do we say that George is merely responsible for choosing to shoot. If his choice and effort to shoot Martha are followed by successfully killing her, we say that George is responsible for choosing to shoot Martha, pulling the trigger, hitting her, and killing her (even though the last few items were consequences of the first). George is responsible for all of the consequences of his action since the bodily movement he caused also led to these consequences.

---

310 The evaluation I employ here is similar to that of Donald Davidson in his paper “Actions, Reasons, and Causes” (1963). Here he argues that a single basic action (what an agent physically does) can be individuated multiple ways based on re-describing the action (here we are using the various consequences of that action)—both intentional and unintentional (see especially section II, pp. 4-5). While Davidson’s
While causal responsibility is certainly not sufficient for moral responsibility, it does seem to be a necessary condition. Imagine that in the above scenario, just seconds prior to George’s presentation of his gun, Martha dies of a heart attack. George cannot be considered responsible for Martha’s death. While Martha is dead, and while George intended for her to be dead, he is not morally responsible for her death. Nothing that George did caused Martha to die. This example suggests a necessary condition of one’s being responsible for an event obtaining is that one cause (or perhaps causally contribute to) that event. If this is correct, then moral responsibility is not only a type of intentional responsibility, but also a type of causal responsibility.

We are responsible for these things because of the causal role that we play in them. But ascriptions of moral responsibility are not apt in all cases of causal responsibility. For example, if Les has a seizure that results in his knocking over a bookshelf at the library, Les is, in some sense, causally responsible for the bookshelf being knocked over. However, he is not morally responsible. He is not even responsible in the sense that Fischer’s battered wife is responsible since it was not Les’s reasons-responsive mechanism that caused his shelf-tipping behavior. He is, at most, merely causally responsible; he is causally responsible in the sense that he was an integral object in the chain of events, without which no shelf-tipping would have occurred.

---

paper is not specifically about being responsible for the many actions that are individuated from a single choice, its subject matter is useful for our purposes. Here I only wish to show that we are responsible for actions and what counts as an action can depend on the nature and number of the consequences brought about by the basic act.
Now, our taxonomy of responsibilities might be illustrated as such:

**Figure 4: Types of Responsibility According to Entailment Views**
Non-actions: Omissions and Negligence

Figures 3 and 4 illustrate how standard accounts of moral responsibility and praise- and blameworthiness like Fischer and Ravizza’s and Pereboom’s consider all instances of praise- and blameworthiness to be examples of causal responsibility. However, there are certain examples that seem to constitute examples of moral responsibility or praise- and blameworthiness that, arguably, do not involve causation at all.\(^\text{311}\) Such instances are cases of *omissions* and *negligence* and they have received a lot of attention from responsibility theorists in recent decades. Most theorists recognize that various instances of omitting a behavior or neglecting to behave in a certain way intuitively seem to count as appropriate scenarios for placing praise, blame, and retrospective responsibility. However, since the majority of views ground praise- and blameworthiness and moral responsibility in actions, it is often difficult to explain how such instances of *non-actions* also qualify as worthy of moral appraisal. I think that my moral attitude account is better equipped than standard action accounts to explain such cases. When one omits to do something that one ought to do, one does not have the right moral attitude towards reasons to perform the act. When one is negligent in a morally sensitive situation, one does not have the right moral attitude towards that situation.

On the other hand, my attitudinal account is not the only view that discounts the value of outward behavior. As an account concerned with acts of will, Arpaly’s targets

---

\(^{311}\) I say that it is arguable that such instances are not causings since, to accommodate pictures like Figures 3 and 4, theorists often consider omissions as a kind of action. I do not think that such categorization is correct. Still, as I shall argue in the next section, I also believe that there are cases where people hold attitudes towards moral reasons that are deserving of praise or blame, and such examples cannot be accommodated by any system that considers causal responsibility necessary for praise- and blameworthiness.
internal factors as grounding praise- and blameworthiness. Therefore, it is worthwhile to see why we should prefer my view to Arpaly’s. In order to show that my view is preferable, let us consider how Arpaly’s view is equipped to handle omissions and negligence and then I shall argue that specific examples of these non-actions are better explained by my view.

On Arpaly’s view, discussed at greater length back in chapter 4, one is blameworthy when one takes the wrong course of action out of lack of good will or out of ill will (the degree of blameworthiness being sensitive to which quality of will leads one to the wrong action and to what degree). Similarly, praiseworthiness is a feature of an agent who takes the right course of action out of good will. Unlike the view I wish to endorse, this view requires that agents are praiseworthy and blameworthy for actions only.

I think that to identify the truthmaker of a property of praise- or blameworthiness as a quality of will is too specific. On a standard understanding of what a will is, it involves some kind of endeavoring towards some action. If one considers such an endeavoring to be a type of action (regardless of its success or whether it is thwarted), then every instance of willing is also an action. Unlike Arpaly, I think there is a more basic attitude that is sufficient for bearing properties like praiseworthiness and blameworthiness that does not automatically entail the presence of an action. The above examples demonstrate why restricting praise- and blameworthiness to willings or acts of will might be too specific. Still, I think that Arpaly’s view is still a move in the right direction away from actions as overt behavior.

Arpaly’s view is better equipped to explain how one can be worthy of praise or blame for omissions as well as for negligence. Let us first consider omissions. I take omissions to be failures to act, where the failure can be voluntary or involuntary. That is, one can fail to rescue a swimmer even though the person tried (i.e., it is an omission due to a lack of success in what the person was trying to do) and one can fail to rescue a swimmer because the person was too busy working on his tan (i.e., the person did not even try to do what he failed to do). It is a widely held view that one can be morally responsible for one’s omissions, barring certain conditions. One such condition might be one’s ability to do other than one does, or, perhaps more importantly, one’s ability to succeed in doing the action the person is being blamed for omitting. Even compatibilists are often compelled by the principle of alternate possibilities when it comes to omissions. For example, John Martin Fischer defends:

…the thesis that there is a basic asymmetry between moral responsibility for actions and moral responsibility for omissions. One can be morally responsible for performing an action, even though one couldn’t have avoided performing it—moral responsibility for action does not require freedom to do otherwise. But one can’t be morally responsible for failing to perform an act, if one couldn’t have performed it—moral responsibility for things we have left undone (omissions) requires freedom to do the thing in question.313

I do not think that this is right. First, I do think that one can be morally responsible for failing to do something that the person could not have successfully done. Second, I think

that, even if it can be shown that there are cases where one cannot be held morally responsible for an omission of a certain type due to her inability to do otherwise, I think we can still find her praise- or blameworthy (again, a finding that is not possible on the standard view that gives primacy to moral responsibility).

In order to illustrate this point, let us consider one example, both Arpaly and Frankfurt’s evaluation of it, and my alteration of it. The following example comes from Peter van Inwagen:

Suppose I look out the window of my house and see a man being robbed and beaten by several powerful-looking assailants. It occurs to me that perhaps I had better call the police. I reach for the telephone and then stop. It crosses my mind that if I do call the police, the robbers might hear of it and wreak their vengeance on me [etc.] … So I decide “not to get involved” … quite unknown to me, there has been some sort of disaster at the telephone exchange, and that every telephone in the city is out of order and will be for several hours.314

Van Inwagen concludes that he is not responsible for failing to call the police. He is not responsible for failing to call since, had he tried, it was not possible for him to succeed. While he does not speak specifically about whether he is blameworthy, it is clear that if he subscribes to the traditional view of moral responsibility (the one that gives primacy to moral responsibility), he would have to declare that he is also not blameworthy. But this conclusion seems very counterintuitive to me. Frankfurt and Arpaly agree.

---

Frankfurt points out that this case fails to take into account the fact that the reason Peter does not call the police is because he did not try to call the police:

If the telephones were working, it might be more appropriate to refer to his failure to call the police; if they were out of order, it might be more appropriate to refer to his failure to try to call the police. But both the quality of the moral judgment and its degree—whether P is blameworthy or praiseworthy, and to what extent—will be exactly the same in both cases.\footnote{Frankfurt (1988f): 99.}

Granted, had Peter tried to call the police, he would have failed. But that is not the point. Like all Frankfurt-style cases, the agent is responsible if what the agent does (in this case, omit calling the police) is done due to the fact that the agent controlled his or her body in such and such a way. So the fact that Peter could not do otherwise does not exempt him from responsibility for what he did do, which includes putting the receiver down and sitting still. If we were to imagine two similar possible worlds, one like the world described by van Inwagen and another quite similar, except for the fact that the phones work, is there more cause for blame in the second world? Both Arpaly and I agree with Frankfurt that there is not. However, we agree for reasons that are incompatible with Frankfurt’s, because he also adds:

But the reason why the moral evaluation of P will be the same in either case is not that he is subject to moral praise or blame exclusively for his psychological characteristics. What a person does is not relevant to moral evaluations of him merely because it is an indicator of his mental state.
People merit praise and blame for what they do, and not just on the basis of what they do.\footnote{Frankfurt (1988f): 100.}

Frankfurt seems to be suggesting that the man in van Inwagen’s example is morally responsible and blameworthy for sitting still. Neither Arpaly nor I have a problem with considering sitting still an overt behavior (even if only a non-movement). It is true that remaining in place can be considered an action. For example, a child playing freeze tag can be taking an action, that of staying frozen, by remaining still. Also, if one agent tells another not to move, that second agent can presumably obey the command of the first by not doing anything. However, it seems natural to consider obeying a command to be an action.\footnote{Both of these examples were given to me by Timothy Schroeder in conversation. In particular, it is compelling to consider and agree to the second case. We can imagine a case where the first agent tells another person who is asleep or in a coma to stay still. It would be absurd to think that the prostrate person is obeying the command. That person is not aware of the command. So, certainly, obeying a command is more than just the overt behavior matching the content of the mandate. It seems like some kind of will (and therefore, action) must be involved.}

What is striking about Frankfurt’s claim here is that, had there been no such voluntary bodily movement, there would be nothing for which Peter would deserve or warrant blame. For example, imagine that Peter never grabs for the receiver. Imagine a similar scenario where Peter is, unbeknownst to him, temporarily paralyzed. He deliberates as if he could reach over and try to call the police. He chooses not to. All the while, he made no movements. He thinks he has willfully remained still; but, unbeknownst to him, his stillness is in fact a result of his paralysis. I still think that Peter is blameworthy for his attitude towards calling the police and I think it is only because of this feature of his “psychological characteristics” that he is blameworthy. I am not the only one; I think that Arpaly would (and has already preemptively) agree(d) with me.
Arpaly’s own assessment of this case is that, regardless of whether Peter is responsible for not calling the police (or even for not trying to call the police), he is blameworthy. She writes:

I agree that as long as it is true that calling the police is impossible, the citizen in question is not responsible for not calling the police, even if it is true that he did not try to do so. On the other hand, whether or not calling the police was possible—even in a case in which trying to call the police was ruled out by the person being tied effectively to a chair—the mere fact that the citizen would not have called the police if he could have speaks ill of him, makes him blameworthy. The fact that the citizen was unable to (try to) call the police excuses her from having to compensate the victim’s family. It does not, however, make me unjustified in feeling disgusted by her…

Here Arpaly would seem to be suggesting something more like the account I want to advance; for she is blaming something about the agent that is more fundamental than what the agent does. However, she still seems to be focused on what willings the agent did not or would not have had (that is, even if she could have had them). I argue that the willings alone are not what is at issue, for someone could will to call the police in a similar circumstances and still have a blameworthy moral attitude. For example, imagine an alternate scenario involving Paul and the potential police call. In this scenario, like Peter, Paul does not want to call the police. He is about to decide, like Peter did, “not to get involved” when another thought occurs to him: what if it is later proven that Paul did

not try to call the police and he is found guilty of violating a good Samaritan law?

Frustrated at his lose-lose situation, Paul begrudgingly grasps the receiver. Much to his relief, there is no dial tone. Something is wrong with the phone. He cannot be legally charged for breaking the good Samaritan law, and he does not have to get involved. He gleefully returns to watching his daytime television shows.

Now, as the case of Paul shows, we can have a person willfully do the right thing for the wrong reasons. In other words, it is not the case that Paul would have failed to call the police if the phones were working. Now, this is not to say that Arpaly is not equipped to deal with Paul’s situation. I imagine she would say that Paul is still blameworthy because he was not responding to the moral reason to call. She would agree with me that Paul is still blameworthy for failing to have the right motive to call the police. However, I suggest that the reason that we can say this has more to do with Paul’s moral attitude towards the right reason that underlies how he responds to (or neglects to respond to) relevant features of his situation. Therefore, while I agree with Arpaly that Paul is blameworthy and also that Paul’s motive is a poor response to a situation in which moral reasons were present, I would also add that the reason why Arpaly and I can make this claim is that Paul harbors either an indifferent or negligent attitude towards moral reasons to help others and to be concerned for their welfare.

On my account, an agent can be praiseworthy or blameworthy concerning a morally relevant attitude, where apathy and indifference count as examples of such attitudes. On Arpaly’s account, indifference to morally relevant reasons is blameworthy. However, such blameworthiness must be a property borne by the agent for some volition. I do not think that this is correct. I think that blameworthiness is grounded by a moral
attitude, which could move one to have a volition, though it need not do so. We can imagine indifference in cases such as van Inwagen’s where the agent considers and then rules out a possible action. So indifference does sometimes manifest as willful omissions, and I imagine that these are the kinds of cases Arpaly has in mind. Still, what my view includes and Arpaly’s requires revision to successfully include is insensitivity. On my account, an agent can be blamed for not even considering a moral reason. For example, consider a third version of van Inwagen’s scenario—Phil. Like Peter and Paul, Phil looks out his window to see “a man being robbed and beaten by several powerful-looking assailants…” Unlike Peter, it DOES NOT “occur to [him] that perhaps [he] had better call the police.” Instead, Phil thinks to himself “that’s strange,” and casually sits back down. It does not occur to Phil that he should even deliberate over whether to call the police. While Arpaly might think that such examples qualify on her account, I do not see how this could be the case. For her account requires that an agent be praise- or blameworthy for a willing. However, if a person is so insensitive to moral reasons that he does not realize that he ought to consider a particular moral consideration, how can that person be said to willfully disregard it?

There seems to be another type of omission that is not willful or deliberate at all but might still be a candidate for praise- and blameworthiness. Angela Smith describes such situations in her example of forgetting a friend’s birthday:

I forgot a close friend’s birthday last year. A few days after the fact, I realized that this important date had come and gone without my so much as sending a card or giving her a call. I was mortified. What kind of friend
could forget such a thing? Within minutes I was on the phone to her, 
acknowledging my fault and offering my apologies.\textsuperscript{319}

Unlike van Inwagen’s case, Smith:

Did not consciously choose to forget this special day or deliberately 
\textit{decide} to ignore it. I did not \textit{intend} to hurt my friend’s feelings or even 
\textit{foresee} that my conduct would have this effect…And yet, despite the 
apparent involuntariness of this failure, there was no doubt in either of our 
minds that I was, indeed, responsible for it.\textsuperscript{320}

In Smith’s case, she is to blame for something she did not willfully do. In fact, that is the 
point. Her friend’s birthday was, in a sense, “off the radar” of the will, and it should not 
have been.

Smith, like others, considers moral responsibility and praise- and 
blameworthiness to be intimately related where being responsible for something \textit{just is} 
the agent’s being “\textit{open} to moral appraisal on account of it.”\textsuperscript{321} Her verdict as to why she 
is to blame for forgetting her friend’s birthday is that failing to notice or neglecting 
something that one ought to notice and attend to is indirectly connected to our judgments 
about what is important or significant. Since we can always be asked to account for 
reasons supporting our judgments, we can also be held responsible such that we must 
give an explanation for intentional mental states that are so indirectly linked to our 
judgments.\textsuperscript{322}

\textsuperscript{319} Smith (2005): 236.
\textsuperscript{320} ibid.
\textsuperscript{321} Smith (2005): 238.
\textsuperscript{322} Smith (2005): 270.
I agree with Smith in many ways. For example, I think that negligence can be blamed even if it is not voluntary or chosen; I believe that this is the case even if the negligence involves no particular will that is a response to a reason. However, I do not think that the moral attitudes we are to be praised or blamed for are parasitic on our judgments. That is, I think there are moral attitudes that are praiseworthy or blameworthy and are distinct from judgments. For example, a frustrated mother might truly judge that physically harming her children is morally unacceptable. However, it makes sense for her to feel guilt if, when the children misbehave, she desires to beat them severely. If we were to learn of this desire we would be horrified and reproach her for feeling that way. She might even agree with our assessment, since she cognitively judges that her children are innocent and do not deserve any such treatment.\footnote{This example is an adaptation of Gary Watson’s example, originally intended to demonstrate the difference between valuing and desiring: (2004): 19.} Judgments might be one of the types of moral attitudes for which we deserve praise or blame. However, I do not think that all such attitudes are reflections of judgments of what is important or significant.

Similarly, Arpaly has argued that one can be akratic in a virtuous way. If one’s judgment about what one ought to do is flawed, and one does something against that judgment in response to something that is actually a good moral reason to recognize and respond to, then that person is worthy of praise. Her example is that of Huck Finn. Huck reasons that he ought to turn his friend, Jim, in to the authorities. Jim is a slave and is therefore property. When one finds missing property, one ought to return it. However, Huck simply cannot bring himself to do it. He loves his friend and simply wants him to have what will make him happy—freedom. He allows his moral attitude towards Jim to override his judgment about what he is obligated to do. I agree with Arpaly that, despite
his flawed judgment, Huck is praiseworthy because he did the right thing by responding to an actual moral reason.\textsuperscript{324}

**Blameworthiness, Praiseworthiness and Causal Responsibility**

Figure 4 indicates that moral responsibility requires either praiseworthiness or blameworthiness and that every instance of moral responsibility is also an instance of intentional and causal responsibility. In this section, I shall argue that blameworthiness and praiseworthiness are independent from causal responsibility. If correct, then \textit{a fortiori}, praiseworthiness and blameworthiness are independent from moral responsibility, and we need not look for necessary conditions of moral responsibility when considering instances of praise- or blameworthiness.

In this final chapter, I aim to argue for my theory of praise- and blameworthiness as one that is centered on moral attitudes and, therefore, distinct from causal responsibility as well as its sub-species of intentional and moral responsibility. After presenting my theory of praise- and blameworthiness, I shall sketch my view of moral responsibility. I hold that moral responsibility is derivative of praise- or blameworthiness in conjunction with intentional responsibility.

In order to illustrate why I consider praise- and blameworthiness distinct from all species of causal responsibility, let us recall and assess George in each of the two cases that we considered earlier: in the first case, he successfully shoots and kills Martha and in the second, Martha dies of natural causes just before George shoots. As we concluded,

\textsuperscript{324} Arpaly (2002): 228-231.
George is not responsible for Martha’s death in the second scenario. He cannot be responsible for something that he does not cause. However, does this mean that his worthiness of blame is different? To my mind, George is equally blameworthy in each situation. Whether he successfully kills Martha in the proposed scenarios is a matter of luck. However, whether he is deserving of blame is a feature of his ill will and malicious attitude. Regardless of how successful his will is in these scenarios, George is blameworthy. And if my assessment is correct, it indicates that blameworthiness can and does exist independently of moral responsibility. For it is not the case that George is less blameworthy in the second scenario. And if George is equally blameworthy in each scenario, but not equally morally responsible, how is it that blame is a feature of moral responsibility? How is it that the same level of blame is sustained in the second scenario when it is not supported by the same degree of moral responsibility?

In addition to the asymmetrical relationship between degrees of responsibility and blameworthiness, this point can be further illustrated if we imagine a third and fourth George. Third George never tries to kill Martha, but often plans it without following through. He is still blameworthy for plotting to kill Martha even if he does not actually try to kill her. This is evident when we compare third George with fourth George. Fourth George loves and cares for Martha, and he never plots to kill her or anyone else, for that matter. We can even imagine that, on the outside, third and fourth Georges live seemingly identical lives, with every overt behavior in common. I am still inclined to

---

325 Incidentally, this claim has the consequence of shifting traditional problems of moral luck from assessments of an agent’s guilt to assessments of an agent’s responsibility. There is still an element of luck, but it suggests that the luck happens after the agent is blameworthy. This might have interesting consequences for the problem of moral luck. I am comfortable with these results because I have never been moved by objections concerning moral luck that suggest two agents with similar intentions could be assessed differently based on lucky consequences. This might be an issue worth pursuing in future projects.
blame third George for features of his internal life and, in comparison, judge fourth George to be a far more praiseworthy individual (both on any given occasion when third George is plotting and scheming as well as over the course of their lifetimes). If neither George causes any harm to Martha and neither intends to do so (although third George does come close), then how is it that I can say that one is more blameworthy than the other? It certainly cannot be the case that one is more blameworthy in virtue of actions or effects that one causes. It must be that one can be blameworthy in virtue of certain attitudes one holds, regardless of their causal efficacy and the control the agent has over them.

To see how attitudes and actions can come apart and have importantly different values concerning moral responsibility and praise- and blameworthiness, let us consider a different case in which someone is blameworthy, even though the person had no control over relevant events. For example, imagine that Terry and his child-hating friend, Bob, are driving carefully down a road one day when a toddler jumps out from between two parked cars into the street in front of Terry’s vehicle before he is able to slam on the breaks. The child dies. It can hardly be said that Terry (or Bob) is morally responsible for killing the child. Even though he is causally responsible for it, Terry is not intentionally responsible for the death of the child. As our figures have indicated, intentional responsibility is necessary for moral responsibility. However, when Terry discovers that Bob was silently chuckling inside at the thought that they had “rid the world of one brat,” Terry blames him for taking pleasure in this tragedy. Now, Bob was not causally responsible for the death of the child. Also, his amusement is not under his control. We can imagine that Bob is trying with all of his might to stifle the involuntary laughter that
sneaks out from behind his lips. Therefore, his callous reaction is similarly not under his control. That being said, a blaming attitude towards Bob seems apt under the circumstances. If blameworthiness depends on moral responsibility and moral responsibility depends on causal responsibility, then Bob cannot be blameworthy for his attitude towards the child’s death. Given the traditional assumptions about the relationship between moral responsibility and blameworthiness coupled with the conclusion that Bob had no control over the death of the child or his reaction to it, Bob is, evidently, not worthy of blame in this situation.

This situation leaves us with two options. Either, we can cleave steadfastly to our previous picture and accept that we have no right to blame Bob, or, we can adjust the picture to include blaming people in similar circumstances. I am inclined to take the second option. Not only do I believe that Bob is blameworthy in this case, I also believe that there are analogous instances of praiseworthiness that we would have trouble explaining if the traditionally held assumptions were true.

For example, imagine that Megan’s daughter was brutally murdered by a man just convicted of the crime. At sentencing, Megan and other family members are allowed to address the murderer, and he is also allowed to address the court. Megan, like most of us I shall presume, is very angry with this man. When her turn to speak comes, Megan’s attitude changes. She sees the murderer’s mother, visibly distraught. She tells the court that she had prepared to give a wrathful speech; but just now, her heart is softened by the pain of the other mother. As sad as she is to lose her daughter, she no longer feels the need to harbor anger and resentment for the man. Indeed, she believes that she forgives him for what he has done. Even to her own surprise, Megan is sincere in speaking these
words. She truly does feel differently towards the man. To my mind, this new attitude of Megan’s is praiseworthy. I believe that she has every right to be furious with this man and to never forgive him. Such forgiveness is supererogatory; the man does not deserve her forgiveness and Megan is therefore not obligated to forgive him. However, such forgiveness is, in a way, *morally beautiful*. Since forgiveness is not really an act, but rather a change in how we regard another, Megan’s praiseworthiness would seem to be grounded by an attitude and not an act of will.

Now we are in a position to summarize the proposed account. Taking a modified version of Arpaly’s assessment of praiseworthiness and blameworthiness, let us consider one praiseworthy in a morally relevant situation for having a morally appropriate (or perhaps, supererogatory) attitude towards a morally relevant aspect of a situation and let us consider one blameworthy in a situation for having a morally indifferent or bad attitude towards a morally relevant aspect of a situation. Now a complete picture of the relationship between praise- and blameworthiness and responsibility can be illustrated as such:

---

326 I realize that there could be dispute over whether there is a class of attitudes that are appropriate but “par for the course” and, to deem them praiseworthy might result in watering down the nature of this property. For example, I leave it to individual theorists to decide whether having the attitude that murder is wrong in general is par for the course and one is only praiseworthy when one has strong motive to kill someone and still respects that person’s right to live.
Praise- and blameworthiness (Independent of causation)

Causal Responsibility

Intentional Responsibility Mere Causal Responsibility

Moral Responsibility Mere Intentional Responsibility

Praiseworthy Blameworthy

Figure 5: The Moral Attitude View
In short, I hold that there is a relationship between moral responsibility and praiseworthiness and blameworthiness on a realist account. However, it is not a relationship that gives primacy to moral responsibility. We are morally responsible in situations that are morally relevant, where we exhibit some type of overt behavior that has morally relevant causal ramifications, and when such behavior is issued intentionally by a volition resulting from an underlying attitude towards the morally relevant aspects of the situation. In addition to this picture of moral responsibility, I have argued that we should consider the worthiness of praise or blame to be of primary significance when making ascriptions of moral responsibility. In fact, I think that the above analysis has shown that whether one is worthy of praise or blame is of the utmost importance to ethical concerns. Moral responsibility is more of a byproduct of such properties’ manifesting in a web of cause and effect. Recognizing moral responsibility as subordinate to the worthiness of praise and blame can even prove to be helpful in our dealings with other conceptual difficulties in ethics.

**Where Do We Go From Here?**

Now that we have considered the positive theory put forth in thesis, let us review what we have done thus far, so that we can consider where we should go from here. If my theory is correct, many old problems in moral psychology should be revisited and, most likely, will be ripe for new solutions.

We have seen how many contemporary theories favor either sourcehood or reasons-responsiveness as integral to moral responsibility and praise- and
blameworthiness. I have argued that we are praise- and blameworthy for our moral attitudes in relation to moral reasons. Therefore, I contend that a reasons-responsiveness approach to praise- and blameworthiness is appropriate, since praise and blame are justified given our attitudinal reactions to moral reasons. However, unlike contemporary reasons-responsiveness theorists, I do not think that the only moral attitudes that matter are willings or other attitudes resulting in action. However, actions and willings are relevant to retrospective moral responsibility, since it involves causation and consequences. Therefore, I think a future project might be to build upon the account offered here of praise- and blameworthiness by constructing one about moral responsibility. I have already suggested that retrospective moral responsibility would depend on an agent’s praise- or blameworthiness. However, notions of causation, liability, and accountability would need to be investigated and fleshed out more than I have been able to do thus far in this thesis. That is, since we are morally responsible for outcomes, an adequate theory of moral responsibility would need to consider how we cause things, whether degree of causation should be taken into account, how credit or liability factor in, and how we may pragmatically approach social practices reflective of these results. My theory only suggests how praise and blame may be justified. My theory is not helpful in determining how we ought to punish and hold others accountable. This is a task for future work.

Another question raised by my view is what is a moral attitude? I have suggested some examples such as canonical propositional attitudes (e.g., beliefs, emotions, desires, hopes, etc.). However, it might be interesting to refine a notion of moral attitude, so as to bar the view from potential criticisms based on a misunderstanding of what a moral
attitude is. For example, we can imagine that a critic might complain that my account yields the consequence of considering everyone praise- and blameworthy for an exorbitant number of mental states. My initial inclination is to consider defending a notion of sincerity—some mental states are fleeting and alien to our values, selves, and goals. For example, I might simply think, as I walk across a high bridge, “I technically could jump off of this thing!” However, when such a thought pops into my head, I do not sincerely have a positive attitude towards doing so. I am not even responding to any moral reasons when such a thought crosses my mind. I am simply thinking. Therefore, a point of future investigation might be to consider the nature of sincerity in an attitude and decide whether it can be illustrated that there is a threshold of sincerity a mental state must reach to qualify as a moral attitude deserving of praise or blame.

Such an investigation might have an interesting consequence. While this clarification does limit the number of mental states in virtue of which we are praise- and blameworthy, it does allow for the possibility that one could be praiseworthy and blameworthy at the same time. That is to say, I might have conflicting, yet sincere, moral attitudes. However, I do not think that this consequence is necessarily a bad one, since it is likely the case that we are all praiseworthy for some attitudes and blameworthy for others even when we maintain two such attitudes simultaneously. For example, when struggling with a moral decision, we are deserving of some praise for recognizing the

---

327 As a matter of fact, it is *likely* that many of us will be praiseworthy and blameworthy at the same time all the time. I say that it is likely for two reasons. First, I do not require that one be consciously attending to a moral attitude in order to be deserving of praise or blame for it. For instance, one’s racist attitude is always blameworthy; it is not blameworthy only when the agent is thinking about it. Second, I believe that many of us have conflicting attitudes. Consistency might be something towards which we strive, but it is hardly a ubiquitous characteristic of human minds. Therefore, it is likely that many of us are praiseworthy and blameworthy all of the time. However, I am not uncomfortable with this consequence of my view. I think it is simply a fact that we are often praiseworthy and blameworthy at the same time. I consider this a fact of the human condition. Any discomfort one has with this consequence might be reducible to what existentialists considered typical human angst.
moral reasons, but also deserving of some blame in virtue of attitudes that prevent us from responding to the moral reasons immediately. I say that this is not problematic, since I believe that a nice clean division of praise- and blameworthiness within a person is only important when we aim to judge a person as a whole, such as when we are faced with the task of considering how to dole out punishment.

On the other hand, punishment is distinct from blameworthiness. I suspect that we are sometimes blameworthy to a degree that does not meet the threshold necessary to warrant punishment. For instance, if Megan were to desire to see the murderer of her child tortured severely, but she does not do anything to bring such a situation about, I do not think that she deserves to be punished. I suspect that the desert of punishment is more closely related to moral responsibility and action than the desert of blame. Still, as the question of the justification of punishment is one of moral desert, it is likely that it is related to praise- and blameworthiness. While this thesis does not address desert of punishment, if and how it might be related to praise- and blameworthiness or moral responsibility is an interesting question, and such an inquiry might make for an interesting possible future project.

In addition, acceptance of my view might require us to revisit canonical problems in moral psychology and ethics. For example, the problem of moral luck might be worth reconsidering. For example, Joel Feinberg has proposed one theory of collective responsibility based on the idea of moral luck (i.e., the worry that multiple people could be more or less blameworthy/guilty based on factors beyond their control even if all
actions within their control were performed in an identical manner). Feinberg asks us to imagine that we are among a group of individuals who drink to impairment, yet still drive vehicles home anyway. Perhaps we do not drink to excess, yet we drink beyond the legal limit to drive. Some of us hit and injure others, and some of us are “lucky” and do not. Feinberg explains that under the arrangement he calls, *Group Liability with Contributory and Noncontributory Fault*, “Most of us are ‘guilty’ of this practice, although only the motorist actually involved in the accident is guilty of the resultant injury.” Philosophers such as Elizabeth Beardsley have since agreed and held that we are all *blameworthy* for such behavior, but only the few involved in accidents are *responsible*. Given what we have seen in the earlier chapters of this thesis, most contemporary theories do not support such an analysis—a person cannot, these theories have maintained, be blameworthy *without* being responsible. My attitudinal account supports Feinberg and Beardsley’s intuitions and even does so with attitudes and not just risky actions. Indeed, I do think that the added significance of moral responsibility might involve a degree of moral luck. However, I also think that if the luck is now shifted to causation and not to what grounds one’s desert of blame, we are in a better position to revisit this age-old problem of ethics.

Feinberg considers this “theory” an *arrangement* of collective responsibility. That is to say, it is not entirely clear that he endorses this view as the way to see collective moral responsibility. He is, more accurately, illustrating how some might come to view collective moral responsibility.


Beardsley (1979).

Bernard Williams tries an analogous shift of the realm of luck in moral thinking. He suggests that whatever is moral must not have any luck involved. In order for one to deserve moral credit or blame, one must not deserve so as a matter of luck. However, he distinguishes the moral from the ethical. For Williams, something is ethical if it is rational. Something is rational if it is in the agent’s interest. However, sometimes a thing can be in an agent’s interest in a way that is not predictable and might require a certain amount of luck. Williams’ own example is that of the famous painter, Paul Gauguin. Gauguin abandons his family to move to Tahiti because he believed that living in a more primitive society would inspire him to
Finally, it might be worth reconsidering the original worries of source incompatibilists in light of my findings. That is to say, would a world in which my account of praise- and blameworthiness and moral responsibility that is also a determined world be any different, morally speaking, from a similar world with indeterminism or agent causes? At the moment, I am tempted to concede to them that certain metaphysical views might result in our being more accountable for our attitudes and behaviors than others. That is, in a determined world, we might almost never be responsible for being blameworthy or praiseworthy. In an indetermined world, it is possible that we could be blameworthy and praiseworthy for our attitudes and be morally responsible for having such attitudes. Still, I am not entirely sure how such a difference would matter. For, in either case, praise and blame are justified. Punishment, too, is justified in cases of action. So, I suppose, one could deserve both blame and punishment for being the kind of person that one is. However, I am reluctant to say that our society is interested in (or could be better off if we began) punishing people for who they are. In any case, I am satisfied in concluding that every metaphysical picture supports sufficient justification for praising and blaming agents for their moral attitudes.

paint great works of art. Morally speaking, abandoning one’s family is simply wrong and there is no amount of luck involved. Rationally speaking, it might be in Gauguin’s interest to move to Tahiti if it turns out that his prediction is correct. However, whether “it turns out” this way or not is a matter of luck. Essentially, Williams’ aim is to discredit Kantian views of morality that equate the moral with the rational. On his view, what is moral is under our control and does not involve luck, whereas what is rational (though, still a matter of value) is not necessarily so under our control. C.f. Williams (1981).
Bibliography


Cambridge, MA: MIT Press.


_Ancient Philosophy_ 10.1: 81-103.


Fischer, John Martin and Mark Ravizza. (1993a) _Perspectives on Moral Responsibility_.

221


http://www.iep.utm.edu/responsi/.


