Moral Responsibility

A Constructivist Account

By

Maria Luz Crebay

Thesis presented for the degree of

MASTER OF PHILOSOPHY

Supervised by Professor Christel Fricke

Department of Philosophy, Classic, History of Art and Ideas
Faculty of Humanities

UNIVERSITY OF OSLO

Autumn 2019

Moral Responsibility

A Constructivist Account

Maria Luz Crebay

Abstract

In his influential piece, *Freedom and Resentment*, P.F. Strawson suggested a new approach to the old debate between freedom and determinism. He proposed understanding agential responsibility by analyzing psychological and sociological questions relating to our interactive practices of praise and blame. Strawson did not give an account of moral responsibility by defending the agent´s exercise of his free will (libertarian account). Neither did he justify our practices of moral responsibility based on their utility (consequentialist account). Strawson´s compatibilism argued instead that ascriptions of responsibility are best understood as "expressions of sentiment" (Shabo 2012:131). Despite Strawson´s significant influence among contemporary accounts on moral responsibility, his theory failed putting an end to the metaphysical debates about agential responsibility.

Influenced by Strawson, Gary Watson famously claimed that our practices of holding responsible are "incipient forms of communication" (2004:230). What does the communication between a wrongdoer and his victim imply about the wrongdoer's responsibility? Does a victim´s resentment (and thus her holding the wrongdoer responsible) contribute to the constitution of the wrongdoers´ responsibility, as an expressivist account would hold? Or does the victim, by virtue of resenting the wrongdoer, merely track a quality of will which he has independently of his being held responsible, as a realist would defend? Is there a third option, namely, a middle ground between expressivism and realism? I think there is — Mckenna´s Conversational Theory (2012). In this thesis, I defend this view. Agential responsibility is best understood as an unfolding conversation between those in the moral community holding wrongdoers responsible and those being held responsible. The topic of such a conversation will be the wrongdoers´ quality of will. It is in virtue of such quality of will that responsible agents open up the possibility of a conversation about the moral values of their actions. The analogy between conversation and responsibility is illuminating, but how far can McKenna stretch this analogy? In this thesis, I discuss the limits of McKenna´s view and defend it against an objection. McKenna's theory presupposes an exchange between the wrongdoer and the victim. However, the pervasive phenomenon of private blame and blame in the absence of the wrongdoer threatens to render his account inadequate. Should we follow McKenna in building a theory of moral responsibility based on an interpersonal exchange between blamer and blamed? I contend that we should do so.

Acknowledgments

"I should have based my judgment upon deeds and not words"
(Antoine de Saint-Exupéry, Le Petit Prince ([1944] 1995 :38).

IX

x

Table of Contents

1 Introduction: Moral Responsibility as a Philosophical Subject

This thesis explores the nature of moral responsibility as it is discussed in contemporary philosophy.[1] Moral responsibility[2] is a deep philosophical issue, both challenging and with a very long history. The following introduction is meant to limit the scope of this thesis and provide some necessary conceptual groundwork. After reading this introduction, the reader should be able to recognize the philosophical problem this thesis aims to engage with, as well as the specific question it will try to answer.

For the sake of clarity, I will start by distinguishing between two kinds of questions about the notion of moral responsibility. To begin with, moral responsibility can be addressed by analyzing the following questions: "what is moral responsibility?" "What is the nature of an agent´s responsibility?" "What does it mean for an agent to *be* responsible?" Secondly, moral responsibility explores the limits of an agents´ *presupposed* responsibility. If an agent is responsible at all, how should we limit her responsibility? For instance, to what extent is an agent responsible for the lack of knowledge regarding the consequences that his deliberate choice brought about? At the core, this kind of questions explores how much responsibility an agent has and what restricts or limits this amount.

The first kinds of questions are metaphysical, which means their main concern is with the nature of moral responsibility and its possibility. The answers to these questions make different metaphysical assumptions. For instance, while some argue that in order to be responsible, an agent must be in possession of a certain kind of control, such as the agent´s *free will;* Others simply contend that being a responsible agent amounts to being an appropriate target of certain emotional responses from those holding the agent responsible. In contrast to the metaphysical kind of questions, the second kinds of questions focus instead on the limits of an agent´s responsibility. There is widespread disagreement about what a responsible agent is responsible for, even if we agree on what the nature of moral agency is.

---

[1] For a thorough and recent treatment on contemporary moral responsibility, see Talbert (2016).
[2] For ease of expression I may sometimes use the word "responsibility" instead of "moral responsibility".

Among these, we can think about the well-known cases of non-culpable ignorance or omissions.

This thesis explores metaphysical accounts of moral responsibility and hence, focuses first and foremost on the first kinds of questions (though presumably both issues eventually overlap). My point of departure is P. F. Strawson´s 1962 influential essay, *Freedom and Resentment*. This work has been a touchstone for most of the philosophical debate on contemporary moral responsibility. Briefly, Strawson aimed to reconcile compatibilist forward-looking (consequentialist) accounts of moral responsibility and incompatibilist accounts of moral responsibility. He proposed understanding agential responsibility by analyzing psychological and sociological questions relating to our interactive practices of praise and blame. Strawson did not give an account of moral responsibility by defending the agent´s exercise of his free will (libertarian account). Neither did he justify the practices based on their utility (consequentialist account). Strawson´s compatibilism argued instead that ascriptions of responsibility are best understood "as expressions of sentiment" (Shabo 2012:131). Strawson claimed that we should leave metaphysical questions about moral responsibility behind and focus instead on our practices of praise and blame. Just by analyzing the moral sentiments and their place in our interpersonal relationships, we can justify agential responsibility without going beyond the "facts as we know them" (2008:22). Or so he claimed.

Given Strawson´s aim in changing the debate about moral responsibility by leaving the metaphysical debate aside, my choice of discussing *metaphysical* questions about the nature of moral responsibility might come as a surprise. After all, Strawson rejected the metaphysical treatment of the free will problem. He argued instead that justifications of our moral responsibility practices are *internal* to these practices (2008:40). However, contemporary discussions on the nature of moral responsibility, make conspicuous that Strawson´s enterprise failed in changing the topic entirely. Even if we focus on our practices of holding responsible, metaphysical questions concerning the agents involved in these practices seem to pull us back into the free will debate. i.e., "Strawson's celebrated proposal to construe freedom and responsibility as constitutive of human society failed to convince enough of us that metaphysical issues cannot have a bearing on the attitudes and perhaps even the practices associated with these notions" (Berofsky; 2000:135).

The traditional way of analyzing the nature of moral responsibility is to focus on the status of the morally responsible agent, namely: What does it mean for an agent *to be* morally responsible? In this traditional sense, in order to be responsible, an agent must meet specific objective criteria. I take it that what Strawson (and Strawsonian views after him) intend to do is to shift the focus from the agent´s status as a responsible agent to the practices of holding people responsible. Instead of answering the question of what it means to be a responsible agent, these accounts try to answer first, "what does it mean to hold an agent responsible". Once answered this question, they contend, it is possible to formulate an adequate theory of agential responsibility. Presumably, by adopting this "reversal"[3] in the order of explanation (Watson, 2004:222), questions on the metaphysical nature of the responsible agent would somehow dissipate. However, the truth is, as Watson famously claimed, "the practice turns out to be less philosophically innocent than Strawson supposes" (2004:221).

Influenced by Strawson, Gary Watson famously claimed that our practices of holding responsible and the expressions of our moral emotions within these practices are "incipient forms of communication, which make sense only on the assumption that the other can comprehend the message" (2004:230). What does the communication between a wrongdoer and her victim imply about the wrongdoer's responsibility? Does the victim´s resentment (and thus her holding the wrongdoer responsible) contribute to the constitution of the wrongdoers´ responsibility, as an expressivist account would hold? Or does the victim, by virtue of resenting the wrongdoer, merely track a quality of will which he has independently of his being held responsible, as a realist account of moral responsibility would defend? Is there a third option, a middle ground between expressivism and realism? I think there is -McKenna's constructivist account.

In this thesis, I examine and critically discuss three *strawsonian* accounts of moral responsibility, an "expressivist," a "realist," and a "constructivist" account.[4] I shall raise criticisms against all three accounts, and try to elucidate which interpretation results in a more

---

[3] The reversal was formulated by Watson´a infleuential reading of Strawson: "In Strawson´s view, there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around: It is not that we hold people responsible because they *are* responsible; rather, the idea (*our* idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another" (2004:222).

[4] Given that all three accounts are metaphysical in kind, this demonstrates, in part, that Strawson failed in putting an end to metaphysical debates about agential responsibility.

plausible theory of moral responsibility. Ultimately, I defend a constructivist account, Mckenna´s Conversational Theory (2012). Agential responsibility is best understood as an unfolding conversation between those in the moral community holding responsible and those being held responsible. The topic of such a conversation will be the wrongdoers´ quality of will. It is in virtue of such quality of will that responsible agents open up the possibility of a conversation about the moral values of their actions. I believe that McKenna´s account represents the more nuanced, original, and plausible theory of moral responsibility among contemporary accounts. In the last chapter of this thesis, I defend McKenna's conversational theory from an important objection.

The structure of my investigation is as follows: Chapter 2, "P. F. Strawson´s Freedom and Resentment," is concerned with Strawson´s Theory of Moral Responsibility. I start (2.1) by laying out what Strawson takes to be the problem with consequentialist and incompatibilist accounts of moral responsibility. Further, I analyze Strawson's theory, following Russell (1992), by presenting his Rationalistic Argument (section 2.2) and his Naturalistic Argument (2.3). Lastly (2.4), I close this chapter by raising criticisms toward both arguments.

In chapter 3, "Responsibility Expressivism and Responsibility Realism", I examine and critically discuss two opposed interpretations of Strawson's original argument: Expressivism (3.1) and Realism (3.2). Within expressivism, I explore the accounts of Gary Watson (2004) and R. Jay Wallace (1994). I raise three problems to the expressivist account: (i) the Non-cognitivist Objection (ii) The problem of Evil, and (ii) the charge of circularity. In section 3.1.1, I examine Watson´s (2004) noncognitivist expressivist account. I argue that this view faces an important problem: what I call the non-cognitivist objection. To solve this problem, I contend that an expressivist account should follow Wallace´s normative account (1994) and adopt a cognitivist construal of the reactive attitudes. Further, I maintain that Wallace´s normative account faces a charge of circularity: the conditions for responsible agency cannot be exclusively understood from the normative standpoint of holding responsible. This is because those standards presuppose a specific kind of agency to which those standards should answer. We cannot define moral responsibility by answering the question of when it is appropriate to hold responsible, namely because the appropriateness rests on the specific kind of agency of the agents involved in the practices themselves. Therefore, Wallace exposes himself to a charge of circularity in his argument. I explore the charge of circularity by discussing (section 3.1.2.) Watson´s paradigmatic case of Robert Harris (2004:225). I propose

4

looking into this example and see whether an expressivist reading of Strawson, can help us determine whether an agent is morally responsible or exempted from our practices, or if in the end, some other account is better fit to the task. Further, I discuss the problem of Evil as discussed in Watson (2004), and in section 3.1.3, I offer a solution to the problem by drawing on McKenna (1998). I then turn to a Realist account (3.2). I critically discuss Brink and Nelkin´s (2013) rational abilities view and examine two main factors that can mitigate or incapacitate an agent´s responsibility: normative competence (3.2.1) and situational control (3.2.2). Further (3.2.3), I problematize the realist account by discussing why the realist remains "hostage to traditional worries about freedom of the will" (Brink and Nelkin 2013:288).[5] Finally, in section 3.3, I turn to the charge of circularity. I reject the expressivist metaphysical and explanatory primacy of our practices. Then, drawing on McKenna (2012), I contend that one way to overcome the vicious circle is to deny that holding people responsible or being responsible should be explanatory or metaphysically more basic; both are mutually dependent.

In chapter four, "Conversation and Responsibility," I explore McKenna's conversational account of moral responsibility (2012). First, I present McKenna´s contribution by pointing out both his affinities and differences with Strawson and Strawsonian accounts of moral responsibility. I argue that despite the similarities between McKenna and Strawsonian accounts, McKenna is a "Strawsonian of a different stripe" (Vargas; 2016:225). In section 4.2, I analyze McKenna´s three-stage structure: The *Moral Responsibility Exchange*. In section 4.3, I discuss McKenna's constructivism by examining the "Interdependence Thesis" as well as McKenna´s quality of will condition (2012:60). In section 4.4, I discuss "the quality of will condition" and defend it against opposing views. Further, I discuss an "apparent" "dis-analogy" in McKenna´s conversational analog. Finally, I problematize McKenna's analogy between conversation and responsibility by raising questions about how to interpret an agent's quality of will, and most importantly, who determines its meaning.

In the fifth and last chapter, entitled "Defending the Conversational Theory: Vargas and Gossip", I discuss an objection raised against the Conversational Theory, and I reject it. Cases of blaming in the absence of the blamed are pervasive. However, Mckenna insists on analyzing responsibility by focusing on the paradigmatic case of directed blame (blame in the

---

[5] An important question to raise in this respect is whether McKenna's view is also hostage to these worries, more on this on chapter 4.

presence of the blamed). Naturally, this issue raises concerns as to how far McKenna can stretch his analogy between communication and responsibility. It is in this context that Manuel Vargas (2016) offers an alternative proposal for the conversational model, the model of gossip. First (5.1), I present Vargas´ criticism of McKenna's theory and his alternative proposal for the conversational model, the model of gossip. Next, I present what I consider to be the main characteristics of gossip (5.2). With a definition of gossip in place, I discuss whether gossip, as Vargas contends, manages to "capture much of what seems appealing about the conversational approach" while providing "an even more social" (2016:238) understanding of our practices of blaming. I disagree. To ground my rejection, I explore the communicative aspect of blame by drawing on recent work from Miranda Fricker, David Shoemaker, and Coleen Macnamara. I argue that the communicative aspect of blame is best captured if we focus on cases of directed blame instead of gossip. Next, I offer four different, albeit related functions of the practice of gossip (5.3). I contend that none of these functions manages to capture what is at the core of moral blame. More importantly, gossip fails to capture the moral sentiments that Strawson so vigorously claimed should be at the center of ascriptions of moral responsibility (5.4). Against Vargas´s intentions of introducing gossip, I defend the strawsonian claim that the reactive attitudes with which we hold people responsible are not "tools" for regulating behavior in socially desirable ways, but rather expressions of human nature, i.e., the reactive attitudes "do not exploit our natures, they express them" (Strawson, 2008:27).

1.1 The Conditions for Moral Responsibility

Although this thesis takes as a starting point Strawson´s seminal paper *Freedom and Resentment* (1962), discussions on the nature of moral responsibility (and its relationship to the free will debate) can be traced back to Aristotle´s *Nichomachean Ethics*. In Book III of the NE, Aristotle explores the topic of moral responsibility by analyzing the nature of human deliberation. He argues that we can discriminate actions as "voluntary" or "involuntary". What is significant about voluntary actions is that only these can be the object of praise and blame (1110a). Voluntary actions, according to Aristotle, are "up to the agent", namely actions in which "the starting point is internal to the agent" (1110a) (1114b). Involuntary actions are actions that, up to some extent, are not up to the agent. For example, acting under compulsion or acting out of ignorance. In the case of the former, some external force (external to the agent) moves the person to act. In the latter, the action is brought about by the agent,

but the agent is not aware of the consequences his action might produce. Imagine, for instance, the case of an agent who presses a button but does not know that doing so will lead to the death of another person. In this case, although the agent´s actions caused the killing of a person, there is a sense in which the act of murder was not "brought" about by the agent since he was acting from ignorance.

Aristotle´s main interest was focused on how personal choices and habits can contribute to forming character (virtuous or vicious). Therefore, it is with no surprise that a person´s choices are so important. After all, these actions seem to play a substantial role in how human beings form their character traits. Since, according to Aristotle, only voluntary actions can be the object of praise and blame (1110a), from his discussions of voluntary actions, we can draw two conditions for moral responsibility. The so-called *control* and *epistemic* condition (also known as the freedom and knowledge condition, respectively). Although Aristotle is recognized as the pioneer of these conditions, there is widespread disagreement about how he meant these conditions should be understood. I now turn to explain what these conditions entail.[6]

Suppose you push someone to the floor. Philosophers writing on moral responsibility agree that for you to be morally responsible for that action, you must meet certain conditions. To begin with, when performing the action, you must have possessed a certain degree of control. This is the control (or freedom) condition. This condition raises the following question: were you freely performing the action of pushing someone to the floor? Suppose it turned out that you were coerced or compelled to perform the action. In this case, you would have failed to meet the control (or freedom) condition necessary for moral responsibility. Despite sharp disagreements with regards to *how much* or the *specific kind* of control an agent must have to meet the control condition, everyone agrees that at least *some* control is necessary in order to meet the conditions for being a morally responsible agent.

Further, in order to be morally responsible, an agent must meet an *epistemic* condition. Suppose this time I pushed you to the floor. However, as it turns out, I am a blind person who just happened to walk into what I thought was an empty room. Most people would agree that I

---

[6] It is important to remark that these conditions though distinct, overlap significantly. For instance, ignorance can limit control. See Mele (2010). In a similar vein, McKenna claims that "it is probably a distortion to think of them (epistemic and control condition) as entirely distinct, since how an agent controls her conduct will be in part a function of her epistemic resources" (2012:13).

was not morally responsible for pushing you over. The epistemic condition considers that the responsible agent must be aware of performing the action in question and of the moral significance the action comprises. As Aristotle puts it, the agent must be aware of what it is she is doing or bringing about (1110a-1114b). Some philosophers argue that the epistemic condition requires knowledge (that the agent knew that it was an actual wrong). Others argue it only requires the belief that the action was wrong.

It is easy to see how both conditions are highly controversial, understood in the broad sense above; there is plenty of room left for further debate. Ever since Aristotle discussed these conditions, philosophers have disagreed on how to interpret them. Further, during the past decades, philosophers have come to distinguish additional conditions.[7]

1.2 The Freedom Condition and the Threat of Determinism

As mentioned earlier, I shall be exploring metaphysical accounts of moral responsibility. What is distinctive about metaphysical questions on the nature of moral responsibility is their strong relationship to the free will debate. Briefly, the free will debate is concerned with the well-known problem of reconciling the notion of free will with the thesis of determinism.[8] I will say more about this in what follows.

The freedom condition for moral responsibility is very much implicated in free will debate. Recalling Aristotle, voluntary actions are actions in which "the starting point is internal to the agent" (1110a-1114b). However, how should we evaluate whether an action is "up to the agent"? to say that an action is up to the agent entails that the agent must exercise a certain kind of control over it. However, the characterization of such control can take very different forms. For instance, some philosophers maintain that the control necessary to freely perform an action requires the agent´s free will. Moreover, these philosophers characterize the notion of free will as the agent´s ability to choose from a range of possibilities unimpeded. If the agent cannot freely choose among alternative courses of action, then he cannot be acting

---

[7] For instance, this thesis explores McKenna's Conversational account, here he proposes three conditions for moral responsibility: control, epistemic and a "quality of will condition" (McKenna, 2012). I discuss the quality of will condition in chapter four.

[8] There are many ways of understanding what "determinism" implies, I shall follow van Inwagen (1983). He argues that the theory of determinism entails that at any instant there is one and only one physically possible future. i.e., that only one future is possible, given the actual past and the laws of nature (1983:3).

freely. The problem with this characterization is that the truth of determinism would undermine such "free" acts, making freedom and determinism incompatible.

As human agents living in the modern world, we see ourselves as capable of influencing our environment, our world through our actions and decisions. We choose things daily; some of the choices we make are easy to forget. Others require more time and may affect us over long periods of our lives, perhaps all our life. To be able to do this, we presuppose a range of alternatives that, in some sense, lie "open to us," alternatives from which, we believe, we can (up to some extent) freely choose. This range of possibilities we believe lies open to us, is deeply related to a further belief, what Robert Kane, paraphrasing Aristotle, calls "up to us-ness" (2005:5). When an action is "up to us" what we usually mean is that we believe ourselves to be the origin of the action, in other words, we believe that we possess the control necessary to do or to refrain from doing things. Naturally, this belief conflicts with seeing our actions as a result of Fate or God´s foreknowledge, even as the product of Nature.

The traditional conception of the agent´s Free Will can be summarized by the following features: (i) that to choose among alternatives is up to us and (ii) that we are the origins or begetters of ours actions. Throughout the past centuries, both features have been highly contested. The question underlying such criticisms is whether free actions (understood as actions performed by the agent´s free will) are compatible with a deterministic metaphysics of the world.[9] Indeed, "the problem of free will arises when humans reach a certain higher stage of self-consciousness about how profoundly the world may influence their behavior in ways of which they were unaware" (Kane; 1996: 95-6). Becoming aware of such influences affects our understanding of moral responsibility profoundly. For if moral responsibility requires the agent´s free will, then either determinism is false and has no bearing on the freedom condition, or it is true, which would thus make responsibility impossible. Further, some argue that we should instead revise the control condition. Perhaps we have set the bar "too high" without need, perhaps free human actions can be compatible with the truth of determinism.

The threat of determinism has been a significant theme during both Ancient and Medieval periods of history (although in the latter, the problem was considered as a theological

---

[9] For an overview of the different versions of physical determinism, see Lucas (1993:20).

problem).[10] Given the main features that characterize the modern worldview[11], it is with no surprise that a deterministic metaphysics of the world would become both appealing and highly intuitive. Take, for instance, people´s pre-theoretical presuppositions about moral responsibility. It has become increasingly common to justify an agent´s action or decision (even in the self-reflective case) based upon considerations which lie in some sense "outside" the agent´s control. For instance, people often appeal to the agent´s genetics or social upbringing in order to explain an act of wrongdoing. Consider the following example,

> After hearing the rumors circulating about how Jacob had (once more) beaten his girlfriend in a public event, and given his previous involvement in a case of sexual assault, a very close friend of his, made the following remark: "How can I blame him? If you knew half of what I know about his parents and the dreadful place in which he was raised, you would feel pity and compassion for him rather than anger and indignation.

As the example above suggests, the belief in the truth of determinism raises the following perplexing question concerning an agent´s responsibility: What would happen to the meaning we attribute to our choices and decisions if they were to be explained by past events which started even before the agent himself was born? This last question raises the even more difficult question regarding our understanding of moral responsibility for wrongdoing: Can I reasonably be held responsible for what is the product of my genes, my childhood experiences, or my cultural environment? i.e., how much of our choices do we think is the product of nurture and nature, and how much do we believe is the product of deliberate choice? And lastly, how does moral luck affect our thoughts about moral responsibility?[12] As for the example above, do we think there is a sense in which at least, up to some extent, Jacob "brought about" the kind of agent he is? Further, would this be enough to hold him accountable for his actions? Finally, does the idea that we *could* have become just like him were we to have been exposed to the same circumstances makes us unable to cast blame on him?

---

[10] Philosophers such as Augustine or Thomas Aquinas were also concerned with these questions, though they related determinism to a theological problem. If God is omniscient and therefore knows about every single aspect of existence, this entails that we are in some sense not free to act otherwise than what God already knows. Moreover, there is the problem of evil. Why does God, as an omniscient being, allow evil to exist among us? Does this imply that God is responsible for evil?

[11] Namely: rationality, physicalism, individualism, and critical thinking.

[12] For a skeptical account of moral luck which shows how our actions are explained by factors beyond our control see, Levy, Neil "Hard Luck" (2011).

The problem of whether free human actions can exist in a deterministic world and hence whether the possibility of compatibilism between free will and determinism, is indeed a profound philosophical challenge. One should not underestimate its influence in our everyday lives. Philosophers who deny the possibility of free will within a deterministic world are known as "Incompatibilists". To understand this position, it might be useful to present Peter van Inwagen´s famous "Consequence Argument,"[13]

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us (1983:56)

Incompatibilism comes in two main and opposite forms, libertarianism, and hard determinism. Both accounts agree on the incompatibility of free will and determinism, but they disagree on the matter of facts. While the libertarian argues that human beings do possess free will and thereby reject the truth of determinism, hard determinists argue that given the deterministic world in which we live in, the kind of free will presupposed by libertarians is impossible. The libertarian notion of free will defends the claim that human beings have an open range of possibilities from which they can freely choose. Thus, the libertarian defends the claim that an agent can choose to act otherwise than he does at the moment of action (this is also known under the name of Principle of Alternate Possibilities famously coined by Harry Frankfurt 1969). The libertarian appeals to phenomenological considerations, such as the "feeling" human beings have of having different alternatives of action. This last argument is not a knock-down argument, of course, but according to the libertarian, a prerequisite for free will is that we have such feeling.[14]

Despite incompatibilism´s popularity and attractiveness, the contemporary debate on moral responsibility has been mostly discussed within a compatibilist framework. Compatibilism has a long history. Some of its significant contributions can be traced back to Thomas Hobbes and David Hume. Compatibilists (also called soft determinists)[15] believe that the truth of

---

[13] For recent formulations of the Consequence argument, see Ginet (1966), Lamb (1977), and van Inwagen (1983).

[14] As Samuel Johnson famously put it, "all theory is against free will; all experience for it" (Boswell, James [1791] 2012).

[15] Fischer and Ravizza defend what they call a semi-compatibilist position. In which they deny the PAP (Principle of Alternate Possibilities) for moral responsibility, and thus accept that moral responsibility and

determinism does not necessarily conflict with human freedom. Compatibilist theories aim at reconciling both determinism and free will by revising or weakening the concept of free will. To the classical compatibilist,[16] the problem between determinism and free human action does not amount to a real problem. For instance, according to Moritz Schlick "freedom means the opposite of compulsion; a man is *free* if he does not act under *compulsion*, and he is compelled or unfree when he is hindered from without…when he is locked up, or chained, or when someone forces him at the point of a gun to do what otherwise he would not do. (1966:59). Thus, he contends, the problem of free will is nothing more than a "pseudo-problem" (1966:54).[17]

However, quite independently of whether one endorses the belief in metaphysical determinism or the metaphysical reality of freedom, according to the moral responsibility skeptic, none of these views provides a convincing account of an agent's responsibility. For philosophers such as Galen Strawson, who strongly reject the idea that human beings can be truly originators of their actions, the whole discussion on moral responsibility is based on the false assumption, namely, the assumption that beings like us *can* be morally responsible. Hence, G. Strawson argues that ultimate responsibility is impossible independently of whether determinism is true or not. In his Basic argument for the "impossibility of moral responsibility" (1994), G. Strawson argues as follows,

(1) Nothing can be causa sui - nothing can be the cause of itself.

(2) In order to be truly morally responsible for one's actions, one would have to be causa sui, at least in certain crucial mental respects.

(3) Therefore, nothing can be truly morally responsible.[18]

---

determinism are compatible, but they argue freedom does require PAP and hence is incompatible with determinism, see Fischer 1994, Ravizza 1994, Fischer and Ravizza 1998.

[16] Classical compatibilism is often associated with the claim that the word *freedom* in the expression *freedom of will* only modifies a condition of *action* and not the agent's will. Schlick is a clear example of this, he speaks in terms of *freedom of action* (Schlick1930 [1966]). Classical compatibilism goes back all the way to Hobbes, Hume and J. S. Mill. New compatibilists include Gary Watson, Susan R. Wolf, P. F. Strawson, and R. Jay Wallace.

[17] Although Strawson intended to leave the old free will debate behind, he was a compatibilist. The realist and expressivist accounts explored in chapter three, are also placed within compatibilism. As for McKenna he defends compatibilism elsewhere (2009) but his Conversational Account (McKenna 2012) and thus his constructivist view is to be regarded as neutral within the free will debate, and thus is supposed to be compatible with both compatibilists and incompatibilists accounts.

[18] G Strawson, (1994:5).

Furthermore, some philosophers that share G. Strawson's view have come to argue that our practices of holding people responsible (such as blaming practices) together with many of our beliefs about moral responsibility should be disregarded and "eliminated" from our human interactions. This "revisionist" account of moral responsibility comes in different degrees. In a weaker sense, revisionism entails that our understanding of moral responsibility should be modified at least up to some extent. In a stronger sense, like G. Strawson´s view above, revisionism entails that the concept, practices, and attitudes we developed around the concept of moral responsibility should be eliminated. It is this latter kind of revisionism, the one usually associated with moral responsibility skepticism.[19]

1.3 From Free Will towards Moral Responsibility

As mentioned earlier at the beginning of this introduction, the traditional way of analyzing the nature of moral responsibility is to focus on the status of the morally responsible agent, namely: What does it mean for an agent *to be* morally responsible? The answer to this question usually focuses on a specific set of objective criteria an agent must meet in order *to be* responsible. For instance, a control and an epistemic condition. Given strong disagreement on how to understand the freedom condition, discussions on the topic of moral responsibility tend to focus on the free will debate. Further, this tendency makes discussions on the nature of moral responsibility remain "trapped" in metaphysical discussions regarding the kind of control an agent must exercise in order to be responsible. The problem with this is that "an account devoted to one condition for moral responsibility is not adequate as an account of moral responsibility itself" (McKenna, 2012: iv).

Strawson was fully aware of this problem. Thus, his aim in *Freedom and Resentment* was to change the topic of the debate: from the free will debate towards moral responsibility. In order to do this, Strawson took a different tack to answer the question of what the nature of moral responsibility is. Strawson´s theory of moral responsibility (which I discuss in the next chapter) focuses first and foremost on the responsive sentiments that naturally arise to people involved in the practices of holding responsible. Whenever we say that we hold people responsible for their actions, we are not only focusing on the agent´s moral status (freedom and epistemic condition) but also *our* reactions towards them and their doings. Perhaps if we

---

[19] See, Pereboom 2001, G. Strawson 1994. For a full defense of skepticism about moral responsibility, see Rosen (2004). An important philosopher considered as a predecessor of this kind of skepticism is thought to be Spinoza (see Spinoza 1677 (1992)).

just focus on these reactions, their internal workings, and their place within our interpersonal relationships, we are able to provide an adequate theory of moral responsibility without incurring into the libertarian´s "panicky metaphysics" (2008:27). In doing so, Strawson thought we would be moving away from the free will debate and towards moral responsibility. According to Strawson, the answer to questions such as "what does it mean to hold a person responsible?" rests on psychological and sociological questions relating to our interactive practices of praise and blame that do not bear in metaphysical issues concerning the agent´s freedom. Conceivably if we start from the "central commonplaces" (2008:5) of our practices, we can account for the nature of moral responsibility while avoiding the unresolved metaphysical disputes concerning freedom of the will.

1.4 Terminological clarifications

Before I go on to Strawson´s Theory of moral responsibility, it might be helpful to make a few clarifications regarding the use of the term "responsibility". Further, I would like to clarify a custom in contemporary literature, namely, the tendency to focus on responsibility for blameworthy actions over praiseworthy ones.

The use of the term "responsibility" in our language is pervasive. In order to distinguish among the different varieties of responsibilities, it is now common to distinguish among prospective and retrospective responsibility, namely responsibility regarding future or past events.[20] Prospective responsibility corresponds to duties; for instance, parents are responsible for their children´s safety. This is to say, parents have a duty to take care of their safety. In this sense, when a parent is said to be "a responsible parent", this means, in other words, that he takes the duty of taking care of his child seriously. Retrospective responsibility is the responsibility for past events. For instance, when I blame my husband for not watering the plants while I was away, I take him to be responsible. This last example represents a case of responsibility in a retrospective sense. Retrospective responsibility can be moral or non-moral.

---

[20] For a discussion on the prospective sense of responsibility, see: Smith, Angela, "On Being and Holding Responsible" (2007:468).

In this thesis, I will be focusing on moral retrospective responsibility.[21] Contemporary work on moral responsibility, influenced by Strawson and strawsonian theories of moral responsibility, focus on the nature of moral responsibility by analyzing the ways in which we hold each other responsible. Holding people responsible is an essential part of our interpersonal practices. Further, holding people responsible for past actions comprises both praiseworthy and blameworthy actions.

This thesis focuses on blameworthy actions instead of praiseworthy actions, on what Rosen has called "the dark side of responsibility" (Rosen; 2015:68). The focus on moral responsibility for blameworthy actions is also manifest in the contemporary literature. But why is moral responsibility for blameworthy (rather than praiseworthy actions) that which captures most interest?[22] Philosophers give different reasons for this. A plausible reason for this is that there seems to be more at stake when it comes to blaming a person than praising her. i.e., "part of the explanation is that blaming tends to be a much more serious affair; reputation, liberty and even life can be at stake, and understandably we are more concerned with the conditions of adverse treatment than with those of favorable treatment" (Watson, 2004:283).[23] Furthermore, holding people responsible for blameworthy actions seems to be more important for sustaining the order of a moral society than praiseworthy actions are.

---

[21]Within retrospective moral responsibility, some have argued that there exist different kinds of moral responsibility: Watson was the first to distinguish among responsibility as attributability and responsibility as accountability. Furthermore, David Shoemaker has accounted for a tripartite theory of moral responsibility, he distinguishes among attributability, accountability and answerability (Shoemaker, 2017). Still, some contend there is only one kind, namely, responsibility as answerability (Smith:2015).

[22] Some accounts allow for morally responsible behavior that is neither praiseworthy nor blameworthy: see McKenna (2012:16–17) and Zimmerman (1988:61–62).

[23] Some suggest that there are reasons to believe moral praise and blame are symmetrical, and thus what applies to praiseworthy acts do so as well for blameworthy ones. However, not everyone agrees on this point, (see Nelkin 2008, 2011; Wolf 1980, 1990).

2 P.F. Strawson´s Freedom and Resentment

In this chapter, I start by introducing Peter Strawson's landmark essay *Freedom and Resentment*. This essay has been for decades now, a compass in contemporary moral responsibility. In this work, Strawson attempts to reconcile two opposing views at stake in the free will debate: incompatibilism and compatibilism. The discussion will proceed as follows. First, I will start by giving an account of both traditional views and what Strawson identifies as the flaws in their arguments. Next, I present Strawson's attempt of reconciliation, namely, his theory of moral responsibility. Following Russell (1992), I characterize Strawson´s view by distinguishing two main arguments in his theory, the "Rationalistic Argument" and the "Naturalistic Argument". Finally, I raise criticisms against both.

2.1 Pessimists and optimists

Strawson´s aim in *Freedom and Resentment* is to reconcile two opposing parties regarding the problem of free will and determinism, namely, "optimists" and "pessimists" (2008:1) accounts of moral responsibility. Both parties agree that free will is required in order to justify our practices of praise and blame. One of these parties claims that given the truth of determinism, our practices of moral condemnation and punishment would still preserve its meaning; namely, these practices would still be useful as modes of regulating behavior in socially desirable ways. This is Strawson´s optimist. The optimist holds a compatibilist and consequentialist account of moral responsibility. Contrary to the optimist, Strawson´s pessimist finds the utilitarian justification inadequate. The pessimist argues that in order to justify our practices of moral condemnation and punishment, people must truly *deserve* being blamed or praised. Deserving praise or blame, according to the pessimist, requires the agent´s free will, which is undermined by the truth of determinism. Hence, Strawson's pessimist is an incompatibilist.

Strawson believes that both optimists and pessimists are looking in the wrong direction. Indeed, both accounts are misguided.[24] The problem with these positions, according to Strawson, is that by appealing to external considerations (external to the practices) such as the metaphysics of the agent or the overall utility of the practices as such, these accounts lose sight of the "central commonplaces" (2008:5) in which these practices take place. These

---

[24] Although Strawson admits being more sympathetic to the optimist (2008:27).

accounts overlook the justifications that are "internal" to these practices. Thus, by looking for external justification of these practices, both accounts "overintellectualize the facts as we know them" (2008:25). How so? Strawson notes:

> The optimist's style of over-intellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism. He seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight (perhaps wishes to lose sight) of the human attitudes of which these practices are, in part, the expression. The pessimist does not lose sight of these attitudes, but is unable to accept the fact that it is just these attitudes themselves which fill the gap in the optimist's account (2008: 25).

By taking an "objective" and external standpoint, optimists and pessimists offer the wrong kind of reasons to justify these practices. Strawson claims that both accounts fail to recognize the practical import of our practices of holding responsible. According to Strawson, it is possible to account for moral responsibility, without incurring into the libertarian´s "panicky metaphysics" (2008:27) or the utilitarian calculus. How? By focusing instead in the moral sentiments and their role in holding people responsible.

Strawson´s account of moral responsibility focuses first and foremost on the responsive sentiments, or as he famously named them, the "reactive attitudes" (2008:5). These moral sentiments, such as gratitude, resentment, and forgiveness, to mention some, come to people naturally. These feelings represent "the very great importance that we attach to the attitudes and intentions toward us of other human beings and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions" (2008:5). The reactive attitudes arise in response to a person´s regard or lack of regard for another, namely, a person´s "quality of will" (2008:15). The reactive attitudes are an expression of the interpersonal demands that members of a moral community place upon each other.

To picture what Strawson has in mind, imagine you are on your way to work on a very crowded bus when suddenly you get hit in your face. You realize that a very tall man standing in the middle of the corridor had hit you with his backpack. Startled at first, soon you become angry. Your feeling becomes even more pronounced when the man doesn't seem to care about what has happened. You become resentful and blame him. You think he owes you an apology.

This example above describes one of three kinds of reactive attitudes Strawson analyses. The feeling of anger or resentment expressed in the case above is what Strawson called "personal reactive attitudes" (2008:16); they comprise our feelings towards another person on our own behalf. Strawson distinguishes two more types, "vicarious" or "self-reactive attitudes" these comprise attitudes felt on behalf of others (indignation) and feelings directed towards oneself (guilt), respectively (2008:15). Further, "all these three types of attitudes are humanly connected. One who manifested the personal reactive attitudes in a high degree but showed no inclination at all to their vicarious analogs would appear as an abnormal case of moral egocentricity, as a kind of moral solipsist" (2008:9). Or if we could imagine a human being that could experience one or two of these attitudes alone, "then we imagine something far below or far above the level of our common humanity—a moral idiot or a saint" (2008:9).

Consider, for instance, the case of resentment. This reactive attitude occurs in response to a fact: a person has flouted your expectation of reasonable regard. This injury, "constitutes a prima-facie appearance" that your expectation has been flouted or unfulfilled (2008:17). Further, this attitude has a motivational feature: resentment motivates the offended party to address the wrongdoer with a complaint; in other words, it motivates the agent to hold the other morally responsible. In this way, the reactive attitudes track both belief and motivation. The belief that the person has shown disregard and the motivation to address the other with a complaint.

Both pessimists and optimists, Strawson contends, overlook these attitudes. For them to reconcile, he claims,they must recognize the place of the reactive attitudes in our interpersonal relationships within the moral community. These attitudes are distinctive features of our human form of life. Trying to justify them by appealing to the utilitarian calculus or the libertarians "panicky metaphysics" (2008:27), is mistaken.

I will now present Strawson´s account of agential responsibility. I will do so, following Russell (1992), by presenting two main arguments, the Rationalistic Argument, and the Naturalistic Argument. I shall start with the former.

2.2 The Rationalistic Argument

The Rationalistic Argument allows Strawson to distinguish between responsible and non-responsible agents without incurring into the truth or falsity of determinism. Strawson argues it is possible to account for "all we mean, when, speaking the language of morals, we speak of

desert, responsibility, guilt, condemnation, and justice" without going beyond "the facts as we know them" (2008:24). The way to do this, according to Strawson, is by focusing on description and analysis of the reactive attitudes. In the following, I will explain how this argument works.

Strawson does not only point toward the responsive sentiments as solid ground to build a theory of responsibility, he also raises two critical questions concerning how they work: (i) When is it appropriate to target someone with the reactive attitudes? (ii) Are there any situations in which the reactive attitudes ought to be suspended? The reactive attitudes, Strawson argues, can be modified or suspended. To explain this, he introduces two groups of pleas.

Consider once more the example of the bus mentioned earlier. This time you are not only hit in the face abruptly, but you are also pushed to the floor by some stranger. Once more, you resent and blame the person who pushed you. Strawson recognizes two different reasons that might inhibit or suspend your response in this particular case. Suppose the person who pushed you did so by accident and tried to apologize accordingly, you then accept the apology and perhaps even apologize for blaming him in the first place. What this case illustrates is that the specific injury (an accident), inhibits your anger and makes you reconsider the appropriateness of your response.

Another way in which your response might be inhibited, Strawson argued, is if you were to find out that the *person* who pushed you was "only a child", a "hopeless schizophrenic" or someone who had undergone "a great strain recently" (2008:8). In these cases, it is not the injury that inhibits your response but the *person* whom your reactive attitudes are directed to. These agents are not the kind of agents to whom you would normally respond with reactive attitudes. In these cases, we suspend the reactive attitudes either because the agent is currently undergoing a particular circumstance that made him an inappropriate target of your response (e.g., someone who had gone through a stressful situation recently), or because generally speaking the agent is not a person whom you can make interpersonal demands, either temporarily (children or some temporary abnormal cognitive capacity) or permanently (mentally disabled).

Gary Watson (2004:223) refers to the exculpatory conditions explored above as "excuses" and "exemptions," respectively. "Excuses" are concerned with how we might suspend our

reactive attitudes at the time of the injury. What is particular about excuses is that we still see the agent as an appropriate target of the reactive attitudes, it is only the specific *injury* that inhibits our response. Thus, excuses present us with a reinterpretation of the agents´ conduct. The reinterpretation explains how the agent himself did not express disregard or ill will.

Exemptions, instead, inhibit the reactive attitudes because the agent in question is either at the time of the injury (stress or manipulation) or all the time (child schizophrenic), not an appropriate target of the reactive attitudes. While in the first case we still see the agent as a fully morally responsible agent, we just don't attribute responsibility to him on this particular injury, the second case, in Strawson's words, "invite us to view the agent himself in a different light from the light in which we should normally view one who has acted as he has acted" (2008:9). We cease to see the agent as capable of interpersonal relationships. Most importantly, in these cases, Strawson argues: "all our reactive attitudes tend to be profoundly modified" (2008:9). In these cases, we adopt what Strawson calls the "objective attitude" (2008:9) as opposed to the participant stance. The participant stance includes the whole range of reactive attitudes. In contrast, to adopt the objective attitude toward another person

> is to see him, perhaps, as an object of social policy, as a subject for what, in a wide range sense, might be called treatment, as something to be taken account, perhaps precautionary account, of, to be managed or handled or cured or trained, perhaps simply to be avoided (2008:9).

Strawson argues that even though we may be prone to certain emotions when we engage in the objective stance, this attitude does not include the whole range of reactive attitudes he discusses in his paper, not gratitude nor resentment. It is important to distinguish between seeing an action objectively and taking the agent himself objectively. When we view a person objectively, we see her as someone to be treated or managed. In the latter case, we are not able to resent the person, namely because of our objectivity of attitude. Thus, Strawson believes that if we take the objective attitude towards a person permanently, then we cannot engage with her in a personal mature and adult relationship: "if your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him even negotiate with him, you cannot reason with him" (2008:10). Being involved in interpersonal relationships includes both the "objective" and the "participant" attitude. "They are not altogether exclusive of each other, but they are, profoundly opposed to each other" (2008:9). The participant attitude includes the whole range of reactive attitudes that we usually display when we engage with normal adult people. It is

our exposure to these attitudes that make us the kind of human beings that we are. Strawson is here providing us with a descriptive account of our moral psychology.

Strawson has so far accounted for a theory of moral responsibility, which arguably could ground the responsibility or non-responsibility of an agent, based on how our responsive sentiments are inhibited or modified. Strawson´s rationalistic argument thus proposes that the internal analysis of the reactive attitudes is all we need to provide an epistemologically adequate theory of moral responsibility. When responding to different human actors, we make certain evaluations that happen to be metaphysical in kind: we decide whether our resentment is appropriate according to how the different actors reflect self-control. None of these reactions or inhibitions, respond to the truth of determinism, which in turn renders determinism irrelevant to an account of moral responsibility. All we need, Strawson grants is a proper analysis of the reactive attitudes and the cases in which we suspend them or modify them within the participant stance.

Most importantly, Strawson claims that both optimists and pessimists are mistakenly applying the objective attitude universally and permanently. For instance, according to the pessimist, if determinism is true, then all agents should be exempted. However, if we look at our practices, we find out that we usually exempt agents who either temporarily or permanently exhibit an abnormal ability to engage in interpersonal relationships. Why should we consider determinism as an additional consideration to exempt agents? Even if we should accept it as an added consideration, Strawson argues, this would be self-contradictory since this would imply that abnormality should be understood as the "universal condition" (2008:12).

Although we may concede that the descriptive analysis that Strawson presents might be sufficient to base our responsibility practices as they are, the pessimist might still claim that to justify our practices of praise and blame, we need to explain why the truth of determinism would not exempt all humans. In the end, they might criticize Strawson´s theory for offering an incomplete analysis. To explain Strawson´s attempt to answer this criticism, I now turn to his second argument, the Naturalistic Argument. Before this, I formulate what I take to represent Strawson's rationalist argument:

(RA): an analysis of the reactive attitudes and their inhibition in cases of excuse and exemption provides us with a theory of moral responsibility that can distinguish between responsible and non-responsible agents.

2.3 The Naturalistic Argument

According to Strawson, adult interpersonal relationships include the range of reactive attitudes Strawson described to be part of what he called the participant stance. This stance is characterized by our natural commitment to the reactive attitudes. This natural commitment provides us reasons to believe that these attitudes are "part of the general framework of human life" (2008:14). Strawson thus defends that part of our nature is characterized by participating in normal adult relationships. Participation in these relationships includes the range of attitudes with which we hold each other responsible. The truth of determinism, in this case, would require us to take this objective standpoint universally and permanently, and this, in turn, would imply a complete withdrawal from our commitments to interpersonal relationships, at least as we know them. The reactive attitudes are our natural commitment to these attitudes as well as part of the "general framework of human life", and not "something that can come up for review" (2008:14). It is by being exposed to these attitudes that we became the kind of agents we are, in the first place. Giving up these attitudes because of the truth of determinism would imply, so to speak, to give up our humanity. [25] Strawson admits we can sometimes leave "the participant stance" and adopt "the objective attitude" even in cases were normal mature adults are involved "as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity" (2008:19). What we cannot do is to adopt the objective stance permanently while at the same time preserve our interpersonal relationships.

Strawson's naturalistic argument reinforces his rationalist claim, namely, that what matters to our approbative and punitive practices is the agent´s ill or goodwill and not whether the will has been determined. When agents are exempted from moral responsibility, it is not because they are "determined," but instead, because (i) they haven't matured their sensitivity towards others as well as to themselves (the case of children) or because (ii) they do not longer possess this sensitivity (mentally disabled) or they never developed these capacities altogether (cognitive abnormality). In other words, there is no apparent reason to think that one should consider the truth of determinism as an aggregate condition for what excuses or exempts people in our ordinary interpersonal practices of holding people responsible. Strawson believes that there is no need to justify these practices because they are part of our psychological makeup and not something that we can choose to engage with or not. From the

---

[25] For a similar view, see also Wolf (1981,1990).

discussion above, I present what I consider to be the main argument in the naturalistic argument:

(NA): justification of our practices of holding people responsible does not need to be external to the reactive attitudes; their justification lies internal to our natural commitment to these attitudes themselves.

Strawson´s Naturalistic and Rationalistic arguments represent Strawson's theory of moral responsibility. Strawson argues that only by focusing on the reactive attitudes, we can distinguish among responsible and non-responsible agents. The rationalistic argument explains why we excuse or exempt agents based on the analysis of the reactive attitudes within the participant stance. The Naturalistic Argument claims that these attitudes are a natural commitment to interpersonal relationships and not something that we could decide to abandon, since abandoning them would imply losing our humanity.

Strawson is aiming at reconciling pessimists and optimists by suggesting both parties to stop "overintellectualizing the facts as we know them" (2008:25) and analyze instead what happens within the practices as such. It is just by analyzing the participant stance that we can manage to construct an adequate theory of moral responsibility.

To summarize, since determinism is irrelevant to explaining what we consider as explanations and justifications of our responses within the participant stance (Rationalistic Argument) and since it is not a practical possibility for us to adopt the objective stance towards everyone all the time (Naturalistic Argument), we should turn our eye from external justifications to the internal working of these attitudes.

2.4 Criticisms

I will now raise criticisms against the Naturalistic and the Rationalistic Argument. I shall start with the former. One problem with this argument is that Strawson didn't say much about what sort of incapacity would exempt agents for adult interpersonal relationships. This is problematic since given that Strawson never specified what is it that makes an agent exempt, he cannot explain why determinism would not exempt all individuals. For instance, the responsibility skeptic might argue that the reason why we should exempt everyone is because we are not the kind of agents that fulfill the criteria in a sense presupposed by our practices. Likewise, the libertarian might argue that the reason we exempt some individuals is that if

determinism is true, then these agents fail to meet the freedom condition, an ability to do otherwise, which is undermined by determinism.

Gary Watson has argued that when analyzing the considerations that inform us whether we should hold people responsible, the reactive attitudes can and indeed conflict with respect to specific agents. How can expressivism adequately distinguish morally responsible agents from exempt agents in these paradigmatic cases? If the reactive attitudes cannot determine whether we should hold an agent responsible or not, then Strawson´s theory fails to provide an adequate theory of moral responsibility. Absent this explanation, Strawson's rationalistic argument fails.

Another significant criticism is one raised against the Naturalistic Argument. Strawson argued that it would be a practical impossibility (psychologically speaking) for us to abandon the reactive attitudes without losing our interpersonal relationships as we know them. For some, this argument is not only false, but most importantly, it misses the point. According to the incompatibilist, the fact that we cannot help but treating people as if they were responsible, does not necessarily mean that the agents in question are genuinely responsible. As Downie famously put it,

> Strawson is confusing practical beliefs with metaphysical beliefs. We may all find it difficult in practice to think of the taste of cheese as being other than in the cheese, and eating might be poorer if we succeeded. But it does not follow that when we shut the dining room and open the study door, the practical belief should continue to dominate; a theorist wants to know what is true, and not simply what he cannot help believing in practice (1966:36).

Further, Strawson's argument that it is a practical impossibility for us to abandon our reactive attitudes and still have adult interpersonal relationships has been highly contested. Some philosophers have argued that we *can* and in fact, *should* call these practices to substantial revision. As mentioned in the introduction, revisionism has become a significant voice in the contemporary discussion on moral responsibility. When Strawson argues that it is practically speaking, impossible for us to abandon these practices, a revisionist might reply that the personal reactive attitudes, as well as their vicarious analogs, can and should be modified. And that even after such revision, we would still have what Strawson called interpersonal normal and adult relationships and nothing like an "impoverished" picture of our lives (Milam, 2016:118). For instance, according to Tamler Sommers, "our commitment to

attitudes that presuppose moral responsibility can soften and fade, often without our noticing it" (2007:321).[26]

Another line of argument against Strawson's naturalistic premise focuses on the historical figures of Gandhi and Martin Luther King. These figures illustrate how it is possible to give up at least the negative reactive attitudes and nevertheless manage to confront wrongdoers with a demand, thereby holding them responsible (Watson 2004:255-8 and Nelkin 2011:42-5).

In the next chapter, I will engage with criticism raised against the Rationalistic Argument. I contend that these criticisms make evident that Strawson´s contribution failed in putting an end to metaphysical debates of agential responsibility.

---

[26] For a defense of this view, see Pereboom (2001), especially chapter 7. For criticism against revisionism, see Seth Shabo (2012)

3 Responsibility Expressivism and Responsibility Realism

Last chapter I presented P. F. Strawson´s influential contribution in *Freedom and Resentment*. Strawson's 1962 paper has been widely discussed, starting a significant debate that still shapes our contemporary discussion on moral responsibility.[27] Despite Strawson´s novel and influential defense of compatibilism, his theory remained incomplete in various ways, leading to at least two different and opposed interpretations of his original argument: "Responsibility Expressivism" and "Responsibility Realism" (from now onwards, Expressivism and Realism).[28] This chapter examines and critically discusses the realist and expressivist positions, respectively, and raises objections against both accounts.

Before I discuss both views, it is important to observe that the distinction between expressivism and realism is a metaphysical one. However, this distinction differs from the distinction between compatibilism and incompatibilism. Compatibilists and incompatibilists disagree about how the truth of determinism would affect our practices of blame and praise. They did so, recalling Strawson by appealing to "external" justifications of the practices (such as agent causation or social utility). Expressivist and realist accounts, following Strawson, shift from "external" justifications of moral responsibility to justifications based on the analysis of the practical import of our practices of praise and blame. Thus, both accounts try to analyze moral responsibility by answering the following question: "When is it appropriate to hold an agent responsible? The concern underlying these approaches is whether metaphysical issues related to the free will debate arise in this new context too. Both accounts, following Strawson, try to provide a normative conception of moral responsibility. However, as I shall discuss, both views face difficulties.

3.1 Responsibility Expressivism

As discussed in the second chapter, Strawson rejected "external" justifications of moral responsibility proposed by optimists and pessimists (agent causation or social utility). Similarly, the Expressivist account rejects the existence of external or independent justifications of our attributions of responsibility. Indeed, "there is no such independent notion

---

[27] For criticism see, Strawson´s son Galen Strawson (1993), Russell (1992), Pereboom, (2001); Rosen (2004) Smilansky (2000); Sommers (2008).

[28] I shall also refer to realism and expressivism, as response-independent and response-dependent theories respectively.

of moral responsibility that explains the propriety of the reactive attitudes" (Watson, 2004:222). Ascriptions of responsibility are to be accounted solely by attending to the practices of praise and blame and the responsive sentiments involved in them. It is only by analyzing these that the expressivist can offer an adequate epistemological theory of moral responsibility. Within expressivism, I shall discuss the accounts of Gary Watson (2004) and R. Jay Wallace (1994).

3.1.1. The non-cognitivist objection

In his highly influential reading of Strawson, *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme* ([1987], 2004), Gary Watson sympathizes with, as he calls it "the expressive theory of responsibility" (2004:226). In this contribution, Watson argues the following: "Strawson´s radical claim is that these `reactive attitudes` (as he calls them) are *constitutive* of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions" (2004:220). Watson (2004) and others (e.g., Vargas 2005:406, Shoemaker 2017:494), argue that Strawson favors an expressivist account of moral responsibility in the traditional non-cognitivist sense. A non-cognitivist holds that the judgments, in this case, judgments regarding a person´s responsibility, have no truth value. For those holding a non-cognitivist expressivist view about moral responsibility, being a morally responsible agent amounts to nothing more than the expression of the judging agent: "to regard oneself or another as responsible *just is* the proneness to react to them in these kinds of ways under certain conditions" (Watson, 2004:220, my italics). Jonathan Bennett´s account of Strawson (1980) also embraced this noncognitivist interpretation. To him, it is the emotions one feels towards the wrongdoer and not some further belief about the responsible agent that justifies our practices, "the reactive attitudes are essentially spontaneous, adopted without the guidance of a telos" (Bennett; 1980:68). The non-cognitivist construal of the reactive attitudes implicit in Watson´s as well as Bennett´s view above can be formulated as follows:

(D): S is morally responsible (for action x) if and only if we are *disposed*, under favorable conditions, to hold S morally responsible (for action x).[29]

---

[29] Wallace uses this biconditional to represent what he refers to as "the dispositional schema" See (1994:89).

The above reading of Strawson, namely that people are responsible because we are *disposed* or prone to hold them responsible, faces an important objection. Expressivism is supposed to analyze ascriptions of responsibility based solely on the reactive attitudes. However, how would expressivism determine whether an agent is, in fact, responsible if different agents present contradictory reactive attitudes towards the same agent? In these cases, a noncognitivist expressivist interpretation of Strawson cannot determine whether the agent is, in fact, responsible or not. Hence, the dispositional account is unable to distinguish between responsible and non-responsible agents. Note that this problem would render Strawson´s rationalistic argument implausible since the theory wouldn't be able to distinguish among responsible and non-responsible agents.

In a sense, it is undoubtedly true that the ways in which we hold each other accountable by being the target of the reactive attitudes are part of what it is to be a responsible agent. However, describing how we target each other without any further account of the beliefs concerning the objects of these attitudes seems to fall short to the task. Hence, an important problem with the dispositional account is that it cannot explain why, even if we did (or were disposed to) blame the severely mentally ill and small children, they nevertheless would not be blameworthy.[30]

I contend that the problem with dispositional accounts of moral responsibility comes from conflating two interacting, although distinctive phenomena: that of "being responsible" and "holding responsible". Even if it is in our nature to hold *some* people morally responsible for *some* of their actions by targeting them with the reactive attitudes, this does not necessarily mean that the targets of these emotions *are* morally responsible.

Patrick Todd explores a parallelism between objections raised against this non-cognitivist reading of Strawson and standard objections to "(unsophisticated versions of) divine command theory. If something is right because God commands it, doesn't this imply, absurdly, that if God commanded murder, then murder would be right? (2016:211). Similarly, just because I am disposed to hold a person responsible, this does not necessarily imply that the person is, in fact, responsible. Further, if Watson meant that Strawson´s "radical" claim (2004:220) was to argue that holding responsible involves a tendency to react to people in

---

[30] This criticism has been raised by Martin Fischer and Ravizza (Fischer and Ravizza; 1993:18) and Patrick Todd (2016:209).

specific (inevitably human) ways, then Strawson´s "radial claim" appears to be "so weak as to be almost trivial" (2016:213).[31]

Strawson intended the reactive attitudes to be at the center of our ascriptions of responsibility, and the noncognitivist interpretation seems to capture this nicely. However, Strawson did not only attend to the reactive attitudes; he also introduced the exculpatory pleas, namely how these attitudes tend to be inhibited or modified. And, excuses and justifications *do* seem to provide us with true or false propositions. The reason why I suspend my resentment in the case of an excuse is, in part, founded in that you, in fact, did *not* show ill will or disregard towards me. The fact that you were pushed and did not mean to hurt me is either true or false. These judgments do not depend on whether I am disposed to react with resentment or not, but rather on whether you are an appropriate target of my resentment in the first place.

In his influential account of moral responsibility, R.J. Wallace (1994) argues that although judgments of moral blame involve an expression of emotions, this "does not entail that such judgments are *exclusively* expressive" (Wallace; 1994:75). The emotions must rest on some further propriety conditions that make them appropriate too. Thus, if we want to understand moral responsibility by analyzing the different ways in which we hold people responsible, we should be able to provide the reactive attitudes with distinctive propositional objects. Holding responsible should be tied to specific emotions that have a privileged connection with the person and the beliefs concerning her responsibility. Wallace´s normative interpretation rests on the appropriateness of the reactive attitudes and not on our disposition to hold people responsible by targeting them with the reactive attitudes. Hence, he avoids the problem of rejecting truth-conditions about the reactive attitudes and their targets, since the former are to be construed in cognitivist terms. Thus, in Wallace's account, an agent may be considered blameworthy and an appropriate target of the reactive attitudes even if others do not explicitly manifest the reactive attitudes toward him. It is the belief that he has violated an obligation that would make it appropriate to feel the reactive attitudes. Wallace´s cognitivist construal of the reactive attitudes can be formulated as follows,

---

[31] When criticizing Watson's view, Todd maintains that "not only do we not have something radical; we seem to have something that no one could reasonably dispute. What justifies us in blaming and resenting people? The fact that those people are involved in a certain form of life. What sort of form of life? The sort on which it is appropriate that they be blamed and resented. So: what justifies us in holding people responsible—blaming and resenting them—is the fact that it is appropriate that they be blamed and resented. What is "unique" to Strawson's theory turns out to be a borderline triviality (2016:216). David Shoemaker calls this dispositional view, response-dependents' "most naïve form" (2017:482).

(W): S is morally responsible (for action x) if and only if it would be *appropriate* to hold S morally responsible (for action x) (Wallace, 1994:91).

Wallace´s "normative interpretation" (1994:1) of Strawson has been of significant influence among contemporary theorists. Although very much inspired by Strawson, Wallace suggested substantial differences to the original strawsonian picture. For instance, when Strawson introduced the reactive attitudes, he included emotions such as hate, love, pride, gratitude, anger, regret, resentment, indignation, and forgiveness (2008:5). However, it is easy to see how some of these emotions are more akin to the practices of blame and praise than others. According to Wallace, this inclusive interpretation of the reactive attitudes makes it difficult to defend the claim that "holding responsible is inextricably a part of this web of attitudes" (1994:11). Thus, he suggests a narrower class of reactive attitudes by taking the paradigm cases of guilt, resentment, and indignation (1994:33).

What is it that unifies the reactive attitudes of guilt, resentment, and indignation? Wallace argues it is their unique connection with expectations. Holding someone to an expectation, according to Wallace, is to "demand of people that they act as we expect them to" (1994:21). To hold someone to an expectation, Wallace says is "essentially to be susceptible to a certain range of emotions in the case that the expectation is not fulfilled, or to believe that the violation of the expectation would make it appropriate for one to be subject to those emotions" (1994:21). In Wallace´s account, only believing that such emotions, *if* felt, would be appropriate is enough to "hold" the person responsible, which means one does not need to actually feel the reactive attitudes, but rather believe they would be appropriate if felt.

Strongly influenced by Strawson Wallace argued that it is difficult "to make sense of the idea of a prior and thoroughly independent realm of moral responsibility facts" that is separate from our practices and yet to which our practices must answer (1994: 88). However, he claimed, this does not mean we should give up on facts about moral responsibility. Instead, "we must interpret the relevant facts [about responsibility] as somehow dependent on our practices of holding people responsible" (1994: 89).

According to Wallace, when we start by analyzing our practices of holding people responsible, these practices disclose a fundamental commitment to fairness (1994:101). Thus, he maintains that moral responsibility rests on questions of fairness. To Wallace, the question of when we should hold people responsible should be replaced by the question of when it is

fair to hold people responsible. If we focus on the normative understanding of our practices, Wallace contends, we discover that accountability does not require freedom of the will. This is because, the fairness of holding people responsible depends on moral competence, namely the power to grasp an apply moral reasons and to control one's behavior accordingly. Moral competence, he argues, is compatible with determinism. The normative interpretation can be now specified as follows,

(F) "S is morally responsible (for action x) if and only if it would be *fair* to hold s morally responsible (for action x).[32]

3.1.2 The Charge of Circularity

Michael McKenna (2012) argued that to understand whether a person is responsible, we usually start by looking for evidence present in the practices of holding responsible. He suggests we see this search as an epistemic move. For instance, to understand the nature of an electron, we might look at the activity in a cloud chamber. However, we still think "it is the electron that is more fundamental or basic in the order of being as in comparison with the goings-on in the cloud chamber" (2012:40). In an equal manner, affirming that an epistemic move is necessary to understand the nature of moral responsibility does not imply that the practical stance of holding responsible is more fundamental or basic than the agent´s responsibility. Therefore, when we inquire whether a person is responsible by asking whether it would be appropriate to hold him responsible, we are making, paraphrasing McKenna, an epistemic move.

However, when Watson notes that "Strawson´s radical claim is that these `reactive attitudes` (as he calls them) are *constitutive* of moral responsibility" (2004:220) he is defending a metaphysical claim, one that responds to "the order of being" (2012:40) rather than to "the order of explanation" (2012:47) as such. Hence, by focusing exclusively on the conditions for holding responsible, expressivist readings of Strawson such as Watson´s or Wallace´s are not suggesting an "epistemic move," they are establishing a "metaphysical priority or basicness" (2012:40).

McKenna criticizes Watson´s contention that holding responsible is "primitive" and that it should "not to be justified by, for example, offering a metaphysical thesis about the nature of

---

[32] Wallace (1994:91).

being responsible" (2012:41). To McKenna, arguing that these attitudes are "primitive", is also offering a metaphysical thesis, "a thesis that, in the case at hand, treats as nonrational, naturalistic, and primitive our preparedness to regard others as responsible, is a kind of metaphysical view, it's one that treats holding morally responsible as more basic than holding responsible" (2012:41).

Further, the former criticism seems to apply to Strawson as well. When criticizing optimists and pessimists of looking for external justifications of the practices, Strawson argued that "questions of justification are internal to the structure" (2012:40). Thus, according to McKenna, Strawson has then also "shifted to a thesis about the order of being by adopting a deflationary metaphysical thesis. Our practices of holding morally responsible are what comes out to be more fundamental or basic" (2012:40). In this way, both Strawson and expressivist accounts defend the claim that the practices of holding people responsible are more fundamental than an agent´s responsibility. Nevertheless, "if we rely upon certain explanatory priority in the order of knowledge, it need not follow that we are entitled to draw symmetrical conclusions about the order of being" (2012:40). Watson and Strawson seem to be doing exactly this, namely moving from an epistemic priority (first holding responsible) to justifying these practices from this context too.

Although Wallace´s normative interpretation differs from Watson´s in significant ways (for instance, he understands that what is at stake is the "appropriateness" of our attitudes rather than our "dispositions" to target others with them), his view faces similar challenges. When defending his view, he contrasts his normative interpretation to what he calls the "metaphysical interpretation" (1994:88). The metaphysical interpretation, according to him, claims that there are facts that are prior and independent of our practices of holding responsible. He argues that this interpretation is problematic and unpromising. The reason he gives for this is that he cannot see "how to make sense of the idea of a prior and independent realm of moral responsibility facts" (1994:88). To avoid this interpretation, he argues, "we must interpret the relevant facts as somehow dependent on our practices of holding responsible" (1994:89).To Wallace, facts about whether people are morally responsible "are *fixed* by the answer to the question of when it is appropriate to hold people responsible" (1994:93, my italics). Wallace is then, alike Watson, claiming that the facts of moral responsibility are somehow "fixed" by the practices. Justifying the practice of holding responsible by interpreting the facts as dependent on our practices, makes the "normative

standards" for holding people responsible metaphysically basic. Therefore, Watson´s and Wallace´s expressivist views move from "an innocuous point about the context of discovery" to "a highly contentious one about the context of justification" (2012:40).

Although both Watson and Wallace face a similar objection, in Wallace´s view, it is easier to appreciate how moving from an epistemic claim to a metaphysical thesis, faces a charge of circularity. According to Wallace, facts about moral responsibility "depend" on the practices of holding responsible, and not on some "independent realm of moral responsibility facts" (1994:88). However, he adds, "this is not to deny that we want our practices of holding responsible to answer to the facts, including the facts about what is it to be morally responsible (if there are such)" (1994:88). Wallace claims that the facts of moral responsibility depend on the practices of holding responsible, but he contends that the practices need to answer to the facts about what it is to be morally responsible in the first place. Isn´t this argument circular? By trying to justify our practices from the standpoint of holding responsible, Wallace faces the following problem. Those practices presuppose facts about agency, namely the kind of agencies involved in these practices. Therefore, it is impossible to understand the conditions for responsible agency by merely laying out the standards we employ in holding responsible. This is because "those standards presuppose that the agency at issue has a standing to which they – the standards- must answer" (2012:50).[33] Is there a way for the expressivist to avoid this circularity?

Before I turn to McKenna for a solution (Section 3.3), I suggest exploring the charge of circularity, by analyzing a paradigmatic case presented by Watson (2004:225). I propose looking into this example and see whether an expressivist reading of Strawson (one which considers the relevant facts about moral responsibility somehow "fixed" by our practices of holding people responsible (Wallace, 1994:93)) can help us determine whether or not an agent is morally responsible, or if in the end, some other account is better fit to the task.

---

[33] Mckenna objects to Wallace's charge of circularity too (2012: 50). For a similar view see Arpaly (2006:6-28)

3.1.3 Moral Community, Moral Luck, and the threat of determinism

"*no matter how he came by them*, his evil beliefs are part of who he is, morally, and make him a fitting object of reproach." (Adams, 1985:19, my italics)

(…) in his last years Harris either remained, or became once again, capable of friendship and remorse. His crimes were monstrous, but he was not a monster. He was one of us" (Watson, 2004:259).

In *Two Faces of Responsibility,* Watson famously claimed that the reactive attitudes are "incipient forms of communication" (2004:230).[34] Influenced by Strawson, he argued that targeting an agent with the reactive attitudes is appropriate only if the agent can be morally addressed, or as he puts it, only if the agent can "comprehend the message" (2004:230). If an agent cannot be addressed, then it follows that the agent cannot be presented with the demand for reasonable regard. Those who are eligible targets of the reactive attitudes are members of the so-called moral community. To be a member of the moral community, an agent must possess the relevant capacities to be an eligible moral interlocutor.

The conditions for moral responsibility in the strawsonian theory are the two categories of exculpatory pleas: excuses and exemptions. Excuses are appropriate when a particular behavior which would generally represent ill will or disregard, does not in the present case, actually represent disregard. Examples of excuses are: "I didn't mean to" or "I was pushed." An agent who is excused remains a participant in the moral community. In contrast, cases of exemption present us with an agent who is incapacitated for ordinary interpersonal relationships. In the cases of exemptions, since the agent cannot be addressed as a member of the moral community, the agent does not remain a participant in the moral community. The examples of exemptions given by Strawson include being a "hopeless schizophrenic," a person "under great strain," "being a child," and being "unfortunate under formative circumstances" (2008:8-10). But since "a child can be malicious, a psychotic can be hostile, a sociopath indifferent, a person under great strain can be rude, a woman or man `unfortunate under formative circumstances` can be cruel" (2004:228), how can the expressivist explain whether these agents are to be held responsible or exempted? Our reactive attitudes towards

---

[34] Seth Shabo has also argued in favor of a communicative approach: "in so far as resentment and indignation are analogous to `incipient forms of moral address`, we can view their expressions as forms of moral communication whose purpose goes beyond influencing the target´s behavior" (Shabo; 2012:132 fn3).

these agents would commit us to hold them responsible; however, these agents are incapacitated for moral responsibility, which means that they should be exempted. How can expressivism adequately distinguish morally responsible agents from exempt agents? Absent this explanation, Strawson's rationalistic argument fails.

Watson tries to complete the strawsonian theory by construing the exempting conditions "as indications of the constraints on intelligible moral demand" (2004:228-9). He explains the case of "being a child" and "being under great strain" by offering the following arguments. Children "do not yet (fully) grasp the moral concepts in such a way that they can fully engage in moral communication, and so be qualified members of the moral community" (2004:229). Thus, if the demand for reasonable regard is supposed to be addressed "to one who is capable of understanding the demand" (2004:230), then children cannot grasp the terms of these demands, which explains why they should be exempted. Something similar applies to agents under great strain. In this case, the agent can understand the moral demand. However, given that these agents have acted "uncharacteristically" (2004: 231),[35] and that their conduct does not reflect their true "moral self" (2004:226), Watson concludes, we excuse them. For instance, if a person acted rudely but was under great strain lately, we exempt her but not because she wasn't rude toward us but rather because her present state provides "an explanation of the rudeness" (2004:232).

But what about the case of "being unfortunate under formative circumstances"? These agents, Strawson argued, are incapacitated for membership in the adult interpersonal community. However, Strawson didn't say much about what sort of incapacity would exempt agents for adult interpersonal relations. This incompleteness in Strawson´s theory leaves open the possibility that what makes some agents incapacitated for these kinds of relationships is that they are determined. Therefore, "([u]ntil we understand the relevant capacities, we cannot maintain that if determinism were true, these capacities would not be undermined" (McKenna 1998:126).

In his 1987 paper, Watson problematized Strawson´s theory by analyzing a case of being "unfortunate under formative circumstances". He concludes that Strawson´s theory would commit us to the following paradox: cases of extreme evil would count as its own excuse. In

---

[35] Here Watson raises a problem with this argument, namely "why our responses under stress are not reflections of our moral selves?" I will not pursue this point here, but it is worth mentioning that Watson agrees that "the explanation requires development" (2004:233).

the following, I explain what this paradox amounts to and why it would undermine expressivism. Next, I consider a solution to the problem.

Watson examines the famous 1982 case of Robert Alton Harris (2004:233-57), who viciously murdered two young boys while trying to use their car in a bank robbery. After promising not to hurt them and allowing them to leave, Harris shot one of them in the back. He soon followed the other boy and shot him four times. Shortly after this, he ate the young boys´ lunch and told his brother about dressing up as a police officer to tell the parents of the youths about their death.

By virtue of his actions, Harris is a clear candidate for moral blame. His "heartlessness and viciousness" is met by the rest of us, "with moral outrage and loathing" (2004:238). However, Watson argues that our reactive attitudes toward his vicious character can be affected by other considerations. Harris's history was indeed a tragic one. Ever since he was in his mothers´ womb, Harris was the cause of his parents´ anger and despair. He had an abusive father, and his mother neglected him. As a young child, he was still rejected and despised by his parents; his mother would not allow him so much as to touch her. Her sister reported the following, "He'd come up to my mother and just try to rub his little hands on her leg or her arm. He just never got touched at all. She'd just push him away or kick him. One time she bloodied his nose when he was trying to get close to her" (Watson 2004:241). He spent many years of his adolescence in prisons for juvenile offenders, and as an adult, he was imprisoned and held in penitentiaries.

How does Harris´ history affects our feeling of indignation towards his vicious crimes? Do we feel sympathy toward him after we come to know about how he came to be the way he is? Does this feeling of sympathy suspend our negative reactive attitudes toward him? Watson does not think so. The way we seem to be affected by Harris´ history, Watson thinks, is by creating a feeling of overall ambivalence. We react with indignation towards the crimes he committed, but we react with sympathy toward the child he was, the problem then results in that "the sympathy toward the boy he was is at odds with outrage toward the man he is" (2004:244).

Moreover, this feeling of ambivalence makes us look at his rejection of the moral community and his incapacity to interpersonal relationships in a different way. In some sense, Harris´ incapacity and rejection of the moral community becomes "understandable" outcomes of his

tragic life. It is important to remark that our responses to Harris (our indignation) are not a total suspension of the reactive attitudes (otherwise expressivism would consider Harris as a clear case of excuse, and the case wouldn't present a problem to expressivism). What seems to be problematic is that Harris´ past creates an overall sense of ambivalence. As Shoemaker puts this ambivalence, "consists in our suspending one subset of responsibility responses (…) without suspending others (2015:87).

To understand how the following example can be problematic for the strawsonian theory and how it would undermine expressivism, some background is necessary. As aforementioned, being a responsible agent, according to Watson, means being a member of the moral community. To be a member of the moral community requires the possibility of moral address. To be addressed, an agent must be capable of understanding the demands and the reactions of other members in the moral community. For Watson, this understanding requires a shared framework of values as well. Since Harris doesn't share this framework of values, he argues, Harris is not part of the moral community. The problem then results in that people such as Harris, who in virtue of their behavior forswear the values of the moral community, and thereby fail to be appropriate moral interlocutors, cannot be part of the moral community and hence cannot be responsible agents. The paradox then results in that extreme evil disqualifies for blame.

But the problem doesn't end here. Harris´ case puts into question whether skeptical issues arising from incompatibilism might undermine Strawson´s compatibilism (and thus expressivism). This is due to historical considerations and how they affect our responses toward Harris. To begin with, Harris's history reminds us of how "lucky" we are to have the past histories that we have had. Secondly, this explicit appeal to "moral luck" makes us feel in no position to cast blame because, for all we know, if we had been placed in those circumstances, we might have also ended up the way he did. Further, if according to determinism, evil is the joint product of nature and nurture, then facts about Harris's constitution and his environment can explain why Harris committed the wrongs he committed. This raises the following question: can "evil be the object of unequivocal reactive attitudes only when it is inexplicable?" (2004:247). Moreover, skepticism toward evil agents such as Harris can be applied to lesser wrongs as well. This would take us to feel "ambivalence" and skepticism toward holding people responsible for any kind of evil, extreme, or even moderate.

Assuming the historical considerations about Harris's life are *nonevidential,* which basically means that they do not provide evidence of an actual incapacity,[36] how can expressivism explain why the reactive should or shouldn't be sensitive to considerations which suggest a skeptical diagnosis? And most importantly, what does expressivism tells us about Harris´ status as a responsible agent, should we exempt him? The problem with answering these questions is that if the reason why we exempt Harris is because of his abusive history (which caused his rejection of the moral community), then some independent theoretical conviction ends up justifying why he should be exempted. This explanation would automatically leave the rest of the agents on equal footing, suggesting that extreme evil or any wrong should be excused. The optimist and pessimist can appeal to external criteria to determine whether to exempt these agents or not. But this is not available to expressivism since expressivism was supposed to account for responsibility without incurring into external criteria (recall Watson and Wallace's account discussed earlier in this chapter). As Watson noted, "the practice is less philosophically innocent than Strawson supposes" (2004:221).

Watson´s paradox of Evil raises skepticism of the reactive attitudes and skepticism of the reactive attitudes to the expressivist, amounts to skepticism of moral responsibility in general. The pessimist might argue that what individuals such as Harris lack is his "incapacity to do otherwise" (2004:219). Might there be a solution to expressivism available? I think there is. There is a way to argue that this problem rests on a false assumption. Further, there are ways to respond to the skeptical challenges raised as well. I turn first to the problem of exemption and then to skeptical challenges.

3.1.4 The problem of exemption and a possible solution

In his paper, *Limits of Evil and The Role of Moral Address*, McKenna attempts to defend Strawson's compatibilism against the problem raised by Watson, namely that Strawson´s theory would take us to treat cases of extreme evil as its own excuse. While McKenna has great sympathy for Watson's strawsonian account of moral responsibility, he believes the problem he raises can still be accounted for from a strawsonian theory of moral responsibility. All that is needed is "a slight modification" to Strawson´s theory (1998:129). This

---

[36] See Watson (2004:243). For some the story of Harris, does provides evidence of such incapacity, but for the sake of the argument let's assume it doesn't, in order to analyze how nonevidential historical considerations could affect our reactive attitudes.

modification, McKenna argues, will help us explain how people such as Harris can still be regarded as responsible agents.

The problem lies in the requisites for being a member of the moral community and a morally responsible agent (an appropriate target of the reactive attitudes). There is a clear sense in which Harris is not a member of the moral community; after all, he "declares himself as a moral outlaw" (2004:239). Further, Watson argues that by flouting and rejecting the moral demand, Harris somehow puts himself outside the moral community. But if Harris isn't a member of the moral community, could it be possible to argue that membership in the moral community is not necessary for an agent to be morally responsible? Clearly, suggesting this would be problematic for expressivism. This is because claiming that membership in the moral community is not necessary for moral responsibility, would depart from "Strawson's naturalist attempt to explain moral responsibility purely in terms of the attitudes from within the framework of the interpersonal community" (1998:128).[37] How should the expressivist account proceed?

Following McKenna, I argue that we should consider morally responsible agents as those who have the *capacity* to *participate* within the community (capacity for membership) instead of an "*actual* participation" (McKenna;1998:129). Being a member of the moral community requires actual participation and interaction with other members, but those who are not members still have the capacity for membership. Further, although agents might "disagree" to the shared framework of values in virtue of their wrongful acts, they do not need to adopt the shared framework of values to be morally responsible agents; it suffices that they understand that there "is" one.

The fact that Harris doesn't share our framework of values doesn't mean he is "*immune*" to our demands (1998:131). Harris doesn't lack an understanding of the moral demands placed upon him. It is because he understands the demands that he is willing to defy them. When we confront Harris, "it is not as if Harris is unable to appreciate our point of view", in fact, "more than likely he anticipates it" (1998:132), and this anticipation requires some sort of

---

[37] Here McKenna argues that preserving the membership would fit Strawson's expressed intentions best, see (1998:128).

understanding of moral demands.[38] And this Harris doesn´t lack. Although he "refuses dialogue" (Watson; 2004:239) by his contempt towards the moral order and his unwillingness to be part of the community, he is engaging in a form of moral communication after all. Unlike the child or the mentally deranged, Harris seems to understand the basic demand. Therefore, whether he is responsible or not "will be decided by his *capacity* for membership in the moral community" (1998:130) and not by actual participation in it. This modification doesn't break with Strawson's naturalism since it is still an ability to participate within the moral community that, in the end, determines the status of a morally responsible agent.

I am very sympathetic to McKenna´s defense of the strawsonian theory against Watson´s objection that extreme evil exempts. Interestingly, I find it very much alike Wallace´s normative interpretation. What matters to responsible agency, McKenna argues, is "not only *knowledge that* moral demands are made, but also *knowledge of how* to guide one´s conduct in light of those demands, and *knowledge of why* one should do so" (1998:131). However, some may argue that what Harris lacks is this second kind of knowledge, namely knowledge on how to guide his conduct in light of the demands made upon him. McKenna grants the possibility, but in the end, he rejects it because of being too premature: "[w]e cannot move from the *fact* that Harris did not guide his conduct by the moral community´s demands, to the conclusion that Harris *lacked the knowledge of how and why* to do so" (1998:131). If Harris lacked this knowledge, then he would be a clear case of exemption and would not be a problem for the strawsonian theory. The fact that he was unwilling to engage in a moral conversation with us doesn't mean that he lacked knowledge on how to guide his conduct. And the fact that he failed to adopt the shared framework of values doesn't make Harris incapable of moral address.

3.1.5 What about the skeptical challenges?

Harris, although not part of the moral community, is still addressable because he has the capacity to engage in a moral conversation with us. The fact that he defies us and hence is unwilling to have such conversation doesn't take him off the hook. However, we have still to account for the other problem with Harris´ case. How can a strawsonian answer to the threat of skepticism from historical considerations? McKenna recognizes four skeptical challenges

---

[38] I will come to this specific requirement, that of "understanding the moral demands" in the next chapter, when I discuss McKenna's quality of will condition.

in Watson´s argument. I will here focus on one of them, namely, the threat of determinism. I focus on this challenge, especially since I take this skeptical challenge to be the most relevant to the present work.[39]

How does the threat of determinism influence our judgments on Harris? As Watson puts it, if determinism is true, then "evil is the joint product of nature and nurture" (2004:247). If this is so, then it is easy to see what the skeptical challenge amounts to. If facts about an agent´s constitution and environment can explain what goes wrong when an agent does wrong, then this would lead to universal ambivalence towards all persons we hold responsible for any wrong (extreme or not). Thomas Nagel has very eloquently expressed this problem in his famous book, "The View from Nowhere",

> When we first consider the possibility that all human actions are determined by heredity and environment, it threatens to defuse our reactive attitudes as effectively as does the information that a particular action was caused by the effects of a drug — despite all the differences between the two suppositions. . . . Some of the externally imposed limitations . . . on our actions are evident to us. When we discover others, internal and less evident, our reactive attitudes toward the affected action tend to be defused, for [the action] seems no longer attributable in the required way to the person who must be the target of those attitudes (1986:125)

One way to avoid the skeptical challenge is to deny that our sympathetic response influences our antipathetic response toward Harris. The sympathetic response is a response to Harris´ past, not to how Harris is today. But this argument is unsatisfactory. This is because, in the present case, Harris´s past is "intended to, an actually do affect the original antipathetic response itself" (1998:137). The problem expressivism then seems to be facing once more is that of distinguishing what elicits the reactive attitudes, from what makes these attitudes appropriate reactions. In the case of Harris, although we may be prone to react with ambivalence towards him, it is not all too clear that we should think that it is inappropriate to hold him responsible (granting he is not subject to an incapacity). In one sense, the influence of our sympathetic response toward him can be thought of as "a psychologically unavoidable effect of learning of Harris´s past" (1998:138). However, some might argue that our sympathetic response toward Harris is not only unavoidable but an appropriate response (even grating that this case does not represent an incapacity).

---

[39] For a full account of all four challenges, se McKenna (1998:136).

According to Wallace, histories such as Harris´s provide "mitigating" factors to a person´s moral responsibility. If the impairments are not extreme (as it is in the case of psychopaths), what these histories might do is affect the agent by impairing his "power of reflective self-control without altogether depriving the adult of those powers" (1994:233). Therefore, when a person has been subject to unfortunate formative circumstances, what happens is that he might find it "extremely difficult" to comply with the norms of morality. Hence, an agent´s unfortunate past works as a mitigating factor.

Further, Wallace sees these mitigating factors not as "determinant" factors but rather as explanations for how an agent´s capacities have been diminished. For instance, "childhood deprivation affects the adult´s responsibility only insofar as it leaves continuing traces in the adult's psychological life" (1994:233). Answering to the skeptical diagnosis, Wallace argues that it is not that the agent´s attitudes were inevitable (determined) but rather, that his past presents him with difficulties to comply with the moral demand. In a way, these factors can make the agent´s responsibility in some sense "psychologically intelligible" (1994:233n8). Thus, it is not that the agent "was determined" to become evil, but rather that his past makes intelligible why it is more difficult to comply with the demands. Therefore, "unfortunate formative circumstances" do not show an incapacity; instead, they clarify why it is "overwhelmingly burdensome" for the agent to exercise this capacity. A capacity which the agent *does* have. Thus, unless it is shown that every person´s history represents a difficulty in complying with the moral demands, then there is no reason to draw universal skepticism of the reactive attitudes in cases were ambivalent responses arise. In specific agents with particular unfortunate pasts, "it seems that we ought to take their social and developmental circumstances into account, recognizing those circumstances to be mitigating factors when the responses of blame and moral sanction are in question" (1994:233). For instance, an example he mentions is that of people raised in poor and violent modern houses product of industrialization. These people, he argues, "are apt to find it much more difficult to take morality seriously than those whose formative circumstances and life prospects are more fortunate" (1994:233).

As discussed in the previous section (3.2.2), the expressivist account faces a charge of circularity: the standpoint of holding responsible presupposes facts about agency, namely the kind of agencies involved in these practices. Therefore, it is impossible to understand the conditions for responsible agency by merely laying out the standards we employ in holding

responsible. This is because "those standards presuppose that the agency at issue has a standing to which they – the standards- must answer" (2012:50).

The accounts proposed by Wallace and Mckenna above regarding cases such as Harris´ turn to the agents' capacities to communicate within the moral community. Indeed, both McKenna and Wallace point towards an understanding of the relevant capacities an agent must possess to be an appropriate target of our reactive attitudes. Therefore, one way to respond to the charge of circularity is to argue that the practices of holding responsible depend on the agent´s capacities, and that (in contrast to expressivism) it is the agents presupposed responsibility that which is more fundamental. To examine these capacities more thoroughly, I suggest that we turn to a realist reading of Strawson (Brink and Nelkin, 2013; Nelkin 2011). Perhaps there are reasons to support the claim that some "elements in Wallace's account [and McKenna´s] fit better with an alternative realist reading" (Brink and Nelkin: 2013:314).[40]

3.2 Realist account of Agential Responsibility

In this section, I critically discuss Brink and Nelkin´s (2013) rational abilities view. I examine and discuss two main factors that can mitigate or incapacitate an agent´s responsibility: normative competence and situational control. Further, I shall problematize the realist account by considering why the realist remains "hostage to traditional worries about freedom of the will" (Brink and Nelkin 2013:288).[41] Finally, at the end of this section, I return to the charge of circularity. While the realist contends that the conditions for moral responsibility are independent of our practice and thus that the metaphysical status of the responsible agent is more basic than our practices of holding responsible; the expressivist argues instead that it is the practices that contribute to the wrongdoers´ responsibility and that in this sense, the status of the morally responsible agent is fixed by these practices which turn out to be more fundamental. First, I reject the expressivist metaphysical and explanatory primacy of our practices. Then, I contend that another way to overcome the charge of circularity is to deny

---

[40] In the next chapter (section 4.6) I suggest that McKenna´s reading of Strawson (fully developed in McKenna (2012), which happens to be strongly influenced by Wallace´s, could also fit better with an alternative realist interpretation.

[41] An important question to raise in this respect is whether McKenna's view is also hostage to these worries, more on this on chapter 4.

that holding people responsible or being responsible should be explanatory or metaphysically more basic; both are mutually dependent (McKenna 2012).

Inspired by Strawson, expressivists claim responsibility should be accounted for by looking at the conditions that make it appropriate to hold people responsible. The expressivist endorses what Watson referred to as "the reversal" of the traditional order of explanation (2004:222). This reversal entails that holding people responsible is explanatory basic. For the expressivist, it is not that we hold people responsible because they are responsible. Instead, we should reverse the order of explanation: the idea that people are responsible "is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another" (2004:222).[42] In recent work, Watson argued that few have taken "the truly strawsonian turn" (Watson, 2014:16), namely the reversal.

However, even if we start by looking at our practices of holding responsible, why should we think the debate will only be a *normative* one? Although expressivists claim that the reactive attitudes are "constitutive" elements of moral responsibility and that the propriety of our judgments of moral responsibility should be based on the standpoint of holding responsible, this appropriateness (as discussed in the above section) seem to rest on the agent´s capacities that makes him an eligible target (of moral address) in the first place. Thus, the expressivist seems to face a charge of circularity in his argument.[43]i.e., it is impossible to understand what being responsible is by attending *exclusively* to the normative standards uncovered in our practices, this is because "the application of those standards presupposes that the agency at issue has a standing to which they -the standards- must answer" (McKenna; 2012:50). But then why should we understand our ascriptions of responsibility as "dependent" or "fixed" by our practices? For instance, if the answer to whether Harris is a morally responsible agent will be decided by *his* capacity for membership in the moral community" (McKenna; 1998:130, my italics), then why should we claim that responsibility can only be accounted for in terms of the reactive attitudes, instead of claiming that "the appropriateness of the attitudes depends on whether people really are responsible?" (Nelkin; 2011:29).

---

[42] For criticism of the strawsonian reversal, see Todd (2016).

[43] Mckenna mentions specifically this charge of circularity in Wallace (2012:50). For a similar objection to Wallace´s account, see Vargas (2004:225).

The realist reading of Strawson argues that the reactive attitudes track facts about the agent´s responsibility, and hence that the reactive attitudes only make sense when we presuppose the responsibility of the target. The realist claims that facts about the responsible agents will "uncover an independent conception of responsibility that can support the reactive attitudes" (Brink and Nelkin; 2013:291). Naturally, by focusing on the agent's rational capacities, the realist argues that it is the responsible agent "that makes the reactive attitudes toward them fitting or appropriate" (Brink-Nelkin; 2013:288) and not the other way around, as expressivists claim.

The problem Wallace´s expressivist account faces, according to realism, is "the well-known difficulty of explaining how we are to discern the conditions for `appropriate`[44] responses without appealing to independent features of the agent and situation in question, such as her rational capacities and opportunities" (Nelkin, 2011:29).

The realist accepts that in interpretative terms, the expressivist reading of Strawson or what they call a "response-dependent" reading of Strawson is probably the most accurate reading. However, they argue that as a systematic rather than as an interpretative matter, the realist approach, what they call a response-independent account, is better equipped to account for the nature of moral responsibility (Brink and Nelkin, 2013:287).

Instead of focusing on the reactive attitudes per se, the realist argues that responsible agents are beings with certain capacities and that it is in virtue of these capacities that they are appropriate targets of the reactive emotions. It is the agents´ presupposed responsibility, which makes the reactive attitudes appropriate, not the other way around. Thus, in the realist account, the reactive attitudes play an *evidential* role, and as such, do not constitute responsibility (Brink and Nelkin, 2013:287). Whether it is appropriate to target others with the reactive attitudes depends on the agent´s capacity at the time of action and whether he is nterfered by external factors that prevent him from exercising these capacities. Since these capacities do not dependent on how other members in the moral community hold the agent responsible, these are independent conditions. According to the realist, there exist two main

---

[44] In a similar vein, Todd claims that "strawsonians struggle with the word appropriate" (2016:209).

kinds of factors that can mitigate or incapacitate an agent´s responsibility: normative competence and situational control. I start by analyzing the former.

3.2.1 Normative Competence

Normatively competent agents can "act on their strongest desires" and are also "capable of stepping back from their desires and act on good reasons" (Brink and Nelkin, 2013:292).[45] Since normative competence presupposes recognizing and responding to reasons for action, this competence presupposes a certain amount of reasons-responsiveness.

Reasons responsiveness entails both cognitive as well as volitional capacities: cognitive capacities are needed in order to distinguish right from wrong. Volitional capacities are needed in order to conform one´s conduct to that normative knowledge (2013:292). According to the realist, "an agent is more or less responsive to reason depending on how well her judgments about what she ought to do and her choices would track her reason for action" (Brink and Nelkin, 2013:292-4). For instance, the agent´s reasons-responsiveness can help us "distinguish between the person who, weak-willed, freely drinks the beer from the person who is compelled by her addiction to do so (…) the former is responsive to a wider range of reasons for not drinking than is the latter" (McKenna, 2013:153). Further, responsiveness to reasons can be analyzed in degrees. Which, in turn, raises questions of what the adequate amount of reason responsiveness should be. Indeed, how responsive does an agent need to be, so as to be recognized as a responsible agent?[46] The realist opts for an intermediate degree of reasons responsiveness; this entails that "where there is sufficient reason for the agent to act, she regularly recognizes the reasons and conforms her behavior to it" (Brink and Nelkin, 2013:294).

"Being a responsible agent is not merely having the capacity to tell right from wrong but also requires the capacity to regulate one´s actions in accordance with this normative knowledge" (Brink and Nelkin, 2013:297). Therefore, the volitional dimension of normative competence is crucial since it enables agents to act based on their previous judgments. The volitional

---

[45] For views who deny the normative competence as a condition of moral responsibility see, Scanlon (1998,2008), Angela Smith (2005, 2008 2012) and Mathew Talbert (2008, 2012). For these theorists judging and acting on reasons only is enough for being an appropriate target of the reactive attitudes. For instance, Talbert argues that psychopaths may be blameworthy when they hurt us (even though they might be blind to the moral reasons against hurting us).

[46] Brink and Nelkin´s realism defend a reasons responsiveness that is agent based rather than mechanism based such as Fischer and Ravizza's. in this the realist finds similarities with McKenna's communicative approach he too bases his view on agent-based reason responsiveness.

dimension consists of emotional and appetitive capacities that enable agents to form their own intentions, making it possible to "execute these intentions over time, despite distraction and temptation" (Brink and Nelkin, 2013:297). Familiar difficulties in volitional competence include clinical depression, irresistible desires, or paralyzing fears, among others.

From the above description, we can conclude that the paradigm case of responsible agents for the realist amounts to "normal mature adults with certain sorts of capacities" (Brink and Nelkin 2013:292). Being a responsible agent according to the realist amounts to the mere possession of these capacities, and not whether agents properly exercise them. If an agent fails to exercise his capacities, he can still be regarded as a responsible agent. The realist doesn't "excuse the weak-willed or the willful wrongdoer for failing to recognize or respond appropriately to reasons" (Brink and Nelkin 2013:293) but does exclude the mentally deranged or young children for not having the relevant capacities. This realist interpretation is compatible with our judgments about young children and their gradual acquisition of these rational abilities. Something similar also applies to the severely mentally impaired.

Recalling Harris´ case (assuming once more, for the sake of argument that the case itself is nonevidential), according to the realist, whether Harris is a responsible agent or not would be in part, decided in terms of normative competence. What will help us make this judgment is whether he possessed an ability to recognize wrongdoing and an ability to conform his will to this normative understanding (Brink and Nelkin, 2013:293), and not whether he "actually" exercised this capacity or not.[47] Further, other factors that can influence our ascriptions of moral responsibility are situational factors. To these, I now turn.

3.2.2 Situational factors

The realist believes that there is more to being morally responsible than having normative competence (cognitive and volitional capacities). There are indeed external or situational factors that might compromise the agent's rational abilities and thus work as mitigating factors. The basic idea is that an agent may be normatively competent, but there might be factors "outside" of him that undermine his ability to "act for good reasons" (Nelkin 2011:7).

---

[47] In the criminal law for instance, insanity is used as an excuse because the agent lacks the capacity "to discriminate right from wrong at the time of action" (Brink and Nelkin, 2013:295) Here brink and Nelkin specify the difference between criminal and moral wrongdoing, while the cognitive abilities to recognize these two different kinds of norms might be different, the result might be that "it might be possible to be criminally responsible without being morally responsible and vice versa" (2013:296)

Prototypical cases of this impairment include duress and coercion. For instance, imagine a ship´s captain in the middle of a storm unable to control the course of the ship, thereby risking every man on board´s life (by no fault of his own).[48] External factors compromise the agent´s abilities, not because the agent lacks normative competence but because it raises the question of whether the agent is, all things considered, responsible for the tragedy. Other cases include instances of justifying the lesser of two evils. In these cases, the situational factors do not erase the fact that a person has committed an act of wrongdoing (if this is the case), but rather it provides excuses for the wrongdoing.[49] For instance, take the following example from Robert Kane´s book, "The significance of Free Will",

> A man is involved in an auto accident in which a young bicyclist has been killed. He reports to the police that the accident was caused by circumstances over which he had no control. The sunlight was bright in his face, reflecting off the front window, and he did not see the bicyclist (1999:83).

If the accident was, in part, the product of factors that somehow lay "beyond" the agents' control, then these factors could be understood as mitigating factors. This, unless the agent was responsible for these (as for instance the case of drunken drivers, the alcohol might affect your capacities at the time of the accident; nevertheless, you are responsible for the previous decision of drinking and driving, or at least you should have prevented this situation in some way).

Both normative competence and situational control are independent factors, but they might potentially interact with each other. Both factors correspond to "the architecture of responsibility" and thus lay out the conditions for responsibility under a single and unified independent principle: "the principle of fair opportunity to avoid wrongdoing" (2013:284). If an agent either lacks normative competence or situational control, the agent does not have a fair opportunity to avoid wrongdoing. Thus, these agents cannot be responsible. This principle is independent because it does not depend on people's disposition to react with responsive sentiments. The reactive attitudes do not explain by themselves why normative competence and situational control are required for moral responsibility. They only provide us with evidence, not with an explanation for why we hold people responsible.

---

[48] This case resembles Aristotle's example in NE (11108-9).

[49] This is indeed an interesting case. Since as we will discuss McKenna will grant that this case represents of an agent who did not act form an objectionable quality of will, which means that he is responsible but not blameworthy.

It is worth noticing that the way in which the realist "uncovers" this independent principle is by carefully examining the cases of excuses and exemptions. Exemptions work as general impairments, while excuses are case-specific, they involve otherwise normal and responsible agents but agents who acted inadvertently. The realist analyses both excuses and exemptions by focusing on three different dimensions of the obstacles involved in the agent´s culpability (scope, duration, and location). For instance, immaturity has a wide scope; it lasts longer, and it is located "within" the agent. Unintentionally stepping into someone on a crowded bus, has a narrow scope, it doesn't last long, and can be referred to "external" conditions such as a reduced amount of space.

3.2.3 Realism and Freedom of the Will

So far, the realist seems to avoid the problems I raised against expressivism, namely, the noncognitivist objection and the charge of circularity. The realist avoids the charge of circularity by using the reactive attitudes to uncover an independent principle. However, the realist hasn't yet explained the source of these independent conditions, which raises the following question: why are normative competence and situational control conditions for moral responsibility? The reactive attitudes provide us with evidence, but they themselves do not explain why normative competence or situational control should be independent conditions of moral responsibility. Then why should we agree that normative competence and situation control are the conditions of moral responsibility?[50]

Further, the realist seems to face another problem. Namely, skeptical challenges related to the threat of determinism. The realist relies on a specific notion of "control," which grounds the fairness of holding an agent responsible. But why is the notion of "control" so important to the realist? The realist argues, "control seems important, at least in part, because it seems unfair to blame agents for outcomes that are outside their control" (2013:305). Thus, the realist argues that the principle of fair opportunity to avoid wrongdoing presupposes the agent´s ability to do otherwise (alternative possibilities), which in turn explains why the

---

[50] This is what Rosen calls the "explanatory question", namely why are the conditions of blameworthiness as they are? (2015:65).

realist remains "hostage to traditional worries about freedom of the will" (Brink and Nelkin 2013:288).[51]

By appealing to the notion of control, the realist must answer to skepticism toward responsibility that appeals to the incompatibility between responsibility and determinism. The realist argues that "because our practices track forms of normative competence and situational control, rather than the truth of determinism, they promise to ground a compatibilist conception of moral responsibility" (2013:288). To support this view, they argue that the threats to our ability to avoid wrongdoing do not arise "from the fact that our actions are caused but rather from specific impairments of cognitive and volitional capacities and specific kinds of external interference with our ability to conform to the relevant norms" (2013:309).

Thus, the realist view must defend a compatibilist account. The realist argues such compatibility can be defended by grounding the general ability to do otherwise in the agent´s normative cognitive and volitional capacities and opportunities "at the time of the action" (2013:208). What this implies is that if "nothing impairs these capacities and opportunities in a relevant way" (2013:309), then the agent is free to do otherwise (Frankfurt-style cases and deterministic cases satisfy the condition above according to the realist).[52] For present purposes, I will not pursue this issue any further.[53] What is important to remark is that the realist view remains hostage to traditional worries about freedom of the will.

After having presented and critically discussed both expressivist and realist accounts, I have given solutions to the noncognitivist objection, and I have offered a possible solution to Watson´s problem of evil. But there is one problem that remains unanswered, namely the charge of circularity. Is there a way to solve this without appealing to an "independent" account of moral responsibility, such as the realist´s approach? To answer this, I suggest looking into McKenna´s constructivist view (2012). In the following, I shall give an account

---

[51] An important question to raise in this respect is whether McKenna's view is also hostage to these worries, this will be discussed in chapter 4, section 4.6.

[52] Answering the skeptical challenge that the deterministic past might prevent the agent from exercising the relevant capacities, Nelkin argues: "the fact that the past is deterministic does not make it interference or prevention in the relevant sense any more than is the counterfactual intervener in a Frankfurt scenario. Note Brink and Nelkin 2013:309).

[53] For a full defense of the realist defense of compatibilism, see Nelkin 2011, especially Chapter 5 "A Compatibilist Account of Agent Causation".

of what I think the problem of circularity amounts to and offer a solution, McKenna's Interdependence Thesis (2012:80).

3.3 Explanatory primacy: Rejecting the primacy of "holding people responsible" and answering the charge of circularity

In her paper, *On being and holding responsible (2007)*, Angela Smith criticizes theories of moral responsibility (such as Wallace´s) that understand responsible agency as the conditions that make it fair to hold people responsible. This way of theorizing on the nature of moral responsibility, by appealing to the conditions that make it appropriate to hold people responsible, is, according to Smith, "fundamentally misguided" (Smith; 2007:465). Her criticism, as well as McKenna´s criticism, explored earlier, target the expressivist´ claim that our practices are metaphysically and explanatory more basic than an agent´s responsibility.

She claims that while there may exist reasons not to "hold" a person responsible, yet we can still regard the person as responsible. Hence, even if a person is responsible for some act of wrongdoing, it doesn't necessarily follow from it that it would be appropriate for others to hold her responsible. Judgments of culpability are prior to blaming attitudes or expressions of blame.

For instance, suppose a friend of yours makes a racist remark. You might decide not to blame him for different reasons; it may be that you do not like pointing out people´s faults in public or that you know that the other person is undergoing a difficult period in his life. You might think, "I know what he did was wrong, but I do not blame him" (2007:477). However, even if you do not hold him responsible by expressing your negative emotions, this doesn't undermine the fact that his comment was racist. As Smith contends, "[s]ince the justification for blaming attitudes and responses depends upon factors that go beyond judgments of culpability, it does not follow directly from the judgment that a person is responsible and culpable for something that it would be appropriate to adopt or express blaming attitudes toward her for that thing" (2007:467). She might be at fault, but there may be other considerations that might lead us not to overtly blame her, this is, to target her with the reactive attitudes. Smith explores these considerations and underlines three of them. First, our standing to express these attitudes (it might be meddling or hypocritical of me); secondly, the significance of the fault (as the example above suggests) and thirdly the agent´s response, for

instance, it could be that your friend already refrained from making such the comment before you blamed him.

Smith´s discussion is more focused on the practices of blaming and what blame is than on our practices of "holding each other responsible", however, I grant both practices are overlapping phenomena, and either of them leads to considerations regarding the other. Usually, questions about agential responsibility arise, when people address each other, such as in the practices of blaming. Recalling our main interest in this thesis, namely, what is responsible agency and how should we analyze an agent´s responsibility, I believe Smith has something important to say in this respect. I take her to defend the claim that if we are to find an answer to whether a person is responsible or not, doing so by analyzing *exclusively* from the standpoint of those holding her responsible, may lead to a false conclusion.[54]

I agree with Smith´s criticism and thereby reject the expressivist claim that moral responsibility should be analyzed *exclusively* form the conditions that govern our practices, making the practices explanatory and metaphysically more basic than being responsible. In doing this, I will align myself with McKenna´s view in Conversation and Responsibility (2012). Here, McKenna argues that a person´s status as a morally responsible agent "has a metaphysical standing to which considerations about holding morally responsible must answer. If so, then holding morally responsible cannot unequivocally be regarded as more basic than being morally responsible" (2012:3).

At first glance, it may look as if McKenna holds a realist account of moral responsibility, yet he denies this. To understand why it is that McKenna doesn't align himself with responsibility realism, I must explain his concept of "agent meaning" and his "quality of will" condition for moral responsibility. Although McKenna rejects the primacy of holding responsible, he doesn't concede that being responsible is more basic instead. If there is a vicious circle, he argues that a way out of it is to deny that either of these (being or holding) is more basic. Mckenna defends a constructivist account in which, neither *being* responsible nor *holding* responsible is metaphysically nor explanatory more fundamental than the other. Instead of opting for an exclusive one-way explanatory direction, he defines the relationship "between responsible agency and the conditions for holding responsible as mutually supporting and

---

[54] For instance, consider Nelkin´s example: "a friend might release me from a promise to repay a debt to him, but I might still have an obligation to make the repayment, nonetheless. Similarly, a friend might forgive me for not repaying a debt, but that does not mean that I should not compensate him. Even if he does not `hold me to it, it does not follow that I should not repay him" (2011:47).

interdependent" (2012:54). When an agent performs an act, he is engaging "in a set of practices that presupposes considerations about the nature of holding morally responsible" (2012:49). Therefore, the agent´s initial action and, most importantly the meaning of his action depends on the agent´s own appreciation of how other´s might react in response to his action. This, McKenna argues, explains why being responsible and holding responsible are interdependent. To this conception of moral responsibility, I now turn.

4 Moral Responsibility and Conversation: A Constructivist Account

"I should have based my judgment upon deeds and not words."
(Antoine de Saint-Exupéry, *Le Petit Prince ([1944] 1995 :38)*

In this chapter, I explore McKenna's conversational account of moral responsibility (2012). I start by giving an overview of McKenna´s contribution and by pointing out both his affinities and differences with Strawson and strawsonian accounts of moral responsibility. Ultimately, I claim that despite his strong strawsonian commitments, McKenna is a "Strawsonian of a different stripe" (Vargas; 2016:225). Further, I present McKenna´s three-stage structure: The *Moral Responsibility Exchange*. Once having laid out McKenna's conversational analog, I explore McKenna´s "Interdependence Thesis" and McKenna's quality of will condition (2012:60). The agent's quality of will plays a prominent role in McKenna´s theory. Non-Responsible agents do not have such a quality of will. In this chapter, I discuss the quality of will condition and defend it against opposing views. After discussing McKenna's quality of will conditions, I problematize his view by discussing an apparent "dis-analogy" in the conversational analog. When people engage in a linguistic conversation, they usually intend to communicate something to their interlocutor. However, in the case of moral responsibility, wrongdoers usually prefer acting in secrecy and therefore do not wish to communicate anything to their victims, rather quite the opposite, some of them are quite explicit in their intentions to remain "undercover". How can McKenna accommodate this apparent dis-analogy? Far from being an objection, I believe the dis-analogy helps McKenna illustrate the validity of his "quality of will" condition. Further, I raise the question of how an agent's quality is of will is to be interpreted, and most importantly, who determines its meaning. To close this chapter, I raise criticism against McKenna's account. From his discussion, I draw similarities between McKenna's constructivism and responsibility realism.

4.1 McKenna and the Conversational Theory

In *Conversation and Responsibility* (2012), McKenna discusses the nature of moral responsibility by emphasizing the communicative aspect of our practices.[55] He does so by providing a conversational theory of moral responsibility, namely, an account that stresses the similarity between our practices of holding people responsible and linguistic exchange between two competent speakers of a common language. I will start this section by giving an overview of McKenna´s contribution and by pointing out both his affinities and differences with Strawson (1962) and strawsonian accounts of moral responsibility. Despite McKenna´s strong strawsonian commitments, I contend that McKenna is a strawsonian of a particular sort.

Following Strawson, McKenna embraces a "conative-affective" (2012:45) theory of moral responsibility. First, holding responsible is more than merely judging responsible; holding people responsible involves "adopting a practical attitude" (2012:22). For instance, you can judge *that* a person is morally responsible and blameworthy for some act. However, this doesn't necessarily imply that you *hold* the person morally responsible for it (McKenna 2012:22). When you hold someone responsible, you would consider it appropriate for others (or yourself) to adopt certain attitudes towards the responsible agent. This motivational function of the practices of holding responsible is illustrated in McKenna's example:

> Evil Sam can congratulate Sinister Cindy for her deliciously violent crime. Recognizing that Cindy acted freely in knowingly doing moral wrong, Sam fully recognizes *that* Cindy is blameworthy. But since Sam is delighted with Cindy's evil ways, he does not hold Cindy morally responsible for her crimes. Thus, while it is plausible to think that "holding morally responsible for *x* requires judging morally responsible for *x*, mere judgment is not sufficient for holding. (McKenna, 2012:22)

Holding responsible, according to the strawsonian view (and McKenna´s), should be explained by the morally reactive attitudes and the practices in which these emotions arise (2012:56). Further, McKenna endorses an interpersonal theory, one that presupposes a conversational exchange between members of a moral community. I will explore the characteristics of such exchange in what follows.

McKenna claims that considerations bearing on the practices of holding responsible are essential for judging that a person is responsible, however, a person´s status as a morally

---

[55] Joel Feinberg's ''The Expressive Function of Punishment'' (1965) is an example of a communicate theory of moral responsibility.

responsible agent "has a metaphysical standing to which considerations about holding morally responsible must answer" (2012:3). He suggests that none of these terms are prior to the other explanatorily or metaphysically. Instead, he argues for the mutual dependency of being and holding responsible. By focusing on the interpersonal nature of moral responsibility, McKenna is thereby led to his defense of the interdependence thesis, the mutual dependence of being, and holding responsible.

Regarding the free-will discussion, his theory is meant to remain neutral among compatibilist and incompatibilist accounts, and hence can be endorsed "independently of any commitment to the free-will controversy" (2012:2). This last point has been subject to criticisms.[56] Incompatibilists find it hard to accept McKenna´s impartiality with regards to the free will debate, particularly given McKenna´s strawsonian commitments. They criticize McKenna for not addressing the issue more straightforwardly, considering the importance of the role of determinism and its possible threat to the concept of moral responsibility (something even Strawson discusses in his 1962 piece).[57] I believe that part of the reason why McKenna does not take these topics more straightforwardly rests in his apparent intention of building a theory of moral responsibly that does not merely discuss the freedom condition for moral responsibility. I will have more to say about this intention and its criticisms by the end of this chapter.

Despite McKenna's affinities with Strawson, I argue that it is in how he departs from the strawsonian framework that his original account comes forward. As I already stressed, McKenna is indeed a "Strawsonian of a different stripe" (Vargas; 2016:225).[58] In support of this, I give two examples of how he departs from the strawsonian mainstream. The first one is his rejection of the explanatory and metaphysical primacy of holding responsible. McKenna argues that "a person´s status as a morally responsible agent has a metaphysical standing to which considerations about holding morally responsible must answer" (2012:3). When rejecting the primacy of holding responsible, McKenna distances himself from Wallace and likewise strawsonian accounts of responsibility (expressivist accounts).

Contrary to these views, he argues that the theory "needs to be reoriented so as to fix more directly on the agent who is responsible rather than on considerations about the nature of

---

[56] For criticism of this point, see Smilansky 2000; Russell 2012; Vargas 2016.

[57] I shall discuss this point later in section 4.6 when I raise objections to McKenna's account.

[58] In a reply to Vargas (2016), McKenna writes "I am not a typical strawsonian" (2016:260).

holding responsible" (2012:31).[59] To McKenna, holding responsible is not in some sense "prior" to being responsible, but rather "holding responsible is meant to be a response to an agent who is responsible" (2012:47). These arguments may sound as if McKenna holds a realist account, one that in terms of explanatory primacy, favors the primacy of being responsible. However, McKenna rejects the primacy of being responsible as well, and thus explicitly distances himself from realist accounts such as Nelkin´s and Zimmerman´s (2012:49). Instead of choosing a one-way explanatory direction, McKenna seeks a middle ground. His novel approach rests in the mutual dependency of being and holding responsible, namely, his Interdependence Thesis (2012:53).

The second way in which McKenna's account departs from other strawsonian accounts is the role he ascribes to the reactive attitudes and, most importantly, the central role he assigns to what Strawson characterized as "the agent´s quality of will" (2012, chapter 3). In contrast to Wallace´s expressivist account, which places a considerable amount of emphasis on the reactive attitudes, McKenna gives prominence to the agent´s "quality of will" (2012:57). Since the reactive attitudes are further responses to an agent's quality of will, McKenna argues we should place more emphasis on the latter rather than the former. In this way, McKenna´s theory reinserts a somewhat forgotten theme (especially within expressivism) from Strawson´s original account, namely the importance human beings attribute to manifestations of regard (or disregard) from other towards himself or other members of the moral community. The reactive attitudes are thus further responses to an agent´s manifestation of good or ill will as revealed through his conduct (Strawson, 2008:15).

But to this last point, one might object, was not Strawson´s original account already stressing the agent´s quality of will? It is true that Strawson´s view already stressed the importance of the agent´s regard for others, namely, the agent´s quality of will as revealed through her actions. However, so-called "strawsonian" accounts of moral responsibility, especially the expressivist account explored in the last chapter, have mainly focused on the reactive attitudes instead of the agent´s quality of will (see, for instance, Wallace 1994) thus giving less prominence to the agent´s quality of will. In contrast to these theories, McKenna somehow

---

[59] Here Mckenna shows affinities with Fisher and Ravizza (1993) and Arpaly (2006).

reinserts Strawson´s original insistence on the agent´s quality of will instead of focusing primarily on the reactive attitudes.[60]

McKenna rejects the explanatory and metaphysical primacy of being responsible. Moreover, he denies that agential responsibility can be accounted for independently from the standpoint of holding people responsible. Still, when stressing the importance of the agent´s quality of will, McKenna argues that "the nature of holding morally responsible and the underlying rationale for excuses, justifications, and exemptions are to be understood by reference to the agent who is responsible and, most crucially, the moral quality of the will with which she acts" (2012:78). By giving more prominence to the agent´s quality of will, rather than the various ways in which we hold the agent responsible through the reactive attitudes, McKenna focuses primarily on the nature of the responsible agent rather than considerations about the practices of holding responsible.

The underlying question, given McKenna´s emphasis on the agent´s quality of will, is whether it resembles and thereby favors a realist interpretation of Strawson. McKenna denies this. In McKenna´s account "the moral quality of an agent's will, as well as a person´s standing as a morally responsible agent is *dependent* upon the agent´s appreciation of expectation of due regard for others as revealed in the practices of those holding responsible" (2012:58 my italics). To McKenna, theorizing about the standpoint of the morally responsible agent inevitably leads us to considerations regarding our practices of holding responsible (something Wallace's account emphasizes). Similarly, our practices of holding responsible, are further responses to the agent´s quality of will, which naturally depends on considerations that pertain to the responsible agent (something Nelkin´s account stresses). Thus, he rejects an exclusive one-way explanatory and metaphysical direction, favoring instead the mutual dependency of being and holding responsible.

In summary, (1) McKenna´s rejection of the explanatory and metaphysical primacy of the standpoint of holding responsible and (2) his "reinsertion" of the agent´s quality of will as central to his account, distances McKenna's account from Strawson and strawsonians

---

[60] In another work (2006), McKenna claims the following "On the alternative Strawsonian account that I defend, morally responsible agency is to be explained in terms of the moral quality of will. The reactive attitudes are important insofar as they cast light on the importance we place upon the quality of an agent's will, as well as the conditions in which we think that an agent's will fails to meet the requirements for competent agency (2006:27).

accounts such as the expressivist account (Watson and Wallace). Both features are also essential to Mckenna´s defense of the Interdependence Thesis.

4.2 The Moral Responsibility Exchange

"(…) words are like the air: they belong to everybody. Words are not the problem; it's the tone, the context, where those words are aimed, and in whose company, they are uttered."
Juan Gelman

In this section, I shall start by pointing out three key ingredients to McKenna's theory, these are: the agent´s quality of will, the strawsonian reactive attitudes[61], and the pleas offered by responsible agents, such as excuses, exemptions or apologies (McKenna, 2012, chapter 3). The three key ingredients are already introduced by Strawson in his 1962 paper. What is original about McKenna´s account is how he presents them, namely, as part of his analogy between conversational competency and "the competency required for morally responsible agency" (2012:80). After presenting the Moral Responsibility Exchange, I will focus on the agent´s quality of will. The quality of will plays a prominent role in McKenna´s theory. I will discuss the quality of will condition and defend it against opposing views.

McKenna´s conversational theory is modeled on an analogy between our practices of holding people responsible and a linguistic conversation. The different ways in which we hold people responsible can be thought of as analogous to the different ways in which speakers of a common language engage in a conversation. For effective communication to take place, speakers of a common language must possess both; skills that allow them to communicate with others and interpretative skills to understand those that may answer back to them. Dialogue is thus possible only when all members share these skills.

According to McKenna, the same applies to the competence required for moral responsibility. To him, "acting skills and holding responsible skills are enmeshed" (2012:86). In order to communicate within the moral community, and to be a morally responsible agent, an agent must both (1)appreciate the meaning of his actions which McKenna understands as the quality of will revealed through them, and (2) possess the ability to interpret other agents´ quality of will. Consider the following case which illustrates McKenna´s conversational analog,

---

[61] McKenna follows Wallace in narrowing the class of moral reactive attitudes, namely resentment indignation and guilt. Nelkin does so too (Nelkin; 2013:286).

> Suppose you are standing outside a cafeteria waiting for a friend to arrive when you overhear a conversation. Two men are talking to each other saying what is perceived by you as an offensive, racist remark. Since you happen to be standing next to them, you can confirm that their jokes are racist and that they are targeting you. After some thought, you decide to blame them (you verbally express your feelings) openly. One of them denies having done anything wrong, while the other recognizes the offense and honestly tries to apologize.

Once an agent performs an action, that particular action is already open to the various interpretations from other members in the moral community, members who may request a justification and thereby members that may hold the agent responsible for that particular action. Actions, in a similar way to verbal expressions, serve as communicative tools that we use to express ourselves within the moral community. Most importantly, actions of responsible agents, are an expression of their "quality of will", namely, the agent´s regard for others or lack thereof. As the case above illustrates, the act of making a racist remark is revealing in the sense that it expresses a person´s regard for the others, i.e., it expresses the person´s quality of will.

McKenna defines the agent´s quality of will, as "the value or worth of an agent´s regard for others" (2012:59). But how can we identify in the agent, his regard for another? According to McKenna, the regard one has for others can be expressed in the agent´s intentions, choices, or her reasons for acting, but this "depends on context" (2012:59). Someone´s regard for others may not be revealed by her intentions or reasons for action, "sometimes it is what she fails to attend to that is revealing" (McKenna; 2012:59). For instance, if Sophie forgets to congratulate her best friend Sara with her child´s birth, even though she did not have any wrongful intentions towards her friend nor any bad reasons for acting (one could say that she simply forgot to do so), the fact that she failed to attend to what both consider a significant life event is what it is revealing about her regard for her friend. As Angela Smith puts it, "our tendency to notice and remember things are themselves a reflection of what we judge to be important or significant in some way" (2010: 523, fn4). What McKenna wants to make clear is that the regard one has for others is not a specific property that the responsible agent has; instead, it is the "*value" (2012:59)* or worth of the person´s will as revealed in her actions.

Actions performed by responsible agents have, according to McKenna, "agent meaning" (2012:4). The meaning actions have work similarly to words in a linguistic conversation. Both actions and words can represent a person´s intentions, motivations, and reasons for acting. Moreover, actions are revealing of the agent´s regard or lack of regard towards another

person. In order to decipher what their meaning is, a proper interpretation of them is required. It is by way of interpretation that we understand what words and actions reveal about the agent performing/pronouncing them. McKenna´s interpersonal theory suggests that members in the moral community stand ready to hold others responsible, based on the interpretations of their actions. Of course, interpretations can be erroneous, and more than often, misinterpretations occur. Therefore, the use of pleas, such as excuses or justifications, plays an essential role in the moral conversation that takes place between wrongdoers and victims.

Besides the interpretation of other members of the moral community, it is most importantly the agent *herself* who knows that her actions will be subject to interpretation what makes her a responsible agent. In this sense, being able to appreciate the meaning her actions might take is what grants her the (full) status of a morally responsible agent. To McKenna, in order to be a responsible agent, an agent must be able to anticipate how others may interpret his actions as well as the kind of responses his action might bring about. Without this capacity, an agent is not fully aware of the meaning his actions might have. Therefore, in McKenna's account, the morally responsible agent should be able to, up to some extent, *anticipate* the ways in which others might hold her responsible. To understand the importance of anticipating the various ways in which we might be held responsible for our actions, a deeper understanding of the role of the agent´s quality of will is needed.

According to McKenna, an agent can display an "ill quality of will," *only* if she can appreciate the meaning of her own wrongful actions (2012:94). This entails the agent´s ability to anticipate the various ways in which other members of the moral community might hold her responsible. Absent this appreciation, she cannot be held responsible for her actions since she is not acting from an objectionable quality of will, after all. For instance, we do not blame very young children, since we do not think they can (fully) understand the wrongfulness of their actions, the quality of their own will as revealed through their actions, nor the harm involved in them. McKenna´s responsibility theory fits with these intuitions we have towards children, namely, in McKenna´s theory, displaying an objectionable quality of will is a necessary condition for responsible agency. To act from a quality of will that can be morally assessed, the agent must have open before him the various ways in which he might be held accountable for his actions, very young children, or those with severe mental disabilities seem to lack this appreciation.

McKenna´s three key ingredients are built into a three-stage structure, namely the *Moral Responsibility Exchange.* The first stage of the moral conversation, the "Moral Contribution", unfolds with the agent´s blameworthy action. It is in this stage where the main theme of the moral conversation appears, namely the agent´s quality of will. It is the action and the quality of will of the morally responsible agent, which "opens up the possibility of a conversation" (2012:88).

The first stage is then followed by the stage of "Moral Address," which corresponds to the further responses from the victim (blaming practices). In this stage, the victim openly blames the wrongdoer and asks for a justification or an apology from him. There are many ways in which "blame" can take place, and many "pathologies of blame"[62] are to be avoided. Indeed "like most things in life, our practice of blame is susceptible to the vices of being done from the wrong sort of motive, in the wrong degree, in the wrong way, or with the wrong sort of object" (2016:168). One should always keep in mind that whenever an agent holds another responsible, her judgment is based on a mere suspicion. After all, all things considered, her judgment could be misplaced. Finally, the third stage, the stage of the "Moral Account". Here, the wrongdoer responds either by excusing or justifying himself or by asking for forgiveness.

It is important to notice that, in McKenna´s account, the pleas are explained and accommodated by the agent´s "quality of will", these offer "reinterpretations of an agent´s conduct (2012:75). For instance, excuses and justifications provide evidence to support the claim than an agent was neither responsible nor blameworthy since she wasn't acting from a morally objectionable quality of will (2012:78). In the case of exemptions, the agent doesn't have the status of a morally responsible agent, since he is incapable of acting from a will "that could have the relevant moral quality" (2012:78). In McKenna´s conversational account, mentally disabled people or young children are not morally equipped with what is needed for a felicitous conversation and thus are thereby exempted. In other words: "moral dialogue" with them is not possible (2012:77).

---

[62] Fricker (2016) I will draw from this work later in chapter five.

4.3 A further condition for Moral Responsibility: The quality of will condition

The agent´s quality of will, as revealed through his actions, plays an essential role in McKenna´s theory. This is made explicit in McKenna´s sufficient conditions for an agent to be morally responsible and blameworthy:

> a person is morally responsible and blameworthy for her action x if (and only if) she knows that x is morally wrong, she performs x freely, and in x-ing she acts from a morally objectionable quality of will (McKenna; 2016:61).

In the introduction to this thesis, I explored two widely recognized necessary conditions for moral responsibility, namely, the "epistemic condition" (knowingly doing wrong), and the "control condition" (freely doing wrong). As introduced in the first section, discussions about these conditions can be traced back to Aristotle´s Nichomachean Ethics (see section I). What is novel about McKenna´s theory of moral responsibility, is that these conditions do not "exhaust the consideration bearing on the quality of will" of an agent (2012:60). Therefore, he distinguishes an *additional* condition to moral responsibility, namely, "the quality of will condition" (2012:60). He argues that a further "quality of will" condition should be distinguished from epistemic and control conditions. In McKenna´s account, for an agent to be morally responsible and blameworthy, apart from "knowingly and freely" doing wrong, an agent must act from an *objectionable quality of will*.

Not everyone finds McKenna´s quality of will condition convincing. In *Who knew? Responsibility without awareness,* George Sher claims that agents "are often responsible for acts whose wrongness they are unaware" (2006:245), he argues agents can "be responsible and blameworthy for acts that they perform with wills whose defects they do not recognize" (2006:245). He strongly rejects McKenna´s claim that only agents who *know* how others might respond to their wrongful actions can act from an objectionable quality of will. For instance, to be able to recognize the wrongness of lying, Sher claims, it is not required for the agent to know how it would be appropriate to respond to liars. To prove him wrong, McKenna must somehow show that the practices of holding people responsible and the meaning it confers to the actions of responsible agents "penetrate the meaning of the stage at which a morally responsible agent acts" (McKenna, 2012:74). He must somehow provide the arguments to defend his claim that "a person´s inability to participate in the expressive and communicative practices of holding responsible" is "an impediment to her being responsible"

(2012:80). And perhaps most importantly, he must explain what these capacities or abilities amount to and specify their temporal dimension (Vargas 2016), this is, whether these capacities should be analyzed at the moment of evaluation of an action or whether they should be analyzed at the moment of action. To this discussion, I now turn.

Acting from an objectionable quality of will, as I discussed above, presupposes the agent´s ability to appreciate the significance of his actions, and most importantly, of the quality of will revealed by them. This further entails that the agent can anticipate the different ways in which other members in the moral community might hold him responsible were he to act in specific ways (recall Harris, though he declared himself a moral outlaw Harris defies the moral law and thus does not show lack of understanding, but instead as McKenna puts it, "it is not as if Harris is unable to appreciate our point of view. More than likely he anticipates it" (1998:132). Lacking this ability makes an agent incapable of *being* a morally responsible agent, in other words, the agent is impaired in a way that excludes him from being able to communicate within the practices of holding responsible with other members of the moral community. Therefore, according to McKenna´s quality of will condition, what makes an agent unable to be a morally responsible agent, is that she is unable to understand the quality of will from which he acts, in other words, "she is unable to express herself competently within these practices" (2012:79).

McKenna builds upon Watson´s notion of "moral address" (2004: 228). Recall that Watson argued agents are to be exempted if they aren't a proper target of moral address. What McKenna adds to these arguments is that it is an impossibility for the agent to participate within the practices that also makes the agent, not an eligible candidate. It is by focusing on this capacity that McKenna discloses the mutual dependency of being and holding responsible. Because the responsible agent must be aware of the various ways in which she might be held responsible, he argues, responsible agency depends on the standpoint of holding responsible.[63] Likewise, holding responsible is a response to an agent who is responsible. This explains, according to McKenna, their mutual interdependency.

Sher is not alone in criticizing McKenna's claim that for an agent to be blameworthy, he must, in fact, understand how others would react to his blameworthy action (See Holly Smith

---

[63] See Russell´s account in (2004) which offers a very similar reading. He expresses it somehow differently by giving prominence to the reactive attitudes "a responsible agent must be capable of holding herself and others responsible" (2004:288).

2014, Vargas 2016). Responding to some of these criticisms (McKenna, 2013), McKenna argues that "an agent who lacks an ability to understand the forms of communication expressed in the practices and norms of holding morally responsible is shielded from a rich range of reasons, and this adversely affects the sorts of meaning her actions can take on" (2013:75). To understand the "meaning" of an agent´s action, it will be helpful to introduce McKenna's distinction between what he calls the "moral quality of an *action"* and the "moral quality of an *agent´s will"* concerning her regard for others, as expressed in her actions (2012:18).

Consider the following example from McKenna: suppose a woman has an obligation to save a group of people from a certain death. She is also aware of the possibility of saving her child from the same harm. She is not able to save both the group and her child, which means that she must choose between either saving the group *or* saving her child. With great discomfort, she decides to save her child instead of saving the group. McKenna argues that although "she knows she is violating" a "moral obligation" and she does so freely and knowingly, "there is nothing negative revealed in the quality of her will in terms of her regard for other´s morally relevant interests" (2012:19). Therefore, it can be argued that she was morally responsible, but she was not blameworthy for her decision.[64] The woman "had great concern for moral considerations, and so was pained to make the choice she made" (2012:61). This example illustrates how the moral quality of an "action" and the moral quality of an "agent´s will" can come apart in significant ways. While the former might ground moral responsibility, the latter, McKenna claims, is a necessary condition for blameworthiness.

Another way to explore the role of McKenna's quality of will condition is to evaluate whether the agent´s quality of will affects the *degree* of an agent´s blameworthiness. For instance, consider two agents who freely and knowingly stole a loaf of bread. While one of them was delighted in doing so, the other did so out of necessity. Would the agent´s regard for others (namely the agent´s quality of will) affect our judgments of blameworthiness? McKenna thinks so (2012:62). Hence, if the case above provides us with grounds for distinguishing between degrees of blameworthiness, namely, if we would consider an agent *more* blameworthy than the other based on the quality of will displayed in their action, then there is

---

[64] Here we can consider the case presented by Aristotle which also resembles the case of a moral dilemma as McKenna presents it, "actions done because of fear for greater evils or because of something noble, for example, if a tyrant with control over your parents orders you to do something shameful, and if you do it, they will be saved, whereas if you do not do it, they will be put into death" this says Aristotle "gives rise to disputes about whether the actions are voluntary or involuntary" 1110a, NE book III

reason to claim, against skepticism of the further quality of will condition, that the quality of will condition does affect our judgments of moral blameworthiness after all.

One important reason why an agent who is incapable of acting from an "objectionable quality of will" is not morally responsible nor blameworthy is that the agent, according to McKenna, has no access to the so-called "Second Personal Reasons" (2012:83). McKenna´s theory draws significantly from this darwallian concept. Darwall distinguishes "Second Personal Reasons" from reasons of utility or general duty. When you refrain from stepping on someone else´s toes your reason for doing so, is not only because it causes unnecessary pain or because you recognize the duty not to harm others unnecessary, but simply because you acknowledge that the other person "expects that you´ll see the fact that he is asking you to do so as itself a reason" (2012:84). Therefore, when regretting her conduct, a morally responsible agent is able to recognize another person with a moral standing to make such demands. Being able to grasp second personal reasons is a necessary condition for being morally responsible. The agent must be able to recognize a face-to-face demand that comes from another person, as well as to acknowledge the other person as "one entitled to make such demands" (2012:83).

The existence of second personal reasons relies on the "presupposed authority and accountability relations between persons" (Darwall; 2006:8). Therefore, if an agent cannot appreciate second-personal reasons, McKenna argues, he cannot be an eligible target for moral responsibility since he is unable to recognize the "vast array of reasons presented to morally responsible agents" (2012:84).

Recalling the further quality of will condition and why we should consider this condition as necessary for blameworthiness, I can now formulate the following question: Why isn't it possible for an agent to be morally responsible and blameworthy after performing a morally impermissible action without showing ill quality of will or lack of regard? Or put differently, are there reasons that make actions wrong or impermissible that aren't reasons of disregard? Questions such as these make evident that there is still be room for further discussion; however, with further developments, McKenna's quality of will condition strikes me as right in many ways.[65]

The conclusion McKenna draws from this discussion is that if an agent is not able to understand the reactions his actions might elicit in others, the agent himself cannot appreciate

---

[65] For further criticisms, see Vargas (2016:233-235).

to its full the moral status of his own actions and thus is incapable of action from an objectionable quality of will. An agent

> who is incapable of participating in the complex social practices through which her reactive attitudes can be expressed is not only impaired in her ability both to appreciate the challenges put to her (…) she is also handicapped in her ability to be morally responsible, to act as a morally responsible agent (…) she is incapable of acting from a will with a moral quality that could be a candidate for assessment from the standpoint of holding responsible (McKenna; 2012:78).

To McKenna, the agent´s ability to be responsible depends on her ability to hold herself and others responsible (2012:88). The quality of will condition puts forward McKenna´s claim that to understand the nature of being responsible; we must also consider the agent´s ability to hold herself and others responsible.

4.4 Quality of will and the Interdependence thesis

The "quality of will condition" is highly connected to McKenna´s IT. What lies implicit in the quality of will condition, is McKenna's claim that moral responsibility presupposes the practices of holding responsible, establishing a mutual dependency between being and holding responsible (IT). To put this point more sharply, just as a competent speaker of a language must be able to interpret the meaning of both his words and his interlocutor's in order to have a felicitous linguistic communication; a responsible agent must be able to interpret both the meaning of his own actions as well as the actions of other members of the moral community. Therefore, McKenna claims that understanding the meaning of her own actions presupposes the agent´s awareness of the responses her actions might elicit in others, namely, the different ways in which she might be held responsible for them. To appreciate the communicative and deeply interpersonal aspect of moral responsibility highlighted in McKenna's conversational account, consider this famous poem by Valerie Cox titled "The Cookie thief" (1993),

> A woman was waiting at an airport one night, with several long hours before her flight. She hunted for a book in the airport shops, bought a bag of cookies and found a place to drop.

> She was engrossed in her book but happened to see, that the man sitting beside her, as bold as could be. . .grabbed a cookie or two from the bag in between, which she tried to ignore to avoid a scene.

So she munched the cookies and watched the clock, as the gutsy cookie thief diminished her stock. She was getting more irritated as the minutes ticked by, thinking, "If I wasn't so nice, I would blacken his eye."

With each cookie she took, he took one too, when only one was left, she wondered what he would do. With a smile on his face, and a nervous laugh, he took the last cookie and broke it in half.

He offered her half, as he ate the other, she snatched it from him and thought… oooh, brother. This guy has some nerve and he's also rude, why he didn't even show any gratitude!

She had never known when she had been so galled and sighed with relief when her flight was called. She gathered her belongings and headed to the gate, refusing to look back at the thieving ingrate.

She boarded the plane, and sank in her seat, then she sought her book, which was almost complete. As she reached in her baggage, she gasped with surprise, there was her bag of cookies, in front of her eyes.

If mine are here, she moaned in despair, the others were his, and he tried to share. Too late to apologize, she realized with grief, that she was the rude one, the ingrate, the thief.

Although the narrative in this poem is based on a "misunderstanding" and in a sense, does not represent a `successful` conversational exchange, it still helps us explore some aspects of moral responsibility I believe are central to McKenna´s theory. McKenna claims that "our moral relationships with others are built around our appreciation of how our conduct might be interpreted by them as revealing our regard for them" (2012:87). For McKenna, a morally responsible agent has "acting skills" and "holding-responsible skills" that are "enmeshed" (2012: 86).

Being a member of the moral community requires the ability to interpret the actions of other members as conveying different meanings. What is more, these interpretative skills apply to our own actions as well. As members of the moral community, we must choose our actions wisely. Hence, before an agent acts, he acts aware (up to some extent) of the different interpretations his actions might elicit. Therefore, responsible agents carefully choose their actions in order to avoid any misunderstandings about what others believe they mean by them. This is a fundamental capacity for responsible agency. For instance, by regretting her conduct, we can argue that the woman in the story can "appreciate the significance of her own actions, what they might or might not convey about her quality of will" (2012:86).

Consider instead the case of CarelessLucy. Once having realized it was someone else´s cookies she ate, she reacts with total indifference. She simply couldn't care any less. She has no interest in other´s moral assessment of her actions. However, if being able to appreciate the significance of her own actions as to what they convey about her quality of will is fundamental to responsible agency, would not then this imply that CarelessLucy, is in fact, impaired of responsible agency? The answer is no. In McKenna´s words, "it is one thing for such a speaker not to care, another for her to be unable to understand what she conventionally implies" (2012:87). Therefore, even if CarelessLucy is not concerned about the quality of will revealed through her actions, this doesn't imply that she is somehow impaired of responsible agency. No matter how little or inexistent her regard for others is, she remains a member of the moral community and is accountable for her behavior all the same.

In a famous passage from *Lying in Politics*, Hannah Arendt writes, "the trouble with lying and deceiving is that their efficiency depends entirely upon a clear notion of the truth that the liar and deceiver wishes to hide" (1969:34) Arendt´s remark helps us explore McKenna´s analogy between responsibility and communication. Firstly, it illustrates McKenna´s quality of will condition. For an agent to act from an objectionable quality of will, he must act aware of the meaningfulness of his actions as well as the different interpretations these might convey. Consider the parallel between wrongdoing and cheating. When playing a game, a person´s ability to cheat, as Arendt puts it, depends "upon a clear notion of the truth". Similarly, for a person to act from an objectionable quality of will, he must act aware of the quality of will displayed in his actions and aware of how others could respond to his wrongdoing. Otherwise, he cannot act from an "objectionable" quality of will.

Cheating requires understanding the "rules of the game" and wrongdoing requires understanding the different ways in which we hold each other accountable. Suppose that instead of cheating a person "accidentally" and "inadvertently" misread the rules of the game, one wouldn't consider him blameworthy of cheating. Cheating requires a particular understanding. Similarly, blameworthy conduct requires understanding the quality of will of one´s conduct as revealed in one´s actions. In McKenna´s own words, "appreciating how others would react to us allows us to make vivid to ourselves the meaning of our action" (2013:78).

Another feature of McKenna's account implicit in Arendt´s quote concerns the wrongdoer´s and his "wishes to hide". When people engage in a linguistic conversation, they usually intend

to communicate something to their interlocutor. However, wrongdoers usually prefer acting in secrecy and therefore do not wish to communicate anything to their victims, rather quite the opposite, some of them are quite explicit in their intentions to remain "undercover". This seems to be at odds with McKenna´s analogy between communication and responsibility. How can McKenna accommodate this apparent "dis-analogy" in his theory?

To discuss this, it might be helpful to understand McKenna´s notion of "meaning" implicit in an agent´s action. McKenna draws his understating of "meaning" from Paul Grice (2012:96-100). In his well-known contribution, "Meaning" (1957), Grice explains that "[f]or A to mean something by x… A must intend to induce by x a belief in an audience, and he must also intend his utterance to be recognized as so intended" (1957:219). He then distinguishes between "speaker meaning" and "sentence meaning", explaining how both meanings may come apart. Speaker meaning is what the speaker intends; sentence meaning represents the literal meaning of the words. For instance, if I tell my young son, "I was so hungry that I even ate the plate", he would most probably interpret my words literally rather than figuratively. In this sense, since both meanings come apart, the respective audience must manage to recognize the intention behind the utterance. The intended meaning would be that I am really hungry, while the sentence meaning would be that I actually ate the plate.

Following Grice, McKenna argues that just as Grice distinguishes between "speaker meaning" and "sentence meaning," it is possible to distinguish between what he calls "action meaning" and "agent meaning". The meaning of an action works similarly to the meaning of a sentence, and hence the meaning of an agent´s action and the meaning of an action can also come apart. However, the question underlying this communicative aspect of the analogy lies in the "intention", of the speaker. Regarding this, McKenna argues, "the agent who, at the state Moral Contribution, performs a blameworthy (or praiseworthy) act need not be thought of as intending to engage in something analogous to a conversation about the quality of her will as revealed in her action" (2012:90).

However, if the responsible agent, according to McKenna, "opens up the conversation" but at the same time has no specific intentions to communicate, why does McKenna insist on the analogy between conversation and responsibility despite its apparent dis-analogy? One possible "solution" to the dis-analogy is to reconsider the stages of the moral conversation. Perhaps McKenna´s conversational model should start with the reactive stance, namely, with the victim who holds the wrongdoer responsible instead of the wrongdoer's supposedly

blameworthy action. In this way, it is easy to see how the victim has clear intentions of addressing the wrongdoer. McKenna considers this possibility as a plausible alternative (2012:93-4). However, in the end, he argues that the Moral Responsibility Exchange should be accounted as starting at the stage of the Moral Address (2012:94).

McKenna emphasizes the importance of "agent meaning", which presupposes the idea that the actions of responsible agents are subject to the interpretation of other members of the moral community. Further, the agent performing the action is aware of the publicity of his action and that others stand ready to interpret them. When a wrongdoer intends to conceal his action, part of the reason why he does so is because he knows and understands the meaning his action bears in the first place. In other words, "that the source of the action, the agent, understands that her actions could take on that kind of meaning, is precisely why, for example, she sometimes acts in secret" (2012:98). Far from being an objection, I believe the dis-analogy helps McKenna illustrate the validity of his "quality of will" condition. Even if wrongdoers often act undercover, this doesn't erase the fact that they are able to anticipate the various ways in which they might be held responsible. As a matter of fact, the fact that they try to conceal their actions from others explains that they understand the meaning their actions convey, as Arendt puts it "they have a clear notion of the truth". Even if one tries to conceal the action, and in some way, tries not to communicate, one is actually communicating that one tries to hide the meaning of the action.

Mckenna´s quality of will condition and his analogy between communication and responsibility can be further explored by considering the difference between acts and deeds. One possible way to explore their difference is to say that while acts are related to an agent in a causal sense as when we say "Maria violently pushed me to the floor", deeds, in contrast, relate to, in McKenna´s own terms, the agent´s "quality of will" as revealed through his actions; as when we say, "Maria violently pushed me to the floor, in order to save me from a car accident". Therefore, when judging the blameworthiness of an agent, judgments should be based not merely upon acts but deeds. Actions should be part of our judgments, but only as long as we can interpret them as representing the person´s quality of will. In a similar way, this explains McKenna´s Gricean understanding of agent meaning, namely the difference between speaker and sentence meaning. The intention and the quality of will with which an agent acts can strongly differ from the words or the "sentence meaning," which the speaker

uses. Consider, for example patting a friend in the back; the conventional meaning is that you do so as a caring attitude rather than being rude.

When treated in isolation, actions can be interpreted as conveying different meanings, but with regards to an agent´s responsible agency, it is the agent´s regard for others reflected in them what matters. Freely and knowingly performing an act, doesn't reveal the agent´s regard for others, this is why McKenna defends a further quality of will condition;

> when interpreting a speaker, our interest is not directly in the meaning of the sentences which a speaker uses. Rather, it is in what the speaker means to convey in exploiting sentence meanings (sometimes in nonstandard ways) to express herself. Similarly, when determining questions of an agent´s responsibility for her actions, we are not directly concerned with the significance of the type of action an agent performs. That is, we are not concerned with what such an action typically signals about the quality of any persons´ will; rather we are concerned with the quality of the will with which an agent act (McKenna, 2012:96).

Understanding the agent´s quality of will requires some interpretation (2012:94). Similarly, deeds (in opposition to an action) requires an interpretation of the agent´s quality of will. Hence the difference between acts and deeds: while the latter seems to require some interpretation, the former can be understood without it. Actions of responsible agents in these senses are to be thought of as analogous to deeds. And not merely acts.

4.5 What kind of Conversation? Who interprets whom?

I have now arrived at the core of the conversational account: actions in a similar way than verbal expressions can be interpreted as revealing the agent´s quality of will. The agent´s quality of will is the main topic of the conversation that shall take place between the wrongdoer and the offended party. Further, as becomes clear from the above discussion, the agent´s quality of will is not only represented by the "facts", namely, what happened between the wrongdoer and the wronged party, but these facts should be accompanied by a further interpretation of how they reflect upon the agent´s quality of will. This is McKenna's constructivist account. A middle ground between expressivism and realism.

As we saw from the last chapter McKenna rejects the expressivist account (by rejecting the primacy of holding responsible, 2012:39-49). He argues instead that the victims' resentment is fitting only if the agent is fully responsible: "holding responsible should in the first-place answer to the facts about what it is being responsible" (McKenna 2012:50). Although the victim´s resentment helps us track the agent's quality of will, the resentment doesn't

"constitute" the agents´ responsibility as Watson and others contended. But doesn't this make McKenna a realist? Not really. Mckenna denies this; he argues that his theory is "expressivist in kind (McKenna 2012:57). This is because being a responsible agent and holding responsible are mutually dependent. In this account, "moral responsibility does presuppose an interactive relationship with or a response from those holding morally responsible" (McKenna 2012:53).

A question that may be raised to his account is regarding the agent´s quality of will. W*ho* is then supposed to determine which interpretation of the facts reveals the agent´s "true" quality of will? Most importantly, does McKenna´s defense that it is the wrongdoer´s initial action that starts the conversation imply that it is the wrongdoer who decides the "meaning" of his action? I do not think so. In a sense, it is true that McKenna gives prominence to the agent's quality of will and that the conversation unfolds with the agent's blameworthy action. But this does not imply that the answer to what the meaning of the initial action should be, is already settled by then. I believe it is in the course of the conversation, instead, that this matter should be settled. According to McKenna, this would imply that the meaning should be the result of the conversation, namely, of the conversational exchange between the wrongdoer and the offended party. The reason I believe McKenna focuses on the blameworthy actions as starting the conversational analog is that for the conversation to take place at all, then the initial action and hence, the wrongdoer herself, should be in possession of "holding responsible skills" that will make such conversation possible in the first place. Otherwise, without an objectionable quality of will, which is the main topic of the conversation, no actual conversation would take place between them.

Although wrongdoers act without intending to start a conversation, I think that McKenna is right in defending the wrongdoers' initial action as the first stage of the conversational exchange. One reason for this is that it fits perfectly with McKenna´s view of the responsibility practices as intrinsically social. While some tend to "see the individual as a kind of moral atom, and her relation to others in her moral community as extending outward only when she engages in free acts that obligate her to others", others such as McKenna "see the individual and the moral shape she takes as the product of her organic relation with a moral community" (2012:204). Although it might be true that wrongdoers generally do not act intending to start a conversation about their wrongful acts, I think here McKenna presupposes that morally responsible agents, hold a general commitment toward morality. In

this way, "one who cares about morality desires that relevant others "inform how she and others practically reason and act" (2012:168). Hence it would be reasonable to assume that responsible agents *would* like others to initiate a conversation about their wrongful acts after all since this carries implicit their willingness to commit themselves to the moral community and to morality in general.

An important thing to remark is this. I contend that although McKenna bases his account of moral responsibility as analogous to exchanges between two interlocutors in a linguistic conversation, I don't think he has in mind actual, or more specific, "literal" conversations.[66] The analogy is supposed to work in a broader sense, "blaming and other manifestations of holding morally responsible are a matter of altering highly complex social practices and interpersonal relationships" (2012:137). McKenna´s conversational theory is then made intelligible by the structure in the Moral Responsibility Exchange (Moral Contribution, Moral Address, and Moral Account). These stages correspond to each of the key ingredients of the theory, quality of will, the reactive attitudes, and the pleas, respectively. Although the conversation doesn't involve a "literal" exchange, it does involve an exchange between parties, one being supposedly held responsible for some action by virtue of his objectionable quality of will, and another party holding him responsible. Bear in mind that these exchanges can happen without a "literal" response from either party. A grimace, a shrug, a change of schedule can communicate as much as verbal communication does).

McKenna´s conversational account is modeled on a conversation that takes place between the wrongdoer and the offended party, namely cases of what McKenna calls "directed blame" (2013:121).[67] Naturally, the offender´s presence is crucial to this model, and cases of private blame or blame in the absence of the wrongdoer are to be understood as derivative or extended forms of this original model.

However, McKenna is fully aware that blaming in the absence of the blamed party is pervasive. Nevertheless, he maintains that the conversational theory should be based on directed blame (or as he also calls it the "conversation-as-address" model). Cases of private

---

[66] Vargas is aware of this as well since he mentions "The "conversation" need not be literal" instead he says, quoting McKenna "blaming and other manifestations of holding people responsible are a matter of altering highly complex social practices and interpersonal relationships" (2012:137).

[67] Here he defines directed blame in the following way: "*Directed blame* is a form of overt blame manifested in the presence of the blamed party (2013:121).

blame and blame in the absence of the blamed party, not only shows us the limitations of the conversational analog, most importantly as McKenna claims, these cases "brings into stark relief how infrequently we overtly blame in the presence of the blamed (2012:175). This raises two important questions against the conversational account: (i)"Is there not something problematic about building a theory of moral responsibility around the notion of an interpersonal exchange between blamer and blamed when such exchanges capture only a sliver of the actual extension of episodes of blaming?" (2012:175), and (ii) is the conversational theory simply incapable of explaining blame in this other wide range of cases?

In his contribution, *Responsibility and the Limits of Conversation (2016),* Vargas criticizes McKenna's model of conversation-as-address, he argues that the paradigmatic model should be based instead in conversations between victims and third parties in the *absence* of the wrongdoer, namely "to regard gossip as the paradigmatic conversational form: (2015:221). In the last chapter, I present Vargas´ challenge to the Conversational theory. I shall defend McKenna´s view and reject the gossip model. Ultimately, I contend, together with McKenna, that "moral responsibility is most fundamentally a dynamic, interpersonal affair that relates a person who is morally responsible for what she has done to those who stand prepared to hold her morally responsible by interacting with her in certain ways" (2012:iv).

4.6 The Conversational Theory: objections

Before I turn to defending McKenna´s theory from Vargas´s objection, I turn to an important question. Is McKenna's view immune to the objections raised against expressivism and realism? Mckenna´s construes the reactive attitudes as cognitive (2012:35), hence he avoids the problem of noncognitivism. Moreover, as I discussed at the end of chapter three, McKenna resolves the charge of circularity by rejecting the expressivists' primacy of holding responsible. He argues instead that being and holding responsible are mutually dependent. When it comes to the problem of evil, as I have discussed in chapter three, McKenna offers a solution by modifying the conditions of moral address as the capacities to participate within the moral community (and not an actual participation).

Mckenna´s rejection of the metaphysical primacy of "being responsible" differs significantly from Strawson and other strawsonian views. For instance, recall Wallace´s account. As discussed in the last chapter, Wallace claims that appealing to the metaphysics of free will

comes from misunderstanding what the normative demands involved in the practices of holding responsible are. McKenna takes a different path.

To begin with, McKenna (alike Strawson and Wallace) aims to account for the nature of moral responsibility without engaging in the free will debate. At the outset of his book, McKenna writes,

> Philosophical treatments of moral responsibility, even in books with the words "moral responsibility" in the title, are mostly eclipsed by attention to the free will topic, when the latter is regarded as fully satisfying the control condition for the former. But an account devoted to one condition for moral responsibility is not adequate as an account of moral responsibility itself. With few exceptions, most philosophers writing on the free will problem do attend a bit to the nature of moral responsibility as it figures into their preferred theory of freedom, although what is offered as a characterization of moral responsibility usually falls shy of a comprehensive theory on the subject (2012:2-3).

As the quote above illustrates, McKenna has clear intentions in construing a theory of moral responsibility, without discussing the free will debate (2012:2). In this sense, I contend that his theory comes closer to fulfilling Strawson´s overall project, namely a move towards reconciliation among compatibilists and incompatibilists. I thus take it as a merit of his theory that it intends to put into practice Strawson´s conciliatory plan by changing the topic of the agenda: from free will towards moral responsibility. However, despite McKenna's intentions, there are reasons that suggest that he doesn't succeed completely. To this issue, I now turn.

McKenna´s conversational theory supposes moral agents with specific capacities that enable them to be appropriate targets of moral address. In doing so, McKenna appeals to the notion of reasons responsiveness (see McKenna 2006:25). For instance, when answering the problem of evil, McKenna argued that we must focus on the capacity to *participate* in the moral community. According to his conversational account, an agent is morally responsible only "if he is able to be responsible and to hold herself and others responsible" (2012:88). Mckenna thus follows Watson and argues that "the person incapacitated for adult interpersonal relations- is so because she is incapable of understanding the complexities of the practices and reactive emotions of those in the moral community holding responsible" (2012:77). Further, what McKenna adds to Watson´s explanation of exempting conditions is that the agent is not only incapable of understanding the moral demand, but he is also incapable of participating in them in the first place (2012:77).

However, McKenna's talk about "capacity" is very little explored and specified in his book. For instance, while the realist draws on normative competence and situational control as conditions for moral responsibility, McKenna leaves these capacities in some sense "open". Further, McKenna does not explain what it would mean for an agent to have an unexercised capacity.[68] This, in turn, raises suspicions that providing an account of what amounts to these capacities might take as back to the free will debate in the same way that Nelkin´s realism did in the last chapter: when accounting for the capacities needed to reflect self-control, Dana Nelkin also appeals to the notion of reasons-responsiveness in order to explain the relevant capacities. Because of this, the realist remained hostage to traditional worries of freedom of the will. Therefore, if we explore what the capacities required for participating in the moral community are, this would probably pull us back into the free will debate. Mckenna would thus need to offer a defense of his compatibilist position, just as Nelkin did. Indeed, as Levy noted, McKenna´s "theory is hard to accept without embracing compatibilism" (2013:580). Therefore, I contend that although when it comes to the free will debate, he intends to remain neutral, in the end, his theory remains hostage to traditional worries about freedom of the will.

Further, the above objection makes it evident that McKenna´s account shares strong similarities with responsibility realism. One important criticism to raise in this respect is whether McKenna's constructivism is another version of responsibility realism. When it comes to McKenna's rejection of the metaphysical primacy of being responsible, why doesn't McKenna embrace this primacy if, after all, "holding responsible is meant to be a response to an agent who is responsible" (2012:48)?

Furthermore, when the realist claims that "being morally responsible" is metaphysically more basic, this does not necessarily imply that being morally responsible can be accounted for without considering our practices of holding each other responsible. I contend that it is one thing to defend the metaphysical primacy of being responsible (what Wallace called an extreme metaphysical view), and it is another thing to say that being a responsible agent can be accounted for independently of our practices of holding responsible. Recall McKenna´s example of the electron and the cloud chamber. When discussing the agent´s quality of will, Mckenna argues, "if we examine in more detail the internal workings of the conditions for Leslie´s being responsible, we will find, buried in them, substantive features of agency that

---

[68] For a thorough objection on this point see Vargas (2016) here he discusses the notion of the capacities required in the conversational account and raises difficulties for the overall account. For a reply from McKenna, see McKenna 2016.

concern the abilities involved in holding morally responsible" (2012:81). The cloud chamber is as much part of the electron as the practices of holding responsible are to a person's responsibility. I do not think it is possible to study the electron without reference to what is going on in the cloud chamber, analogously, I do not think that it is possible to analyze the concept of moral responsibility independently of our practices of holding responsible. However, this wouldn't contradict the claim that, in the end, it is the electron that is more basic.

In a previous work, while distancing himself from Strawson and strawsonian accounts, McKenna writes the following:

> The point on which I part company with Watson and other Strawsonians is in how this communication-based theory is developed. The moral conversations that are the "glue" of our moral responsibility practices on my account are initiated by morally responsible agents, and the place to begin in theorizing is with these agents, not with those in the morally community holding these agents responsible via the morally reactive attitudes (2006:28).

As the above quote illustrates, McKenna's account and Nelkin´s are highly compatible. Even if, when it comes to the explanatory and metaphysical primacy, Nelkin defends the primacy of being rather than holding responsible; she still acknowledges that moral responsibility presupposes our social practices, this is why, as discussed in section 3.2, she draws on the reactive attitudes and the different kind of pleas to help her elucidate the agent´s responsibility. Although other realist views such as Zimmerman´s defend the primacy of being responsible as well, there exists an essential difference between them: while for Zimmerman, moral responsibility is not inherently interpersonal in Nelkin´s account of moral responsibility the reactive attitudes play an essential role.

In her comments to McKenna´s reply to her criticisms, Nelkin writes, "ultimately, it is facts about the agent and her particular action that make it appropriate, when it is, to hold her responsible for her action" (Nelkin 2013:66). "Ultimately" here, I think, plays an important role. She then continues, "It can be true that being responsible for an action is prior—explanatory and metaphysically—to the appropriateness of holding responsible, and yet it can also be true that one condition of being responsible (at least for humans, or most humans) is a general capacity to understand what it is to hold responsible" (2013:65). I cannot see how this contradicts McKenna´s theory of moral responsibility. Therefore, the main question is: why is

McKenna so explicit in his rejection of the primacy of being responsible and in distancing from the realist´s and Nelkin´s account?

One reason, or at least part of the reason why he rejects Nelkin´s realism, I believe, is that while McKenna wants to take distance from realists accounts such as Nelkin's´ and Smith´s, he also distances himself from ledger theorists such as Zimmerman, and in doing so, it seems as if he is addressing problems that may not apply to both realists and ledger theorists equally. He considers, for example, the problem with an "intrapersonal" theory such as Zimmerman´s and argues it is misguided since "thinking of the moral responsibility facts in isolation from the practices of holding responsible, leaves out vital elements of the nature of responsibility itself" (2012:49). I argue Nelkin´s realism does not consider responsibility facts "in isolation" from our social practices. McKenna´s critical stance towards Zimmerman (and others such as Pereboom) certainly cannot apply to realists' positions such as Nelkin´s who consider the reactive attitudes and the strawsonian use of pleas as central to her account.

The above similarities raise questions as to whether McKenna´s view is just another version of responsibility realism.[69] But also it raises the preliminary objection that by appealing to the notion of reason-responsiveness, his view also seems hostage to traditional worries of freedom of the will just as much as the realist does.

---

[69]See Nelkin´s paper, "Moral Responsibility, Conversation and Desert; comments on Michael McKenna´s Conversation and responsibility" (2014).

5 Defending McKenna´s Conversational Theory: Vargas and gossip

 "I find that when I am gossiping about my friends as well as my enemies, I am deeply conscious of performing a social duty; but that when I hear they gossip viciously about me, I am rightfully filled with righteous indignation" Gluckman (1963:315).

In this chapter, I take up an objection raised against the conversational theory and reject it. The chapter will be structured as follows. First, I present Vargas´ criticism of McKenna's theory and his alternative proposal for the conversational model, the paradigm case of gossip. Next, I present what I consider to be the main characteristics of gossip. With a definition of gossip in place, I discuss whether gossip, as Vargas contends, manages to "capture much of what seems appealing about the conversational approach" while providing "an even more social" (2016:238) understanding of our practices of blaming. I disagree.

To ground my rejection, I offer four different, albeit related functions of the practice of gossip. I contend that none of these functions manages to capture one of the main features of moral responsibility, namely its communicative aspect. I argue that in the conversational model, the reactive attitudes are a specific kind of *communicative* entity. When we blame, our emotions have a specific addressee: the wrongdoer. Namely, "both blame and punishment have an expressive or communicative role; they would be pointless if they were meant to go unrecognized by *the person blamed*" (2012:135, my italics). The communicative aspect of blame[70], I contend, is best captured if we focus on public cases of directed blame, namely, overt blame in the presence of the blamed. Any other practice, such as gossip, is "second best" compared to directed blame.[71] Finally, I criticize Vargas´ intention of introducing gossip. Vargas sees the practice of gossip as promoting some kind of social advantage. I reject this account of blame. I argue instead that this utilitarian account of blame fails to capture the moral sentiments that Strawson so vigorously claimed should be at the center of ascriptions of moral responsibility. Recalling Strawson's famous passage, "the reactive attitudes do not merely exploit our nature; they express them (2008:27).

---

[70] For criticism against the argument from communication see, T. M. Scanlon (1998, 2008), Angela Smith (2013), and Matthew Talbert (2008, 2012).

[71] I will use "directed blame" and moral responsibility-as-address as interchangeable terms. Both suppose a conversation in which a victim overtly blames the wrongdoer in his presence.

5.1 Vargas´s Objection

In his paper, *Responsibility and the Limits of Conversation*, Vargas offers an alternative proposal to the conversational model. Namely, "gossip as the paradigmatic conversational form" (2015:221). In the paradigm case of gossip, the conversation "is *about* the offender"; further, the "offender need not be present, and indeed, frequently is not" (2016:238). Vargas argues that in this kind of conversation, we "express our concerns to the larger group and invite a shared commitment to judgments about what is permissible in order to induce or affirm norms of shared conduct" (2016:238). In gossip, he claims, "the chief insight of the Strawsonian approach is our disposition to socially significant violations of norms" (2015:238).

It is important to emphasize that part of the appeal of this new model, comes from Vargas´s objection to the conversational theory, namely that the theory struggles with accommodating "communicative" cases of private blame, and cases in which the offender is absent (either because he is unavailable at the time or dead).[72] These cases indeed challenge the main analogy between communication and responsibility, since these cases do not present us with an `actual` exchange or conversation between both parties.[73] Aware of this problem, McKenna offers a solution to justify these cases of blame (2012:173-178). He supposes a counterfactual theory in which blaming the dead would be understood in terms of "what would be fitting or reasonable ways to manifest blame as a means of engaging the one who is blamed, were she present" (2012:223).

Drawing on this difficulty, Vargas claims that the gossip model can help us solve the problem. Gossip, he claims, doesn't require the presence of the offender: people gossip about those not present, and people gossip about the dead. Hence, he claims this model "allows for the kind of dialogue McKenna emphasizes, but does not require it (…) What we need mainly, is the idea that the dead and absent violated a norm we accept or think others should accept" (2016:238). Ultimately, Vargas embraces the conversational account in general terms but rejects McKenna´s conversational model-as-address. According to Vargas, gossip not only "captures the main features of moral responsibility" (2016:221) but also comprises "an even

---

[72] For an account of this objection, see Vargas case of DEAD and GONE (2016:230). For further criticism of this point see Driver (2016) where she discusses the specific case of private blame.

[73] Bear in mind that the problem isn't that the conversation isn't literal, but rather "actual", it doesn't, so to speak, "happen".

more social conception of moral responsibility" (2016:222), which the conversational-as-address model doesn't. I disagree. In the following, I explain why.

5.2 What is Gossip?

Before I give an account of the reasons for grounding my rejection, I shall first clarify what I consider to be the main characteristics of gossip. What is it about the practice of gossip that makes us recognize it as such? In the following, I outline five characteristics that, together, amount to what I believe defines the practice of gossip. With a proper definition of gossip in place,[74] I shall discuss whether the conversational account should follow Vargas´s suggestion.

As a first characteristic, gossip is something that we do with *other* people; we do not gossip alone (or in one´s head, so to speak). Secondly, gossip is talking about someone who is *necessarily* absent (or at least about someone whom we believe not to be present). For instance, John Sabini and Maury Silver claim talk isn't gossiping if "we wouldn't be embarrassed by the subject's overhearing us" (1982:97). Thirdly, gossip is usually "idle" as opposed to oriented toward some explicit goal. However, there is a distinction in the literature that is worth mentioning here. The practice of gossip can be distinguished between "idle gossip" and "evaluative gossip". Idle gossip has simply no specific purpose or clear intention, as when, for instance, two friends gossip about how much money their boss makes. Evaluative gossip is a specific response to a norm violation. Thus, a fourth characteristic of gossip is that it is generally evaluative. Namely, there is always an implicit subjective evaluative judgment about the facts one gossips about. Finally, as a fifth characteristic, evaluative gossip is intended to articulate and confirm our evaluative view of the world. Further, it targets a specific and identifiable *person or group* and not merely a specific act of wrongdoing. This last characteristic explains why people usually gossip with those whom they share their values and personal tastes.

With a definition of gossip in place, we can now raise the question of whether gossip captures "much of what seems appealing about the conversational approach," and provides "an even more social," (2016:238) understanding of our practices of blaming.

---

[74] It is important to remark, that Vargas doesn't offer a definition of gossip. If this conception of gossip laid out here, doesn't share some characteristics with what is at stake in Vargas´ discussion, then I am inclined to think that it is not "gossip" what it at issue here, but rather some other practice that might share *some* characteristics with gossip.

5.3 Gossip, a "second best"

Suppose you have been recently involved in heated debates among your peers at work. After one of these intense discussions, just before you were about to leave, one of your colleagues verbally harasses you. Stunned, you cannot manage to respond to his rude manners. After this episode, you come back home, greet your partner, and before you could even realize, you are telling him about the incident. You feel resentful and talk "about the offender" (2016:239). You argue that he is responsible and blameworthy and that he owes you an apology. Further, you invite your partner, paraphrasing Vargas, to endorse your normative commitments and to express solidarity with you. After all, you know that your partner commits to the same norms himself: he would have never talked to a colleague of his in that way. Thus, both of you identify the offender as one who has acted in violation of group norms.

Your gossip did not start with your husband and did not end there either. On your way home, you met a friend in the subway and could not keep yourself from telling him. Further, that same evening, you gathered with your friends and could not help but ventilate your anger once more. Your gossip, in some sense, confirms that you hold him responsible for his rudeness, but also, even after gossiping with both your partner and your friends, this way of holding the wrongdoer responsible does not seem to help you get rid of your resentment.

Does this way of "blaming" the wrongdoer captures "the main features of moral responsibility" in a way that directed blame doesn't, as Vargas contends? As I see it, the practice of gossip misses an essential feature of moral responsibility, one which happens to be at the center of the conversational model.

Firstly, there is a question on what "the main features of moral responsibility" (2012:221) are. In this, it might be that Vargas and McKenna disagree. However, since Vargas considers that "the conversational model has considerable appeal" (2015:238) (he disagrees with the conversational *theory*), I take it that he does think that one of the main features of moral responsibility is its *communicative* aspect. The reason for believing this rests, in part, in why he offered the gossip model in the first place: because the "communicative" aspect of directed blame loses its point in blaming the absent or the dead. David Shoemaker (2015), Miranda Fricker (2016), and Coleen Macnamara (2016) each took a cue and underlined the importance of the communicative aspect of blame. Further, all of them took directed blame as their paradigm case. By drawing on their work, I intend to show how directed blame manages to

capture the main features of moral responsibility, precisely its communicative aspect, in a way that gossip does not.

Macnamara argues that the reactive attitudes seek "the uptake of the wrongdoer. A reactive attitude is a communicative entity with an addressee, and the addressee is the wrongdoer" (214:2016). Blame then "is not, like a radio correspondent's broadcast, for just anyone able and willing to give it uptake" (214:2016). Instead, blame is a demand directed at a specific agent. When I demand my husband to clean our apartment if his mother cleans his apartment instead, then "my demand has not achieved the response it sought" (214:2016). Fricker emphasizes the communicative aspect of blame by analyzing what she calls the paradigm case of "communicative blame". Communicative blame, she argues, is achieved when "the blamer *accuses the wrongdoer of fault*" (2016:171).

There is a clear sense in which blaming practices require a specific "target" of the responsive sentiments, whom they are directed to, so to speak. Further, those analyzing the communicative aspect of blame claim that the blamer, wishes the wrongdoer to recognize his protest and acknowledge what has happened. This "communication" continues with further responses from the parties involved: apologies, excuses, and justifications then enter the scene. Since blaming intends to be recognized by the wrongdoer, without the wrongdoers' acknowledgment, it is as if blame "loses its point". Blame is a sort of speech act " whose distinctive feature is that" it "cannot be fully successfully performed without the uptake of the hearer—that is, without the hearer recognizing the speaker's intention to perform just that speech act" (Fricker; 2016:173).

Mckenna´s understating of blame focuses on its communicative aspect too. As explored earlier in chapter four, the response of the victim at the stage of Moral Account, reveals a clear intention to address the wrongdoer with a complaint. Indeed, "both blame and punishment have an expressive or communicative role; they would be pointless if they were meant to go unrecognized by *the person blamed*" (2012:135, my italics).[75]

---

[75] Further, this communicative aspect of blame helps with mainting the bonds of the moral community, namely "If the community is going to stay together, these rifts need to be repaired. Repair happens when wrongdoers inhabit their fault, i.e., feel guilt and give it its natural expression in apology and amends. Thus, insofar as resentment and indignation elicit guilt in wrongdoers, they catalyze the reparative work that is essential to keeping the moral community together and healthy. (Macnamara 2016:221).

The above example of gossip is intended to illustrate its place in our blaming practices. Gossip can help us share our *protest* with others. Moreover, gossip can help us find support in others, as Vargas puts it in gossip, we can "build coalitions". In this way, in gossip, we hold wrongdoers responsible for their acts. However, nothing involved in this practice entails the acknowledgment of the wrongdoer. Indeed, all of it happens without the intention of communicating with the wrongdoer at all. This aspect of gossip should not be a surprise since gossip is meant to remain "secret" "behind the backs" of the wrongdoer. The problem with this picture of blame is that it misses the communicative aspect of blame in which the reactive attitudes are intended to be recognized by the wrongdoer. When blaming, we want to make the wrongdoer fully aware of what he has done, more specifically, that he *acknowledges* what was done. As Shoemaker observed, "if the offender (…) fails to recognize and understand one's protest *as such*, agential anger feels frustrated; it "aspires" to be more than just public noise. Mere protest, where only the protestor or non-slighting third parties make up the entire listening audience, is second best. In its fullest expression, agential anger has a confrontational, communicative motivational component (2015:105-6).[76]

Mckenna´s conversational account explores, among other things, our commitment to being part of a moral community. Being a morally responsible agent entails "being treated as a person whose actions are meaningful, and so can be taken to manifest the quality of her own will" (2012:163). Thus, the responsible agents' action has, what McKenna calls, Agent meaning.

Further, being a morally responsible agent, and hence an appropriate target of the reactive attitudes, exposes agents "to the risk of blaming practices" (2012:162). What does this risk involve? In the practice of blaming one´s interests usually, tend to be negatively altered. In blame, "a person´s moral standing is challenged, and sometimes drastically altered […] interpersonal relationships are adversely affected," and one is "is also sometimes not considered a candidate for various sorts of future relationships or exchanges" (2012:138). Furthermore, "it is emotionally unsettling to learn that others are angry or disappointed in us.

---

[76] Here Shoemaker gives the following example: "suppose I cause your downfall without you knowing I had anything to do with it" now suppose instead "I cause your downfall, but I make it known to you" (2017:105). What these cases show, according to Shoemaker, is that "agential anger with successful, but uncommunicated, revenge feels incomplete whereas agential anger with successful, but non-vengeful, communication does not" (2017:105).

But it is especially so to learn that they are angry and disappointed in us in the ways that are distinctive of blaming" (2012:139).

Another reason McKenna gives to support his conversational account of responsibility-as-address is related to the risks involved in the blaming practices. When we blame, he argues, blame "is directed at a specified wrongdoer" and admittedly "overt blame in the absence of the blamed could harm other parties" (2012:176). In directed blame, the wrongdoer is given the opportunity to offer an apology "a plea, or even a defiant rejection of the blamer's standing to blame" (2012:176). Blame is thus a "moral transaction" that "can only be featured when the blamed is on the scene" (2012:176). Therefore, I contend that the gossip model seems to lack the resources to explain the use of pleas such as justifications, apologies, or excuses. What is more, if Vargas´s gossip-based model can act as a free-standing thesis: Would then this imply that cases of directed blame are to be understood as degenerative cases of gossip? How would this work?[77]

Further, and perhaps one of the most familiar problems with gossip is that the wrongdoers´ absence jeopardizes the fairness of the practice itself. When a victim engages in a talk with a third party about a particular event or attitude, she is not only merely engaging in an explanation of the facts; she is already interpreting those facts. Therefore, were she to blame the wrongdoer in his absence this fact would naturally cast doubts as to how fair the blaming really is. As McKenna noted, "there is some sense of justice in the thought that a person deserves to be addressed directly if others are going to have her pay (social) costs for what she is alleged to deserve" (2015:262).

To this last point, it might be objected that confronting the wrongdoer may be more than a victim can manage; after all, the victim was victimized by the wrongdoer, which reveals an unequal power relation. Margaret Urban Walker (2006:134-5) made this a major point in her account of blame.[78] She suggests that the victim turns to the moral community instead, asking

---

[77] McKenna raises the following concern with the gossip model: it would treat in an extremely odd way what happens in those core cases in which an agent does directly morally address another in expressing her resentment and indignation. This is to be accounted for in terms of gossiping with the blamed agent about her culpable deed? Or are we to now treat these as the parasitic or degenerate cases? That is implausible on its face (2015:262).

[78] Walker noted that the reactive attitudes "invite confirmation from others that we have competently judged a normative violation, and that others share our interest in affirming the norms we hold, in showing disapproval of the conduct out of bounds, and perhaps in seeking redress of violations" (2006:26).

for support in addressing the wrongdoer. However, Walkers' [79] point here differs significantly from what Vargas has in mind when he talks about 'gossip'. Turning to third parties does not necessarily imply the intention of blaming in the absence of the blamed party, which lies at the heart of gossip.

Another concern with gossip is its scope. According to Fricker, "blame should be appropriately contained in its proper remit, both temporally and in terms of the relationship(s) it affects" (2016:169). This, in turn, implies that the expression of blame, "should not be allowed to go on too long and should not be allowed to migrate into regions of the relationship where it does not belong, or indeed to wantonly damage other relationships into the bargain. Blame should be allowed neither to fester nor to spread" (2016:169). This aspect of communicative blame immediately raises concerns towards the gossip model. Gossip, as it is usually well-known, spreads fast, which in turn creates the risk that blame might migrate into other social groups and hence damage other relationships in its process.

In arguing against the different "pathologies of Blame" (2016:168), Fricker mentions, among other things, the problem of blaming in the absence of the blamed party. Here she raises the following question, "in whom am I trying to inspire remorse as miles from the scene I chat casually about which party I blame? No one. With whom am I upset? No one. Whom am I addressing? Not the wrongdoer for sure" (2016:177). In the conversational analog, blame is communicative. One of the main features of communicative blame is that it is not "merely expressive (as if one needed to get the resentment off one´s chest) but rather transformative: the illocutionary point is to bring the wrongdoer to remorse for what they have done so that they come to be appropriately moved by new shared reasons incorporating the point of view of the wronged party" (2016:176). This last quote resembles McKenna´s understanding of blaming practices. According to McKenna, these intend to foster moral dialogue and aim at reconciliation. The communicative purpose I claim the model of Conversation-as-address has is meant to increase the alignment of the moral understandings of both the wrongdoer and victim, and "increasing alignment will be properly achieved dialogically" (2016:173). The dialogue should include both the victim and the wrongdoer.

Given the above reasons, and how gossip fails to fulfill the "communicative" aspect of blame, why does Vargas think that gossip "captures the main features of moral responsibility" and that

---

[79] Walker emphasizes the communicative aspect of blame herself too, blame, she argues "both expresses a sense of wrong and calls out to others for recognition and a reparative response" (2006:136).

we would "de better to regard gossip as the paradigmatic conversational form"(2016:221). As I see it, Vargas has a specific intention in offering gossip as an alternative model, an intention which I believe, generates problems with the conversational theory. When accounting for the practice of gossip, Vargas suggests that the practice of gossip can be understood "in teleological terms, as oriented around the attainment or support of some outcome" (2016:239). As I see it, Vargas sees the gossip model as playing a similar role to that of a moral calibrator. A social tool, one that could be used for "coalition-building, norm affirming, social status-establishing behavior" (2016:238). This idea may fit nicely with Vargas's agency cultivation model (2013), but how does this idea fit McKenna´s conversational account?

In what follows, I distinguish among different "functions" of gossip and point out which of these, I believe, rests behind Vargas´ intention of introducing gossip. After this, I argue why the specific function Vargas has in mind when introducing gossip generates problems with the conversational model.

5.4 The functions of Gossip

I recognize four functions of gossip: gossip can be informative, entertaining, therapeutic, and most importantly, it can be a powerful tool of influence.[80]

Firstly, gossip can be informative. Thanks to gossip, people can get to know information about other members within the moral community, which otherwise, they would have never known. Further, according to some gossip may painlessly gather information (Suls 1977). Although the targets of gossip could be *potentially* harmed. The kind of information that usually flows around by the practice of gossip helps us compare our own ideas and normative judgments about how to live to that of the peer group. Thus, gossip can help us get a better grasp of how to live and participate in a specific moral community. In this sense, gossip helps us build a social map of a specific cultural environment. This function is closely connected to what I see as gossip's second function.

Secondly, gossip is a powerful tool of influence. I contend that this function has a special connection with Vargas's intention to follow the "social regulation or moral influence style view" (2016:239). Gossip can strongly influence people to act and behave without trespassing the norms accepted within a moral community. Hence, it acts both as a mechanism of

---

[80] Indeed, gossip can "powerfully inform, strongly influence and thoroughly entertain" (Fine and Rosnow, 1978:167).

socialization and social control. And a compelling one. Take, for instance, how gossip affects the community of a local school or the environment of a specific workplace. Gossip, in this sense, can be a critical factor when it comes to a person´s membership within a specific group. In this sense, it is also true that gossip can also be used as a vengeful practice to protect ourselves from others. When it is aggressive enough, it may lead to ostracism. Gossip, in this sense, warns or encourages people about certain others, and influences people in either becoming or not, alike those subjects whom they gossip about. For instance, according to Gluckman (1963), gossip contributes not only to unity but to harmony of opinion feelings and interests of the people within a group.

Considering both positive and negative aspects, we can conclude that gossip can be a very ambiguous tool. In one sense, it is constructive. It avoids conflict and aims at resolution. In another sense, gossip tends to be destructive since it may degenerate, creating conflict and disruption that may lead to ostracism.

Thirdly, gossip can be a great source of entertainment. Many people admit that engaging in gossip with one's intimate partners can be very fun and distressing. The intimacy of a relationship seems to allow for such kind of dialogue too. As mentioned earlier, one of the main characteristics of gossip is that those gossiping usually share the same values and tastes.

Finally, gossip can also have a therapeutic function. When we gossip about how others violated a norm, we accept and think others should accept as well; we can manage to, so to speak, to express our strong feelings and emotions without open confrontation. Hence, gossip can help a person deal with his emotions while doing so without the risks or fear of confronting the wrongdoer.

As we can see, gossip can have different functions. I have explored that gossip can be informative, entertaining, even therapeutic. Moreover, I have stressed that gossip can be a useful source to influence people in specific ways. However, I contend that none of these functions manages to capture what is at the center of blame: its communicative aspect. As Macnamara noted, "a hammer gathering dust on the workbench has the function of pounding in nails; just so, the resentment that remains buried in one's heart has the function of eliciting uptake of its representational content" (562:2015). In the following, I discuss why Vargas's intention of introducing gossip fails to account for the communicative aspect of moral blame.

5.5 "The reactive attitudes do not exploit our natures, they express them" (2008:27)

According to the conversational account, blame is a dynamic rather than a static enterprise. It is a process that promotes moral dialogue and has the specific aim of "resolution and reconciliation", "it is the practice that sustains the bonds of the moral community" (2012:169).[81] This process, McKenna believes,

> is one that concerns the non-instrumental value of a process that begins at one end with a wrong done, that then conversationally answers that wrong by way of some blaming practice, and that invites an extension of the unfolding conversation in a manner that values sustained bonds of moral community. That can be understood as an end in itself, and not as something to value in the service of further considerations" (2012:169).

Mckenna sees the process of blaming itself as "non-instrumentally good" (2012:170)[82], a process that involves, among other things, the wrongdoers´ call to repentance. McKenna´s view of seeing the process as "an end in itself" and not an instrument for some further purpose, seems to be at odds with Vargas's intention of introducing gossip. When accounting for the practice of gossip, Vargas suggests that gossip can be understood "in teleological terms, as oriented around the attainment or support of some outcome" (2016:239). As I see it, Vargas´ intention of introducing the gossip model rests in it playing a similar role to that of a moral calibrator.[83]

McKenna's conversational account, as explored in chapter four, is strongly influenced by Strawson. As we can recall from chapter two, Strawson aimed to discuss the "central commonplaces" (2008:5) of our punitive and approbative practices within our interpersonal lives, the reactive attitudes. McKenna, following Strawson, argued that holding people responsible means "something in practice" (2012:67). The reactive attitudes affect our interpersonal lives by altering how we usually interact with people. For instance, when people feel angry or resentful with some closer acquaintance, they usually express their anger by modifying their conduct in some specific way. Some people might do this by showing less

---

[81] For a similar view see, Macnamara (560-564:2015).

[82] I will not defend this claim here. For McKenna's defense of the non-instrumental goodness of blaming practices see (2012:134-146).

[83] For instance, Vargas draws from Boehm. Boehm claims that gossip "functions as a system through which the groups idea of what should be morally acceptable or unacceptable is continuously rehashed and refreshed (1984:84).

interest in the other person. Others might blame them privately. Still, others might find it helpful to talk about it with their friends and families. However, underlying all these ways of blaming lies the assumption that the reactive attitudes seek the uptake of the wrongdoer, and most importantly his sincere acknowledgment of the wrong done. In blaming what our "resentment seeks is not getting back, but the other's acknowledgment of having wrongfully injured one and the other's taking responsibility for what he has done" (Darwall, 2011, 331).

The different practices of blaming can be (more or less) fitting. Nevertheless, what is important to remark is that Strawson´s intention of introducing the reactive attitudes was to emphasize the backward-looking character of the practices of blame (and praise). Which in turn, explains why the reactive attitudes are reactions to a past incident. As Strawson and others contended (Wallace 1994, 56), this backward-looking character cannot be captured by utilitarian reasons alone. This idea can be found in Strawson's memorable words: "the reactive attitudes do not exploit our natures, they express them" (2008:27).[84] Resentment is communicative because it tries to bring the wrongdoer to acknowledge the harm done and to see things from the victim´s perspective. If we substitute directed blame with gossip, then we give up this communicate standpoint and consider the social utility of these practices instead. The problem with this view is that "our resentment and indignation is not something whose direct aim is anything like promoting social advantage" (2016:262).[85]Thus, I believe that this strawsonian insight is best captured by cases of directed blame rather than gossip. Which, in turn, makes sense of understanding the other cases as degenerate forms of this model.

The analysis provided above, by no means, overrides the fact that gossip has been and still is an essential tool for socializing, as Vargas points out (2015:239). Rejecting Vargas´ proposal of gossip as a substitute of address does not imply that we should reject the practice altogether. Indeed, the gossip model "could easily be constructed as a friendly supplement to explain features of our practices", namely those which are "not easily accounted for by the conversational model, such as the way in which blame, even in the absence of the blamed, can function as a form of moral protest (see, e.g., Hieronymi 2001; Smith 2013; Talbert 2012)"

---

[84] Recall Strawson when answering to the optimist who defends a consequentialist account: "But the only reason you have given for the practices of moral condemnation and punishment (…) is the efficacy of these practices in regulating behavior in socially desirable ways. But this is not a sufficient basis, it is not even the right sort of basis, for these practices as we understand them" (2008:4).

[85] In a similar fashion, T. M. Scanlon argues that "the usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility which we now employ" (1988:160).

(McKenna, 2016:262). However, substituting directed blame with gossip fails to capture what Strawson already contended is at stake in moral blame, i.e., that "to speak in terms of social utility alone is to leave out something vital in our conception of these practices" (Strawson, 2008:24). The reactive attitudes implicit in blaming "do not exploit our natures, they express them" (Strawson, 2008:27).

When thinking about the paradigmatic cases of blaming, we are thinking about the specific relationship between "blame" and "holding responsible". Moreover, as McKenna contended, moral emotions such as resentment "are vehicles whereby we hold morally responsible" (2012:2). Understanding gossip as the paradigmatic case of the conversational analog, in other words, that holding people responsible should be understood by analyzing the paradigm case of gossip, would entail building a theory that dismisses the wrongdoers´ response (in the way of apology, excuse or justification). Gossip thus puts at risk the fairness of the practices by allowing the victim to blame in the absence of the wrongdoer. Instead, the paradigm case of directed blame presents us with a dialogic conversation, namely, one that presupposes a "cooperative" conversation between those holding responsible and those being held responsible. I contend that *this* is the "kind of dialogue that McKenna emphasizes" (2015:238) and not the kind of dialogue implicit in gossip. If blame is a way of holding agents responsible, and if holding responsible presupposes the relation between those being responsible and those held responsible, then one of the main features of blame and of moral responsibility itself is that it is relational and communicative. As I argued in this chapter, this communicative aspect of moral blame is best captured if we focus on the paradigm case of directed blame.

Conclusion

Moral responsibility is a fascinating subject within contemporary philosophy. Not only because of the significant contributions its rapid growth in the last decades has accomplished, but also because of the personal impact it has on our social and interpersonal lives.

When inquiring about any subject, such as the nature of moral responsibility, it is reasonable to think that our primary focus should lie in finding the right answers to the questions we investigate. Nevertheless, it is crucial to remember that finding the right answers to the question we investigate is as important as raising the right questions. For the past five decades, work on the topic of moral responsibility has expanded substantially. I contend that part of the reason behind this growth is owed to Strawson's contribution, *Freedom and Resentment*. Strawson did not give a fully developed account of moral responsibility, nor did he put an end to the metaphysical discussions about moral responsibility. Quite the contrary. His contribution was a source of inspiration for a considerable amount of incompatibilist positions, such as revisionist approaches. And most of its critics object that his theory was highly underdeveloped. However, what is notable about his contribution, I believe, does not lie in his answer to the question of what the nature of moral responsibility is. Rather the merit of his contribution lies in raising a new and essential question to our traditional understanding of moral responsibility. Instead of asking "what does it mean to be a responsible agent?", Strawson sought to answer "what does it mean to be an appropriate target of a specific class of emotions, the reactive attitudes". Although many have criticized this new approach, I believe that by raising a different question, Strawson has pushed us to explore and discover a deeper understanding of moral responsibility as well as the relational and interpersonal nature of ourselves as moral agents.

In this thesis, I explored metaphysical questions regarding the nature of moral responsibility. I started by presenting two conditions for moral responsibly, freely and knowingly doing wrong, drawn from Aristotle´s work in NE. The freedom condition, which is very much implicated in the free will debate, has often been at the center of discussions of moral responsibility. In this way, discussions on moral responsibility tend to focus on this specific condition for its possibility, namely, the freedom condition. Thus, discussions on the nature of moral responsibility remain anchored in the free will debate, i.e., since we cannot resolve the metaphysical issues at stake in the freedom condition, we cannot pursue a complete characterization of moral responsibility. Strawson suggested that we leave the metaphysical

discussions behind and instead focus on the practices of holding responsible, especially in the moral emotions at stake in them.

In this thesis, I presented and critically discussed Strawson´s theory. I raised criticisms against it, but most importantly, I contended that despite Strawson's strikingly influential theory, his theory failed in putting an end to metaphysical debates of agential responsibility. Influenced by Strawson, Gary Watson famously claimed that our practices of holding responsible are "incipient forms of communication" (2004:230). What does the communication between a wrongdoer and his victim imply about the wrongdoer's responsibility? I explored two main answers to this question. Watson and Wallace argued that Strawson defends an expressivist account of moral responsibility. To them, holding people responsible is a constitutive element of the wrongdoer's responsibility. Moreover, these practices are, in some sense, "primitive" and do not seek any other kind of "external" justification. Brink and Nelkin offered a different reading of Strawson, a realist reading. The realist argues that the wrongdoer's responsibility is independent of whether the victim reacts with responsive sentiments toward him. The reactive attitudes serve as evidence about when to hold people responsible, but it is the agent´s responsibility that makes these attitudes fitting or appropriate (Nelkin, 2013:287). Thus, the reactive attitudes track independent features (normative competence and situational control), independent of our practices.

I raised objections to both accounts. Whereas the realist contends that the conditions for moral responsibility are independent of our practice and thus that the metaphysical status of the responsible agent is more fundamental than our practices of holding responsible. The expressivist argues instead that it is the practices that contribute to the wrongdoer's responsibility and that in this sense, the status of the morally responsible agent is fixed by these practices, and thus the practices turn out to be more fundamental. The problem with this last view is that the conditions for responsible agency cannot be exclusively understood from the normative standpoint of holding responsible; this is because those standards presuppose a specific kind of agency to which those standards should answer. We cannot define moral responsibility by answering the question of when it is appropriate to hold responsible. This is because the appropriateness rests on the specific kind of agency of the agents involved in the practices themselves. Therefore, I argued that expressivist accounts meet a charge of circularity in their argument. I rejected the expressivist metaphysical and explanatory primacy of our practices. Then, drawing on McKenna (2012), I argued that one way to overcome the

94

vicious circle is to deny that holding people responsible or being responsible should be explanatory or metaphysically more basic; both are mutually dependent. This is McKenna's constructivist account.

In the last two chapters of this thesis, I explored McKenna's constructivism. In the end, although I raised an important criticism against his account, namely that his view remains hostage to traditional worries about freedom of will, I contend that from the accounts explored in this thesis, McKenna's comes closer to fulfilling Strawson´s overall project. McKenna's intention in developing the conversational account was to explore the nature of moral responsibility by analyzing it independently of the free will debate. This aim demonstrates how McKenna took up Strawson´s ambition in *Freedom and Resentment* and tried to put it into practice. He did so by leaving the metaphysical discussions about free will and determinism and focusing instead on the interpersonal and expressive nature of our practices. To do this, he relied on an illuminating analogy between conversation and responsibility. By relying on this analogy, McKenna managed to accommodate the main features of moral responsibility.

In the last chapter of this thesis, I defended The Conversational Theory from an objection. McKenna's expressive and interpersonal account presupposes the actual exchange between the wrongdoer and the victim. However, the pervasive phenomenon of private blame and blame in the absence of the wrongdoer threatens to render his account inadequate. Vargas tried to solve this inadequacy by providing a more social conception of our practices, the gossip model. I rejected Vargas´ suggestion by defending the communicative aspect of the conversational account. I argued that in the conversational model, the reactive attitudes are a specific kind of *communicative* entity. When we blame, our emotions have a specific addressee: the wrongdoer. Recalling David Shoemaker, blame is meant to "be more than just public noise" (2015:105). Blame seeks the wrongdoer, most specifically the wrongdoer´s acknowledgment of the wrong done. If blame is inherently communicative in this specific way, then I argued, it is best captured by the paradigm case of directed blame. "[I]n its fullest expression, agential anger has a confrontational, communicative motivational component (2015:106). Therefore, I contended that any other practice of blame, such as gossip, is "second best" compared to directed blame. As Darwall puts it, blame (and praise as well) come with an RSVP (2006:40). Such a request is directed to the wrongdoer specifically.

References

Adams, Robert Merrihew. 1985, "Involuntary Sins", *The Philosophical Review*, 94(1): 3–31.

Arendt, H. 1969, "Lying in Politics" In Arendt, H. (Ed.), Crises of the Republic: 1–48. New York: Harcourt, Brace and Company.

Aristotle. 1985, *Nichomachean Ethics*, trans. Terrence Irwin. Indianapolis, IN: Hackett.

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

    2006, *Merit, Meaning, and Human Bondage*. Princeton, NJ: Princeton University Press.

Bell, Macalester. 2013, "The Standing to Blame: A Critique" in D. Coates and Tognazzini (eds) *Blame: Its Nature and Norms*. New York: Oxford University Press, pp 66-83.

Bennett, Cristopher. 2002, "The Varieties of Retributive Experience." *Philosophical Quarterly* 52 (207): 145-163).

    2013, "The Expressive Function of Blame", in D. Coates and N. Tognazzini (eds), *Blame: Its Nature and Norms*. New York: Oxford University Press, pp.263-81.

Bennett, Jonathan. 2008, "Accountability (II)." In McKenna and Russell: 47-68.

Berofsky, Bernard. 1966, *Free will and Determinism*. New York: Harper & Row.

    1995, *Liberation from Self*. Cambridge. Cambridge University Press.

    2000, Ultimate Responsibility in a Deterministic World, *Philosophy and Phenomenological Research,* 1:135-140.

Boehm, Christopher. 2012, Moral Origins: *The Evolution of Virtue, Altruism and Shame.* New York: Basic Books.

Boswell, James. [1791] 2012, James Boswell's Life of Johnson. New Haven, CT: Yale University Press.

Brink, David O. and Dana K. Nelkin. 2013, "Fairness and the Architecture of Responsibility", in Shoemaker 2013: 284–314.

Clarke, Randolph. 2003, *Libertarian Accounts of Free Will.* New York: Oxford University Press.

Coates, D. Justin and Neal A. Tognazzini. 2013, "The Contours of Blame," in D. Coates and A. Tognazzini (eds), *Blame: its Nature and Norms*. New York: Oxford University Press, pp 3-26.

Coates, D. Justin and Neal A. Tognazzini. 2013, *Blame: its Nature and Norms*. New York: Oxford University Press.

Cox V. The cookie thief. In Canfield J & Hansen MV: Chicken Soup for the Soul. Deerfield Beach, FL: Health Communication Inc.: 1993

Darwall, Stephen. 2006, *The Second-Person Standpoint: Morality, Respect, Accountability.* Cambridge, MA: Harvard University Press.

Downie, R.S. 1966, "Objective and Reactive Attitudes" *Analysis*, 27:32-39.

Driver, Julia. 2016 Private Blame, *Criminal Law and Philosophy*, 10:215:220.

Feinberg, Joel. 1986, *Harms to Others*. New York: Oxford University Press.

Fine, G. A., & Rosnow, R. L. 1978, Gossip, Gossipers, Gossiping. *Personality and Social Psychology Bulletin*, 4(1), 161–168.

Fischer, John Martin. 1994, *The Metaphysics of Free Will*. Oxford: Blackwell.

Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. 2007, *Four views on Free Will.* Malden, MA: Blackwell.

Fischer, John Martin, and Mark Ravizza. Fischer, John Martin, and Mark Ravizza. Eds., 1993. *Perspectives on Moral Responsibility*, Ithaca: Cornell University Press.

    1998, *Responsibility and Control: An Essay on Moral Responsibility.* Cambridge: Cambridge University Press.

Frankfurt, Harry. 1969, "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829-839.

1971, "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68:5-20.

Fricker, Miranda. 2016, "What's the Point of Blame? A Paradigm Based Explanation", *Noûs*, 50(1): 165–183.

Ginet, Carl. 1966, "Might We Have No Choice?", in *Freedom and Determinism*, Keith Lehrer (ed.), New York: Random House, pp. 87–104.

1990, *On Action*. Cambridge: Cambridge University Press.

Glover, Jonathan. 1970, *Responsibility*. New York: Oxford University Press.

Grice, H.P. 1957, "Meaning, *Philosophical Review 66*. Reprinted in Grice, 1989: 213-223.

Gluckman, M. 1963, "Gossip and Scandal" *Current Anthropology,* 4:307-316.

Haji, Ishtiyaque, 1998, *Moral Appraisability*. New York: Oxford University Press.

Hart, H. L. A. 1968, "Postscript: Responsibility and Retribution," in Hart, *Punishment and Responsibility: Essays in the Philosophy of Law.* Oxford: Oxford University Press, pp.210-37.

Hieronymi, Pamela. 2004, "The Force and Fairness of Blame," *Philosophical Topics* 18:115-48.

Hume, David. 1978, *An Enquiry concerning Human Understanding*, ed P: H. Niditch. Oxford: Clarendon.

Kane, Robert. 1999, *The significance of Free Will*, Oxford: Oxford University Press.

1999, "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism", *Journal of Philosophy*, 96(5): 217–240.

2005, A Contemporary Introduction to Free Will. New York: Oxford University Press.

Lamb, James W. 1977, "On a Proof of Incompatibilism", *The Philosophical Review*, 86: 20–35.

Levy, N. 2005, "The Good, the Bad and the Blameworthy", *Journal of Ethics and Social Philosophy* 1(2): 1-16.

2011, Hard Luck: How Luck undermines Free Will and Moral Responsibility. New York: Oxford University Press.

2013, Conversation and Responsibility (Book Review), *Mind*, 486:578-582.

Lucas, R. J. 1993, *Responsibility*, New York: Oxford University Press.

Macnamara, Coleen. 2013, "Taking Demands out of Blame," in D. Coates and Neal A. Tognazzini (eds), *Blame its Nature and Norms*. New York: Oxford University Press, pp. 141-61.

Mason, Elinor. 2015, "Moral Ignorance and Blameworthiness," Philosophical Studies.

McGeer, Victoria. 2013, "Civilization Blame," in D. Coates and N. Tognazzini (eds), *Blame its Nature and Norms*. New York: Oxford University Press.

McKenna, M. 1998, "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism." Journal of Ethics 2: 123-142.

2000, "Toward a speaker meaning Theory of Moral Responsibility." In Ton van den Beld, Ed., Responsibility &Ontology. Dordrecht, Netherlands: Kluwer Academic: 247-258.

2005, "Where Strawson and Frankfurt meet," *Midwest Studies in Philosophy* 29: 163-180.

2006, "Collective Responsibility and an Agent Meaning Theory." *Midwest Studies in Philosophy* 30: 16-34.

2009, "Compatibilism and Desert," *Philosophical Studies* 144(1):3-13.

2012, *Conversation and Responsibility,* Oxford University Press.

2013, "Directed Blame and Conversation", in D. Coates and N. Tognazzini (eds), *Blame its Nature and Norms*. New York: Oxford University Press.

2014, "Defending conversation and responsibility: reply to Dana Nelkin and Holly Smith", *Philosophical Studies* 171:73-84.

2016, "Quality of Will, Private Blame and Conversation: Reply to Driver, Shoemaker, and Vargas" Criminal Law and Philosophy, Vol10(2), pp.243:263.

McKenna, Michael, and Paul Russell. eds.2008, *Free Will and Reactive Attitudes: Perspectives on P. F. Strawson′s "Freedom and Resentment,"* Aldershot, UK: Ashgate.

Mele, Alfred. 2006, *Free Will and Luck.* New York: Oxford University Press.

2010, "Moral Responsibility for actions: Epistemic and Freedom Conditions" *Philosophical Explorations* 13:101-11.

Milam, Per-Erik. 2016, "Reactive Attitudes and Personal Relationships", *Canadian Journal of Philosophy*, 46(1): 102–122.

Nagel, Thomas. 1976, *"Moral Luck." Proceedings of the Aristotelian Society* 50: 137-151.

1986. *The View from Nowhere*. New York: Oxford University Press.

Nelkin, Dana Kay. 2008, "Responsibility and Rational Abilities: Defending an Asymmetrical View", *Pacific Philosophical Quarterly*, 89(4): 497–515.

2011, *Making Sense of Free Will and Responsibility*, New York: Oxford University Press.

2014, "Moral Responsibility, conversation, and desert: comments on Michael McKenna′s conversation and responsibility," *Philosophical Studies* 171: 63-72.

2015, "Psychopaths, Incorrigible Racists, and the Faces of Responsibility", *Ethics*, 125(2): 357–390.

Oshana, Marina, 2004. "Moral Accountability." Philosophical Topics 32:255-274.

Pereboom, Derk. 2001, *Living without Free Will.* Cambridge: Cambridge University Press.

Rosen, Gideon, 2004 Skepticism about Moral Responsibility, *Philosophical Perspectives*, 18:295-313.

2015. "The Alethic Conception of Moral Responsibility" In Clarke, Randolph, Michael McKenna and Angela smith (eds.), *The Nature of Moral Responsibility. New essays,* Oxford: Oxford University Press.

Russell, Paul. 1992, Strawson´s Way of Naturalizing Responsibility. *Ethics* 102: 287-302.

Sabini, John, and Maury Silver. 1982, Moralities of Everyday Life. New York: Oxford.

Saint-Exupéry, Antoine. 1995, *The Little Prince,* Hertfordshire: Wordsworth Classics.

Scanlon, T.M. 1988, "The significance of choice." In The tanner lectures on Human values. Ed. S. McMurrin. Cambridge: Cambridge University Press.

1998, *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

2008, Moral Dimensions: Permissibility, Meaning, Blame. Cambridge, MA: Belknap Harvard Press.

Shabo, Seth. 2012, "Incompatibilism and Personal relationships: another look at Strawson´s objective attitude," *Australasian Journal of Philosophy*, 90:1, 131-147

Schlick, Moritz, 1930 [1966], "When is a Man Responsible?", in his *Fragen der Ethik*, Vienna: Verlag von Julius Springer. Translated in his *Problems of Ethics*, David Rynin (trans.), New York: Prentice-Hall, 1939. Reprinted in Bernard Berofsky (ed.), *Free Will and Determinism*, New York: Harper & Row, 1966, 54–63.

Sher, George. 2006, *In Praise and Blame*. New York: Oxford University Press.

2015, Responsibility, Conversation and Communication, The Nature of Moral Responsibility: New Essays, edited by Randolph Clarke, Michael McKenna and Angela Smith

Shoemaker, David. 2007, "Moral Address, Moral Responsibility, and the Boundaries of Moral Community." Ethics 118:70-108.

(ed.) 2013, *Oxford Studies in Agency and Responsibility Volume 1*, New York: Oxford University Press.

2015, *Responsibility from the Margins*, New York: Oxford University Press.

2017, "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame", *Philosophical Review*, 126(4): 481–527.

Smart, J.J.C. 1963. "Free Will, Praise and Blame." Mind, 70: 291-306

Smilansky, Saul. 2000 Free Will and Illusion. Oxford: Clarendon.

Smith, Angela M., 2005, "Responsibility for Attitudes: Activity and Passivity in Mental Life", *Ethics*, 115(2): 236–271. doi:10.1086/426957

2007. `On Being and Holding Responsible, ` Journal of Ethics 11: 465-484

2008, "Control, Responsibility, and Moral Assessment", *Philosophical Studies*, 138(3): 367–392.

2013, "Moral Blame and Moral Protest", in Coates and Tognazzini 2013b: 27–48.

2015, "Responsibility as Answerability", *Inquiry*, 58(2): 99–126.

Smith, Holly M. 2014, Does Being Responsible Depend on the Ability to Hold Morally Responsible? *Philosophical Studies* 171:51-62

Sommers, Tamler, 2007 "The Objective Attitude", philosophical Quarterly 57(28): 321-342-

Strawson, Galen. 1994, "The impossibility of Moral Responsibility". Philosophical Studies. 75: 5-24.

Strawson, Peter. 2008. Freedom and Resentment and other Essays, Routledge, London.

Suls, J. M. 1977 Gossip and Social Comparison, *Journal of Communication,* 26:164-168.

Talbert, Matthew. 2008, "Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy?" *Pacific Philosophical Quarterly* 89: 516:35.

2012, "Moral Competence, Moral Blame, and Protest", *The Journal of Ethics*, 16(1): 89–109.

Todd, Patrick. 2016, "Strawson, Moral Responsibility, and the 'Order of Explanation': An Intervention", *Ethics*, 127(1): 208–240.

Van Inwagen, Peter.1983, *An Essay on Free Will*. Oxford: Clarendon.

Vargas, Manuel. 2005, "The Trouble with Tracing", *Midwest Studies in Philosophy*, 29: 269–291.

2013, *Building Better Beings: A Theory of Moral Responsibility*, New York: Oxford University Press.

2016, "Responsibility and the Limits of Conversation," *Criminal Law and Philosophy 10*:221–240.

Walker, Margaret Urban. 2006. Moral repair. *Reconstructing Moral Relations after Wrongdoing*. Cambridge: CUP.

Wallace, R. Jay. 1994, *Responsibility and the Moral Sentiments*. Cambridge. MA.: Harvard University Press.

Watson, Gary. 1987, "Responsibility and The Limits of Evil: Variations on a Strawsonian Theme." In Schoeman: 256-286. Reprinted in Fischer and Ravizza, 1993.

1996, "Two Faces of Responsibility". Philosophical Topics 24 (2): 227-248.

2004, *Agency and Answerability*. New York: Oxford University Press.

2014, "Peter Strawson on Responsibility and Sociality," *Oxford Studies in Agency and Responsibility,* vol. 2, ed. David Shoemaker and Neal Tognazzini, Oxford: Oxford University Press.

Wolf, Susan. 1980, "Asymmetrical Freedom", *The Journal of Philosophy*, 77(3): 151–166.

1990, *Freedom Within Reason*, New York: Oxford University Press.

Zimmerman, Michael J. 1988, *An Essay on Moral Responsibility*, Totowa, NJ: Rowman and Littlefield.

2015, "The Varieties of Responsibility" In Clarke, Randolph, Michael McKenna and Angela smith (eds.), *The nature of Moral Responsibility. New essays,* Oxford: Oxford University Press.