

Georgia State University

ScholarWorks @ Georgia State University

---

Philosophy Theses

Department of Philosophy

---

8-12-2016

## Agency and Resentment: Reinterpreting Strawson's Compatibilism

Bobby Bingle

Follow this and additional works at: [https://scholarworks.gsu.edu/philosophy\\_theses](https://scholarworks.gsu.edu/philosophy_theses)

---

### Recommended Citation

Bingle, Bobby, "Agency and Resentment: Reinterpreting Strawson's Compatibilism." Thesis, Georgia State University, 2016.

doi: <https://doi.org/10.57709/8605870>

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

# AGENCY AND RESENTMENT: REINTERPRETING STRAWSON'S COMPATIBILISM

by

BOBBY BINGLE

Under the direction of Eddy Nahmias, PhD and Nicole Vincent, PhD

## ABSTRACT

In his influential work, "Freedom and Resentment," P.F. Strawson argues that the truth of determinism would be irrelevant to our moral responsibility practices, since our commitment to these practices is somehow connected to both our reactive attitudes—e.g. resentment, gratitude, and love—and participation in interpersonal relationships. However, some of the moves made by Strawson in his work remain unclear. In this paper, I address three prominent attempts to explain these moves, and contend that none of them captures his view because they all focus too narrowly on notions of moral demands and feelings of resentment. I argue that the correct understanding of Strawson's compatibilism needs to take seriously his close connection between the proneness to feel all of the reactive attitudes and our seeing ourselves as agents (as opposed to mere objects). In light of this argument, I conclude by offering a more plausible reading of Strawson's compatibilism.

INDEX WORDS: Strawson, Moral responsibility, Compatibilism, Resentment, Determinism

AGENCY AND RESENTMENT: REINTERPRETING STRAWSON'S COMPATIBILISM

by

BOBBY BINGLE

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2016

Copyright by  
Bobby Bingle  
2016

AGENCY AND RESENTMENT: REINTERPRETING STRAWSON'S COMPATIBILISM

by

BOBBY BINGLE

Committee Chairs: Eddy Nahmias

Nicole Vincent

Committee: Tim O'Keefe

Sandra Dwyer

Electronic Version Approved:

Office of Graduate Students

College of Arts and Sciences

Georgia State University

May 2016

**DEDICATION**

For Grandma and Grandpa Boyd

## ACKNOWLEDGEMENTS

First and foremost, I thank my thesis directors, Eddy Nahmias and Nicole Vincent. Both have been generous with their time and happily helped me at each stage of this thesis's development. I'm grateful for their feedback, support, and sage words of advice.

Second, I thank my thesis committee, Sandra Dwyer and Tim O'Keefe. Despite being incredibly busy, both have always had the time to support me here at Georgia State University.

Third, I'm thankful for the entire philosophical community here at GSU. I'm uniquely grateful for my (former) fellow graduate student, Jennifer Daigle, and her willingness to talk about all things Strawson with me.

Finally, I'm uniquely grateful for the love and support of my wife, Amanda Bingle, and my parents, Bob and Ruth Bingle. I couldn't have done any of this without them.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....		v
1 INTRODUCTION.....		1
2 ATTITUDES AND RELATIONSHIPS.....		3
<i>2.1. The Communicative Interpretation</i> .....		3
<i>2.2. The Sentimentalist Interpretation</i> .....		6
3 UNITY OF THE REACTIVE ATTITUDES.....		11
<i>3.1. The Resentment Interpretation</i> .....		11
<i>3.2. The Agency View</i> .....		14
4 STRAWSON'S NOTION OF AGENCY.....		19
<i>4.1. Normative Competence Interpretation</i> .....		19
<i>4.2. The Disappearing Agent Problem</i> .....		21
<i>4.3. The Relativizing Response</i> .....		23
<i>4.4. Strawson's Compatibilism, Revised</i> .....		27
5 CONCLUSION.....		31
WORKS CITED .....		34



## 1 INTRODUCTION

In his widely influential work, “Freedom and Resentment” (1962), P.F. Strawson argues that the truth of determinism would be irrelevant to our moral responsibility practices, since our commitment to these practices is somehow connected to both our reactive attitudes—e.g. resentment and indignation—and interpersonal relationships. However, some of the moves made by Strawson remain unclear, and philosophers have offered three prominent (and not necessarily competing) interpretations of what Strawson intends:

1. *The Communicative Interpretation* - A commitment to moral desert is required for the outward expression of the reactive attitudes, and these expressions communicate information vital for maintaining healthy interpersonal relationships.
2. *The Proneness to Resentment Interpretation* - The proneness to feel resentment and indignation towards others is, in some way, psychologically necessary for feeling positive reactive attitudes towards them. (In other words, if *P* cannot feel resentment towards *Q*, then *P* also cannot feel mutual love, gratitude, or forgiveness towards *Q*.)
3. *The Normative Competence Interpretation* – For someone to regard another as a morally responsible agent means that she regards the other as receptive to moral demands or moral reasons; and the reactive attitudes are, somehow, essentially linked with these moral demands or moral reasons.

According to philosophers who favor the *Communicative Interpretation* or *Proneness to Resentment Interpretation*, Strawson is thought to argue that our commitment to healthy interpersonal relationships or positive reactive attitudes (e.g. love, gratitude, and forgiveness) precludes us from taking seriously any threat posed to moral responsibility by determinism. Likewise, under the *Normative Competence Interpretation*, Strawson is thought to argue that since determinism in no way jeopardizes our ability to understand and follow moral demands, and since possessing such abilities is all that is required for our reactive attitudes to be legitimate, it cannot threaten the legitimacy of our reactive attitudes.

Many important figures in the free will debate adopt one (or more) of these interpretations, and several have even taken them as starting points to develop their own interesting and nuanced theories of compatibilist moral responsibility. Even so, in this paper, I cast doubt on attributing *any* of these

three interpretations to Strawson. Specifically, I aim to critically evaluate each interpretation, and offer a more plausible competing interpretation of Strawson's compatibilism.

To be clear, I will not argue that the contents of the interpretations themselves are false. For instance, I will not argue that the expression of our reactive attitudes does not perform some crucial healthy relationship-sustaining function. Likewise, I will not argue against any of compatibilist theories that take these interpretations as their basis. (Indeed, I am sympathetic to many of them.) Rather, my aim is to suggest that these three interpretations do not adequately capture how Strawson addresses the problem of free will and determinism.

This thesis proceeds as follows: In section 2, I address the Communicative Interpretation. Specifically, I summarize the version taken on by free will skeptic Derk Pereboom (*subsection 2.1*). Afterwards, I raise doubts and offer a more plausible reading of Strawson, the *Sentimentalist Interpretation* (*subsection 2.2*). My discussion of the Sentimentalist Interpretation leads naturally into section 3, wherein I critically investigate the version of the Proneness to Resentment Interpretation proposed by Gary Watson (*subsection 3.1*). I suggest that this interpretation is based on a misreading of a passage found in "Freedom and Resentment." Instead, I propose an alternative interpretation, which is based in a close reading of two of Strawson's later works, 1985's "Morality and Perception" and 1992's "Freedom and Necessity" (*subsection 3.2*). Under my reading, the Agency View, Strawson intends that the reactive attitudes are unified as reactions to another's agency, and that this notion of agency is at stake in the problem of free will and determinism. Nevertheless, my analysis also leaves open important questions about the precise nature of this sense of agency. I take on these questions in section 4, and begin by addressing the answers offered by the widely influential Normative Competence Interpretation, as proposed by philosophers like Watson and R. Jay Wallace (*subsection 4.1*). To determine if the Normative Competence Interpretation correctly captures Strawson's view of agency, I contend that we must first examine how Strawson understands the supposed threat that

determinism poses to free will, called the *Disappearing Agent Problem* (subsection 4.2). Finally, I argue that Strawson would deny the Normative Competence Interpretation, as he claims that any attempt to offer a positive conception of agency as a counter to the Disappearing Agent Problem is both unnecessary and doomed to failure. Instead, I argue that Strawson offers a different way to address the Disappearing Agent Problem, called the *Relativizing Move* (subsection 4.3). Taken together, the Sentimentalist Interpretation, Agency View, and Relativizing Move comprise a more accurate interpretation of Strawson's compatibilism, and I conclude by offering two additional remarks about it (subsection 4.4).

## 2 ATTITUDES AND RELATIONSHIPS

### 2.1 *The Communicative Interpretation*

Pereboom reads Strawson as maintaining that outward expressions of our reactive attitudes—specifically resentment and indignation—are essential for healthy interpersonal relationships. For example, Pereboom writes:

[Some claim that] taking the skeptical view would threaten our personal relationships and the fulfillment in life that they provide. P. F. Strawson (1962) has developed a philosophical elaboration of this reaction. For him a skeptical conviction, supposing it was psychologically possible for us, would undermine expressions of the other-directed reactive attitudes essential to good personal relationships... (2014, 175, also see 179)

Under Pereboom's reading, the expressions of the reactive attitudes play an essential *communicative role* in personal and societal relationships (180). Although Pereboom does not go into detail about what the reactive attitudes communicate, he appears to offer, as three possibilities, moral statements or demands (184, also see 147-151), interpersonal care and concern (183), and emotional vulnerability (183).<sup>1</sup> It is in playing this communicative role that the reactive attitudes derive their significance, and

---

<sup>1</sup> Pereboom also seems willing to accept that expressions of resentment and indignation communicate these three pieces of information; as will be seen below, however, he thinks they do so poorly.

the worry he thinks Strawson has is that “if we were to strive to modify or eliminate these attitudes, such relationships might well be damaged” (180).

The notion of responsibility Pereboom connects to some of the reactive attitudes is what he calls *basic desert responsibility*, which he explains as follows: “For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary” (127). Holding an agent morally responsible in the basic desert sense, then, means believing that she deserves to be praised or blamed for what she has done irrespective of any forward-looking considerations. The attitudes whose expression presupposes basic desert responsibility include resentment and indignation, and are therefore categorically rejected by Pereboom’s responsibility skepticism (see 127 and 179).<sup>2</sup>

With these considerations in hand, we might summarize Pereboom’s interpretation of Strawson’s *Argument Based on the Communicative Interpretation* as follows:

- C1. If *P* is in a healthy interpersonal relationship with *Q*, then *P* must be able to communicate her moral claims or demands, care and concern, etc., to *Q* in response to *Q*’s wrongdoing.
- C2. The expression of resentment and indignation is how we communicate moral claims or demands, care and concern, etc., in response to some types of wrongdoing,
- C3. Thus, if *P* is in a healthy interpersonal relationship with *Q*, then *P* must be able to express her resentment and indignation towards *Q* in response to *Q*’s wrongdoing. [From (C1) and (C2)]
- C4. If *P* expresses her resentment and indignation towards *Q* in response to *Q*’s wrongdoing, then *P* must believe that *Q* is morally responsible.
- C5. Thus, if *P* is in a healthy interpersonal relationship with *Q*, then *P* must believe that *Q* is morally responsible. [From (C3) and (C4)]
- C6. Therefore, if *P* does not believe that *Q* is morally responsible, then *P*’s interpersonal relationship with *Q* will be damaged. [Contrapositive of (C5)]

Pereboom provides two objections to (C2). First, while conceding that expressions of resentment and indignation communicate moral demands, care and concern, etc., Pereboom nevertheless contends

---

<sup>2</sup> Importantly, other attitudes, including love, forgiveness, hurt feelings, and moral disappointment, are not rejected by the skeptic.

that they are “suboptimal as modes of communication” (2014, 179). What he means is that, in addition to failing to contribute to the well-being of their target, expressions of resentment and indignation are often meant to cause physical and emotional pain, potentially sparking destructive resistance rather than reconciliation (180). Second, Pereboom contends that other attitudes which do not presuppose basic desert can serve the same communicative function. These replacement attitudes are called *moral analogues*, and their expression crucially does not presuppose a belief that their target deserves blame. As Pereboom writes: “When we are wronged in our relationships there are other emotions present or available not challenged by the skeptical view, whose expression can also convey the relevant information... [including] feeling hurt or shocked or disappointed at what the offending agent has done” (180).

Here’s a helpful example of how moral analogues might work. Suppose that *A* promises to pick *B* up from the airport, but at the last minute bails on *B* leaving her stranded. *B* rightly believes that her relationship with *A* is compromised (since it appears that *A* is a lousy friend), and needs to communicate this fact to *A*. Under Pereboom’s understanding of Strawson, *B* communicates to *A* that the relationship is jeopardized by expressing her resentment at *A* (“How dare you do that to me, *A*! What kind of crappy friend are you?”). In addition to presupposing basic desert, Pereboom also believes that *B* expressing her resentment is not likely to repair the relationship. Instead, *A* is likely to get defensive, and, ultimately, the relationship will end or be damaged. Instead, *B* could communicate to *A* that the relationship is damaged by expressing her disappointment (“*A*, I’m very upset that you think so little of me that you can break your promise. Can’t you see how much you hurt my feelings?”). In addition to not presupposing that *A* is morally responsible in the basic desert sense, *B*’s disappointment communicates the same information and in a healthier way. So, by pointing to these alternative expressions, Pereboom takes himself to have removed a crucial obstacle towards the plausibility of a widespread acceptance of responsibility skepticism.

But has the Communicative Argument correctly captured Strawson's argument? I don't think so. In the next subsection, I offer a more plausible interpretation that also has the advantage of avoiding Pereboom's objections to (C2) of the Communicative Argument. Although I believe that my interpretation is closer to Strawson's actual views, it will ultimately leave open important questions about the relationship among the reactive attitudes (which will become the focus of section 3).

## ***2.2 The Sentimentalist Interpretation***

On my interpretation of Strawson's position, the following three ideas play a crucial role: [i] the general proneness to feel *all* of the reactive attitudes is [ii] partly constitutive of our involvement in interpersonal relationships, and [iii] presupposes that the one toward whom we are prone to feel these attitudes is an agent. This interpretation differs from the Communicative Interpretation in three ways: First, it is primarily concerned with the proneness to feel certain ways about one another, rather than the legitimacy of the outward expressions of our emotions. Second, it does not make a claim about what makes for healthy interpersonal relationships, but rather is concerned with what sorts of proneness are constitutive of our involvement in distinctly *human* relationships. Third, it focuses more broadly on the all of the reactive attitudes, rather than simply resentment and indignation.

I begin with [ii]; Strawson describes the reactive attitudes as not merely contingently related to our interpersonal relationships or as serving a crucial sustaining role. Rather, he describes these attitudes as partly constitutive of our involvement in *human* interpersonal relationships. He asserts that the reactive attitudes are "the non-detached attitudes and reactions of people directly involved in transactions with one another" (1962, 75), linked with "involvement or participation in a human relationship" (75), and insists that they share "common roots in our human nature and our membership in human communities" (80). By the same token, Strawson also claims that the absence of the reactive attitudes equates with "human isolation" (78), and that if a group of people permanently ceased feeling the reactive attitudes towards each other, then it would be "doubtful whether [they]

should have anything that we could find intelligible as a system of human relationships, as human society” (93).

Regarding [i], Strawson’s focus is less on the outward expressions of these attitudes, and more on our proneness to *feel* them. He locates his discussion of moral responsibility in “the attitudes and intentions towards us of those who stand in these relationships to us, and of the kinds of reactive attitudes and feelings to which we ourselves are prone” (1962, 76-77). Given that focus, Strawson occasionally refers to the reactive attitudes as “reactive feelings” (see 79, 81, 91), and spends a substantial portion of “Freedom and Resentment” considering situations where we either would “normally be expected to feel [attitudes like] resentment” (77), or are expected to “remove” these feelings altogether (77). Indeed, in his later work, *Skepticism and Naturalism*, Strawson states that “Freedom and Resentment” was primarily focused on the “attitudes and feelings, or ‘sentiments’ as we used to say... which we feel in ourselves and attribute to others” (1985, 31). Thus, on Strawson’s view, if people are participants in *human* relationships, then they must be prone to *feel* the reactive attitudes towards each other.

Finally, regarding point [iii], Strawson links the proneness to feel *all of* the reactive attitudes with regarding others as agents. Unlike the Communicative Interpretation, Strawson does not focus on moral desert and or simply attitudes like resentment and indignation. For instance, when first faced with the alleged conflict between moral responsibility and determinism, Strawson states that we “have to” ask:

What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, and forgiveness; of all reciprocated adult loves; of all the essentially personal antagonisms? (80).

Notice here that Strawson claims that the proneness to feel gratitude, love, and forgiveness is also at stake in the free will debate. Although in the absence of moral responsibility one might feel “repulsion

or fear... pity or even love, though not all kinds of love [towards others],” Strawson writes, “But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships” (79). Thus, on Strawson’s view, if one is prone to feel the reactive attitudes towards others, then one must regard others as morally responsible agents.

With these preliminaries in place, I now offer the *Sentimentalist Interpretation* of Strawson’s argument against responsibility skepticism:

- S1. If  $P$  is in a human interpersonal relationship with  $Q$ , then  $P$  must be prone to feel the reactive attitudes towards  $Q$ .
- S2. If  $P$  is prone to feel the reactive attitudes towards  $Q$ , then  $P$  must regard  $Q$  as a morally responsible agent.<sup>3</sup>
- S3. Suppose that  $P$  does not regard  $Q$  as a morally responsible agent.
- S4. Thus,  $P$  is not prone to feel the reactive attitudes towards  $Q$ . [From (S2) and (S3)]
- S5. Thus,  $P$  is not in a human interpersonal relationship with  $Q$ . [From (S1) and (S4)]
- S6. Therefore, if  $P$  does not regard  $Q$  as a morally responsible agent, then  $P$  is not in a human interpersonal relationship with  $Q$ . [From (S3) – (S5)]

The Sentimentalist Argument has two significant advantages over the Communicative Argument. First, notice that if this is Strawson’s argument, then the objections Pereboom raises fail. Pereboom’s response to the Communicative Argument is that attitudes other than resentment and indignation could perform whatever relationship-sustaining functions resentment and indignation perform. Pereboom understands that his response to the Communicative Argument is conjectural in light of understanding Strawson as proposing an empirical argument about what makes for healthy

---

<sup>3</sup> One issue with (S2) concerns how we ought to interpret its contrapositive. Specifically, by (S2), does Strawson intend:

S2a. If  $P$  does not regard  $Q$  as a morally responsible, then  $P$  *should not* be prone to feel the reactive attitudes towards  $Q$ ; or

S2b. If  $P$  does not regard  $Q$  as a morally responsible, then  $P$  *would not* be prone to feel the reactive attitudes towards  $Q$ ?

The crucial difference between these two is that (S2a) is normative, while (S2b) is descriptive. Although Strawson initially starts off considering both, his discussion of moral responsibility skepticism quickly shifts the focus towards (S2b), since he talks mainly about the sorts of attitudes that skeptics *will* not feel (1962, 78-79, which is later echoed in 1985, 35). Many will find (S2b) troubling, since it seems plausible that sometimes people do not *believe* that another is morally responsible, but can’t help themselves from still *feeling* resentment and indignation towards her. In any case, since my aim in this paper is determine the best interpretation of Strawson, I will limit myself to descriptive interpretation, (S2b), since, as we shall see, it seems to best reflect Strawson’s actual views. Again, I am aware that (S2b) might seem troubling to some, but a defense of it is outside the scope of this paper.



relationships, admitting that “whether living in accord with the skeptical perspective would be on the whole better for us is an empirical issue that philosophical method lacks the resources to settle” (2014, 176). For that reason, both supporters and detractors of the Communicative Argument point to social psychological studies which aim to show whether our interpersonal relationships would suffer or be improved without the expressions of resentment and indignation (see Nichols 2007). But since the Sentimentalist Argument does not base itself upon the purported role of expressions of resentment and indignation, nor any direct empirical claim about what makes relationships healthy or impaired, Pereboom’s response does not apply.<sup>4</sup>

Second, unlike the Sentimentalist Argument, the Communicative Argument does not receive as clear or direct support from Strawson’s own statements. For example, recall that the Communicative Argument is based on what makes for *healthy* interpersonal relationships. However, Strawson never mentions *healthy* relationships, focusing rather on what makes for *human* relationships. Likewise, also recall that the Communicative Argument locates the importance of the reactive attitudes in what their expressions communicate. However, Strawson never talks about the communicative functions of these attitudes. The closest that he comes is in stating that resentment and indignation “express,” “rest on,” and “reflect” a demand for goodwill (1962, 84), but it’s clear from later writings that he didn’t intend too much of this claim. For example, when Jonathan Bennett proposed a tighter and more explicit analysis of how resentment and indignation link up with the demand for goodwill (see Bennett 1980, 39-42), Strawson responds by claiming that such an attempt is misguided (and unnecessary):

Reading [Bennett] and re-reading myself, I am driven to think that he construes some of the phrases that I used as having a greater definitiveness than I intended them to have or, perhaps,

---

<sup>4</sup> To be clear, I am not denying that Strawson makes an empirical claim about what sorts of feelings we must have (in a descriptive sense) in order to enter into the distinctively human relationships with others. Furthermore, I am not denying that Strawson’s account permits some implications about what feelings are necessary to have *healthy* human relationships. Rather, I am arguing here that Strawson’s argument about the irrelevance of determinism to free will and moral responsibility is not based on whatever emotions makes for healthy human relationships.

than the case admits of. But I freely affirm that ‘acknowledgement of a claim [or demand]’ is too weak a phrase; and I freely admit the central importance of that sense of sympathy, and *common* humanity, that underlines... indignation. (Strawson 1980, 266, emphasis original)

Strawson does not say in any precise detail what he intended by these phrases, though he emphasizes that “What I was above all concerned to stress was that our proneness to reactive attitudes is a natural fact, woven into the fabric of our lives, given with the fact of human society as we know it” (265).

In any case, if the Sentimentalist Argument is Strawson’s argument, then it avoids Pereboom’s original criticisms of the Communicative Argument since it is not based upon a purported important role played by the expressions of resentment and indignation, nor any empirical claim about what makes relationships healthy or impaired. Rather, it is based on the proneness to feel certain attitudes in response to one another.

Even so, the Sentimentalist Interpretation leaves important questions about Strawson’s theory unanswered. For example, why does Strawson claim that the general proneness to feel *all* of the reactive attitudes is at stake in the problem of free will and determinism (rather than simply the attitudes associated with moral desert, such as resentment and indignation)? In other words, why would the moral responsibility skeptic, in addition to no longer feeling resentment and indignation, also stop feeling mutual love, hurt feelings, and forgiveness? In the next section, I critically examine the Proneness to Resentment Interpretation, which answers this question by claiming that the proneness to feel resentment and indignation is psychologically required in order to feel other reactive attitudes. Since the proneness to feel resentment and indignation is jeopardized in the free will debate, so too is proneness to feel the other reactive attitudes. I will argue that it is unclear if Strawson intends this interpretation. Instead, I will offer the moral plausible Agency View, under which the reactive attitudes are responses to agency, and furthermore, it is this sense of agency is jeopardized in the problem of free will and determinism.

### 3 UNITY OF THE REACTIVE ATTITUDES

#### *3.1 The Resentment Interpretation*

Several philosophers attribute the Proneness to Resentment Interpretation to Strawson. For example, Daniel Goldman (2014) writes: “Resentment is Strawson’s central example of the attitudinal responses that are essential to participation in interpersonal relationships (1)... [and Strawson states] eliminating the susceptibility to resentment in any individual situation will carry the high cost of exiting a specific interpersonal relationship” (7). Likewise, Gary Watson (1987) writes: “Can such an ideal of the person [one who does not feel resentment] be pursued only at the cost of the attachment to the personal? Must we choose between isolation and animosity? Some of Strawson’s remarks imply that we must” (147).

If the Proneness to Resentment Interpretation is correct, then we might understand Strawson as suggesting the following sort of argument against moral responsibility skepticism:

- R1. If  $P$  does not regard  $Q$  as morally responsible, then  $P$  does not feel resentment and indignation towards  $Q$ .
- R2. If  $P$  does not feel resentment and indignation towards  $Q$ , then  $P$  does not feel the other reactive attitudes towards  $Q$ .
- R3. Therefore, if  $P$  does not regard  $Q$  as morally responsible, then  $P$  does not feel the other reactive attitudes towards  $Q$ . [From (R1) and (R2)]
- R4. If  $P$  does not feel the reactive attitudes towards  $Q$ , then  $P$  cannot be in a human interpersonal relationship with  $Q$ .
- R5. Therefore, if  $P$  does not regard  $Q$  as morally responsible, then  $P$  cannot be in a human interpersonal relationship with  $Q$ . [From (R3) and (R4)]

It is worth noting that many philosophers who attribute the Proneness to Resentment Interpretation to Strawson also give us compelling reasons to reject (R2). For instance, Watson claims that (R2) is at odds with an important ideal of love, exemplified by Jesus, Gandhi, and Martin Luther King, Jr. These exemplars lead “lives [that] do not seem characterized by human isolation: They are often intensely involved in the fray of interpersonal relationships... [and] do such things without vindictiveness or malice” (1987, 148). Similarly, Pereboom (2014) reminds us that there are entire Buddhist and Christian monastic communities, as well as Christian reformation movements, who appear to provide

examples of entire groups of people who have given up their feelings of resentment and indignation, while retaining the ability to feel love and gratitude towards one another (149). If so, then it seems that it is implausible to suppose that the proneness to feel resentment and indignation is “psychologically inseparable” from the proneness to feel the other reactive attitudes (185).

However, there are two reasons to doubt the Proneness to Resentment Interpretation. First, in “Freedom and Resentment,” Strawson states that he will not address the nature of the reactive attitudes. “It is one thing to ask about the general causes of these reactive attitudes I have alluded to,” Strawson writes, “I am not much concerned with [that] question” (1962, 77). Indeed, Strawson never offers an explicit analysis of how the reactive attitudes are interrelated. (As I will show in the next subsection, that question is taken on in his later works.) I suspect it is for that reason that Watson writes Strawson “implies,” “suggests,” or “appears to think” that the proneness to feel resentment and indignation is psychologically required to feel the other reactive attitudes.

Second, Watson appears to be misreading the one passage from “Freedom and Resentment” that he points to as support for his Proneness to Resentment Interpretation. Quoted at length, the passage states:

Indignation, disapprobation, like resentment, tend to inhibit or at least to limit our goodwill towards the object of these attitudes... (*These, of course, are not contingent connections.*)... The holding of them does not... involve as a part of itself viewing their object other than as a member of the moral community. The partial withdrawal of goodwill which these attitudes entail, the modification they entail of the general demand that another should... be spared suffering is... the consequence of continuing to view him as a member of the moral community; only as one who has offended against its demands... So the preparedness to acquiesce in that infliction of suffering on the offender which is an essential part of punishment is all of a piece with this whole range of attitudes of which I have been speaking (Quoted in Watson 1987, 147; originally from Strawson 1962, 90)

Although there’s a lot going on here, nowhere does Strawson state that the proneness to feel resentment and indignation towards a wrongdoer is psychologically necessary in order to feel love, gratitude, or forgiveness towards her. For that reason, I am unsure how this passage can be taken as support of the Proneness to Resentment Interpretation. Instead, what Strawson seems to suggest is

that there's a necessary connection between (a) viewing a criminal offender as a member of the moral community, (b) feeling some measure of indignation or moral disapprobation in response to her wrongdoing, and (c) a willingness to allow her to suffer because of her wrongdoing (specifically, through punishment).

Strawson does not go on to develop (a) – (c) at any great detail, but my reading of this passage as a defense of retributivism receives support from the remarks that Strawson makes before and after that passage. For instance, right before the passage, he writes that we “will confine our attention to the case of offenders” (1962, 90). And immediately afterwards, he writes:

I am not in the least suggesting... that we should be ready to acquiesce in the infliction of injury on offenders in a fashion which we saw to be quite indiscriminate or in accordance with procedures which we knew to be wholly useless. On the contrary, savage or civilized, we have some belief in the utility of practices of condemnation and punishment. (91)

Finally, at the end of that section of the paper, he goes on to conclude:

It is far from wrong to emphasize the efficacy of all those [punishment] practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable... What is wrong is to forget that these practices, and their reception, the reactions to them, really are expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes... (93)

Notice that, in both of these passages, Strawson is offering some defense of our wishing to see (or, at least, willingness to allow) a wrongdoer to suffer through punishment (and admitting that punishment also serves some consequentialist goals).

Likewise, the context of this section of the paper is important to bear in mind. In this section of “Freedom and Resentment,” Strawson is attempting to pinpoint what's wrong with a purely consequentialist theory of punishment (offered by a character Strawson calls the “optimist”). The optimist believes that all of our practices of reward and punishment can be justified on purely consequentialist grounds. Strawson thinks that the optimist makes two errors. (I) she attempts to provide the wrong sort of justification for our punishment practices:

But the only reason you have given for the practices of moral condemnation and punishment in cases where this freedom is present is the efficacy of these practices in regulating behaviour 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. (74, emphasis added)

And (II), she overlooks the significance of these practices—in that they are expressions of our moral attitudes towards one another. Strawson writes:

...to speak in terms of social utility alone is to leave out something vital in our conception of these practices. The vital thing can be restored by attending to that complicated web of attitudes and feelings which form an essential part of the moral life as we know it... Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i. e. of *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice. (91)

So it seems that Strawson is not attempting to offer a psychological claim about the unity of our reactive attitudes, but rather to diagnose what's wrong with a *purely* consequentialist theory of punishment, and to offer (at least) a limited defense of retributivism. In any case, I think that it is a mistake to read this passage as evidence for the Proneness to Resentment Interpretation.

In the next subsection, I offer an alternative interpretation for why the proneness to feel the reactive attitudes is jeopardized in the free will debate, called the Agency View. According to the Agency View, the reactive attitudes are our emotional reactions to our perceptions of an agent's actions. Furthermore, it is this sense of agency that is at stake in the free will debate. As we shall eventually see, the Agency View makes sense in light of Strawson's understanding of the putative threat determinism poses to moral responsibility (subsection 4.2); and helps us better understand some of the more puzzling passages found in "Freedom and Resentment" (subsection 4.4).

### ***3.2 The Agency View***

According to the *Agency View*, the reactive attitudes are our emotional reactions to our perceptions of an agent's actions. Moreover, it is this sense of agency that Strawson claims is at stake in the free will debate, and if determinism successfully undermines free will, then it does so by way of showing that we are not actually agents (see subsection 4.2). An important difference between the Proneness to

Resentment View and the Agency View concerns *how* the reactive attitudes are challenged by moral responsibility skepticism. Under the Proneness to Resentment Interpretation, resentment is directly challenged by moral responsibility skepticism, and the other reactive attitudes are only indirectly challenged (specifically, by way of resentment and indignation being challenged). Under the Agency View, it is agency that is directly challenged by determinism,<sup>5</sup> and *all* reactive attitudes are indirectly challenged (since the reactive attitudes are our emotional reactions to the actions of another agent).

The Agency View is found in two of Strawson's later works, "Freedom and Necessity" (1992) and "Morality and Perception" (1985), and it is offered as part of a larger argument against moral responsibility skepticism. However, the argument itself is detailed and nuanced. Thus, I will first present a standardized version of Strawson's argument, followed by an in-depth analysis of each premise. The argument states:

- A1. If *P* does not regard *Q* as morally responsible, then *P* does not view *Q* as another agent.<sup>6</sup>
- A2. If *P* does not view *Q* as another agent, then *P* views *Q* as something like a natural object.<sup>7</sup>
- A3. If *P* views *Q* as something like a natural object, then *P* only feels towards *Q* the sort of emotions that *P* feels towards natural objects.
- A4. Thus, if *P* does not regard *Q* as morally responsible, then *P* only feels towards *Q* the sort of emotions that *P* feels towards natural objects. [From (A1) – (A3)]
- A5. We do not feel the reactive attitudes towards natural objects.
- A6. Therefore, if *P* does not regard *Q* as morally responsible, then *P* will not feel the reactive attitudes towards *Q*. [From (A4) and (A5)]

Next, let us briefly examine each premise, starting with (A1). According to Strawson, we view each other as morally responsible because of how we experience ourselves. He writes that we "feel" in

---

<sup>5</sup> Just to be sure, Strawson does not believe that the challenge succeeds.

<sup>6</sup> Some readers might claim that (A1) is mistaken. Specifically, it seems that sometimes we think that another isn't morally responsible, but still might be an agent (e.g. children). However, I interpret Strawson as making a claim about how regarding another as an agent *constitutes* seeing them as free and morally responsible. For example, Strawson writes: "[it is] the constantly repeated experiences of agency... [that] contribute to, *perhaps constitute*, the sense of freedom [and responsibility]. Experiencing it ourselves, we attribute it also to others" (1992, 135, emphasis added). If Strawson is making a constitutive claim about agency and moral responsibility, then we might express that claim as:

A0. *P* regards *Q* as morally responsible if and only if *P* regards *Q* as an agent.

And (A1) follows from (A0).

<sup>7</sup> Similarly, some might object to (A2) on the grounds that it doesn't seem that not regarding others as agents immediately implies that we regard them something like objects. However, as I shall shortly argue, Strawson is using 'natural object' in a unique way.

ourselves a “sense of self” and an “experience of agency” (1992, 134-135). What Strawson means is that we have beliefs, desires, and deliberations that we identify with and act upon. Likewise, we perceive others as having the same sense of self and agency. For that reason, we are prone to feel the reactive attitudes towards them. Strawson writes:

We see others as other selves, and we are aware that they so see each other. But this is not a matter of conclusion drawn by reasoning. In a variety of ways, it is inextricably bound up with the facts of mutual human involvement and interaction, we feel towards each other as other selves; and this variety is just the variety of moral and personal reactive attitudes and emotions which we experience towards others and which have their correlates in attitudes and emotions directed towards ourselves (1992, 138).

And viewing others as other agents “contributes to (or perhaps constitutes)” regarding them as morally responsible (135).

Turning to (A2). Strawson claims that if we do not view others as agents, we instead view them “objectively,” or as “natural objects,” or “genetically programmed mechanisms of immense complexity... constantly modified by their own history and responding, in constantly modified ways, to sensory inputs with behavioral outputs” (1992, 139-140). Strawson calls the viewpoint whereby we regard others as natural objects the “objective stance,” and he writes:

To see human beings and human action in this light is to see them simply as objects and events in nature, natural objects and natural events, to be described, analyzed, and causally explained in terms in which moral evaluation has no place; in terms, roughly speaking, of an observational and theoretical vocabulary recognized in the natural and social sciences, including psychology (1985, 40).

To be clear, Strawson is not suggesting that there is an ontological difference between agents and natural objects. For example, he is not saying that agents have souls and mere natural objects do not. Instead, he says that viewing another as an agent or instead as a natural object constitutes two “profoundly opposed” psychological viewpoints (1962, 79).<sup>8</sup> Strawson takes himself to be stating a

---

<sup>8</sup> It is worth noting that this tension is *still* found in contemporary debates about morality, agency, and free will. For example, Valerie Tiberius (2015) has helpfully expressed the tension in the following way:

Here’s one way of seeing the problem. We can see ourselves as the scientist does: as blobs of flesh acted on by causal laws, responding to other blobs in the environment, all of which are made of the same physical stuff. Or, we



fact about human psychology: it is difficult (but not impossible) to view another as both an agent and merely an natural object simultaneously (more on this tension and the possibility of the mixed stance in subsections 4.3 and 4.4).

Next, turning to (A3). Strawson claims that it is the tension between *viewing others as agents* and *viewing others as natural objects* that lies at the heart of the free will debate:

The fundamental thought is that once we see people and their doings (including ourselves and our doings) objectively... namely, as natural objects and happenings in the course of nature—whether causally determined occurrences or chance occurrences—then the veil of illusion cast over them by moral attitudes and reactions must, or should, slip away. What happens in nature may be a matter for rejoicing or regret, but not for gratitude or resentment, for moral approval or blame, or for moral self-approval or remorse. (1985, 32)

Stated otherwise, once we view others as objects, then we emotionally respond to them *as if* they were objects; we view them as things that ought to be “managed or handled or cured or trained; perhaps simply to be avoided” (1962, 79).

Finally, turning to (A5), we are not prone to feel the reactive attitudes towards objects. “To see human behavior as consisting simply of physical movement,” Strawson writes, “would, *of itself*, exclude the attitudes and feelings in question; for it is only in relation to behavior understood, or experienced, as intentional actions that these attitudes and feelings ever arise” (1992, 142).<sup>9</sup> In other words, the

---

can see ourselves from a... humanistic point of view, according to which we have special powers of reasoning and choice that we use to respond not just to the physical world, but also to the world of value and moral principles. These two ways of thinking about us seem incompatible. If we are just physical blobs in a physical universe, then we are not beings with special powers in a rationally structured universe, and vice versa. For people who are particularly taken with science or who have lost their confidence that we are more than physical blobs in a universe that can only be described in scientific terms, it starts to look like morality must just be about our blobby flesh and our psychology, because there isn't anything else for it to be about. (16)

<sup>9</sup> Some might wonder if there's a tension in Strawson's position here. After all, Strawson claims that (a) we do not regard children as morally responsible, but yet (b) we still believe that they act intentionally. Strawson here claims that we often adopt a *mixed stance* towards children, writing that:

...the simple opposition of objective attitudes on the one hand and the [reactive] attitudes which I have opposed to them must seem as grossly crude as it is central. Let me pause to mitigate this crudity a little, and also to strengthen one of my central contentions, by mentioning some things which straddle these contrasted kinds of attitude. Thus parents and others concerned with the care and upbringing of young children cannot have to their charges either kind of attitude in a pure or unqualified form. They are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed human attitudes. (1962, 88)

reactive attitude are our emotional responses to our perceptions of another's agency (sometimes called another's *quality of will*). For example, resentment is a reaction to what we perceive as another's ill will, gratitude is a reaction to what we perceive as another's good will or beneficence, forgiveness is a reaction to what we perceive as another's sincere regret over prior wrongdoing, and so on. If we were to view another as an object, then we would *not* take his actions to express his agency or quality of will. Instead, we would perceive that person and his behavior as the result of a series of complex and impersonal mechanisms. Thus, Strawson claims:

The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. (1962, 79)

With (A1) – (A5) in mind, we can now see an alternative explanation for why the proneness to feel the reactive attitudes is at stake in the free will debate. According to Strawson, regarding another as morally responsible is constituted by viewing them as an agent. And it is in viewing another as an agent that causes us to respond to her with our reactive attitudes. Thus, Strawson also claims that if someone is a deeply committed moral responsibility skeptic, then she would view others (and herself) not as agents, but merely as natural objects or mechanisms. If so, then she is only prone to feel the sorts of attitudes towards others (and herself) that she feels towards objects. But since the reactive attitudes are all attitudes that are felt towards agents, the deeply committed skeptic would not feel those attitudes towards anyone. In fact, Strawson would maintain that suggesting that the deeply committed skeptic feels the reactive attitudes towards others is to suggest that she feels the reactive attitudes towards things that she views as merely natural objects, something which Strawson thinks is not possible (see 1985, footnote 5).

If I am right, and if Strawson claims that this sense of agency is (unsuccessfully) challenged in the free will debate, then it leaves two crucial questions. First, what is the nature of this agency? Second,

why does Strawson argue that this sense of agency is immune to the putative threat posed by determinism? In the next section, I will critically examine the Normative Competence Interpretation as an answer to these questions, which states that Strawson's notion of agency primarily concerns being receptive to moral demands. (And since determinism does not entail that one cannot understand moral demands, determinism does not undermine agency.) I will argue that Strawson would deny the Normative Competence Interpretation, as he claims that any attempt to offer a conception of agency as a counter to the putative threat posed by determinism is (i) unnecessary and (ii) ill-fated. Instead, I will argue that Strawson offers a different way to address the problem of free will and determinism, called the *Relativizing Move*. Ultimately, I will conclude that, when taken together, the Sentimentalist Interpretation, Agency View, and Relativizing Move provide a coherent and more plausible account of Strawson's compatibilism.

## 4 STRAWSON'S NOTION OF AGENCY

### 4.1 Normative Competence Interpretation

Although there are several variations of the Normative Competence Interpretation, they all link morally responsible agency with the ability to understand, react to, and guide one's conduct in response to moral demands or reasons. For instance, Stephen Darwall (2013), summarizes Strawson's view of agency as:

[Strawson states that] we can intelligibly address a demand to someone to regulate her will appropriately only if we suppose that she can so regulate it as a result of recognizing our demand's legitimacy (44)... [and] Strawson's point [is] that when we take someone to be morally obligated to hold him responsible with a reactive attitude, we have to attribute to him whatever psychic competences are necessary to enter into mutually accountable, interpersonal relationships. (46-47)

Similarly, R. Jay Wallace (1994) offers a "Strawsonian" approach to holding people responsible, which means that:

...once we correctly understand the stance of holding people responsible in terms of the reactive emotions, we will be able to see that the condition that makes it fair to adopt this stance is not freedom of the will in the strong sense; rather it is a kind of normative competence in virtue of which one is able to grasp moral reasons and control one's behavior by their light. (16)

Strawson himself never explicitly offers the Normative Competence Interpretation. Rather, it comes from Watson's reading of him. Watson (rightly) claims that Strawson's theory is missing a positive account of what it means to be an agent, and he seeks to "fill in" this incompleteness. He proposes the following strategy: We need to explore the examples that Strawson provides of people that we exempt from our moral responsibility practices, and see if we can come up with a coherent underlying explanation. If so, then that explanation likely suffices as an account of Strawson's notion of agency. For the remainder of this subsection, I will briefly summarize portions of Watson's analysis.

Strawson lists four types of people that we usually do not regard as morally responsible agents: small children, the severely mentally ill, those acting under long-term compulsion (such as hypnotic suggestion), and those under severe emotional distress. Under Watson's analysis, all four cases exemplify people who are unable to understand moral demands. Consider small children. Typically, they lack a clear understanding of the consequences of their actions. Although they might say things like "I shouldn't hurt others," it is unlikely that they fully understand the reasons for this prohibition. Even when children can give some measure of explanation for these rules, Watson writes, they still "lack a notion of reasonable regard or justification" for them (127). For these reasons, Watson concludes, we do not hold children morally responsible for their actions because they do not fully understand the moral demands that are made of them. (He makes a similar argument for the other cases, see 125-129.) For these reasons, Watson concludes that the Normative Competence Interpretation explains Strawson's notion of morally responsible agency (129), as well as being clearly compatibilist: "On this account, the practice of holding responsible does seem metaphysically modest, in that it involves no commitments to which issues about determinism are relevant" (130).<sup>10</sup>

---

<sup>10</sup> Other philosophers have reached a similar conclusion. Wallace (1994, chapter 6) offers a rigorous defense of a version of the Normative Competence Interpretation, as well as an argument for why it should be considered a "Strawsonian" theory of moral responsibility.

Even if the Normative Competence Interpretation explains well the exemption cases offered by Strawson, it is another question whether it correctly captures the notion of agency that Strawson himself had in mind.<sup>11</sup> But how do we answer that question? In the next subsection, I propose that since Strawson insists that agency is at stake in the problem of free will, then we ought to investigate Strawson's own take of the free will debate. In doing so, we will see how Strawson understands the putative threat that determinism poses to agency, and perhaps how his notion of agency is immune to that threat. In doing so, we might get a better understanding of what Strawson had in mind.

#### ***4.2 The Disappearing Agent Problem***

Strawson never offers an explicit characterization of the problem of free will and determinism. For example, in "Freedom and Resentment," he suggests that he cannot do so because he does not believe that anyone has offered a sufficient analysis of the thesis of determinism (1962, 72-73). Later in "Morality and Perception," Strawson defers his characterization to the closing section of Thomas Nagel's "Moral Luck," which Strawson claims "admirably" characterizes the problem (1985, 32). Since Strawson also repeatedly refers back to Nagel's way of framing the problem, we must briefly turn to Nagel's argument.

Nagel (1979) locates the free will problem in two competing viewpoints that humans can adopt towards themselves and each other: the internal viewpoint and the objective (or external) viewpoint. When we view ourselves from the *internal viewpoint*, Nagel maintains, we see ourselves as agents who take part in intentional actions. Nagel himself struggles to give a positive characterization of what it means to be an agent acting intentionally, claiming it "very difficult" (33). Regardless, he claims that we have several ways of describing it. For example, when we (or others) act intentionally as agents, we often say that (a) the act was under our control, (b) we could have done otherwise, or even (c) the

---

<sup>11</sup> I am not suggesting that Watson is unaware of this distinction. He is, and writes that his account of Strawsonian agency is "conjectural" (footnote 3), and admits that his own interpretation might not even be the "best" account available (129).

action stemmed from (or reflected) who we are *as selves*. It is this sense of agency, Nagel claims, that is the target of (or, at least, makes us the appropriate subjects of) our moral and reactive attitudes. “Moral judgement of a person,” Nagel writes, “is judgement not of what happens to him but *of him*... It does not say merely that a certain event or state of affairs was fortunate or unfortunate or even terrible... We are judging *him*, rather than his existence or characteristics” (36).

Under the competing objective viewpoint, we see ourselves from the outside, or as another part of the larger unfolding process of objects and events. Nagel believes that when we view ourselves externally, we struggle to see ourselves as agents or selves. Rather, we see ourselves as objects governed over by deterministic laws.<sup>12</sup> He repeatedly claims that agency and selfhood is “threatened with dissolution” or even “vanishes” when viewed externally, and that we also “lose our grip” on our moral judgements and reactive attitudes (35). Nagel worries that the external viewpoint may well be the correct one, writing that:

I believe that in a sense the problem has no solution, because something in the idea of agency is incompatible with actions being events, or people being things. But as the external determinants of what someone has done are gradually exposed, in their effect on consequences, character, and choice itself, it becomes gradually clear that actions are events and people are things. Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the large sequences of events, which can be deplored or celebrated, but not blamed or praised. (37)

Although this tension appears broader than the question of whether free will is compatible with determinism, Nagel writes that the same tension “is revealed in the appearance that determinism obliterates moral responsibility” (38).

With these passages in mind, I standardize the *Disappearing Agent Problem* as follows:

- D1. If *P* adopts the internal viewpoint of herself, then *P* sees herself as an agent engaging in intentional actions.
- D2. If *P* adopts the external viewpoint of herself, then *P* sees herself as a natural object undergoing events and governed by deterministic laws.

---

<sup>12</sup> He also includes indeterministic laws. However, since this paper deals with the problem of free will, I will limit the discussion to determinism.

- D3. If *P* sees herself as a natural object undergoing events and governed by deterministic laws, then she no longer sees herself as an agent engaging in intentional actions. In other words, viewed externally, *P*'s agency disappears.
- D4. *P*'s external viewpoint is correct (or, at the very least, *P*'s external viewpoint is more accurate than her internal viewpoint).
- D5. Therefore, *P* is correct when she sees herself as a natural object undergoing events and governed by deterministic laws. [From (D2) and (D4)]
- D6. Therefore, *P* is ultimately incorrect when she sees herself as an agent engaging in intentional actions. [From (D1), (D3), and (D5)]

To be clear, Nagel does not necessarily endorse this argument. Instead, he claims that he intends to raise the Disappearing Agent Problem in order to show how it might (eventually) be solved:

The problem... cannot be understood without an account of the internal conception of agency and its special connection with the moral attitudes as opposed to other types of value. I do not have such an account. The degree to which the problem has a solution can be determined only by seeing whether in some degree the incompatibility between this conception and the various ways in which we do not control what we do is only apparent. I have nothing to offer on that topic either. (38)

Nagel appears to suggest here that the way (or, at least, one way) to solve the Disappearing Agent Problem is to reject (D3). Presumably, if a philosopher could offer a robust account of agency, then this account might reveal that agency is in fact compatible with the external viewpoint.

With the Disappearing Agent Problem in place, in the next subsection, I return to the Normative Competence Interpretation and argue that (a) it is not regarded by Strawson as a viable solution to the Disappearing Agent Argument, and instead (b) offer my own interpretation of Strawson's response to that argument.

### ***4.3 The Relativizing Response***

Strawson agrees with Nagel that the supposed threat presented by the truth of determinism is best captured by the Disappearing Agent Problem. For example, he writes:

If it is the viewpoint of participation and involvement [i.e. the internal viewpoint], to which we are so strongly committed by nature and society, which is correct, then some human actions really are morally blameworthy or praiseworthy, hateful or admirable, proper objects of gratitude or resentment ... If, on the other hand, it is only from the so-called "objective" viewpoint that we see things as they really are, then all our moral and quasi-moral reactions and judgments, however natural they may be and however widely shared, are no more than natural human reactions; no

question of their truth or falsity arises, for there is no moral reality for them to represent or misrepresent. (1985, 36-37)

As such, if the Normative Competence Interpretation is Strawson's way of addressing the problem of free will and determinism, then he would use it as a way to solve the Disappearing Agent Problem. However, there are two reasons that we should doubt that the Normative Competence Interpretation is Strawson's account of agency.

First, recall that the Normative Competence Interpretation states that people are agents only if they can understand and abide by moral demands or reasons. If this thesis works as a solution to the Disappearing Agent Problem, then it would show that responsiveness to moral reasons is compatible with the objective viewpoint. In other words, if *P* regards *Q* as merely an object, then does *P* still see *Q* as the sort of thing that understands and responds to moral reasons? Strawson answers negatively, writing:

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*. You can at most pretend to quarrel, or to reason, with him. (1962, 79, emphasis added)

Although Strawson does not explain this passage, it is clear that he believes that viewing another as being capable of engaging in moral reasoning is incompatible with viewing her from the objective viewpoint.

Second, Strawson regularly dismisses any attempt to give a list of necessary (and sufficient) conditions of agency in order to solve the Disappearing Agent Problem.<sup>13</sup> For example, in "Morality and Perception" (1985), he writes:

Attempts to counter such reasoning [the Disappearing Agent Problem] by defending the reality of some special condition of human freedom... which human beings enjoy and which supplies a justifying for our moral attitudes and judgments, have not been notably successful; for *no one* has been able to state intelligibly what such a condition of freedom, supposed to be necessary to

---

<sup>13</sup> To be clear, I am not suggesting that Strawson does claims that we cannot determine *some* of the necessary conditions for agency. Rather, I am saying is that, *insofar as solving the problem of free will and determinism is concerned*, Strawson claims that providing such a list is not the right way to proceed.



ground our moral attitudes and judgments, would actually consist in. *Such attempts at counter-argument are misguided.* (32, emphasis added).

Similarly, in “Freedom and Necessity” (1992), he reiterates:

...attempts to formulate the [conditions for agency] in other terms have never resulted in anything but either high-flown nonsense or psychological descriptions which are in no way incompatible with the thesis of determinism. *No one* has ever been able to state intelligibly what that state of affairs, which has been supposed to be necessary to ground our moral attitudes and judgments, would actually consist in. The question, ‘If we believe in such a condition, what exactly are we believing?’, *remains unanswered and, I think, unanswerable.* (137, emphasis added).

Although Strawson is certainly referring to libertarian attempts (“high-flown nonsense”), he is also referring to compatibilist attempts. For example, he writes that compatibilists have only offered mere “psychological descriptions” that are compatible with determinism. What Strawson means here is not that these compatibilist accounts are false, but rather that they do not solve the problem. (Additionally, Strawson expresses pessimism towards future compatibilist efforts, writing that our understanding of agency “remains a relatively vague and inexact kind of knowledge; and there must be few who suppose that it will ever be anything else” 1992, 139.) For these reasons, it seems unlikely that Strawson would endorse the Normative Competence Interpretation as a solution to the problem of free will and determinism.

But if Strawson does not offer his own account of agency as way to combat the Disappearing Agency Problem, what does he do instead? Strawson rejects (D4), or that the objective viewpoint is *the* correct viewpoint. He writes:

What I want now to suggest is that the error lies not in one side or the other of these two contrasting positions, but in the attempt to force the choice between them. The question was: From which viewpoint do we see things as they really are? and it carried the implication that the answer cannot be: from both. It is this implication that I want to dispute. (1985, 37)

Rather than stating that either viewpoint is erroneous, Strawson maintains that the real error lies in “attempting to force the choice between them” (37). To frame the debate, Strawson introduces the following two characters: the *strict* (sometimes *reductive*) *naturalist*, who argues for the sole legitimacy of

the objective viewpoint; and the *liberal* (sometimes *non-reductive*) *naturalist*, who argues that both viewpoints are legitimate.

According to the liberal naturalist, the fact that we (a) naturally occupy the internal viewpoint, and (b) cannot help but feel the reactive attitudes, suggests that we do not need to give justification for the internal viewpoint. The liberal naturalist also claims that, despite our natural commitment to the internal viewpoint, we can also adopt the objective viewpoint (albeit temporarily). Although these two viewpoints are in tension, Strawson argues that there is no reason to suppose that one is more legitimate than the other.<sup>14</sup> “It is perfectly consistent,” Strawson writes, “with the adoption of the thoroughgoing or non-reductive naturalist’s way with moral skepticism... to allow validity to the [reductively] naturalistic view of human behavior. This can be done... so long as we are prepared to acquiesce in the appropriate relativizations of our conception of the realities of the case” (44). Strawson calls this concession the *Relativizing Move*, which I summarize as:

Let *X* and *Y* be two viewpoints that humans can occupy. Even if *X* and *Y* yield contrary perceptions/judgments/etc., it is possible for both *X* and *Y* to be legitimate.

Based on the Relativizing Move, Strawson cautions that even if the strict naturalist’s sole commitment to the objective viewpoint is mistaken, it does not follow that the internal viewpoint is more accurate than the objective. Returning to the Disappearing Agent Problem, even if it is the case that the internal and objective viewpoints both reach different conclusions about the existence of agency, Strawson’s Relativizing Move nevertheless implies that both viewpoints can be legitimate (hence, he rejects (D4)).

Strawson understands that many will regard the Relativizing Move as an “evasion, rather than a solution” (38). He writes:

But surely, it may be said, two contradictory views cannot both be true; it cannot be the case both that there really is such a thing as moral desert and that there is no such thing, both that some

---

<sup>14</sup> To be sure, Strawson uses the word “valid,” rather than “legitimate.” However, his use of “valid” is confusing, since it’s not clear what sense of validity he is referring to (logical validity, etc.). Instead, I believe that his point can be made clearer with the word “legitimate,” since its usage suggests that we are justified in adopting both of these viewpoints (even if they sometimes reach contrary perceptions or conclusions).

human actions really are morally praiseworthy or blameworthy and that no actions have these properties. (37)

However, Strawson contends that this objection carries with it an unstated assumption:

I want to say that the appearance of contradiction arises only if we assume the existence of some metaphysically absolute viewpoint from which we can judge between the two viewpoints I have been contrasting. But there is no such superior viewpoint... (37-38).

Strawson is actually suggesting two responses to the strict naturalist. First, the only way to assess the legitimacy of the viewpoints would be to have some third superior viewpoint, whereby we could assess both. But, as Strawson argues, there is no such third viewpoint. Therefore, it is not possible to determine if one is more legitimate. We have a natural commitment to the internal viewpoint, and since we have no possible way to assess its legitimacy over the objective viewpoint, we cannot and (should not) have to justify it. Second, Strawson reminds us that although the strict naturalist might wish to say that the two viewpoints are contradictory (and hence it seems that we must choose), he suggests that the strict naturalist is overstating her case. He writes:

[These two] viewpoints and attitudes are not only different, they are profoundly opposed. One cannot be whole-heartedly committed to both at once. It will not do to say that they are mutually exclusive; since we are rarely whole-hearted creatures. But they tend in the limit to mutual exclusion. (1985, 36).

The strict naturalist's point that we must choose one viewpoint over the other might have had the additional force required if the two viewpoints were mutually exclusive. But, since they are *not* mutually exclusive—in the sense that we must choose one viewpoint over the other—we do not need to choose. For these two reasons, Strawson claims that (D4) of the Disappearing Agent Problem is false.

#### ***4.4 Strawson's Compatibilism, Revised***

Taken together, the Sentimentalist Interpretation, Agency View, and Relativizing Move all comprise the correct interpretation of Strawson's compatibilism. Notice that, under these three positions, Strawson's compatibilism does not (a) focus solely on the proneness to feel resentment and indignation or (b) what the outward expressions of these attitudes might communicate. Instead, it

focuses on the connection between our proneness to feel all of the reactive attitudes and how these attitudes are linked with us being agents (rather than us merely being objects). According to Strawson, humans have two viewpoints that they naturally occupy: the internal (or participant) viewpoint and the external (or objective) viewpoint.<sup>15</sup> Under the internal viewpoint, we see ourselves and others as agents, and our actions as expressions of that agency. The internal viewpoint also entails that (i) we view ourselves and others as free and morally responsible; and (ii) we are prone to feel the reactive attitudes in response to the actions of ourselves and others. Strawson also associates the internal viewpoint with the viewpoint that we adopt as social creatures; it is unique to (and constitutive of) human interpersonal relationships that we adopt the internal viewpoint and are prone to feel the reactive attitudes.

Under the external (or objective) viewpoint, we see ourselves and others as something similar to natural objects, and our actions as merely events. As such, we are likely to feel only objective attitudes, or attitudes that one feels towards natural objects. The putative threat posed to free will and moral responsibility by determinism is commonly associated with an exclusive adoption of the external viewpoint; some argue if determinism is true, then we are merely cogs in a larger deterministic machine.

Rather than attempting to argue that the external viewpoint is false, or attempting to offer a more robust conception of agency that would be compatible with the external viewpoint, Strawson instead claims that we do not need to choose between these viewpoints. He lists two reasons. First, although these two viewpoints are opposed, they are not contradictory. It is possible, Strawson claims, to regard each (and the attitudes that one feels under them) as legitimate. Second, we are unable to assess the legitimacy of one viewpoint over the other since we do not have access to a third superior viewpoint.

---

<sup>15</sup> This beginning is similar to what Strawson's own take on his compatibilism offered near the end of his life, that his account is grounded in "the phenomenology of [our] moral life" (1998, 262).

And since we are naturally committed to both viewpoints, and cannot assess the legitimacy of one over the other, we cannot (and should not) choose between them.

Before closing, I will offer two additional points about Strawson's compatibilism. First, these three positions can make sense of two puzzling claims made by Strawson throughout "Freedom and Resentment." (I) Strawson repeatedly insists that our proneness to the reactive attitudes "neither calls for, nor permits, an external 'rational' justification" (1962, 91-92). As we can now see, he means that the reactive attitudes *just are* the sorts of emotional responses that we are prone to when we adopt the internal viewpoint. Since we are naturally committed to the internal viewpoint, and cannot assess its legitimacy over the external viewpoint, our proneness to feel these reactive attitudes (which are the natural consequence of the internal viewpoint) does not require justification. (II) Strawson also compares our proneness to the reactive attitudes with our proneness to inductive reasoning:

Compare the question of the justification of induction. The human commitment to inductive belief-formation is original, natural, non-rational (not irrational), in no way something we choose or could give up. Yet rational criticism and reflection can refine standards and their application, supply 'rules for judging of cause and effect'. Ever since the facts were made clear by Hume, people have been resisting acceptance of them. (Endnote 7)

Notice that Strawson makes five claims about induction: it is (a) original and (b) natural for us, (c) non-rational, (d) something which we cannot choose to give up, and (e) capable of revision and refinement. Now we can see that our commitment to the internal viewpoint and proneness to the reactive attitudes meets the same five criteria. The internal viewpoint is natural for us (meeting (b)), and constitutive of human interpersonal relationships (meeting (a)). Likewise, it is non-rational to us, meaning that we do not need to provide rational justification for it (meeting (c)), and we cannot choose to give up the internal viewpoint for an exclusive adoption of the objective viewpoint (meeting (d)). Finally, Strawson even claims that the sorts of attitudes that we are prone to feel are up for some degree of revision, writing that "Inside the general structure or web of human attitudes and feelings

of which I have been speaking, there is endless room for modification, redirection, criticism, and justification” (93, also meeting (e)).

Second, some might worry about the plausibility of the Strawson’s compatibilism. More specifically, some might worry it is not psychologically possible for humans to regard both the internal and objective viewpoints as legitimate. To be sure, the aim of this paper was not to defend the content of Strawson’s position, but rather offer a plausible interpretation of it. Nevertheless, since the Relativizing Move is essential to the success of Strawson’s compatibilism, I will briefly summarize his response to these worries, which he compares to a similar worry found between (a) viewing perceptual objects from our commonsense viewpoint, and (b) viewing perceptual objects from the so-called scientific viewpoint. Strawson notes that under the *commonsense viewpoint*, the objects of our perception appear to have the phenomenological properties that we experience them as having—e.g. the apple *really is* red, the water *really is* wet, etc. Under the *scientific viewpoint*, objects do not have these phenomenological properties. “This view,” Strawson writes, “credits physical objects only with those properties which are mentioned in physical theory and physical explanation... [and it is an] illusion that physical objects *really are* as we seem to perceive them as being” (1985, 44). But despite yielding contrary judgments (e.g. the apple *really is* and *really isn’t* red), Strawson claims we do not have to choose between these two viewpoints. He writes:

...the impression of irreconcilable antagonism between the two views disappears as soon as we are prepared to recognize, as before, a certain ultimate relativity... in this case, of the real properties of physical objects. Relative to the human perceptual standpoint, commonplace physical objects really are... bearers of phenomenal visual and tactile properties. Relative to the stand-point of physical science (which is also a human standpoint) they really have no properties but those recognized, or to be recognized, in physical theory, and are really constituted in ways which can only be described in what, from the phenomenal point of view, are abstract terms. Once the relativity of these “really”s to different standpoints, to different standards of “the real” is acknowledged, the appearance of contradiction between these positions disappears; the same thing can both be, and not be, phenomenally propertied. (46-47)

Thus, Strawson suggests that since we are comfortable regarding our commonsense and scientific viewpoints as legitimate, so too should we be comfortable regarding our internal and external viewpoints as legitimate.

## 5 CONCLUSION

In this paper, I have critically evaluated three prominent interpretations of Strawson and offered a competing view of Strawson's compatibilism. First, in **section 2**, I examined the Communicative Interpretation, which stated that a commitment to moral responsibility is required for the outward expression of the reactive attitudes, and these expressions communicate important information that is crucial for maintaining healthy interpersonal relationships. In response, I argued that Strawson's position was not about what the reactive attitudes communicate, nor what makes for healthy relationships. Instead, I offered the Sentimentalist Interpretation, or that it is the proneness to feel the reactive attitudes that Strawson claimed constitutes human relationships.

Next, in **section 3**, I examined why Strawson claims that the proneness to feel all of the reactive attitudes is at stake in the free will debate. I began by critically examining the Proneness to Resentment Interpretation, under which Strawson claims that the proneness to feel resentment and indignation towards others is psychologically required for feeling the positive reactive attitudes towards them. I argued that although our proneness to feel resentment plays an important role in Strawson's theorizing, it's not clear that it plays the role purported by the Proneness to Resentment Interpretation. Instead, I argued for the Agency View, or that the reactive attitudes are all reactions to our perceptions of the actions of an agent, and that if we reject that people are morally responsible agents, then Strawson claims that we will view them as merely an object.

However, my analysis did not address the nature of Strawson's conception of agency, nor why Strawson claims that agency is at stake in the free will debate. So, in **section 4**, turned to one influential interpretation of Strawson's notion of agency, the Normative Competence Interpretation.

According to this interpretation, Strawson claims that someone is a morally responsible agent only if she has the ability to understand and abide by moral demands or reasons. I argued that in order to assess whether the Normative Competence Interpretation correctly captured Strawson's view of agency, we needed to understand (a) how Strawson understood the supposed threat that determinism poses to agency, and (b) examine what he has to say about that threat. Concerning (a), I argued that Strawson understood the supposed threat as best captured by Nagel's Disappearing Agent Problem. With that in place, I turned to (b). I offered two reasons why Strawson would not endorse the Normative Competence Interpretation. Afterwards, I examined Strawson's actual response to the Disappearing Agent Problem. According to Strawson, viewing ourselves and others as agents (where we are prone to feel the reactive attitudes) is a direct consequence of adopting the internal viewpoint, whereas seeing ourselves and others as merely objects is a direct consequence of adopting the objective viewpoint. I claimed that Strawson's move was to suggest that these two viewpoints are not mutually exclusive, but rather equally as legitimate. In other words, we do not have to choose between them. For those reasons, Strawson claims that neither a positive conception nor a justification of agency are either permissible or required. Finally, I offered an overall summary of Strawson's compatibilism, and addressed two lingering issues about it.

To be sure, none of what I have offered in this paper should be taken as a defense of Strawson's position. For example, I have not argued (a) that the Sentimentalist Interpretation is true, nor (b) that the reactive attitudes are interconnected reactions to agency, nor (c) that the proper way to handle the problem of free will is by way of the Relativizing Move. Indeed, many, myself included, might find that portions of Strawson's strategy leaves something to be desired. Nevertheless, what I have done in this paper is offer an attempt to better understand Strawson's compatibilism. In doing so, I have shown that many of the views commonly attributed to him are probably mistaken. If I am right, then



the crucial question becomes: *Is the Strawson's compatibilism a successful response to the problem of free will and moral responsibility?*

## WORKS CITED

- Bennett, Jonathan. (1980). "Accountability," in *Philosophical Subjects: Essays Presented to P.F. Strawson*, ed. Zak van Stratten. Oxford: Clarendon Press.
- Darwall, Stephen. (2013). "Moral Obligation: Form and Substance," in *Morality, Authority, and Law: Essays in Second-Personal Ethics*. New York: Oxford University Press, 40-51.
- (2009). *The Second-Person Perspective*. Harvard University Press.
- Fischer, John Martin, and Mark Ravizza. (1998). *A Theory of Moral Responsibility*. Cambridge University Press.
- Macnamara, Coleen. (2015). "Blame, Communication, and Morally Responsible Agency," in *The Nature of Moral Responsibility: New Essays*, eds. Randolph Clark, Michael McKenna, and Angela M. Smith. New York: Oxford University Press, 211-236.
- (2013). "Taking Demands Out of Blame," in *Blame: Its Nature and Norms*, eds. D. Justin Coates and Neal A. Tognazzini. Oxford, University Press, 119-140.
- McKenna, Michael. (2012). *Conversation and Responsibility*. New York: Oxford University Press.
- Nichols, Shaun. (2007). "After Incompatibilism." *Philosophical Perspectives* 21, 497-515.
- Pereboom, Derk. (2014). *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Strawson, P.F. (1992). "Freedom and Necessity," in *Analysis and Metaphysics*. New York: Oxford University Press, 133-142.
- (1985). "Morality and Perception," in *Skepticism and Naturalism: Some Varieties (Woodbridge Lectures)*. New York: Columbia University Press.
- (1980). "Replies," in *Philosophical Subjects: Essays Presented to P.F. Strawson*, ed. Zak van Stratten. Oxford: Clarendon Press.
- (1962). "Freedom and Resentment," in *Free Will: Second Edition*, ed. Gary Watson, 2003. New York: Oxford University Press, 72-93.
- Tiberius, Valerie. (2015). *Moral Psychology: A Contemporary Introduction*. New York: Routledge.
- Wallace, R. Jay. (2011). "Dispassionate Opprobrium. On Blame and the Reactive Sentiments," in *Reasons and Recognition. Essays on the Philosophy of T. M. Scanlon*, eds. R. Jay Wallace, Rahul Kumar, and Samuel Freeman. New York: Oxford University Press, 348-372.
- (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.

Watson, Gary. (1987). "Evil and the Limits of Moral Responsibility: Variations on a Strawsonian Theme," in *Perspectives on Moral Responsibility*, eds. John Martin Fischer and Mark Ravizza (1993), 119-150.