

8-21-2008

# Freedom and Forfeiture: Responding to Galen Strawson's Basic Argument

Eli Benjamin Kelsey

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

---

## Recommended Citation

Kelsey, Eli Benjamin, "Freedom and Forfeiture: Responding to Galen Strawson's Basic Argument." Thesis, Georgia State University, 2008.

[http://scholarworks.gsu.edu/philosophy\\_theses/46](http://scholarworks.gsu.edu/philosophy_theses/46)

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).

FREEDOM AND FORFEITURE: RESPONDING TO GALEN STRAWSON'S BASIC  
ARGUMENT

by

E. BENJAMIN KELSEY

Under the Direction of Eddy Nahmias

ABSTRACT

Galen Strawson's *Basic Argument* is an attempt to prove that no agent can meet the demands for true moral responsibility. The Basic Argument proceeds on the assumption that, in order for an agent to be truly morally responsible for her actions, she must be truly responsible for her *reasons* for performing those actions, which Strawson contends is impossible since it requires an infinite regress of truly responsible decisions to have the reasons one has. In my thesis, I take issue with the Basic Argument. I argue that, contrary to Strawson's claims, the Basic Argument is not persuasive to those who reject that one's reasons *cause* one's actions. For those who are willing to overlook this shortcoming, I then argue that it is possible for an agent to evade the threat of infinite regress, particularly in situations where two simultaneous choices (at least partially) explain each other.

INDEX WORDS: Free will, Moral responsibility, Galen Strawson, Basic Argument

FREEDOM AND FORFEITURE: RESPONDING TO GALEN STRAWSON'S BASIC  
ARGUMENT

by

E. BENJAMIN KELSEY

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

2008

Copyright by  
Eli Benjamin Kelsey  
2008

FREEDOM AND FORFEITURE: RESPONDING TO GALEN STRAWSON'S BASIC  
ARGUMENT

by

E. BENJAMIN KELSEY

Committee Chair: Eddy Nahmias

Committee: Timothy O'Keefe  
Sebastian Rand

Electronic Version Approved:

Office of Graduate Studies  
College of Arts and Sciences  
Georgia State University  
December 2008

## ACKNOWLEDGEMENTS

My sincerest appreciations must be given to Eddy Nahmias, whose consistent feedback on this project has allowed it to rise far and beyond its initial stages. Special thanks must also be given to Timothy O’Keefe and Sebastian Rand, whose comments convinced me of both the strengths and weaknesses of my position, whether or not my writing properly demonstrated the former or managed to overcome the latter. If it is possible for anyone to be *truly* morally responsible for the shortcomings of this work, it is me and me alone.

A final thanks goes to my beautiful wife, Melanie, who remained encouraging of my work, even though it so often took me away from her. I offer this thanks as a small recompense for the many mealtimes and bedtimes that became a philosophical brainstorming session for me rather than a quiet moment of companionship and relaxation between best friends. That you never complained shows that you really are my best friend, and my love and gratitude goes out to you.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
CHAPTER	
1 INTRODUCTION	1
1.1 The Project	1
1.2 Some Preliminary Definitions	4
2 STRAWSON'S BASIC ARGUMENT	7
2.1 The Two Requirements of Freedom	7
2.2 The Impossibility of Self-Determination	8
2.3 The Basic Argument	11
3 THE ROLE OF REASONS IN RATIONAL ACTION	17
3.1 Reasons as Causal Explanations – Strawson's Approach	17
3.2 Reasons as Non-Causal Explanations – Ginet's Approach	23
4 THE FREEDOM OF FORFEITURE	30
4.1 Action-Forfeitures as Actions	30
4.2 Action-Forfeitures as Distinct Choices	33
4.3 Action-Forfeitures and the Evasion of Infinite Regress	36
4.4 Action-Forfeitures and Self-Determination	42
5 FINAL THOUGHTS	45
5.1 Objections	46
5.2 Just the Beginning	50
BIBLIOGRAPHY	52

## 1 Introduction

### 1.1 The Project

Do we, as rational agents, ever act freely? Are we ever morally responsible for our actions and/or choices? Is there a relationship between an agent's acting freely and her being morally responsible? These are some of the basic questions with which the so-called *free will debate* is concerned. The debate is complicated both by what we know and what we do not know about humans, the mind, and the world at large. On the one hand, there is our first-hand, subjective experience—our *phenomenology*—which strongly suggests that a great many things we do are under our direct control. On the other hand, many of us subscribe to what we consider scientific and objective facts about the world, facts that indicate a world in which causal laws may dictate everything that does and does not happen. Are the two views commensurable? Must our assumptions about one or the other be reconsidered or abandoned?

Galen Strawson argues that that free will and moral responsibility do not exist. Interestingly, he also claims that many of the issues with which the free will debate is typically concerned—issues of causation, the relationship of the mind to the body, etc.—need not be resolved before the non-existence of free will and moral responsibility can be proven. Strawson's argument relies on two very simple ideas. The first is that *to act freely* just means *to act in a way for which one can properly be said to be truly morally responsible for one's action*. The second notion is that, to be *truly* morally responsibly for one's action, one must be (in some respects, at least) *truly* responsible for the reasons that lead to performing that action. Strawson's rejection of free will and moral responsibility is tied to this second criterion. The criterion, Strawson contends, is impossible to meet, because it requires that an agent be truly responsible—or *self-determined*, as he sometimes puts it—in either the beliefs or the desires that

make up the agent's reasons for performing the action in question. This in turn is supposedly impossible because it requires an infinite regress of choices to have the reasons one has.

Because the self-determination of reasons is impossible, says Strawson, true moral responsibility is impossible, and by the definition of freedom which Strawson himself provides, no agent ever acts freely.

In what follows, I will argue against Strawson. I will focus first on undermining a key assumption upon which Strawson's entire argument is based—that the reasons an agent has for performing an action *cause* the agent to perform that action. As will be shown, this causal assumption is the primary motivation for Strawson's ultimate rejection of self-determined, and hence truly morally responsible, actions. Once doubt is cast over the causal efficacy of an agent's reasons, I will argue that the first move in Strawson's argument is inadequately supported, and as such, that his entire argument is suspect. Still, for those who are willing to overlook this flaw, I will proceed to tackle the problem of infinite regress that is the heart of Strawson's argument. I will do so by focusing on cases where an agent chooses *not* to take some course of action she considers equally appealing to another course of action she *does* decide to take. In these types of cases, what an agent chooses *not* to do will be explainable primarily by reference to the fact that the agent chose to do something else. This choosing to do something else will serve as a genuine, non-regressive reason for the agent's "forfeiture" of the other course of action under consideration, thereby staving off the infinite regress that underlies Strawson's argument for the impossibility of true moral responsibility.

Here is how my thesis shall unfold. After setting forth some useful definitions at the end of chapter 1, I will in chapter 2 explicate Strawson's argument for the impossibility of freedom and moral responsibility. In chapter 3, I will examine the role that reasons, as the primary

components of rational action, play in explaining an agent's actions. Strawson's belief that an agent's reasons cause her actions will be explored, followed by an alternative view of reasons based largely on the work of Carl Ginet. Siding with Ginet, I will argue that an agent's reasons can rationally explain her actions without assuming that those reasons *cause* those actions. In light of this non-causal account of reasons, I will contend that the first move in Strawson's argument is questionable.

In chapter 4, I will seek to establish that the forfeiture of an action is, in at least some cases, immune to Strawson's concerns of infinite regress. I begin by making it clear that these forfeitures are genuine actions, performed for reasons and susceptible to moral evaluation. I then argue that, although these forfeitures often appear together with a commitment to do something else, the forfeiture and the commitment cannot be regarded as one and the same choice. Consequently, the commitment itself can provide a genuine reason for the forfeiture, and the forfeiture can provide a genuine reason for the commitment. Infinite regress is thereby avoided, since neither the commitment nor the forfeiture can be explained wholly by reasons that precede the agent's decision. With a causal account of reasons having been dismissed in chapter 3 and the problem of infinite regress successfully circumvented, I will then argue that some action forfeitures and commitments should remain eligible candidates for being truly morally responsible actions, since Strawson's argument will no longer pose any obvious threats.

I will conclude this thesis by addressing some potential objections and offering some final thoughts on the meaning of free will and moral responsibility in chapter 5.

## 1.2 Some Preliminary Definitions

In this thesis, the issue of determinism that characteristically dominates the free will debate will largely be eschewed, since Strawson himself believes the truth or falsity of determinism has no bearing on his argument. Time and again, Strawson asserts that his objection to freedom and true moral responsibility holds regardless of one's position in the debate about the compatibility of freedom and determinism, and even regardless of one's theory of mind. Still, it will be useful to provide definitions for several common terms surrounding the traditional debates about determinism, since failing to take these concepts into account would at certain junctures prove irresponsible. The terms to be defined are as follows: *determinism*, *indeterminism*, *compatibilism*, *incompatibilism*, and *libertarianism*.<sup>1</sup>

*Determinism* is the theory that, given whatever natural laws exist (be they physical laws, psychological laws, some other kinds of laws, or any combination of the above) and given the actual history of the world, there is one and only one realizable future. In simpler terms, because of the laws of the universe and because of the way things were and are, whatever happens is necessitated by the past.

*Indeterminism* is, appropriately enough, the theory that, given the way things actually are (including any laws that may hold and the way the world has been up until now), there are at least two genuinely realizable futures. In short, indeterminism states that determinism is false.

*Compatibilism* is the theory that it is possible both for an agent to act freely and for determinism to be true. That is, compatibilism is the theory that the truth of determinism is compatible with an agent's possessing free will. Typically, a philosopher is identified as a compatibilist if she believes that determinism *is* true and that agents *do* possess free will, but this need not be the case. So long as one does not recognize an inherent conflict between

determinism and free will, one is a compatibilist. In theory, then, a compatibilist may believe that determinism is false, that agents are unfree, but that determinism and free will are compatible nonetheless. A compatibilist may even be an agnostic concerning whether or not agents do exist who possess free will, whether determinism is true, or both.

*Incompatibilism* is the theory that determinism and free will cannot co-exist. Ergo, the incompatibilist believes that *if* determinism is true, then there are no agents with free will. Similarly, the incompatibilist must assert that *if* an agent (even just one agent) has free will (or acts freely, even once), then determinism is false. As was the case with compatibilism, this does not obligate the incompatibilist to a particular stance on either free will or determinism, other than to the conditional claims just laid out. In short, then, there are two primary positions the non-agnostic incompatibilist may take: (1) that at least some agents act freely on at least some occasions (and thus that determinism is false), or (2) that determinism is true (and thus that nobody ever acts freely).<sup>2</sup>

*Libertarianism* is the theory occupied by position (1), while *hard determinism* occupies position (2). Libertarians believe that at least some agents do sometimes act freely, and that incompatibilism is true. It follows that all libertarians are indeterminists. It is safe to say that, for libertarians, at least some (if not most) of the indeterminism about the future is based on the fact that agents do have free will. The where, what, when, and even *if* of my eating a meal one year in the future is largely undetermined, primarily because it is genuinely within my power to realize multiple courses of action, both a year from now and in the interim. Even if I can make only one choice at any given moment, and even if that choice is severely limited by

---

<sup>1</sup> Strawson provides a similar account in §1.3 of *Freedom and Belief* (5-7).

<sup>2</sup> The rare incompatibilist may exist who believes that indeterminism is true but that no agent has free will. In such a case, the non-existence of agential freedom will be explained by something other than the truth or falsity of determinism.

circumstances beyond my control, there are situations—perhaps quite a lot of them—wherein I am in no way bound to do any one particular thing. For this reason, libertarianism is sometimes dubbed a “robust” form of free will. Unlike the free will that is sometimes espoused by compatibilists, libertarian free will relies on the notion that an agent is not (or at least not always) determined to choose one way or another.

Hard determinists are incompatibilists who hold that determinism is true, and thus that we do not have free will. Though Strawson may be regarded as a hard determinist, he argues that indeterminism is equally incompatible with free will and moral responsibility. Derk Pereboom subscribes to this position, calling it *hard incompatibilism* (Fischer et al., 2007: 85). But Strawson’s view is even stronger than Pereboom’s in that Strawson denies the very possibility of genuine moral responsibility, whereas Pereboom believes it is possible but not actual. Let it suffice to say that Strawson does not attempt to situate himself within the traditional categories of the debate, identifying himself only as a disbeliever in free will and moral responsibility. He takes no position on whether or not determinism is true, and he does not believe this bears in any way upon the efficacy of his argument.

## 2 Strawson's Basic Argument

### 2.1 The Two Requirements of Freedom and Moral Responsibility

Galen Strawson purports that “to be a free agent is to be capable of being *truly responsible* for one’s actions” (1986: 1). Moreover, he asserts that “to be capable of being truly responsible for one’s actions is to be capable of being truly deserving of praise and blame for them” (ibid.). For Strawson, then, to be free in one’s actions and to be truly morally responsible for those actions are one and the same.<sup>3</sup> In *Freedom and Belief*, he states from the outset that “the word ‘free’ will be used interchangeably with the phrase ‘truly responsible.’ Questions about what freedom is, and about whether or not we are or could be free, will be understood to be questions about what true responsibility is, or might be, and about whether we are or could be truly responsible or truly deserving of praise or blame” (2). Strawson claims this true moral responsibility is unattainable because an agent can never be responsible for the reasons that lead her to act as she does.

Strawson contends that in order for a person to be truly morally responsible for performing some action A, two criteria must be met. First, action A must belong to the class of rational actions—that is, “actions that are performed for a reason (as opposed to ‘reflex’ actions or mindlessly habitual actions)” (2003: 212). Put into more common vernacular, action A must be an action that is done *on purpose*. Strawson does not dwell on this requirement, but it is important to keep in mind since these are the kinds of action that typically fall under the moralistic microscope. When actions are known to be involuntary, we are not inclined to ascribe moral responsibility to those who instantiate them, regardless of the consequences that follow. That is, even if some situation existed wherein my flinching at a fastball that narrowly misses my

---

<sup>3</sup> Hereafter, I will not always include the qualifier “moral” or “morally” when speaking of true moral responsibility. The qualifier may be assumed.

face could bring about cataclysmic results, no reasonable person would hold me accountable for flinching. Similarly, Strawson is concerned only with those actions that an agent performs intentionally, and a key indicator that an action was intentional is that an agent had conscious reasons for performing it.<sup>4</sup>

The second criterion is more complex. Strawson claims that the person performing action *A* must be *self-determined* in her action if she is to be held truly responsible for it, because “true responsibility presupposes true self-determination” (1986: 26). But to be self-determined in one’s *action*, according to Strawson, one must be self-determined in one’s *reasons* for so acting. Hence, it is the origin of one’s reasons that ultimately make or break that person’s freedom. If a person is *self-determined* in her reasons, then she may be regarded as self-determined in the actions that those reasons produce, and she is thereby a candidate for true moral responsibility. If she is *not* self-determined in her reasons, she cannot act in such a way that she is truly morally responsible for her actions. The question, then, is how one can be self-determined in one’s reasons. On Strawson’s account, a person is self-determined if one has “consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and [...] succeeded in bringing it about that one is that way” (2003: 213; cf. 1986: 28). Put more simply, a person is self-determined in her reasons if she is truly responsible, to some extent at least, for having those reasons in the first place.

## 2.2 The Impossibility of Self-Determination

Strawson believes the kind of self-determination requisite for true moral responsibility is impossible to achieve. For Strawson it appears obvious that agents cannot be the authors of their

---

<sup>4</sup> Some might argue that a person can act freely and/or be morally responsible for actions that are done involuntarily, provided that certain conditions exist (e.g. the agent is responsible for being in a position where such an action was

own reasons, those “principles of choice” or “preferences, values, pro-attitudes, ideals” and the like that constitute one’s mental character and bring about the choices one makes (1986: 29; cf. 2003: 213). Furthermore, these principles of choice—which Strawson claims may be broadly lumped into one of two categories: desires and beliefs—seem to give a full account of one’s choices, such that no further explanation is necessary for *why* an agent acted as she did other than pointing to her reasons; an agent’s rational actions are “fully explicable by reference to” her desires and beliefs (1995: 18).<sup>5</sup> Because an agent’s reasons will provide a full explanation of any rational action she performs, the only way an agent can be truly responsible for her actions is if she is self-determined in her beliefs and/or her desires. But Strawson does not think an agent can be responsible for either her beliefs or her desires.

Take beliefs. Strawson does not think anyone should expect, or even desire, to be self-determined in regard to beliefs, since the primary aim of having beliefs is that they correspond with reality—that is, that they be *true*. If we want our beliefs to be *true* beliefs, which seems a given,<sup>6</sup> then we will not want our beliefs to be determined by anything other than reality, not even by “the self.” In fact, even if self-determination *were* possible for beliefs, it would presumably be an unattractive way for any of us to acquire our beliefs. As Strawson notes, “We do not wish to be undetermined by anything, so far as the formation of our beliefs (and therefore their content) is concerned; nor do we wish to be self-determining with regard to the content of

---

likely to occur). I agree, but I am content to work within Strawson’s stricter limits for the time being.

<sup>5</sup> Perhaps it would be best to say that a person’s *choices*, rather than a person’s *actions*, are fully explicable by reference to her desires and beliefs. Strawson does not put it this way, but he does recognize that other, non-mental factors will bear upon an agent’s actions. My ordering vanilla over chocolate ice cream, for example, will require, among many other things, the physical ability to communicate in some way.

<sup>6</sup> I suppose there may be cases in which an agent would rather have a false belief than a true one, such as married persons who would rather live oblivious to their spouses’ extramarital affairs than face the ugly truth of infidelity. But even in these cases, I believe the agents desire first and foremost that their beliefs be true—namely, that their spouses *aren’t* cheating on them. If they learn otherwise, they may want to “block out” the truth and live in self-willed delusion, but I think this is indicative of a fervent desire that things be different than they are, not that their beliefs be rooted in anything other than the truth.

our beliefs; nor do we think we are [...]. Rather, we think (and hope) that what we believe is determined by, and as a result reflects, how things are” (1986: 43). If it were both possible and common for us to self-determine our beliefs, it is doubtful that we would regard beliefs in the same reality-reflecting way that we do.

None of this is to say that we, as agents, do not *influence* our beliefs in some vital sense. Indeed, many of us make choices that directly bear upon the beliefs we come to acquire. To give an example, one’s belief that the notes D, E, F#, G, A, B, and C# make up the scale of D major may hinge upon a past decision to study music theory. But even beliefs whose origins can be traced back to some choice or decision on the part of the agent do not make that agent self-determined in her beliefs. If everything goes according to plan, *all* of our beliefs will be rooted in reality, whether they tie back to some particular decision we made or not. Strawson notes, “We may [...] choose to acquire a lot of beliefs about this or that, but once we are in pursuit of such beliefs we do not wish to be able to choose what their content will be, we just want them to be true” (ibid.).

Desires bring other considerations to the table, though they prove no less problematic in terms of achieving self-determination. Strawson admits that we have a certain amount of sway over our desires, much more so than in the case of beliefs, but not to the extent that we can properly be said to have *self-determined* our desires. The notion of so-called “acquired tastes” will make Strawson’s point clear. Consider a teenager who intentionally drinks beer until he genuinely comes to like it. In some sense, the teenager has intentionally changed his desires regarding beer, but as Strawson would quickly point out, the teenager will have had reasons for seeking to change his beer-disliking disposition in the first place. So unless the teenager is truly

responsible for the reasons that led to his decision to overcome his initial dislike of beer, he will not be truly responsible for developing the acquired taste. According to Strawson:

If one is to be truly responsible for one's actions because one has chosen the desires (values, etc.) which lead one to act as one does, then one must clearly be truly responsible for this choice of desires in turn. And one can be truly responsible for this choice of desires only if one makes it in a reasoned, conscious, intentional fashion. But one cannot do this unless one chooses according to values and preferences one already has in the matter of what desires to have. (1986: 49)

Put simply, even if there is some legitimate way in which an agent can be said to have chosen her desires, the choices that will have determined those desires will themselves have been made in light of desires the agent already had. The agent who seeks self-determination by way of choosing her desires comes up against the same difficulties faced by the agent who seeks self-determination in regard to actions more generally. Neither can claim *true* responsibility for her motives, for the reasons she has for making the choices she does.

### 2.3 The Basic Argument

The preceding considerations give rise to what Strawson terms in (2002) and (2003) *the Basic Argument*. Strawson articulates the argument in a variety of ways, some with much greater detail than others, but the heart of it can be preserved by paraphrasing one of his simpler renditions.<sup>7</sup> The Basic Argument can be stated as follows:

Whenever we choose to perform some action *A*, we do so because of the reasons (made up of desires and beliefs, broadly speaking) that we have for choosing *A* at the time. Because our

---

<sup>7</sup> Strawson offers no less than four variations of the Basic Argument between 443 and 448 of (2003). Other versions appear at 15-16 and 17 of (1995) and 212-213, 219, and 219-20 of (2002).

reasons lead us to choose the way we do, we can only be truly responsible for choosing *A* if we are truly responsible for having those reasons that brought about our choosing *A*. However, to be truly responsible for our reasons for choosing *A*, we must have consciously and explicitly chosen to have those reasons. But even if we did consciously and explicitly choose to have the reasons that led us to choose to do *A*, we will have made *that* choice due to the reasons we had for *choosing to choose* to have the reasons for choosing *A*. In order to be truly responsible for *those* reasons—that is, for the reasons that led us to *choose to choose* to have the reasons for choosing *A*—we would have to have consciously and explicitly chosen to have *them*. We would have to have consciously and explicitly chosen to have the reasons that led us to choose to have the reasons that led us to choose to have the reasons for choosing *A*. The story then repeats *ad infinitum*.

An example may make this clearer. While something as basic as choosing between chocolate and vanilla ice cream would illustrate the point, considering an example of even minor moral import will best preserve the tone of Strawson's argument. (I will utilize this example at various times throughout the thesis.) Consider a woman, Nina, who lives in an apartment complex. One day she arrives home very late at night and notices that the car she recognizes as her neighbor's still has its headlights on. As she walks past the car on the way to her apartment, she sees that nobody is in the car. As would most of us, Nina assumes the headlights have been left on unintentionally. Immediately, Nina thinks of notifying her neighbor, but it's late enough that Nina, who is already quite shy, feels very uneasy about knocking on the door of someone she considers a virtual stranger. However, Nina then notices that the passenger-side door (and only the passenger-side door) is unlocked, and it occurs to Nina that she could very well just open the door, reach in, and turn off the headlights herself. This results in a dilemma. On the

one hand, Nina feels uncomfortable knocking on her neighbor's door when, for all she knows, the neighbor will be more upset about having been awoken than about his headlights still being on. On the other hand, entering her neighbor's vehicle without his knowledge, even to do him a favor, also makes Nina anxious. (After all, what if the neighbor happens to come out as she is climbing into his car? The very thought nearly paralyzes poor Nina!) Of course, Nina could just ignore the situation, figuring a dead car battery is fair recompense for the neighbor's oversight. But as tempting as this is, Nina feels that doing nothing would be immoral. And so, after a moment's deliberation, Nina quickly heads to the passenger-side door, opens it, leans across the seat, and switches off the headlights before hurrying into her apartment and heading to bed.

Many of us would agree that Nina has acted commendably in turning off the headlights of her neighbor's car. However, Strawson would point out that in order for Nina to be *truly* responsible for her actions and thus *truly* deserving of praise, Nina would need somehow to be *truly* responsible for whatever reasons led her to switch off the headlights. The obvious reasons at play, however, do not seem to allow for this. Nina's beliefs that the car headlights were on, that she could turn them off, and that doing so would be the polite thing to do are not beliefs for which Nina can be said to be truly responsible. There is no sense in which Nina was *self-*determined in having those particular beliefs, as they appear to be a byproduct of her situation. Likewise, Nina's *desire* to act courteously is not a product of self-determination. Perhaps Nina made a conscious choice many years ago always to be a courteous person and to do for others any favors that could be done without greatly inconveniencing herself. But even if Nina has acted on the precepts of this conscious choice made in the past (call it *CI*), Strawson would argue that Nina is not *truly* responsible for her actions. Acting in accordance with *CI* is not enough to give Nina *true* responsibility for her actions because Nina will have chosen *CI* based

on some other set of reasons for which she is not truly responsible. She may have chosen *CI*, for example, simply because she believed that adhering to *CI* would make her a praiseworthy person and because she desired to be considered a praiseworthy person. But to be truly responsible for either of those reasons for choosing *CI*, she would have to have consciously chosen to possess that particular reason *and* she would have to be truly self-determined in her deciding to possess it. The story then repeats with a new cast of beliefs and desires, and eventually we will come to some initial beliefs and desires that (incontestably, it seems) are entirely outside of Nina's control, beliefs and desires that stem from Nina's very early upbringing, from her particular genetic makeup, and so on.

It will be useful here to include a simplified version of the Basic Argument. The premises can be given as follows:

- (1) One's actions (at least those for which one might be considered responsible) are determined by one's reasons.
- (2) Thus, to be truly responsible for one's actions, one must be self-determined in one's reasons for performing that action.
- (3) To be self-determined in one's reasons, one must have consciously and explicitly chosen to have (at least some of) those reasons *and* be self-determined in having made the choice to have those reasons.
- (4) To be self-determined in having made the choice referenced in (3), one must have been self-determined in one's reasons for making that choice.<sup>8</sup>

---

<sup>8</sup> One may compare this standardization of Strawson's argument with his own standardization at (1986: 27-28). Though my version is less than half the length of Strawson's, I do not feel it leaves out anything significant.

Premise (4) leads right back to premise (3), supposedly beginning an infinite loop. The only way to break the regress is to concede (eventually) that one has acted on the basis of reasons that were not self-determined.

The Basic Argument relies on two very important moves. The first is the move from premise (1) to premise (2). Strawson could be more explicit about why premise (1) leads to premise (2), but it is clear enough that he sees premise (2) as a consequence of premise (1).<sup>9</sup> One may wish to question this move, however, or at least to come to a clearer understanding of how the word “determined” in premise (1) is being used. It may be that a gap exists between (1) and (2) if certain deterministic presumptions are not met. For instance, it may not be clear that one needs to be self-determined in one’s reasons in order to be truly responsible for one’s actions if one’s reasons do not *cause* one’s actions. As I will argue in chapter 3, that one’s reasons cause one’s actions is an unwarranted assumption which Strawson’s argument requires to proceed. If doubt can be cast over the causal efficacy of reasons, we may be justified in rejecting the Basic Argument altogether, since it will no longer be clear precisely why one must self-determine one’s reasons in order to be truly morally responsible for one’s actions.

The more significant feature of Strawson’s argument, of course, is the threat of infinite regress that comes out in (3) and (4). Even if reasons are non-causal, the kind of actions with which we are presently concerned will all be done *for* reasons, and Strawson will likely contend that to be *truly* responsible, one must avoid the circular trap of (3) and (4) even if the link between premises (1) and (2) has not adequately been established. Though I remain unconvinced that the (potential) infinite regress of one’s reasons threatens true moral responsibility in any *obvious* way, I will respond to this challenge in chapter 4 by arguing that many choices the typical agent makes will be based on reasons that are not subject to the

problem of regress. More specifically, the fact that an agent willingly forfeits certain courses of action will often be explained, at least in part, by the agent's decision to do something else *instead* at that time, a reason that arises simultaneously with (though is not identical to) the agent's decision to forfeit certain courses of action and thus prevents a regress from ever taking place.

---

<sup>9</sup> Compare my "thus" in premise (2) with Strawson's "therefore" in premise (3) of (1986: 28).

### 3 The Role of Reasons in Rational Action

As we have seen, Strawson denies that an agent can be self-determined in her actions. He roots this impossibility in the idea that an agent acts based on reasons for which she cannot be truly responsible, either because she will not have consciously and explicitly chosen to have them or because she will have consciously and explicitly chosen to have them based ultimately on reasons that she *did not* consciously and explicitly choose to have. Because an agent is not responsible for those very things (i.e. beliefs and desires) that determine what she does, Strawson says, the agent cannot be self-determined in her actions. The purpose of this chapter is to show that Strawson's argument is persuasive only if one believes that one's reasons *cause* one's actions. In the first part of this chapter, I will show how Strawson's argument relies on a causal link between one's reasons and one's actions, despite Strawson's attempts to appease those who would disagree with him. In the second part of this chapter, I will draw upon Carl Ginet's work to argue that reasons need *not* be regarded as causal, and as such, that the intuitive force of Strawson's argument is lost. The threat of infinite regress that is the heart of Strawson's argument will then be discussed in chapter 4.

#### 3.1 Reasons as Causal Explanations – Strawson's Approach

The interplay between an agent's reasons for performing some action A and the agent's actual performing of A is of paramount importance to the Basic Argument. Strawson claims that an agent cannot be responsible for her actions if she is not responsible for her reasons, precisely because those reasons bring about those actions. This was evidenced in the move from premise (1) to premise (2) of the simplified version of the Basic Argument given in 2.3. Strawson's way of thinking raises the important question of *how* one's reasons produce one's actions, since the

determination of one's actions by one's reasons forms the foundational premise of Strawson's entire argument.

Strawson believes an agent's reasons are causal. When speaking of actions, the dichotomy he offers is that actions are either "caused (determined) or 'random' (undetermined)" (1986: 8). Thus, when Strawson claims that one's reasons determine one's actions, he presumably means to say that one's reasons *cause* one's actions. This causal assumption permeates Strawson's work. Rather frequently, Strawson goes beyond the notion that an agent will act *in light of* her reasons (whichever course of action she happens to choose) and makes the much stronger claim that the *specific* action the agent undertakes will be the result of factors ultimately beyond the agent's control. "The *particular way* in which one is moved [...]," says Strawson, "and the degree of one's success [...] will be determined by how one already is" (1986: 30, emphasis mine). Elsewhere, he says that, after one deliberates, "which reasons finally weigh with one is *wholly* a matter of one's mental nature [...] which is something for which one cannot be in any way [truly responsible]" (2002: 455, emphasis mine; similar talk can be found in (2003) at 214 and, more extensively, at 223-4). These sentiments are at the heart of the Basic Argument. Reasons bring about an agent's actions, but the agent cannot possibly be truly responsible for those reasons. Hence, the agent cannot be *truly* responsible for her actions.

If one's reasons *cause* one's actions, and if one cannot be responsible for one's reasons, then one cannot be responsible for one's actions. Such an argument holds intuitive sway, even if it is contestable. But it is not so clear that Strawson's argument works if reasons are not presumed to be causal in the first place. Attempting to appease those who would reject the idea that reasons are causal, Strawson offers an alternative definition of "to determine" that does not employ such causal language. He says that "to determine [e.g. one's actions]" may be rendered

“to play a crucial role in whatever process it is that finally determines the nature of [e.g. one’s actions]” (1986: 34-35),<sup>10</sup> and later he states that one’s reasons can be said to *fully* determine (rationally speaking) one’s actions, though this means only that one’s reasons provide “a true and full rational explanation of” one’s actions (1986: 37). Strawson has thus articulated a theory in which one’s reasons can be said to *determine* one’s actions so long as those reasons either explain or significantly factor into one’s actions.

These non-causal definitions are obviously meant to preserve the cogency of Strawson’s argument for the widest audience possible, but in reality they threaten the intuitive force of his argument. Consider the case of Nina introduced in 2.3 in light of the first non-causal definition just offered. Nina’s reasons may indeed “play a crucial role”<sup>11</sup> in her decision to turn off her neighbor’s headlights, but does this make it obvious that Nina needs to be self-determined in those reasons in order to act in a truly responsible way? Not at all; if premise (1) of the Basic Argument is rendered as “one’s reasons *play a crucial role* in determining one’s actions,” then the move to premise (2), “thus, to be truly responsible for one’s actions, one must be self-determined in one’s reasons for performing that action,” seems a bit forced. If (2) is supposed to follow from (1) alone, the motivation for that move should be obvious. But it just is not clear what motivates that move on this non-causal rendition of the Basic Argument, nor is it intuitive (to me at least) that one’s reasons “playing a crucial role” in one’s actions precludes oneself from being the determiner of those actions in some significant sense. If an agent can still be a significant determiner of her actions, even with premise (1) being true, then is it not clear that *true* moral responsibility should also require the agent to be self-determined in her reasons. In

---

<sup>10</sup> Strawson offers this non-causal definition “for the sake of those who are unhappy about the fact that reasons can truly be said to cause actions” (1986: 37). The condescending tone is duly noted.

fact, given the vagueness of Strawson's non-causal definition of "determine," one could easily assert that Nina *does* self-determine her actions. If it is at all appropriate to say that Nina, whatever else Nina is, is the possessor of those reasons and/or the doer of those actions that spring from those reasons, then it seems Nina undoubtedly plays a crucial role in those actions coming about—the actions would not have taken place without her! So long as "determine" just means "play a crucial role," it is not clear why Nina needs to determine her reasons in order to determine her actions.

Another way to look at it: Strawson allows for a definition of "determine" that does not require a causal connection between *X* and *Y* in order for it to be the case that *X* determines *Y*. In turn, Strawson cannot (consistently, at least) claim that Nina needs to *cause* her actions or her reasons in order to be self-determining in those actions or reasons. Surely Nina's very existence "plays a crucial role" in her actions and her reasons coming about, so if "determine" just means "plays a crucial role," I do not see how Nina could *fail* to be self-determining. If Strawson feels that Nina's "playing a crucial role" in her actions is not enough for her to be self-determining, then I am baffled that he should regard Nina's *reasons* "playing a crucial role" in her actions as sufficient for those reasons *determining* those actions, especially to the extent that Nina can no longer take any genuine moral responsibility for what she does.

Strawson's non-causal definition of "fully determine" wreaks equal havoc on the Basic Argument. As noted, Strawson states that an agent's reasons *fully* determine her actions (rationally speaking) so long as those reasons provide a true and full rational explanation of those actions. In other words, an agent's reasons fully determine her actions if, once those reasons are cited, no further explanation for the action is necessary, at least from a rational perspective. If it

---

<sup>11</sup> I have admittedly trimmed down Strawson's non-causal definition of "determine," but I do not believe this affects anything other than making it easier to read. If one wishes, one may substitute Strawson's full non-causal definition

is accurate to say that Nina turned off her neighbor's car headlights because she (a) believed they were on, (b) believed she could turn them off, (c) believed it would be the polite thing to do, and (d) desired to be polite, then assuming there is no other reason (e) that Nina had for turning off the headlights, the reason set comprised of (a)-(d) can be said to have *fully determined* that Nina turned off her neighbor's headlights. But again, it is unclear how the fact that (a)-(d) provides a full rational account of Nina's action would *ipso facto* prohibit her from being truly responsible for performing the action *unless* she is also responsible for having (a)-(d) in the first place.

Suppose someone believes (erroneously, on Strawson's account) that Nina is truly responsible for her actions even though she did not choose to have the desires and beliefs that make up (a)-(d). Now imagine Strawson were to spring from the shadows and confront this person, saying, "Think again, my friend! For Nina's beliefs and desires (a)-(d) provide a *full rational explanation* of Nina's actions!" Strawson's surprise entrance aside, I would expect the confronted person only to be baffled. I would *not* expect this person, in virtue of this new information from Strawson, to suddenly doubt Nina's responsibility. Yet this is the precise kind of thinking on which the first move of the Basic Argument relies—that because one's reasons provide a full rational explanation of one's rational actions, one can only be responsible for those actions if one is self-determined in one's reasons for performing those actions. Thus, altering premise (1) of the Basic Argument to read "one's actions are truly and fully rationally explained by one's reasons" also makes the jump from (1) to (2) questionable.

At this point, Strawson and his supporters are likely to jump to the problem of self-determination. Nina did not self-determine her reasons, nor could she have, for to self-determine one's reasons requires an infinite regress of self-determined choices to have the reasons one has for choosing to have the reasons one has, and this is why she cannot be truly responsible for her

---

of "determine" for any instance where I have abbreviated it.

actions. Such is the primary concern of the Basic Argument as flushed out in chapter 2. Yet the immediate point being discussed is that it is unclear why Strawson would make the self-determination of reasons a necessary condition for true moral responsibility if the causal link between one's reasons and one's actions is not assumed. The connection between premises (1) and (2) of the Basic Argument cannot be explained by appealing to the circularity inherent in premises (3) and (4), so to bypass (1) and go directly into problems of infinite regress is just to assume the truth of (2), which is arguably a contentious premise. Clearly Strawson does not view premise (1) as superfluous to his argument, but the support it offers (2) appears incredibly weak for those who do not share Strawson's belief that reasons are causal. Unfortunately for Strawson, he does not invest more time into defending this move, regarding it (mistakenly) as an unquestionably intuitive move to make.<sup>12</sup>

It could be that Strawson is leaving the most crucial premise of his argument unstated. Perhaps the first premise of the Basic Argument should be something like the following: "If one is to be truly morally responsible for one's actions, one's reasons must cause those actions." This premise would prohibit any non-causal account of reasons from providing a satisfactory response to the Basic Argument. But nothing about the Basic Argument suggests that this premise should be presumed. Strawson does not introduce the premise, of course, but neither is it evident that he takes the premise for granted. In fact, if Strawson does take this premise for granted, it is odd that he should introduce non-causal definitions of "determine" so as to make the Basic Argument tenable for those who reject causal accounts of reasons. Such a step would

---

<sup>12</sup> That Strawson finds the move from (1) to (2) intellectually obvious is evidenced by quotations such as the following: "If one is to be truly responsible for one's actions, then, *clearly*, one must be truly self-determining or truly self-determined in one's actions," and, "Such self-determination may seem evidently impossible. But it can also seem to be *clearly necessary* if one is indeed to be truly responsible for one's actions" (1986: 26-27, emphases mine). These quotes come early in Strawson's explication of his argument and, sadly, he fails ever to explain what makes these connections so "clear."

seem unnecessary if the suggested premise were meant to be part of Strawson's argument, since the premise would already rule out true moral responsibility on any non-causal account of reasons that could be offered. Thus, I reject the idea that Strawson, who has exerted much effort to ensure that his argument is amenable to all, intends either of himself or of his audience for this premise to be presumed.

Evidentially, Strawson goes beyond the mere favoring of a causal definition of "determine" and relies on it to make his argument persuasive. Ultimately, it appears to be the causal role of reasons that is meant to prohibit the self-determination of actions. Our actions are caused by our reasons, so we cannot be self-determined in our actions unless we can somehow self-determine our reasons (which we cannot since this would require a self-determined action in-and-of itself), at least according to Strawson. In the following section of chapter 3, I will take issue with the causal assumptions underlying the Basic Argument. I will argue that sufficient evidence exists for resisting the claim that reasons are causal. Because a non-causal account of reasons will not hinder one's reasons from properly explaining one's actions, there will be no need to posit a causal link between one's reasons and one's actions. The Basic Argument will thus have been shown to rest on a faulty assumption.

### **3.2 Reasons as Non-Causal Explanations – Ginet's Approach**

It is clear that Strawson's argument relies upon the idea that reasons are causal. This reliance is not only implicit in several key passages arguing for the impossibility of true moral responsibility, but it has been shown that the force of Strawson's argument wanes without such a causal assumption in place. Not every philosopher takes it for granted that reasons are causal, however. In defense of indeterminism, Carl Ginet (1995) argues for an account of reasons

explanations that dismisses the notion of reasons as causally efficacious. Ginet's primary motive is to defend indeterminism against the claim that any coherent reasons explanation *must* presuppose the causality of reasons. While I do not wish to argue against determinism in general, Ginet's theory will serve to discredit what Strawson refers to as "reasons/actions determination" (1986: 44-45). With a tenable non-causal account of reasons explanations in place, the impetus of Strawson's argument will founder.

Ginet purports to give what he calls "an *anomic* [i.e. lawless] sufficient condition for a reasons explanation" (81). For Ginet, reasons explanations are meant to capture the motivation behind one's actions, the end or ends that an agent intentionally aims to satisfy by performing some particular action. That is it. Causal claims are unnecessary. According to Ginet, "The only thing *required* for the truth of a reasons explanation of this [anomic] sort, besides the occurrence of the explained action, is that the action have been *accompanied* by an intention with the right sort of content" (ibid.). The accuracy of a reasons explanation does not hinge on the existence of causal laws, then, but rather on the intentional stance of the agent at the time she acts.<sup>13</sup> In other words, if we say that Nina turned off her neighbor's car headlights because she thought it would be nice to turn off the headlights, she believed it was within her means to do so, and she wanted to be nice, then we have fairly and accurately given a reasons explanation for Nina's actions so long as these descriptions match Nina's thoughts as she turns off the headlights. According to Ginet, we need not—and indeed *should not*—go one step further and claim that those reasons *caused* Nina to turn off the headlights. Because Nina's actions are justifiably explained by her reasons without bringing in some kind of causal law requiring that

---

<sup>13</sup> Indeed, it is highly unlikely that most people accept reasons explanations only on the assumption that such causal laws exist. Ginet makes a similar point when he says that an agent need not believe in any such causal law in order to hold the belief that, if a certain condition holds (e.g. the agent's friend does not call by noon), the agent will definitely perform a certain action (e.g. the agent will call the friend) (2002: 400).

she choose the way she does, there is no reason to postulate such a causal law other than to satisfy a pre-existing bias that such a law must exist.

Still, some will insist that a causal link is implicit in the fact that Nina clearly decides to do what she *prefers* to do. Nina's reasons for turning off the headlights are her *strongest* reasons in the situation, as evidenced by the fact that Nina chooses to satisfy those reasons and not any others. If they are her strongest reasons, then they will undoubtedly win out and *cause* Nina to choose as she does. This fact may be obscured when we ask only about the action Nina *does* perform and the reasons behind *it*, but the causal link between reasons and actions becomes obvious once we ask why Nina chooses one course of action *over another*. That is, we must ask not "Why does Nina turn off the headlights?" but "Why does Nina turn off the headlights *rather than* notify her neighbor or ignore the situation?" if we are to bring these causal implications to light. If Nina gives all of these actions consideration and yet chooses to turn off the headlights, then it is clear that Nina's reasons for turning off the headlights outweigh any reasons to do otherwise. The weight of Nina's various reasons is what causes Nina's actions. And so the objection goes.

There is a problem with this objection that becomes apparent when we ask how it is assessed that one set of reasons is "stronger" than another. In hindsight, of course, the assessment is quite easy. If we know that an agent wasn't coerced, that she didn't consider herself to be acting against her will, then we can safely conclude that whatever reasons she acted upon were her strongest. But is this helpful? It seems a rather *ad hoc* theory for determining which of an agent's beliefs and desires are stronger than others. Indeed, there is an inherent circularity to such thinking: Nina turned off the headlights herself because her reasons in favor of doing so were stronger than her reasons for not doing so, and we know they were stronger

because they are reasons on which Nina chose to act. Such an explanation already assumes that one set of reasons *had to be* determinately stronger, and that those stronger reasons, in virtue of being the strongest, would cause the agent to act as she did. As Ginet puts it, we can safely assume that “the strongest motive [i.e. set of reasons] prevails ... only by making it true by definition: ‘the strongest motive’ *means* the motive that prevails” (77). Tautologous definitions of this sort, however, fail to establish a causal connection between reasons and actions.

There may be something else that the objector means when speaking of some reasons being stronger than others, something that really does demonstrate a causal link between one’s reasons and one’s actions. I leave it to the objector to provide that definition. Experience hardly suggests that one type of reason (e.g. pragmatic concerns) always trumps another (e.g. physical gratification). Granted, people report being overcome by their desires to such a degree that, by their own admissions, they could not have done otherwise. But this hardly speaks to all situations, especially to the kind of everyday circumstances with which we are currently concerned. The notion of willpower exists because agents sometimes do what they *do not* feel they want to do most—for example, the dieting man who passes up his favorite dessert. Likewise, weakness of the will can be said to occur when an agent does what she feels *is not* in her best interest—the student who goes to a party the night before her final exams rather than studying, for instance. Keeping these things in mind, the definition of “strongest reasons” becomes elusive if we prohibit references to what an agent *already did* or *will in fact do*. What other properties does a “strongest reason” possess, other than being a reason on which an agent will or did act? The silence is telling.

The causal efficacy of reasons is already looking quite suspect, but there is another reason to be skeptical. Looking again at Nina, there is no obvious way in which Nina’s beliefs or

desires change between the time she begins deliberating about how to handle the headlight situation and the time she decides to turn off the headlights herself. If one's reasons cause one's actions (or, more modestly, one's decisions to act), then it seems an agent who has the beliefs and desires necessary for bringing about a certain action would be moved into action as soon as those reasons come to mind. Put another way, if Nina's beliefs and desires (a)-(d) *cause* her decision to turn off the headlights, then why does she not make that decision immediately upon realizing (a)-(d)? It seems something must explain the delay between Nina's having the sufficient reasons for acting and her actually acting on those reasons, but this is just to cast doubt on those reasons being causal.

Ginet picks up on this very point (2002: 394-5). Using the example of a person deciding to turn on his turn signal to indicate an upcoming turn, Ginet points out that an agent might not act at the very moment the relevant reasons for doing so are in place. Though the driver notices that his turn is coming up, he may not signal immediately, and yet this does not entail that the driver will have different reasons for signaling at the time he turns on the signal than he would have had had he signaled earlier. As Ginet explains,

[The driver] might have been aware that his turn was coming up, and have been intending to signal before getting there, for some time before he actually signaled; and he might have signaled at any of many different moments between that moment and the moment his turn came up, and the reasons explanation of his raising his arm have been the same, namely, that he wanted thereby to let others know that he would be turning at the next intersection and believed that raising his arm would do that. (395)

Ginet's case differs from mine in that the driver is not deliberating about anything—he already knows that he is going to signal, it is just a matter of *when*. Even so, the take-away message is

the same: the fact that an agent need not act on her reasons the moment she has them is a strong indicator that one's reasons do not cause one's actions. Moreover, the fact that an agent's reasons are not causal does not invalidate the reasons explanation that cites those reasons. In the driver case, just as with Nina, the reasons explanation given does not derive its intelligibility from the assumption of causal laws.

Of course, it must be recognized that nothing in this chapter has proven that reasons are *not* causal. But it does appear that we have good reasons for suspending the belief that reasons *are* causal, and as such, we should refrain from endorsing any theory that is coherent only if we take the causal efficacy of reasons for granted. I argued in 3.1 that the Basic Argument derives plausibility from doing just that—presuming that one's reasons *cause* one's actions. Though Strawson attempts to give a non-causal account of the Basic Argument, it was shown that the non-causal definitions of “determine” he offers fail to get his argument off the ground. It has now also been shown that there is strong evidence suggesting that reasons are *not* causal. As such, Strawson's theory has been significantly weakened, relying on a move from premise (1) to (2) that makes intuitive sense only if reasons are causally efficacious, which should not be presumed.

In an effort to preserve Strawson's argument, many will be prepared to accept premise (2), that an agent is truly responsible for her actions only if she is self-determined in her reasons for performing that action, without any supporting argumentation whatsoever. If this is taken for granted, then the rest of the Basic Argument still holds and the problem of infinite regress reemerges. I personally remain unconvinced of (2), but I am also fortunate enough to believe that infinite regress is not the threat Strawson suggests. In chapter 4, I will explain why, arguing that in many cases, an agent's actions are explainable by reasons that were chosen by the agent

but that cannot be traced back in their entirety to still earlier, non-chosen reasons. The result is that an infinite regress never takes place.

## 4 The Freedom of Forfeiture

In chapter 1, I asserted that Strawson's position could be countered by focusing on an agent's decision *not* to take some particular course of action. I call such decisions *action-forfeitures*, and I am now prepared to give an account of how action-forfeitures can avoid the threat of infinite regress posed by the Basic Argument. However, before I begin, I would like to forestall a handful of possible objections to my theory by more closely examining the nature of action-forfeitures. There are two main points about action-forfeitures that I would like to make: (1) action-forfeitures are actions in and of themselves, and thus just as susceptible to moral scrutiny as any other action, and (2) at least some action-forfeitures are unique choices that an agent makes, not identical to any other choice the agent makes.<sup>14</sup>

The first two sections of this chapter will be devoted, respectively, to the two features of action-forfeitures that I have just mentioned. The third section of this chapter will show how action-forfeitures allow us to avoid the problem of infinite regress. The final section of this chapter will then re-examine the notion of self-determination presented by Strawson and argue that, with causal assumptions and worries about infinite regress effectively put to the side, self-determination appears to be a viable possibility.

### 4.1 Action-Forfeitures as Actions

In the previous chapter, there was some discussion of why an agent would choose to do one thing *rather than* another. The very nature of choice<sup>15</sup> demands that there be, or at least seem to be, two or more possible courses of action available to an agent. Hence, every time an

---

<sup>14</sup> It may not prove significant whether we or not we regard action-forfeitures as *actions* per se, so long as we recognize that action-forfeitures are eligible for moral evaluation. However, I follow Strawson when he says that he will "take it to be obvious that choices and decisions can be actions" (1986: 134n2).

<sup>15</sup> For this thesis, I will assume that choices must be consciously made in order to count as choices.

agent makes a choice, she will consciously be choosing one thing *rather than* another. This is true even in situations where the only two options under consideration are (1) performing a certain action *A*, and (2) *abstaining* from performing *A*. To give a rudimentary example, suppose you are driving somewhere in your car and your gas light comes on, indicating that you soon need to refuel. If your destination is not far off, you may consider postponing the filling up of your car's gas tank until after you have completed your business at said destination. In such a situation, you may regard yourself as deliberating over just one thing—namely, stopping for gas—but it is clear that two genuine courses of action appear available to you. You can either stop for gas, or you can continue on to your destination uninterrupted. Because either option entails that the other does not occur, making a choice will require not only a commitment to one course of action, but a “surrendering” of the other. That is, you must “give up” or “forfeit” one possible course of action in order to accomplish the other. If you decide to stop and refuel, you forfeit the option of continuing to your destination uninterrupted, and vice versa.

Whenever an agent consciously and willingly forfeits a certain course of action she believes it is possible for her to take, this may be considered an *action-forfeiture*. In contrast, I will refer to those courses of action that an agent consciously and willingly decides to take (i.e. the opposite of an action-forfeiture) as *action-commitments*. This is to prevent the ambiguity of the more neutral term *choice*. As I see it, we often refer to the deliberating agent who comes to a decision as making a *single* choice, when in actuality the agent typically commits to one or more courses of action while forfeiting other courses of action at precisely one and the same moment—that is, the agent often makes action-commitments and action-forfeitures *simultaneously*. I will explain my reasons for regarding concurrent action-commitments and action-forfeitures as distinct choices in 4.2. For now, it is enough to acknowledge that action-

forfeitures do occur, that an agent does consciously forfeit one or more courses of action whenever she decides between two or more options.

One question that may be asked is whether action-forfeitures should count as actions at all. Lest someone think that action-forfeitures are special cases that Strawson would reject, let me point out that Strawson himself speaks of agents deciding *not* to do things as if they were any other decisions an agent could make. A case in point is when Strawson discusses an agent's decision *not* to eat (1986: 45-46). Trying to combat the notion that an agent must act freely when he chooses to eat because, after all, the agent could have chosen *not* to eat, Strawson contends that a decision *not* to eat will face the same exact difficulties in satisfying the requirements for self-determination that choosing to eat would face:

If [the agent] had chosen not to eat, it would have been because he had *reason* not to despite his hunger and desire to eat. But then we may ask where and how self-determination and true responsibility are supposed to enter the picture, unless [the agent] is somehow responsible for his reasons. It seems clear that what [the agent] does when he acts intentionally is, ultimately, always and necessarily just some more or less complex function of his reasons; and that he cannot therefore be truly self-determining unless he can somehow be self-determining with respect to his reasons [...]. (ibid.)

Notice that Strawson says the agent “acts intentionally,” though the matter at hand is an agent's deciding *not* to eat. Clearly, then, Strawson puts what I have called action-forfeitures and action-commitments on equal ground, in part because action-forfeitures are genuine choices that an agent can make, and in part because action-forfeitures are explainable by reference to an agent's reasons, just like action-commitments.

Classifying action-forfeitures as actions makes even more sense when we realize that action-forfeitures are just as eligible for moral praise and blame as are action-commitments. Consider the man who idly stands by as a child drowns. If it is within reason to suppose that this man could have done something to help the child, the man's passivity is bound to be censured.<sup>16</sup> Likewise, consider the smoker who desperately wants to quit and, in a very concentrated effort to break his habit, refrains from smoking. Undoubtedly, some would find it morally commendable that the smoker withstands his nicotine cravings, even though this praise would stem from the smoker doing "nothing." In both the reluctant smoker and the drowning child cases, the praise or blame we ascribe is likely to be heightened by, if not entirely dependent upon, the fact that the agent in question makes a conscious decision *not* to do something, in one case to smoke a cigarette, in the other case to help an endangered child.

Because action-forfeitures have the feature of being conscious, morally-evaluative choices that an agent makes on the basis of reasons, and because Strawson himself accepts action-forfeitures as actions, I will consider it uncontroversial to regard action-forfeitures as actions while responding to the Basic Argument.

#### **4.2 Action-Forfeitures as Distinct Choices**

So far I have been speaking of action-forfeitures as distinct choices that an agent makes, almost as if they are independent of the agent's decision to pursue some other course of action *instead of* the action being forfeited. As a matter of fact, I do believe that many action-

---

<sup>16</sup> In some cases, the difference between an action-forfeiture and an action-commitment is admittedly slight. We could say that the man makes an action-commitment to stand still, but I believe it depends largely on the man's own thought process. If, in the man's own mind, he is focusing on standing there rather than on the fact that he is not going to help the child, then he may very well be making an action-commitment. Given the situation, I find it hard to believe the man's attention would be on his standing in place so much as on the child. The point is, if the man willingly *and consciously* does not help the child though he believes he could do so, then he is forfeiting that action.

forfeitures can be regarded as distinct choices,<sup>17</sup> distinguishable from action-commitments though the two often occur at the very same instant. Some may not be prepared to accept this claim, however, and so I will here attempt to explicate my position.

As I see it, an action-forfeiture *F* can be considered a distinct choice—that is, distinct from some choice *C*—if the following three conditions are met: (1) *F* results from an intentional decision that the agent makes, (2) the agent makes that decision with the belief that *F* will result from that decision, and (3) it is possible for the agent to renege on *F* (i.e. reconsider and possibly even commit to the action that was forfeited) without affecting the agent’s commitment to *C*. Conditions (1) and (2) do little other than state that the action-forfeiture is indeed a matter of choice, not just an unforeseen consequence of the agent’s decision. What (1) and (2) do not show is that the action-forfeiture in question is a *distinct* choice from *C*, where *C* may very well be an action-commitment concurrent with *F*. It is condition (3) that captures this distinction, showing that the action-forfeiture *F* remains (psychologically, at least) independent of the action-commitment *C*.

Let us test out my theory by way of example. Consider Ronny, a businessman who is about to take a lunch break. On his way out of the office, he deliberates about what to get for lunch. He thinks of the sandwich shop around the corner, a favorite of his, and considers it. He then remembers that a new Chinese eatery about which he has heard very good things has opened up adjacent to the sandwich shop. After a moment’s deliberation, Ronny decides to have Chinese and gives up the idea of eating a sandwich. The result is that Ronny now has an action-

---

It may very well be that he is *also* making an action-commitment to stand in place, but it will depend on his psychological state at the time.

<sup>17</sup> I will stick to the term “distinct,” primarily because action-forfeitures *do* so often appear at the same time as action-commitments and because, as I will argue in 4.3, those concurrent action-commitments may play an important explanatory role in the action-forfeitures themselves. If not for these considerations, I would be comfortable referring to action-forfeitures as “independent” choices.

commitment to eat Chinese food and has made an action-forfeiture of eating a sandwich. I contend that these constitute two distinct choices.

It is clear that both Ronny's action-commitment and Ronny's action-forfeiture satisfy conditions (1) and (2) above. Both the action-commitment and the action-forfeiture were the result of a decision that Ronny intentionally made, and Ronny was well aware of these consequences as he made his decision. However, condition (3) is also met by each choice individually. If the action-commitment and the action-forfeiture were really the same exact choice, then it seems they would have to stand or fall together. That is, it would be impossible for Ronny to, say, maintain his sandwich-eating action-forfeiture and yet reconsider his action-commitment to eating Chinese, for that would mean that Ronny both retains and does not retain his choice at the exact same instant.

As it turns out, the status of Ronny's action-commitment and of his action-forfeiture *can* change independently of one another. It could be that after Ronny commits to the idea of getting Chinese but just before he places his order, he decides he may very well get a sandwich too. Likewise, Ronny could get a sudden craving for pizza before he orders Chinese, and he might then deliberate between Chinese food and pizza while never reconsidering sandwiches. In either instance, it seems correct to say that *one* of Ronny's choices remains intact, even though the other is reevaluated. Ronny does not have to forfeit sandwiches all over again if he gives up his commitment to Chinese, and he does not have to reconsider Chinese food even if he changes his mind about not getting a sandwich. These facts hold true because action-forfeitures and action-commitments, despite so often appearing together, can remain distinct choices.<sup>18</sup>

---

<sup>18</sup> If it is helpful, those who remain unconvinced may regard action-forfeitures and action-commitments as distinct *endorsements*, as it were. That is, Ronny *endorses* or *sanctions* both the idea of not getting a sandwich and the idea of eating Chinese food. This does not strike me as a problematic way of viewing the situation, for it seems clear enough that Ronny can regard these as two distinct consequences, even if they are, under the circumstances,

### 4.3 Action-Forfeitures and the Evasion of Infinite Regress

This chapter has thus far demonstrated that action-forfeitures can be morally-evaluative actions that are not identical to any action-commitments with which they may simultaneously appear. All that remains to bring my argument to fruition is to show that action-forfeitures can provide a tenable account of an action that is not susceptible to infinite regress. This section of chapter 4 will be comprised of just such an account.

As noted in 4.1, Strawson believes an agent can act based on reasons even when she decides *not* to do something. This is important because, going back to chapter 2, Strawson says that *rational* actions—actions for which an accurate reasons explanation can be given—provide the locus for questions about free will and moral responsibility. By Strawson’s own admission, then, action-forfeitures are rational actions, and accordingly, they are eligible for considerations of true moral responsibility.

That action-forfeitures can be rationally-performed actions is obvious enough. The smoker who consciously chooses *not* to smoke a cigarette will be able to point to his reasons for not doing so. He desires to live a long, healthy, and productive life and he believes abstaining from cigarettes will help him to achieve this, for example. The man who decides *not* to stop for gas the moment his gas light comes on will have reasons for not doing so. Perhaps he believes he will make it to his destination without stopping, he believes not stopping will allow him to get to his destination sooner, and he desires to get to his destination as quickly as possible. Nina, the shy woman who discovered that her neighbor’s headlights were on, will also have reasons for *not* knocking on her neighbor’s door to notify him. She believes it would be an awkward situation,

---

mutually entailing. I believe my entire thesis may be rendered in terms of *endorsements* rather than *choices*, if my reader should find it beneficial.

and she believes she can satisfy her desire to solve the headlight problem without knocking on his door.

In many cases of action-forfeiture, however, there is a reason for the action-forfeiture that is unlike any of the reasons cited above. In many cases, a concurrent action-commitment will itself be an important reason—perhaps even the only reason—for the agent's action-forfeiture. That is, the agent makes an action-forfeiture in part or entirely *because* of the action-commitment. This is especially true in cases where an agent is torn between two or more equally appealing courses of action. The agent reaches a decision and forfeits some course of action, but the reason for doing so will consist primarily of the agent's commitment to some other course of action. As a reminder, this does not mean the action-commitment and the action-forfeiture really are one and the same choice. Rather, one provides a reason for the other, even though the two come about simultaneously.

It is important to recognize action-commitments as providing possible reasons explanations for action-forfeitures because it allows us to see how action-forfeitures escape the threat of infinite regress posed by the Basic Argument. This is vital to my argument. Recall that, for Strawson, an action must be both self-determined *and* based on reasons that are, at least in part, self-determined in order for an agent to be truly morally responsible for that action. In chapter 3, it was determined that a primary reason Strawson has for withholding self-determination from an agent's actions is because he presupposes that one's reasons *cause* one's actions. This was shown to be questionable, so the infinite regress now appears to be the only real threat to true moral responsibility that Strawson has offered. It follows that, if an agent's action-forfeiture is done for reasons that are (at least in part) not subject to the problem of infinite regress, then the agent may very well be truly morally responsible for that action-

forfeiture. I assert that, with at least some action-forfeitures, it will be the case that a concurrent action-commitment provides a reason—a very significant and non-regressive reason—for that action-forfeiture.

Recall Ronny, the businessman who deliberates between Chinese food and sandwiches for lunch. Let us change the example slightly and say that Ronny was quite torn about which food to get. On the one hand, he loves sandwiches and wholeheartedly believes he will enjoy his lunch if that is what he decides to get. On the other hand, he wants to try the new Chinese place and trusts that the terrific reviews it has been receiving are indicative of its culinary quality. He believes either option will be equally satisfying. He opts for the Chinese food.

Question: why did Ronny forfeit sandwiches? Answer: he decided to get Chinese food. That seems to be the primary reason, at least, given that Ronny believes Chinese food and sandwiches would be equally satisfying. We might try to say that Ronny *must have* wanted Chinese food *more* than he wanted sandwiches, since he ultimately decided on Chinese food, but this is akin to saying that Ronny's reasons in favor of eating Chinese food were "stronger than" his reasons in favor of getting a sandwich. I argued in chapter 3 that such an explanation is unacceptable. Additionally, Ronny himself needn't believe he wanted Chinese food more than he wanted sandwiches. If we were to ask Ronny, on the basis of evidence, "So, you wanted Chinese food more than you wanted a sandwich?" it is not absurd to think he would answer, "Not really. They both sounded great." Even though I think it *can* be correct to say, in certain situations, that an agent's desiring one course of action more than another explains why the agent commits and forfeits as she does, nothing demands that this is always the case. Sometimes the actions that are forfeited will be desired just as much the actions to which the agent commits.

We then might ask Ronny, “So why did you get *Chinese food*?” and Ronny would reply, “Because it sounded good. I’d heard good things about it. I wanted to try it.” The temptation at this point is to transfer these reasons for Ronny’s getting Chinese food to his sandwich-eating action-forfeiture, assuming that the same set of reasons can explain both choices. To some extent this might work, for it may very well be that those reasons provide *some* of the reasons for Ronny’s not getting a sandwich, if Ronny holds them as such.<sup>19</sup> But they will not provide a *complete* rational explanation for the forfeiture. What must also be included, at the very least, is the fact that Ronny *chose to eat Chinese*. Without this additional reason in place, the rational explanation of Ronny’s sandwich-eating action-forfeiture is lacking. Its absence would suggest that Ronny’s actually choosing to eat Chinese food is irrelevant to his giving up the sandwich, and that is absurd. If Ronny discovered that the Chinese restaurant did not accept credit cards and it just so happened that he did not have any cash on hand, for example, it would still be the case that Ronny thinks Chinese food sounds good, that he holds certain beliefs about the new Chinese restaurant’s quality based on the positive reviews he has heard, and that he wants to try the restaurant someday. But presumably those reasons would not be enough to explain why Ronny does not get a sandwich at *that* point. And that is because the action-commitment to eat Chinese, though itself largely explainable in terms of the Chinese restaurant’s appeal, is an essential part of the reasons explanation for Ronny’s sandwich-eating action-forfeiture and cannot be ignored.

---

<sup>19</sup> Strawson says that a full, rational explanation—one that can be said to *fully determine* an agent’s actions—will be one based on reasons the agent actually has at the time of action, not just reasons that it is possible for an agent to have and that could make the agent’s actions intelligible to someone else (1986: 34n12, 35-37). For example, the health benefits of vitamin C could very well be considered a reason for eating oranges, but it is not true that every person who eats an orange does so for, or is even aware of, this reason. For these agents, the health benefits of vitamin C will not be a part of their reasons explanation for eating oranges.

On the flipside, Ronny's action-commitment to eat Chinese food cannot wholly be explained without reference to his sandwich-eating action-forfeiture. That is, a crucial reason that Ronny commits to eating Chinese food is that he has forfeited the only other food that tempts him—sandwiches. Think of it in terms of beliefs and desires, which are what Strawson generally considers reasons to be. Ronny believes he will not eat a sandwich for lunch. Why not? In part because he believes he will eat Chinese food for lunch. But the opposite is equally true. That is, why does Ronny believe he will eat Chinese food for lunch? Because he now believes he *is not* going to eat a sandwich. While it is true that Ronny would not likely cite his sandwich-eating action-forfeiture if someone were to ask him why he ordered Chinese food, this is only because it would already be understood that Ronny had forfeited everything else under consideration. There just isn't a need to bring up the action-forfeiture in most situations! This is not because Ronny's sandwich-eating action-forfeiture fails to factor into Ronny's action-commitment to eating Chinese food, but because it simply goes without saying, once we know Ronny *has* committed to eating Chinese food, that he has forfeited anything else he had been considering.

I suspect that some of my readers will accept that an action-commitment can explain an action-forfeiture, but not vice versa. Admittedly, claiming that an action-forfeiture can be part of one's reasons for an action-commitment is somewhat counterintuitive. Some will accept only that the forfeiture is, under the circumstances, logically necessary for the commitment. But I propose that some mental state *X* can be regarded as a reason for some action-commitment *C*, if it is true that the agent performs *C* with *X* in mind and, *prior* to committing to *C*, the agent would accept (if the agent were to think about it) the conditional statement, "If I commit to *C*, it will be because *X*." This schema seems to account for Ronny's other reasons for eating Chinese food.

That is, during deliberation, Ronny would presumably accept the statement, “If I commit to eating Chinese food, it will be because Chinese food tastes good, because I believe the good reviews I’ve heard about the new Chinese restaurant nearby,” and so on. But he would also accept the statement, “If I commit to eating Chinese food, it will because I forfeit getting sandwiches for lunch.” As I see it, so long as  $X$  can be regarded as a mental state, whenever  $X$  fits this mold in relation to some action-commitment  $C$ ,  $X$  can be properly regarded as a reason for  $C$ .<sup>20</sup>

At this point, it can be shown that Ronny’s sandwich-eating action-forfeiture (and Chinese-food-eating commitment, for that matter) evades the problem of infinite regress inherent in the Basic Argument. One of the primary reasons that make up Ronny’s rational explanation for not getting a sandwich is Ronny’s action-commitment to eat Chinese food. In order for the regress to start, then, the rational explanation of Ronny’s action-commitment to eat Chinese food would have to be traceable, in its entirety, to reasons Ronny had prior to the action-commitment itself. This would not be enough for an infinite regress, of course, but it is how such a regress would begin. As it turns out, many of the reasons Ronny has for eating Chinese food *do* precede his action-commitment, such as Ronny’s belief that the Chinese food will taste good. But not all of Ronny’s reasons for committing to eat Chinese food appear prior to the commitment itself. Ronny’s action-commitment is also explained by Ronny’s sandwich-eating action-forfeiture. Ronny was not committed to eating Chinese food until he gave up the idea of eating sandwiches, but he did not give up the idea of eating sandwiches until he committed to eating Chinese food. These two distinct choices appear simultaneously, and they both factor into the rational

---

<sup>20</sup> That  $X$  must be a mental state, something like a belief or a desire, is of extreme importance. This is because it will *not* be the case that one of Ronny’s reasons for committing to eat Chinese food is that his credit card does not get rejected, even though Ronny would, hypothetically, accept the conditional statement, “If I commit to eating Chinese

explanation of the other. Thus, a full rational explanation of Ronny's sandwich-eating action-forfeiture (and Chinese-food-eating action-commitment) includes reasons that came into being only at the moment the decision was made. There is no regress.

#### 4.4 Action-Forfeitures and Self-Determination

I have now argued that some cases of action-forfeitures and action-commitments can avoid the Basic Argument's threat of infinite regress. It has also been shown in chapter 3 that reasons should not be assumed to be causal. An obvious question at this point is, what of self-determination? Can it be achieved? Has Ronny achieved it? Given my understanding of Strawson's argument, I no longer see a reason to preclude self-determination from an agent in a situation like Ronny's. Premise (3) of the Basic Argument states, "To be self-determined in one's reasons, one must have consciously and explicitly chosen to have (at least some of) those reasons *and* be self-determined in having made the choice to have those reasons." It is the latter half of this requirement for self-determination that is meant to lead to infinite regress, since to be self-determined in one's choices, one must be self-determined in one's reasons for making those choices. Such was the gist of premise (4), which leads back to premise (3). But there is no regress with Ronny's choices to eat Chinese food and forfeit eating sandwiches, and these are both conscious and explicit choices Ronny makes. So what else must be accomplished for Ronny to achieve self-determination?

To answer this question, it may be helpful to modify premise (3) in such a way that it captures the concerns of premise (4). I believe this can be done by substituting (3\*) for (3):

---

food, it will be because my credit card does not get rejected," and even though Ronny might be aware of this fact at the time he eats his Chinese food.

- (3\*) To be self-determined in one's reasons, one must have consciously and explicitly chosen to have (at least some of) those reasons *and* it must be that this choice cannot wholly be explained by reference to reasons that were either (a) not consciously and explicitly chosen, or (b) traceable back to reasons that were not consciously and explicitly chosen.

This is a verbose premise, but it seems to capture Strawson's overall concerns about infinite regress in one fell swoop. It also makes it easier to see that Ronny has indeed satisfied the requirements for self-determining his reasons. Hence, on the basis of premise (2) of the Basic Argument ("To be truly responsible for one's actions, one must be self-determined in one's reasons for performing that action"), Ronny is truly responsible both for not getting a sandwich and for eating Chinese food. He has made both choices consciously and explicitly, and because each choice is part of the rational explanation for the other, not all of Ronny's reasons can be traced back to beliefs and desires that he had prior to making those choices.

Deciding what to eat for lunch is not ideal for demonstrating moral responsibility, of course, but it is now easy enough to see how other scenarios will pan out. Suppose some agent Mark has promised to take his nephew to the movies. Mark is an avid fan of the movies, and he really enjoys spending time with his nephew, so fulfilling this promise is not something Mark considers to be a burden. However, just before Mark is about to go pick up his nephew, his friends call and invite him to go with *them* to the movie. Mark would like to hang out with his friends, and though he knows he has promised to go with his nephew, he considers going with his friends instead. After a few moments of deliberation, Mark decides to go with his friends. Mark's nephew excitedly awaits the arrival of his uncle, who never shows up.

What are Mark's reasons for not going to the movie with his nephew? One primary reason is that he intends to go to the movie with his friends. But why does Mark intend to go to the movie with his friends? In part because he no longer plans to go to the movie with his nephew. Each distinct choice is explained, at least partially, by the other, preventing an infinite regress of reasons for why Mark has made the choices he has made. Though Mark has reasons both for going with his friends and for going with his nephew that exist prior to his making a decision, neither set of pre-existent reasons sufficiently explain his choices. Mark would not commit to going to the movie with his friends if he were still open to the idea of going with his nephew, and he would not give up on going with his nephew if he did not intend to go with his friends. Nothing about this scenario entails that Mark does not make these choices consciously and explicitly, however, and so it appears that Mark is truly morally responsible for going to the movies with his friends and for breaking his promise to his nephew.

## 5 Final Thoughts

The weaknesses of Strawson's Basic Argument have now been exploited. Strawson argues that, because one's reasons determine one's actions, one needs to be self-determined in one's reasons in order to be self-determined in the actions to which those reasons lead. In chapter 3, I argued that the tenability of such a claim is rooted in the notion that reasons are causal. I then gave support to the idea that reasons are *not* causal and argued that we should not presume that a causal connection between one's reasons and one's actions exist in light of this contrary evidence. Though the arguments in chapter 3 should have undermined the first move in Strawson's argument, I recognized that some would hold to Strawson's concerns about infinite regress even in the absence of the causal assumptions that were ultimately meant to instigate those concerns. Thus, in chapter 4, I provided evidence for Strawson's own belief that what I call action-forfeitures *are* actions, and I then argued that an agent's action-commitments can sometimes play a vital role in the reasons explanation of the agent's action-forfeiture (and vice versa). Because action-commitments and action-forfeitures rationally sustain each other, as it were, the rationally-based infinite regress inherent in Strawson's Basic Argument was circumvented. I concluded chapter 4 by contending that, in the absence of both reasons-causation and infinite regress, some agents appear to meet the criteria for self-determination, and hence for true moral responsibility, that Strawson puts forth.

Undoubtedly, there will be some objections to my account. The first part of this chapter will be dedicated to the entertaining and subsequent dispatching of some of these objections. I will then conclude with a final thought on what this theory leaves open for investigation and why it may be pointing us in the right direction.

## 5.1 Objections

### *Objection #1: True Moral Responsibility is Limited to Cases Where One is Equally Torn*

#### *Between Two or More Courses of Action*

In the case of Ronny which made up the bulk of my argument in chapter 4, I stated that Ronny was equally torn between having sandwiches for lunch and having Chinese food for lunch. Because he viewed the two choices as equally appealing, I said that nothing about one course of action would obviously explain why he gave up the other *without also* bringing in the fact that Ronny had actually committed to that other course of action. Thus, some might suppose that I am limiting true moral responsibility only to cases where one is equally torn between two or more courses of action.

In actuality, the only reason I chose to present a case wherein two courses of action were considered equally appealing was for philosophical convenience. That is, in the case of Ronny, for example, making his options equally appealing prevented Ronny's reasons for forfeiting sandwiches from getting "muddied." Suppose Ronny had believed that the Chinese restaurant was supposed to serve food far better tasting than his favorite sandwiches. If that were the case, then one might suppose that Ronny forfeited sandwiches *because* he believed Chinese food would taste better. This might be true, at least in part, and it's certainly true that when we speak to one another, we often cite only one or two primary reasons for what we do, such that Ronny might have casually listed his belief that Chinese food would taste better as *the* reason he did not get sandwiches.<sup>21</sup> But imagine how strange of a reason this would be if Ronny did *not* also intend to get Chinese food. Because Ronny is considering only sandwiches or Chinese food, a belief that Chinese food will taste better than a sandwich will only explain why Ronny forfeits

sandwiches *if he is already committed to eating Chinese*. It is possible of course for Ronny to believe Chinese food will taste better but not yet be committed to eating it (perhaps because it is more expensive, for example), but in such a case, it will be baffling for Ronny to forfeit the sandwiches without having committed to the only other food under consideration.

The bottom line is that it does not matter if situations exist wherein the reasons in favor of one course of action allow one to rule out another course of action without committing to that first course of action. Perhaps Ronny is considering Chinese food, sandwiches, and pizza, and perhaps he believes the pizza parlor will be the slowest to deliver his food. In such a situation, he may forfeit pizza without explicitly committing either to sandwiches or Chinese food. But the fact remains that, in situations where an action-forfeiture *does* explain an action-commitment and vice versa, infinite regress will be avoided. These need not be cases where two courses of action are considered equally appealing, so long as both courses of action are given consideration. Giving up one for the other, even if it seems like the obvious thing to do, may still allow the action-commitment and the action-forfeiture to partially explain each other. These are the types of cases where the Basic Argument loses efficacy and the possibility of true moral responsibility emerges.

### *Objection #2: This Theory Accommodates Only Libertarians*

Another possible objection is that my argument is inherently libertarian, such that it cannot benefit compatibilists. This objection (although I'm not sure it actually constitutes an objection) might be fueled by the fact that I utilize Ginet, himself a libertarian, for one of the main thrusts of my argument, namely that reasons are not causal. Another reason my argument

---

<sup>21</sup> Compare this to the person who explains why he opted for steak over lobster by saying, "The steak is cheaper." Presumably he did not order steak *merely* because it was cheaper than lobster. Presumably, it will also be the case

may be considered libertarian is that, when an agent freely performs an action-forfeiture, I am ultimately claiming that her reasons for doing so may appear *at the very same moment she performs it*. This in and of itself does not make my theory a libertarian one, but it may be construed as a bit too “magical” for some people’s liking. (Libertarianism is often accused of being inherently mysterious.)

I can understand these concerns, yet, like Strawson, I have tried to keep my argument “non-partisan,” such that it works whether determinism or indeterminism is true. Above all else, I have aimed for accuracy, and I truly believe that agents often forfeit one course of action in part because they are committed to another, that agents often commit to one course of action in part because they have forfeited any others, and that these decisions can happen simultaneously even though they genuinely explain each other. I see nothing about this that is inconsistent with determinism. What the objector may really be wondering is if a non-causal account of reasons is compatible with a deterministic universe. I resolutely answer *yes*. I believe that, if determinism is true, our actions are likely caused by the same (presumably physical) states and laws that cause us to believe we have certain reasons for performing those actions. This type of determinism does not mean that the reasons themselves cause the actions, of course. It just means that we are incredibly fortunate that, most of the time at least, our actions make sense to us in light of our reasons. An advantage of this type of determinism is that it could also explain those times when we *don’t* feel like our actions make sense even to ourselves, such as when we feel out of control, irrational, and/or simply weak-willed.

In the end, it does not matter if this theory of determinism is correct. What matters is that the theory shows one possible way in which a non-causal account of reasons is compatible with

---

that he thought steak would taste good, that he believed he could afford the steak, etc.

determinism, and hence one possible way in which my thesis can be regarded as a compatibilist one.

*Objection #3: This Theory Makes Action-Commitments and Action-Forfeitures Appear Random.*

Some might insist that, if Ronny (or Mark, or any other agent in a relevantly similar situation) cannot give a reasons explanation for why he opted for one course of action *over another* (and not just for why he opted for a particular course of action and, apart from that, why he forfeited some other course of action), his choice must be random, at least from a rational standpoint, which is what we have been concerned with thus far.

The only way I know how to respond to this type of objection is to ask for a more precise definition of “random.” Time and again, Strawson contrasts “random” with “determined,” suggesting that the two are mere opposites of one another (see, for instance, 1986: 8, 25-26). But, in his effort to appease one and all kinds of philosophers, he has given such varied definitions of “determined” that it makes the precise interpretation of “random” somewhat difficult to come by. On the one hand, Strawson clearly thinks of “determined” as meaning “caused,” so if “random” is just the opposite of “caused,” then uncaused actions are, by Strawson’s definition, random. But nothing in my theory relies on one’s actions being uncaused. I merely argue that *reasons* should not be *presumed* to cause one’s actions. If that alone is enough to make an agent’s actions random (as far as her reasons are concerned), so be it. Granting this to Strawson will not prevent the fact that an agent sometimes (a) consciously knows what she is choosing, (b) has reasons for choosing what she is choosing, (c) is aware of those reasons for choosing what she is choosing, and (d) believes she is choosing as she is

choosing because of those reasons. If Strawson wants to call this type of choice “random,” that is his prerogative. To me, it seems a poor word choice.

As it turns out, I do not think Strawson would, should, or consistently could consider these choices to be random. Strawson says an agent’s actions are (rationally speaking) *fully determined* if the agent has a full, rational explanation for that action. As has already been shown, Ronny’s and Mark’s choices *do* have full, rational explanations. If their actions are thereby *fully determined*, then surely they cannot be random if “random” is just the opposite of “determined.” Funny enough, it is Strawson’s account of how reasons *fully* determine one’s actions that most strongly indicates that Ronny’s and Mark’s (and Nina’s, etc.) actions are anything but random. Because these actions satisfy the requirements for being *fully* determined, I suppose they would also have to be fully *not* random. If Strawson himself has articulated a theory wherein these agents are not acting randomly, then I do not see how claiming these actions *are* random will help Strawson. Insisting that these actions are random will merely undermine Strawson as well.

## **5.2 Just the Beginning**

I have now spent several chapters arguing that freedom and moral responsibility may be linked to an agent’s conscious decision *not* to do something. The purpose of this thesis was to refute Galen Strawson’s Basic Argument, but it is my hope that I have accomplished more than that. Humbly, I hope that I have opened the door to a more comprehensive examination of agential freedom, one that looks not only at what an agent consciously commits herself to doing, but at what she knowingly gives up in the process. I believe there is something important to be discovered here. Consider the fact that what I have called action-forfeitures are seemingly

unique to mentally-mature human beings. We do not generally suppose that small children, those with certain mental impairments, and non-human animals are consciously aware of what they *do not* do. For instance, the lion pouncing on a gazelle is not typically presumed to be cognizant of the fact, as he pounces, that he has forfeited the option of attacking other gazelles. Interestingly, these groups are also those that are often denied the status of acting freely and morally responsibly. Something tells me this is not a coincidence, and that a great deal is to be discovered in looking at the twofold nature of choice that I have only begun to touch upon in this thesis. It is my hopes that we, as philosophers, will not forfeit the opportunity to study these issues further.

### Bibliography

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

Ginet, Carl. "Reasons Explanation of Action: An Incompatibilist Account." O'Connor 69-93.

---. "Reasons Explanations of Action: Causalist Versus Noncausalist Accounts." Kane 386-405.

Kane, Robert, ed. *The Oxford Handbook of Free Will*. New York: Oxford University Press, 2002.

O'Connor, Timothy, ed. *Agents, Causes, & Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press, 1995.

Strawson, Galen. "The Bounds of Freedom." Kane 441-60.

---. *Freedom and Belief*. New York: Oxford University Press, 1986.

---. "The Impossibility of Moral Responsibility." Watson 212-28.

---. "Libertarianism, Action, and Self-Determination." O'Connor 13-31.

Watson, Gary, ed. *Free Will*. New York: Oxford University Press, 2003.