

# G. Strawson and Aristotle on Moral Responsibility and Punishment

By

© 2020

Matthew Klados

B.Arts., Villanova University, 2014

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the  
University of Kansas in partial fulfillment of the requirements  
for the degree of Master of Arts.

---

Chair: Tom Tuozzo

---

Sarah Robins

---

Brad Cokelet

Date Defended: 27 July 2020

The thesis committee for Matthew Klados certifies that this is the approved version of the following thesis:

## G. Strawson and Aristotle on Moral Responsibility and Punishment

---

Chair

Date Approved: 27 July 2020

## Abstract

In this paper, I bring G. Strawson's famous 1994 paper "The Impossibility of Moral Responsibility" and Aristotle into conversation. I argue that his Basic Argument is best taken as targeting solely desert-based moral responsibility. I then present Aristotle's discussion of the causal role that humans play in their actions, with a focus on morally significant actions. I argue that, contrary to Strawson's speculation, Aristotle did not believe us to have the desert-based moral responsibility that the Basic Argument purports to prove impossible. Instead, Aristotle's causal account presents us as having *attributional* moral responsibility, and comments he makes elsewhere show him to believe we also have *accountability* moral responsibility. The paper concludes with some consequences of accepting both author's accounts, and also considers the question of why so many believe themselves to have desert-based responsibility, when a simple argument shows that to be impossible.

## Table of Contents

Chapter 1 .....	1
Chapter 2 .....	1
Chapter 3 .....	7
Chapter 4 .....	9
Chapter 5 .....	10
Chapter 6 .....	12
Chapter 7 .....	14
Chapter 8 .....	20
Chapter 9 .....	22
Chapter 10 .....	24
Chapter 11 .....	25

## Chapter 1

In his famous 1994 paper, Galen Strawson argues for the soundness of the Basic Argument, which purports to prove that no human being can be morally responsible in the way that many suppose. After presenting and analyzing the Basic Argument, I will present Aristotle's discussion on the causal role that people play in their morally significant actions. My reading of Strawson will show him to only be proving that people are not morally responsible for their actions in a desert-based sense. In addition, while Strawson says it is natural to assume that Aristotle believed people to have this desert-based moral responsibility, I will show that Aristotle could not have believed us to have it, and instead, that the kinds of moral responsibility that emerge from Aristotle's discussion is attributional and accountability. My final remarks will reflect on the consequences of accepting both accounts, and a possible explanation for why so many believe that we have strong moral responsibility, despite the fact that a simple argument shows it to be impossible.

## Chapter 2

Strawson presents the Basic Argument in a number of forms. He first states it in its most basic form:

1. "Nothing can be *causa sui* – nothing can be the cause of itself.
2. In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects
3. Therefore, nothing can be truly morally responsible."<sup>1</sup>

---

<sup>1</sup> Strawson, pg. 5

The first detailed version of the argument that he offers, which is the one I will be referring to in this paper, is as follows:

1. “Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to ‘reflex’ actions or mindlessly habitual actions).
2. When one acts for a reason, what one does is a function of how one is, mentally speaking. (It is also a function of one’s height, one’s strength, one’s place and time, and so on. But the mental factors are crucial when moral responsibility is in question.)
3. So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking – at least in certain respects. (from 2)
4. But to be truly responsible for how one is [in certain mental respects], one must have brought it about that one is the way one is [in these respects]. And it is not merely that one must have caused oneself to be the way one is...one must have consciously and explicitly chosen to be the way one is [in these respects], and one must have succeeded in bringing it about that one is that way.
5. But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is [in any mental respect at all], unless one already exists, mentally speaking, already equipped with some principles of choice, ‘P1’ – preferences, values, pro-attitudes, ideals – in the light of which one chooses how to be.

6. But then to be truly responsible, on account of having chosen to be the way one is [in those mental respects], one must be truly responsible for one's having the principles of choice P1 in the light of which one chose how to be.
7. But for this to be so, one must have chosen P1, in a reasoned, conscious, intentional fashion.
8. But for this [i.e. for premise 7] to be so, one must already have had some principles of choice P2, in the light of which one chose P1. (from 5)
9. And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice.
10. So true moral responsibility is impossible, because it requires self-determination, as noted in [step 3]."<sup>2</sup> (from 9)

There are some things to note in this formulation of the argument. First, he focuses his argument on our responsibility for actions taken *for a reason*. He is not discussing actions like biting one's nails out of habit, or reflexively swatting a fly near one's ear. Second, he states that actions taken for reasons are the result of the way the person is at the time the action was taken. In addition, for one to be morally responsible for such an action, the way one is – in certain mental respects – must be part of the cause of one's action. He never states what these mental respects are, but taking from one of the objections he later considers, one's character, personality, and motivational structure<sup>3</sup> seem to be a good fit; for we commonly think that something like these are the cause of an agent's consciously chosen actions. Third, given the nature of choice, we

---

<sup>2</sup> Ibid pg. 6-7; brackets replace more cumbersome wording that has the same meaning

<sup>3</sup> Ibid pg. 20

must already have some principles of choice in order to make a choice in a given situation.

It is also important to determine what Strawson takes himself to be doing in this argument. While it may seem that he claims to be disproving moral responsibility simply, evidence suggests otherwise. For example, he takes the anthropological suggestion that most societies can be divided into “shame” and “guilt” cultures as possible evidence that “a conception of moral responsibility *similar to our own* is a natural part of the human moral-conceptual repertoire”, our own being the true moral responsibility disproved in the argument.<sup>4</sup> In this statement, we see he is open to other conceptions of moral responsibility. Earlier on, he also says “the kind of absolute moral responsibility that [the Basic Argument] shows to be impossible has for a long time been central to the Western...tradition”, which I take to mean that the kind of moral responsibility at issue is simply one among multiple possible legitimate conceptions of it.<sup>5</sup>

So, what kind of moral responsibility has been widely believed in among Westerners? Looking at premises 6, 7 and 8 express, we see that by “truly morally responsible” he means ‘being the ultimate cause’ of an action, such that one’s conscious and explicit choice is the first, originating cause in the chain of causation.<sup>6</sup> In other words, the person’s present action must ultimately lead back to such a choice on the person’s part; for among these premises, being truly responsible is said to require, not only being responsible for the cause of one’s action (i.e. for one’s character at the time of

---

<sup>4</sup> Ibid pg. 9, emphasis mine

<sup>5</sup> I am not alone in reading Strawson this way. Matthew Talbert writes “Since the argument targets ‘ultimate’ moral responsibility, it does not necessarily exclude other forms, such as forward-looking responsibility and, on some understandings, responsibility-as-attributability.” (*Moral Responsibility*, Stanford Encyclopedia of Philosophy)

<sup>6</sup> In addition, he often uses the term ‘ultimate’ instead of ‘truly’ throughout his paper



choice), but also that one be responsible for the cause of *that* (i.e. for one's principles for choosing a character), and in turn, that one be responsible for the *previous* principles of choice in the light of which one chose to have *those* principles for choosing character, and so on. Therefore, we can see that the regress is only generated because he means 'true moral responsibility' as ultimate responsibility.

We can get an even better understanding of the subject of the Basic Argument through his use of the story of heaven and hell. Strawson writes "As I understand it, true moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven."<sup>7</sup> If there is eternal punishment in the afterlife, then the punishment could not be justified on the basis of any future-looking reasons<sup>8</sup>: it could not be justified because it will bring about moral change (because it's eternal, it will continue even after moral growth has occurred), nor because it will keep others from doing the same evils (because none of the living see them suffer), nor because it will keep others safe from the agent committing future evils (because the agent is already dead). So if human beings do have ultimate moral responsibility, then eternal punishment would be justified on desert-based grounds-- that is, because the person deserves to suffer in return for what they themselves caused to happen in this life. And taking the earlier analysis of the argument, we see that this would be so because we would be the ultimate cause of the things we do. It would be fair to eternally punish someone for something(s) only if it is *they* who ultimately

---

<sup>7</sup> Ibid pg. 9; emphasis in original

<sup>8</sup> This is an observation taken from Gregg Caruso: "The purpose of invoking the notion of a divine judge in the afterlife is to instill the idea that any rewards or punishments issued after death will have no further utility—be it positive or negative." (*Skepticism about Moral Responsibility*, Stanford Encyclopedia of Philosophy)

determined it. Otherwise, it would be analogous to punishing someone for stealing a car, when the cause of their doing so was being hypnotized and being commanded to do it – since the agent was only an intermediate cause of the outcome, not the ultimate cause, it seems fair to punish the hypnotist rather than the agent.

Having considered all this, the argument is paraphrasable in the following way: one cannot make a choice without already having the things necessary to make a choice: one's principles of choice. True moral responsibility for one's actions requires that one chose to be the way one is at any given moment without already having had any principles of choice beforehand. It requires this because any sequence of choices terminates backward in something given, not chosen, and so to avoid this, a choice must be made not on the basis of any already held principles of choice. Since this is impossible, it follows that true moral responsibility for actions is impossible.

Let's see how one could challenge the Basic Argument. The logic of the argument is solid, so the argument is valid. But is it sound? Premise 1 is true, and it states the focus of his argument. Premise 2 seems to be an accurate statement regarding the kind of actions for which one is held morally responsible: the relevant cause of the action is certain mental features of the person. We think that a person is morally responsible for an action just to the extent that the action is traced back to themselves, and specifically, to certain mental things about them (their beliefs, their character, their personality, etc.). So, to be morally responsible for something, the way that one is mentally at a given time must be a cause of the action. If premise 2 is true, and if "truly morally responsible" means 'ultimately morally responsible', then step 3 follows. Premise 4 states what is logically required for true moral responsibility for anything. Premise 5 states a truth about how one makes choices. Premises 6 and 7 express the notion of

“true” moral responsibility at play. Step 8 follows from premise 5. From these, an infinite regress is produced, as step 9 states: for true moral responsibility to be possible, there would have to be a beginning and end of an infinite chain of choices of principles of choice. And this is impossible, for nothing infinite can have both a beginning and an end. Thus, the conclusion follows from both step 9 and an implicit premise about the nature of infinitude.

### Chapter 3

Despite its apparent soundness, there have been challenges made against the argument. Strawson notes that the only premise that could reasonably be rejected is premise 3,<sup>9</sup> and if this premise is rejected, the conclusion does not follow. He considers three main responses to the argument, all of which attack this premise.

The first response is the *compatibilist* response. A compatibilist may respond that we are truly morally responsible for our actions that are done in the absence of certain things (force, obsessional neuroses, etc.)<sup>10</sup>, even if we live in a deterministic world, i.e. even if the way we are is ultimately determined by things outside our control. But Strawson replies “One does what one does entirely because of the way one is, and one is in no way ultimately responsible for the way one is. So how can one be justly punished for anything one does?”<sup>11</sup> Although the agent is a proximate cause of her action, she is still not ultimately responsible for it; and since it is only by being the ultimate cause that punishment would be her just deserts, desert-based punishment

---

<sup>9</sup> He actually considers objections to a more loosely formulated version of the Basic Argument, and premise 3 in the version I discuss corresponds to premise 2 in the looser version, which reads: “To be truly morally responsible for what you do, you must be truly responsible for the way you are—at least in certain mental respects.” (Strawson Pg. 13)

<sup>10</sup> Ibid pg. 16

<sup>11</sup> Ibid pg. 17

would not be justified. For this reason, “compatibilist responsibility famously fails to amount to any sort of true *moral* responsibility, given the natural, strong understanding of the notion of true moral responsibility”<sup>12</sup>.

The second response is the *libertarian* response. A libertarian would say that true moral responsibility and determinism are incompatible, and they accept the former while denying the latter. They think that indeterminism in our decision-making would make us ultimately responsible for our actions. But the problem is that indeterminism in either character formation or in particular actions would not make one responsible. By definition, an event that is totally undetermined or without explanation is not caused by anything. Hence, nothing and no one, including the agent, is responsible for it, and therefore nothing and no one is responsible for anything else that is, in turn, caused by it.

The third response he considers I will call the *independent self* response. This response states that, while we are not ultimately responsible for our character, preferences, and motivational structure (CPM), we are so for our actions, because the self is independent of these. While the self makes its decisions in light of them, it decides independently of them. Thus, its proponents claim, we can have ultimate responsibility for our actions without being ultimately responsible for the way we are at any moment. The problem with this response, Strawson notes, is that it puts us in the same problem we originally faced: the self decides to act the way it does at any given moment because of the way it is at that time. And since the way the self is at any moment is ultimately determined by things other than the self, one still is not ultimately responsible for one’s

---

<sup>12</sup> Ibid pg. 17; emphasis in original

actions. The objector has simply shifted the cause of actions from CPM to a self that is distinct from them. One still needs to be responsible for the way one is to be ultimately responsible for one's actions, it's just now one must be responsible for how one's *self* is instead of for how one's *CPM* is. And since we cannot be so responsible, we can't have ultimate responsibility for actions. As Strawson points out, none of the three responses successfully invalidates the Basic Argument, and so none secures the kind of ultimate moral responsibility in the face which reward and punishment are just in a desert-based sense.

#### **Chapter 4**

Now we will look at Aristotle's view on moral responsibility for action. To begin, we will first look at his account of causal responsibility for action.

In his *Physics*, Aristotle states that there are four kinds of cause: form, matter, origin of motion, and end. In brief, a thing's form is what it is, a thing's matter is what it is made of, a thing's origin of motion is "that from which the first beginning of change or of rest is"<sup>13</sup> and a thing's end is that for the sake of which it is. As is evident, what he means by the word 'cause' (αἰτία) is different than the ordinary use of the term in English. When we give the αἰτία of a thing, we are stating what is responsible for that thing in some respect. Thus, these are all the ways in which something can be said to be responsible for something else. All things natural are explained completely only when the student of nature has traced the thing back to each of the four kinds of cause.<sup>14</sup>

---

<sup>13</sup> *Physics* II.3, 194b30

<sup>14</sup> *Ibid* II.7, 198a20-198b5

Next we will look at his account of action. For Aristotle, human action is a change or lack of change whose origin is a human being.<sup>15</sup> Since the soul moves the body, the soul of a human being is causally responsible for the actions of that human being as the origin of change.<sup>16</sup> <sup>17</sup> What are the causes of the action in the other categories? It seems that the end and the form of the act are also the soul of the human being; the end of the action is the specific thing for the sake of which the action is taken<sup>18</sup>, which is thought by the person at the time of action to be part of the full, active life of the human soul. The form of the action, for its part, is the human soul, in that the action is the soul itself at work.<sup>19</sup> On the other hand, the material cause is the body of the human being.

But what about *moral* responsibility for an action? For example, we would say that someone who unknowingly gives their friend a cup of poison is causally responsible for that act, but we wouldn't hold him morally responsible, given certain other conditions. Likewise, someone may play an instrument skillfully, but we wouldn't say they are morally responsible for playing it well, but simply causally responsible. Aristotle doesn't use the expression 'moral responsibility'<sup>20</sup>, but he speaks about voluntariness, praise and blame, and what is noble and base, and together we can see him speaking about the same thing.

## Chapter 5

There are different kinds of changes that a person can originate. For Aristotle, one condition for moral responsibility is that the change originated be *voluntary*. There

---

<sup>15</sup> EE II.6, 1223a1-16

<sup>16</sup> De Anima III.10, 433b10-20

<sup>17</sup> Ibid II.4, 415b10

<sup>18</sup> NE I.1, 1094a1-5; I.7, 1097a30-35

<sup>19</sup> NE IX.6, 1168a5-9

<sup>20</sup> Bobzien, pg. 3

are two conditions for an act to be voluntary: 1) the act must be a change whose source is internal to the person in question, not external. In other words, the change must have been in our power to bring about or refrain from bringing about;<sup>21</sup> and 2) this person must know the relevant particular circumstances in which the action takes place.<sup>22</sup> It is only for voluntary actions that people are praised and blamed.<sup>23</sup> If any of these criteria is not met, then the action is either not-voluntary (where no regret is felt for doing it) or involuntary (where regret is felt).<sup>24</sup>

The second step to getting to moral responsibility in Aristotle is the noble and the base. To begin, there are many different kinds of possible human activities, and each can have multiple aspects. For example, take a tennis player competing in a final match. We can ask about how effective his form is, and how good his strategy is in the match, etc. But we can also ask about whether he is competing fairly, and how he carries himself when he wins or loses that match. Only these latter actions are related to what is morally good and bad, or in Aristotle's language, *noble* and *base*.<sup>25</sup> <sup>26</sup> While playing skillfully and being respectful are both things for which a person may be praised, I think Aristotle would agree that the person would receive moral praise for the latter, and nonmoral praise for the former. And since one only receives moral praise and blame if the action pertaining to the noble and base is *voluntary*, moral praise and blame tracks those acts for which we would say we are 'morally responsible'; for, both moral praise and blame

---

<sup>21</sup> EE II.6, 1223a1-20;

<sup>22</sup> NE III.1, 1111a23-25

<sup>23</sup> Ibid. III.1, 1109b30

<sup>24</sup> Ibid. III.1, 1110b18-25

<sup>25</sup> In *NE* I, 12.1101b14-18, Aristotle puts in separate groups justice and bravery, and being skilled at a sport.

<sup>26</sup> 'Noble' and 'base' translate to καλόν and το αισχρόν, respectively.

and moral responsibility only attach to actions that are both voluntary and relate to the noble and base. In other words, to be the source of a voluntary action related to the noble and base is to be morally responsible for that action.

## Chapter 6

We have seen in a preliminary way that Aristotle believes that people are morally responsible for their voluntary actions related to the noble. The reason this is justified is that these actions are the ones whose origin is the soul of a human being. But how does one's soul cause them?

Let's take a look at his 'anatomy' of the soul. Aristotle divides up the human soul into different powers: the vegetative, the perceptive, the desiring, the deliberating, and the intellective.<sup>27</sup> There are three kinds of desire: appetite (the desire for pleasure),<sup>28</sup> anger (the desire to exact revenge),<sup>29</sup> and wish (the desire for something we judge to be good).<sup>30</sup> While the latter is for something we rationally judge to be good, the former two are not, and hence the two nonrational desires can be in conflict with the latter, rational desire. Character is the state or condition that one's desiring part is in. If one's wish is for the noble, and one's nonrational desires have been tamed and are in harmony with this wish, then one is of virtuous character. If one has the right wish, but one's nonrational desires are in conflict with it, then one is either restrained or unrestrained: one is restrained if one's decision to do what is a means to one's wished-for end causes action, and one is unrestrained if, instead, a contrary nonrational desire causes action.

---

<sup>27</sup> De Anima III.10, 433b3-4; slight change in translation for readability

<sup>28</sup> Ibid. II.3, 414b5-7

<sup>29</sup> Rhetoric II.2, 1378a31-33

<sup>30</sup> NE III.4, 1113a15-25



Lastly, if one's wish is for the base, and one's nonrational desires are in harmony with this wish, then one is of vicious character.

Let's consider an example to see how one's character causes one's action: You brought donuts for the department, and your officemate walks by and sees them. Your officemate has conflicting desires: he wishes to be healthy, but he also wants to eat the tasty donut in the break room. He reasons very quickly that the healthy thing to do would be to refrain from eating it. If the decision to do the healthy thing overpowers his desire for pleasure, then his decision will cause him to refrain. If his desire for pleasure overpowers his decision, his nonrational desire for pleasure will cause him to eat it. In both cases, the action is caused by a natural part of the person's soul: his desire, and possibly his deliberating also. While this is true in restrained and unrestrained people, it is also true in virtuous and vicious people: their wish for some apparent good causes their reasoning to determine the best means to the end, which produces a decision to do the determined means, and since there are no conflicting nonrational desires, the decision causes the person to do the means. So, in all kinds of people, we see that the actions one takes are caused by the state of one's desiring part, i.e. one's character, and possibly one's deliberation as well.

What kind of cause of action is one's character? It is the origin of motion, and more specifically it is the first moved mover,<sup>31</sup> since it sets the body in motion. But notice that it is the first *moved* mover. There is another first mover, one that is unmoved: the thing desired.<sup>32</sup> For it is this that sets the desiring part in motion, which

---

<sup>31</sup> Aristotle notes that this part of the soul it is changed from being idle to at-work, and only in this sense is it "moved", De Anima III.10, 433b20-22

<sup>32</sup> De Anima III.10, 433b19

ultimately sets the body in motion. So, the human soul, via its desiring power, causes action as the origin of motion, while the desired object causes action as the action's end. Since both the desiring power and the deliberating power are natural parts of the soul, then when they are at work, their desires and reasonings are natural to the person in question, and hence the action caused is voluntary. And if those actions pertain to the noble and base, then for those actions the person is morally responsible.

## Chapter 7

He then raises the question: are we morally responsible for our character, the state of the desiring part of our soul? If not, are we therefore not responsible for our actions? Since our character determines our actions, this is a key question. We see the Basic Argument's answer to this question in premise 3 "...if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking – at least in certain respects.", and these mental respects could plausibly be read as one's character, personality, and motivational structure. But what does Aristotle think about this question? He gives his answer in Book 3 Chapter 5 of his *Nicomachean Ethics*. He considers this question when considering an argument by a hypothetical interlocuter, wherein it is argued that humans must be responsible for their character to be responsible for their actions.

The argument starts like this: "Now someone may say that all people aim at the apparent good, but they are not in control of how things appear, but whatever sort of person each one is, of that sort too does the end appear to anyone."<sup>33</sup> So far, both premises offered here are ones that Aristotle has accepted in the previous section: "what

---

<sup>33</sup> NE III.5, 1114a32-b2

is wished for simply and truly is the good, but for each person the apparent good”<sup>34</sup>; and, “in accordance with each sort of state there are special things that are noble and pleasant”.<sup>35</sup> For example, the desiring part of the virtuous person is such that things that are truly noble and pleasant, appear so to her, and to the vicious person, other things appear so. If the sort of person one is, is one’s character, what does this have to do with how things appear? The being at work of the perceptual part is being pleased and pained, and the work of desire is pursuing pleasure and avoiding pain. The connection between perception and desire is so close that Aristotle says that they are the same, though their being is different.<sup>36</sup> For this reason, Aristotle agrees with the second premise.

The hypothetical interlocutor continues: “If each man is somehow responsible for his state, he will also be himself somehow responsible for the appearance”. He asserts that if someone is responsible for the condition of the perceiving and desiring part of her soul, then she is responsible for what appears good to her. Then, he goes on to argue what will be true if we are *not* responsible for our character:

“if not, no one is responsible for his evildoing, but everyone does evil acts through ignorance of the end, thinking that by these he will get what is best, and the aiming at the end is not self-chosen, but one must be born with an eye, as it were, by which to judge rightly and choose what is truly good...to both men alike, the good and the bad, the end appears and is fixed by nature or however it may be”<sup>37</sup>.

So, we see the interlocutor’s answer to the key question: we must be responsible for our character to be responsible for our actions. What reason is given? What appears

---

<sup>34</sup> Ibid. III.3, 1113a20

<sup>35</sup> Ibid. III.4, 1113a30

<sup>36</sup> De Anima III.7, 431a11-13

<sup>37</sup> NE III.5, 1114b4-16

to a person to be an end (i.e. what is a final cause for a person) is not up to them to decide. In this scenario, the condition of the perceiving and desiring part of the soul is genetic, so that certain things, and not others, will appear good. To paraphrase the argument using our earlier analysis: whatever causes something to appear good to a person is the cause of the actions that will be taken for the sake of that apparent good. Given the earlier premise they agree on, if we are not responsible for the condition of our perceiving and desiring part, i.e. our character, then we are not responsible for our actions.

Aristotle has the opposite conclusion: even if what appears to be good is not determined by us, but rather by nature, nevertheless good and bad actions are voluntary<sup>38</sup>, because “there is that which depends on [oneself] in [one’s] actions, even if not in [one’s] end”<sup>39</sup>. The action depends on oneself in that it is caused by one’s character. Therefore, the action is voluntary. As he says earlier, “we are not able to trace our actions back to any other sources besides those that are in us”, and humans are sources and begetters of actions just as much as of children.<sup>40</sup> The fact that the condition of your soul, and hence the ends you hold, was not caused by you, is irrelevant.

How do we understand this kind of voluntariness and the moral responsibility that would go with it? I think it’s helpful to think about the purpose of the category of voluntariness. Why care about actions that satisfy the two aforementioned requirements – that the source of the act be the internal, natural desires and reasoning of a person,

---

<sup>38</sup> The actual subject of the sentence is ‘virtue and vice’, but given the flow of the argument, he must mean ‘good and bad actions’, on pain of invalidity. In addition, he appears to use the terms in the singular to mean just this in NE 1113b5-10; Both of these things are pointed out by Bobzien on pg. 15 and pg. 17

<sup>39</sup> NE III.5, 1114b19-21

<sup>40</sup> Ibid. III.5, 1113b18-21

and that the person know the relevant particular facts about the situation in which they are about to act? I suggest that the purpose is to delimit those actions that reflect the intention of the person doing them, and hence their character. When any of those conditions isn't met, the person is still the cause of the action, but since it doesn't reflect their character as an individual, they aren't praised and blamed. So, Aristotle seems to be saying, if we are not responsible for our character, our character still causes our voluntary actions, and so praise is deserved for manifestations of good character and blame for manifestations of bad character.

Aristotle goes even further, though. In much of the *Nicomachean Ethics* until this point, he has stated that our character is shaped by our own actions. As is well known, his position is that "...by performing just actions, one becomes a just person".<sup>41</sup> In fact, he takes it as exceedingly obvious that "it is from one's being at work involved in each way of acting that one's states come about".<sup>42</sup> As he explains elsewhere, virtues and vices are certain relations that we have to pleasures and pains, and it is by how we expose ourselves and interact with them that either perfects or spoils the perceiving and desiring part of the soul, i.e. that shapes what things will delight and pain us.<sup>43</sup> So, he goes on to say that, since the that we character come to possess is voluntarily developed, then the perfect state of virtue, and the most imperfect state of vice, are voluntary as well:

"If, then, as is asserted, the virtues are voluntary (for we are ourselves somehow part-causes of our states, and it is by being persons of a certain kind that we set the end to be so and so), the vices also will be voluntary; for the same is true of them."

---

<sup>41</sup> Ibid II.4, 1105b10

<sup>42</sup> Ibid III.5, 1114a9-11

<sup>43</sup> Physics VII.3, 247a1-15

Firstly, the state of virtue and vice are voluntary because we are co-causes of our character state. We are only one of the causes of our character, since the being-at-work of our perceiving/desiring part, which causes voluntary action, is only part of the equation: for example, upbringing also has a huge influence on the shaping of our desiring part,<sup>44</sup> since our elders get us to act in certain ways through threat of punishment, and the source of the upbringing is our parents, not ourselves<sup>45</sup>. Nevertheless, the fact that our voluntary actions, with repetition, determine our character, makes our character voluntarily taken on. It is for this reason that it is praised and blamed. If it were not voluntary, then we wouldn't praise or blame it, but we do, because we think it is.<sup>46</sup> Secondly, since what appears good to someone is caused by one's character, we are indirectly a co-cause of the things we consider to be ends.

Let's state Aristotle's position using his previous analysis. He thinks the condition of our desiring part, our character, causes actions as first moved mover. Actions that are voluntary reflect our character. The first unmoved mover is the thing desired. But, central to his account is that our voluntary actions are causes as origin of motion in the shaping of our desiring part, and thus are origins of motion for the things that we come to desire, our ends. So over time, with repeated voluntary actions of a certain kind, our desiring part takes on a more firm and definite shape, and the unmoved mover of the end is something that we have determined. Since this state of the soul is caused by voluntary actions, and voluntary actions are caused by our desiring part (and thus reflect its condition) we praise and blame character states as we do actions: as reflecting

---

<sup>44</sup> NE II.1, 1103b25

<sup>45</sup> Not mentioned by Aristotle is also abusive environments, which has much evidence for its negative effects on moral development. For a paper on this topic, see *The Impact of Child Abuse on Moral Development*, by Shayla Stogsdill. The origin of motion of the abuse is obviously other than the child itself.

<sup>46</sup> NE III.5, 1114a28-31

the condition of the soul of the person who brought them about. And since the ends we come to have are caused voluntarily, our actions as adults with formed characters are caused by us in two senses: as origin of motion, and as end, in that we are the origin of motions of the ends we come to have.

It may be wondered, “Why does he think we can’t trace back the action to sources other than those in us, so that we are sources of our actions as much as of our own children?” To give a defense of his account, it will help to look at reproduction. Aristotle and the rest of us say that the parents of a child are the cause of the child – they are the ones who brought it into being. But we also know that the parents themselves were brought into being by each of their pair of parents, and so on. Yet we still consider the parents to be the cause. In a similar way, Aristotle here is saying that the cause of an action is one’s character, even though one’s character is only partly caused by the person herself, and ultimately, starting from what nature and parent provide.<sup>47</sup> Why do we say what we do in reproduction? I suggest it is because the being whose work is X, we say caused X. Since making the child is the work of the parents, not the grandparents, we say that they are the cause as origin of motion. Similarly, the being whose work is to act and shape itself is a particular human soul, and more specifically, one or two of its parts. Since it is their work, we say it is the origin of motion of the changes it brings about, even though it does not put itself in the initial state it is in; indeed, this would be impossible, since this would be true self determination, which, as Strawson points out, is impossible. And indeed, we don’t require parents to have given themselves the power to reproduce, or to have given themselves the condition of their reproductive power, in

---

<sup>47</sup> Ibid. X.9, 1179b20-28

order to say they are the source of the resultant child. However, if someone ruins their ability to reproduce by voluntarily doing things that damage it, we blame them for their developed infertility, since they voluntarily brought it about.

## Chapter 8

So, let's compare the two accounts. The Basic Argument, as presented by Strawson, says that if we are not ultimately morally responsible for our character, personality, and motivational structure, then we are not ultimately morally responsible for the actions they cause. Aristotle says that even if we are not responsible for our character in any way, we are responsible for the actions it causes; and in fact, we are partly responsible for our character. As is evident, Aristotle nowhere claims that we are the ultimate cause of our actions or character. In fact, he allows that we not be responsible for our character, and thus for the things we desire, i.e. the final cause of our actions, while we are still the origin of motion of our actions. So clearly, contrary to Strawson's claim<sup>48</sup>, Aristotle does not argue that people are ultimate causes of their actions and so have desert-based moral responsibility.

In addition, it is evident that Aristotle agrees with the Basic Argument that we are not the ultimate cause of our actions. Aristotle asserts that the desiring part of one's soul causes one's action, and although the condition it eventually takes on in adulthood is caused by the actions of the soul, the coming into being of this part is not caused by us ourselves; rather, the part itself, and the condition it is initially in, are caused by one's parents. Therefore, as the Basic Argument notes, Aristotle would agree that the *ultimate* cause of our actions is not ourselves.

---

<sup>48</sup> Strawson pg. 9



To take a closer look at the kind of causal role Aristotle thinks we have, the state of our soul is the cause of our actions as moved origin of motion, but the condition of our soul is, in part, caused things other than our previous actions, and ultimately caused by our genetics, so that something other than our character is the *ultimate* cause of the actions. Nevertheless, we can say that our character is a proximate cause of our actions.

One can compare it to the way Ball 1 moves Ball 2, and Ball 2 hits ball 3: it is true to say 2 hit 3, even though the ultimate cause of 3's movement is 1. While this may seem not to correspond to how humans are, we need to remember that the way we cause our acts is not the way an inanimate object does. We do so with thought, we use our ability to see what things there are, to consider their value to us, and then to think about how to achieve them. Not only that, but the consequences of our actions, among other experiences, can alter how we perceive things. So, we play a much more active role, with our powers, in the things we cause, even if the state of those causes and how we use them is not ultimately up to us. It similar to how a very complex AI robot could be programmed with certain ends and abilities to learn and deliberate: however it is created, things will affect its ends, it will learn (depending on its learning ability) and it will use its calculative ability to try to determine the best means to those ends. In the same way, when it comes to states, they are voluntary, because our voluntary actions, of which we are the source, are the proximate cause of those states, including the ends that partly constitute them.

Given this determinist picture, and given his acceptance of the legitimacy of moral praise and blame for actions, it appears that Aristotle's position is that of the *compatibilist*: one can have some kind of moral responsibility for actions, while the ultimate cause of the character that causes action is something outside our control.

## Chapter 9

We've seen that Strawson shows that it is impossible for people to be the ultimate cause of their actions. He states what he thinks is a consequence of accepting this conclusion: "...if one takes the notion of justice that is central to our intellectual and cultural tradition seriously, then the evident consequence of the Basic Argument is that there is a fundamental sense in which no punishment or reward is ultimately just."<sup>49</sup> As we saw earlier, without ultimate moral responsibility, no punishment or reward is just on the grounds of moral desert, so his observation is correct if we understand the justice of punishment in question to be the justice of desert-based punishment. It would not be fair to apply punishment on the grounds that one deserves bad to happen to one because one is ultimately responsible for one's action, since the argument shows that things outside one's control, rather than oneself, are the ultimate causes of one's actions. On the other hand, there are possible reasons to apply punishment besides retribution; for example, revoking the license of someone who committed a DUI in order to prevent them from committing it again. If this is so, then accepting the impossibility of desert-based moral responsibility implies the injustice of retributive punishment, but says nothing about the justice of punishment that is taken for forward-looking reasons.

On the other hand, what are the consequences of accepting Aristotle's account? Aristotle identifies the origin of motion of an action to be the character of an individual. He recognizes that there are many factors in the development of one's character. The condition of one's perceiving and desiring part of the soul determines what we pursue, and when we are born, its initial condition is given, and the upbringing of our parents is

---

<sup>49</sup> Strawson, pg. 15-16

one crucial part of how it's shaped. Given this, he doesn't assert that a decision on our part is the ultimate cause of our character and resultant actions, so we do not have desert-based responsibility, and therefore neither would desert-based punishment be just. But since our character causes, and is reflected by, our voluntary actions, then by definition, he is saying we have *attributability* moral responsibility. This is what justifies the application of moral praise and blame for noble and base character and voluntary actions: since they are properly attributed (either immediately or with intermediate steps) to the character of the person, the approval and disapproval are given to the person for their character. However, attributability only gets us this far. While we are justified in expressing disapproval and anger at someone for a base character or action, it does not justify punishing someone for it; for that to be justified, more is needed.

Now, while Aristotle's account merely leaves open the possibility of the justice of punishment for forward looking reasons, there is textual evidence to believe that he thought such punishment is justified. He says that punishment is like a medicine for vice<sup>50</sup>, indicating that it is used for moral reform. He also says that everyone honors and punishes in order to discourage or encourage certain behavior,<sup>51</sup> which is also a forward-looking reason. So not only does his account show that we have attributability moral responsibility, but other passages indicate his belief that we have *accountability*-responsibility, meaning that our character and voluntary actions are such that holding

---

<sup>50</sup> NE II.3, 1104b5-20; EE II.1, 1220a29-35

<sup>51</sup> NE III.5, 1113b23-26

each other accountable, via punishments and other means, is justified for forward-looking reasons.<sup>52</sup>

## Chapter 10

I'd like to take this last section to reflect on an implication of accepting the conclusions of both authors, as well as a suggestion for why so many have supposed themselves to have a kind of moral responsibility that a simple argument shows to be impossible. On the first point, observers and victims of wrongdoing often have a strong hatred and desire for revenge on those who do great evils. Sometimes we even wish them eternal suffering in hell. But although the person does have a crucial role in shaping of their character, we have seen that we are moved by whatever seems good to us at the time of action. If Aristotle is right that the highest good for human beings is happiness, and that happiness is living virtuously,<sup>53</sup> then the perceiving and desiring part of those who living unvirtuously is not in the condition it ought to be in, and thus is not working properly. Either things that are not truly good appear good to them, or this isn't the problem, but other pleasant and painful things have an effect on them such that their reasoning about how to achieve the true goods they wish for is disabled to some extent, similar to what happens when one is drunk, or in the throes of raging anger.<sup>54</sup> If this is so, then in order for them to change, they need to be corrected or rehabilitated, a purpose of punishment already noted by Aristotle. Given both Aristotle's account and the Basic Argument, our perspective of those who do wrong cannot be that they are the *causa sui* of their bad character; that they are the sole and ultimate cause of it; that they

---

<sup>52</sup> Of course, this is not a complete argument for the justice of such punishment. This would require investigating the nature of justice itself.

<sup>53</sup> NE I.7, 1098a16

<sup>54</sup> Ibid.VII.3, 1147a10-21

chose, in fully clarity of mind, to become that way. Rather, the agent, like every person, is pursuing what seems best to them, and they are either sorely mistaken, or are able to be disoriented or clouded by pleasure and pain.

The next point naturally arises: If it is so obvious that we cannot be ultimately morally responsible for our actions and character, then why do many of us often feel anger toward those who do wrong as if they were? As if they were inherently bad? As if they chose for their character to be as it is with full clarity of perception? If a person is born such that nothing truly good appears good to them, and it never can, then they are born that way, and we can't act as if they chose to be that way. Regarding the rest of people, they develop bad character through a number of bad things, including bad actions, poor upbringing, abusive environments, etc. I suggest that the reason for their attitude is, ironically, something Aristotle discussed already: the stunning power of our nonrational desires on our thinking. Similar to the way intoxication and anger can cloud our thinking and lead us to take a course of action contrary to our better judgment, anger at those who've caused harm can cloud our thinking, such that we are not able to think about how things truly stand with regard to responsibility and how the agent came to be the way they are. In addition, I suspect that a desire to think that we are above those who do wrong in some inherent way is another contributing cause of the confusion. If we view those who do such things as inherently bad, then we, who do not do them, are inherently good. And the motive is simple: it feels good to seem to ourselves to be superior to others.

## **Chapter 11**

In conclusion, Aristotle does not argue for the kind of moral responsibility that Strawson's paper refutes. Rather, Aristotle would agree that ultimate self-determination

is impossible, and he presents the kinds of moral responsibility that we *do* have.

Strawson is right that we don't have ultimate responsibility for our actions because we don't have ultimate responsibility for the cause of those actions, our character. He also correctly observes that punishment applied because the person is the ultimate cause of their action is unjustified. For his part, Aristotle offers a compatibilist account of our causal role in our actions and character, which presents us as having attributional moral responsibility; and in addition, has independent reasons for believing us to have accountability moral responsibility, such that reward and punishment for forward looking reasons can be justified.

## Bibliography

- Aristotle. *Rhetoric* from *The Works of Aristotle*. Translated by W. Rhys. Roberts. Oxford: Oxford University Press, 1924.
- Aristotle, and Mark Shiffman. *De Anima*. Newburyport, MA: Focus Publishing/R. Pullins Co., 2011.
- Aristotle. *Aristotles Physics: a Guided Study*. Translated by Joe Sachs. New Brunswick, NJ: Rutgers Univ. Press, 2011.
- Aristotle. *The Eudemian Ethics*. Translated by Anthony Kenny. Oxford: Clarendon Press, 2011.
- Bobzien, Susanne. "Choice and Moral Responsibility (NE iii 1–5)." *The Cambridge Companion to Aristotles Nicomachean Ethics*, July 24, 2011, 1–24. <https://doi.org/10.1017/cc09781139022484.005>.
- Caruso, Gregg. "Skepticism About Moral Responsibility." Stanford Encyclopedia of Philosophy. Stanford University, January 18, 2018. <https://plato.stanford.edu/entries/skepticism-moral-responsibility/>.
- Sachs, Joe, trans. *Nicomachean Ethics*. United States of America: Focus Publishing, 2002.
- Stogsdill, Shayla, "The Impact of Childhood Abuse on Moral Development" (2019). Orphans and Vulnerable Children Student Scholarship. 1. <https://pillars.taylor.edu/ovc-student/1>
- Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75, no. 1-2 (1994): 5–24. <https://doi.org/10.1007/bf00989879>.
- Talbert, Matthew. "Moral Responsibility." Stanford Encyclopedia of Philosophy. Stanford University, October 16, 2019. <https://plato.stanford.edu/entries/moral-responsibility/>.