

Morality in the Meat Machine: An Examination of Moral Responsibility and
Neuroscience in Cases of Early Life Stress

Jessica L. McCormack

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Arts
In
Philosophy

Joseph C. Pitt
James C. Klagge
Ted Parent
David Faraci

May 07, 2013
Blacksburg, Virginia

Keywords: Moral responsibility, neuroscience, Strawson, psychological disorders

Morality in the Meat Machine: An Examination of Moral Responsibility and
Neuroscience in Cases of Early Life Stress

Jessica L. McCormack

ABSTRACT

This thesis examines the excusing and exempting conditions in Strawsonian accounts of moral responsibility. More specifically, it notes some concerns for Strawsonian accounts with regards to exempting individuals on the basis of psychological abnormalities, namely that the excusing/exempting distinction is unclear, and more importantly that treating a person's brain as an entity distinct from the person suggests a dualistic picture of the self that is not consistent with neuroscientific accounts of the brain. If we redraw the distinction to be between external/internal features, and focus on brain processes as the responsible entities for any given action, we can avoid these worries and have a more empirically accurate account of responsibility.

TABLE OF CONTENTS

<u>INTRODUCTION</u>	1
<u>I. EARLY LIFE STRESS</u>	3
i. Introduction	3
ii. Physiology, Psychology and ELS	7
iii. Shame and guilt	14
<u>II. MORAL RESPONSIBILITY</u>	21
I. Introduction	21
II. accounts of moral responsibility	
a. alternative possibilities	
b. reasons	
c. strawsonian	22
iii. Excusing and exempting conditions	35
<u>III. EXCUSING/EXEMPTING CONDITIONS</u>	42
<u>IV. RESPONSIBILITY AND DUALISM</u>	49
"My brain made me do it"	51
"Should have known better"	56
"Who, or what, is the agent?"	60

V. TOWARD A NEURO-RESPONSIBILITY ACCOUNT	69
---	-----------

CONCLUSION	77
-------------------	-----------

ACKNOWLEDGEMENTS

I was fortunate to receive much help on this thesis from mentors, friends, and family in the form of feedback and encouragement. I would like to thank foremost my chair, Joe Pitt, for encouragement, advice, and feedback. I would next like to thank my committee members-- Jim Klagge, Ted Parent, and David Faraci-- for their very helpful feedback and discussions over my thesis, as well as their encouragement. I would also like to thank Ben Jantzen and Dan Linford for helpful discussions and resources about my thesis. Lastly, I would like to thank Angela Smith (Washington and Lee) for providing suggestions on important moral responsibility literature related to my topic. And, of course, I would like to thank friends and family (including everyone just mentioned) for encouragement and support.

INTRODUCTION

In this thesis, I will explore the topic of exonerating individuals of moral responsibility for their actions for psychological reasons. In particular, I will focus on a specific case of individuals who suffered troubled childhoods; these individuals can experience Early Life Stress, which can lead to atypical neurological development. My interest is in reconciling our moral judgments in such cases with current neuroscientific understanding of the brain.

The big picture of the argument I will present is the following. In the literature on moral responsibility, there is a distinction drawn between two different kinds of reasons we may find an individual not to be morally responsible for some given action. As P.F. Strawson and Strawsonian theorists suggest, one kind of reason seems to be that an individual may be constrained to act a certain way, as when people are forced to do something (this kind of reason is 'excusing'). The other kind of reason is that the individual in question is simply not the sort of person to be held responsible (the person is 'exempted' from the moral community).

However, this distinction is quite fuzzy. There are cases of psychological abnormality that seem to straddle the line, since a person may act a certain way due to internal features and yet still be a moral agent (so, he is not exempted) (Ch. III). Moreover, our reasons for exempting individuals on psychological bases do not seem consistent with neuroscientific understanding of people's behaviors. When we question whether a person with a neurological defect is morally responsible for an action, we draw

a distinction between the "person" and the "brain", as though our moral character is something separate from the brain. But this is dualistic (Ch. IV).¹

Since the excusing/exempting distinction is already somewhat unclear, we might consider replacing it with another distinction, simply between internal and external explanations for actions. I will focus mostly on the internal explanations in cases of psychological abnormality. This will provide a clearer distinction that is also consistent with neuroscience.²

This thesis has five chapters. The first is a review of psychological and neuroscientific literature on the topic of Early Life Stress. The second is a review of some of the literature on moral responsibility. The third and fourth chapters address the concerns provoked by attempts to justify exoneration of individuals for psychological reasons (as mentioned above). Lastly, in Chapter V, I will consider a modification to P.F. Strawson's account of moral responsibility that can avoid the concerns raised and accommodate current neuroscience.

¹ This will be clarified much more thoroughly in Ch. IV, but I should note a couple clarifications here. First, my target is not only Cartesian dualism but also any picture of the self that suggests a distinction between the self and the brain. Second, my motivation here is that what we discuss about moral responsibility should be consistent with neuroscience, and dualist accounts tend to argue that the materialist explanation of the brain is not complete. That is, dualists tend to hold in some form or other that there is something about the mind that the materialist explanation leaves out. This is not compatible with what I argue for in this paper, but more on this in Ch. IV. My thanks to Jim Klagge and David Faraci for noting this vagueness.

² Many thanks to David Faraci for helping to formulate the big picture characterization given here.

I. EARLY LIFE STRESS

i. Introduction

Early Life Stress is somewhat of an umbrella category that contains different sorts of traumatic experiences occurring in early life; I will focus on studies done on abuse and neglect, but also included as Early Life Stress experiences are bullying, parental separation or divorce, living in a war zone, socioeconomic deprivation, and others. Short-term and long-term effects of Early Life Stress (ELS) have only somewhat recently been studied psychologically and neurobiologically. In fact, we still lack a clear and complete story about ELS in both domains. It remains unclear what exactly are the neurobiological effects both of ELS and intervention (i.e., treatment of the individuals with ELS) (Twardosz and Lutzker, 2009, 66).

Yet despite an incomplete story, there is a substantial number of studies on the effects ELS has on lifespan, metabolism, stress reactivity and regulation, hypertension, drug dependency, suicide, aggression, and other social and pathopsychological disorders. In general, studies alarmingly reveal that ELS in various forms can have lasting negative effects on victims (Hildyard and Wolfe, 2002). Although all of these effects are important to know, I will be discussing only topics relating to stress reactivity (the ways in which a person responds to stress), aggression and anger, and what some researchers have dubbed "moral emotions" (empathy and guilt, as contrasted with shame)³ (Tangney, Stuewig, Mashek, 2007). This chapter will be divided into two sections. The first will deal with

³ This contrast will become clearer in the second section of the chapter. However, it is worth noting here that shame, but not guilt or empathy, is considered a maladaptive moral emotion, at least in the literature that I discuss.

psychological and physiological effects of ELS.⁴ The second will cover some of the literature on the psychology of moral emotions, including some that suggest those who have ELS are more susceptible to shame complexes, which undercut moral behavior. I will demonstrate in this chapter that although we currently lack a complete neuroscientific understanding of ELS, there is sufficient evidence both psychologically and physiologically in these cases to demonstrate that individuals with ELS are statistically more likely to suffer emotional regulative deficits relevant to questions about moral responsibility.

Before reviewing the scientific literature, some disclaimers ought to be made, as researching ELS is complicated. First, there are issues of underdetermination of the scientific evidence. Because of ethical constraints on performing studies with humans, most information comes from either animal studies (normally involving primates and rats), and retrospective human studies. Anu-Katriina Pesonen and Katri Räikkönen note several difficulties with the latter strategy. First, it can be difficult to isolate whence a particular characteristic arose (Pesonen and Räikkönen, 2012, 722). For example, victims of neglect are statistically more likely to have attention problems (Scannapieco and Connell-Carrick, 2005, 133). To what degree is this higher probability due to genetics and to what degree is it due to childhood environment? This difficulty is even more apparent when attempting to discern whether an ailment such as hypertension or drug-addiction is due to genetics or to one's childhood circumstances (both are statistically higher for those who suffered physical abuse, yet genetic predisposition is also a

⁴ Essentially these can be seen as two ways of describing the same phenomena. Yet some of the research I will be discussing is on the macroscopic level--descriptions of behavioral tendencies for children suffering neglect or abuse-- and some is neurobiological-- descriptions of alterations of the hypothalamic-pituitary-adrenal (henceforth HPA) axis. Neuroscientific explanations of phenomena tend to span different levels of scope on the phenomena to provide full explanation (Churchland, 2002, 28-30; Craver, 2007, 10).

contributing factor to such ailments). Second, in studying long-term effects of ELS, individuals are asked to recount their childhood experiences, so the method relies on self-reporting, hence inviting underdetermination by memory loss and bias (Pesonen and Rääkkönen, 2011, 722). Third, as Loman and Gunnar note, human ELS is "messier" than animal studies (Loman and Gunnar, 2009, 868). That is, animal studies normally examine cases of neglectful parenting rather than adverse parenting; yet often human children experience more than one sort of maltreatment. One type of abuse may be accompanied by another (e.g., emotional frequently accompanies physical abuse), or a type of abuse may be accompanied by neglect (this also happens frequently) (Loman and Gunnar, 2009, 868). This point actually presents two disclaimers: it can be difficult to discern which maltreatment should be considered the cause of a given effect, and the use of findings from animal studies to make inferences about human cases of maltreatment may have this grain of salt accompanying it.

The next disclaimer is that discussing ELS as a unified group is somewhat simplistic. As mentioned earlier, ELS is somewhat of an umbrella term for varying sorts of early life experiences non-exhaustively including: physical abuse, sexual abuse, emotional abuse, neglect, parental divorce, bullying, separation from parents, circumstantial disruptions (such as experiencing natural disasters or living in a war zone), death of a close relative. Generally the severity of the ELS experience is correlated with the severity of the effects it has on a child, as one might expect. Yet different ELS experiences yield different effects. For example, higher rates of aggression and suicidal behavior are correlated with experiencing physical abuse more so than for those who experienced neglect, though both have higher rates than those who do not suffer

maltreatment (Scannapieco and Connell-Carrick, 2005, 199-202). These nuances should not have any real bearing on my argument, but they are important to note.

Abuse and neglect are the primary types of ELS that I discuss. The types of abuse discussed are physical, emotional, and sexual. Definitions of these vary across the literature, but the definitions are approximately the same. Physical abuse is willful action on the part of a guardian that results in physical harm (that is severe enough to leave evidence, e.g., marks or injuries) (Swogger et al., 2010, 363; Twardosz and Lutzker, 2009, 59; Heim et al., 2000, 593).⁵ Sexual abuse is unwanted touching of the victim's private areas, or forced touching of the perpetrator's, attempted or completed rape (of any sort), or other sexual acts without consent (Heim et al., 2000, 593). Emotional abuse is a somewhat fuzzier notion than the other two, but is essentially words and actions that make a child feel as though they are worthless, unloved, have no value, etc. (Scannapieco and Connell-Carrick, 2005, 16-17).

Neglect, like abuse, comes in different forms, primarily physical and emotional. Physical neglect is failure on the part of the guardian to insure the child's physical needs (e.g., nurturance, medical well-being) are met, and similarly emotional neglect is failure to meet the child's emotional needs. More specifically, a guardian emotionally neglects a child when she causes, by action or omission of action, the child's emotional well-being to be disrupted. That is, if the child develops behavioral disorders as a result of the

⁵ The inclusion of the provision that the abuse leave marks may seem surprising, and indeed some may argue that any willful action resulting in harm is abuse. Some cases that may come to mind, such as forms of torture that do not leave marks (e.g., sleep deprivation or Chinese water torture), may be better classified a different type of abuse (psychological/emotional). But if a punch is not severe enough to cause a mark, should it not still count as abuse?

There are conceptual problems with defining maltreatment of various sorts, especially if the intention behind the definition is diagnostic (which might explain the provision about physical abuse leaving marks) (Sanchez and Pollak, 2009, 497-499). At least on some accounts, physical abuse is defined without mention of its visual remnants (Cloitre et al., 2006, 1) My thanks to Jim Klagge for noting this objection.

parent's neglect, it is a case of emotional neglect. In the next section, I will detail some of the physiological and psychological narratives accompanying these types of experiences: what the effects of ELS are and how they come about.

ii. Physiology, Psychology and ELS

As noted earlier, ELS can cause diverse and severe effects on an individual. On the macroscopic level,⁶ some of these effects (the focus of this chapter) are emotional and behavioral irregularities; on the biological level, these phenomena may be correlated with alterations to the hypothalamic-pituitary-adrenal axis (hereafter HPA axis)⁷ and even the prefrontal cortex (Sanchez, 2006). I will move from the microscopic level to the macroscopic level in this section; both levels show the degree to which ELS affects individuals in fundamental ways. ELS can importantly affect threat and stress response systems, as well as other cognitive functions such as attention and emotional regulation. These effects are correlated with alterations to chemical levels important to the function of the HPA axis and the prefrontal cortex. Yet it is important first to understand the functions these parts of the brain serve before examining their malfunctioning.

The HPA axis and its related hormones (glucocorticoids and adrenocorticotrophic hormones) are important to the stress- and threat-response systems, among other biological functions. Roughly, the picture is as follows. In a threatening environment, both the HPA axis and the autonomic nervous system (ANS) coordinate a response. The

⁶ 'Macroscopic' is a common term to find in discussions of neuroscience (both in philosophical and neuroscience literature). The common conception is that phenomena in the brain described on a 'macroscopic' level is also describable in terms of mechanisms and processes on the 'microscopic' level. For example, we can speak macroscopically of depression and associated symptoms, or we can speak microscopically of mechanisms that are correlated with depression, e.g., irregularities in certain kinds of neurotransmitters.

⁷ The HPA axis is part of the neuroendocrine system that controls threat and stress responses (as I will review in this section) as well as other important body functions, e.g., digestion.

HPA axis response ensues in the following order. First, the hypothalamus releases corticotropin-releasing hormone (CRH) (produced in the hypothalamus). CRH stimulates the production and release of adrenocorticotropic hormone (ACTH) from the anterior pituitary. The ACTH in turn triggers the adrenal cortex to release glucocorticoids (GCs)⁸, steroid hormones. This stress-response activity of the HPA axis is mediated by the limbic-HPA (L-HPA); in short, the amygdala, namely, activates the HPA response to stress, and the hippocampus and prefrontal cortex (PFC) inhibits the response (Loman and Gunnar, 2010, 869; Sanchez, 2006).

Glucocorticoids play an important role in the response system as they regulate feedback for the HPA axis through glucocorticoid receptors: mineralocorticoid receptors (MR) and glucocorticoid receptors (GR). MRs have high affinity for GCs, important for triggering HPA activity, and GRs have low affinity for GCs, important for negative feedback following stress response (Sanchez, 2006, 624). High levels of these receptors exist in the prefrontal cortex, and so GCs are thought to be important in the regulation of various functions, such as cognition and regulation of mood and social behavior (Loman and Gunnar, 2010, 870; Sanchez, 2006, 624). In fact, the prefrontal cortex has more GR than does the hippocampus; this could explain impairments in executive functioning (such as processing speed) among individuals with ELS (Gould et al, 2012, 505). The rough picture, as Michelle Loman and Megan Gunnar describe it, is that "elevated GCs have become almost synonymous with 'stress'" (869).

Two of the possible biological effects of ELS are hypercortisolism (Gould et al., 2012, 501) and decreased hippocampal GR expression (Arabadzisz et al., 2010, 1106;

⁸ Also referred to as 'cortisol' with respect to primates and humans.

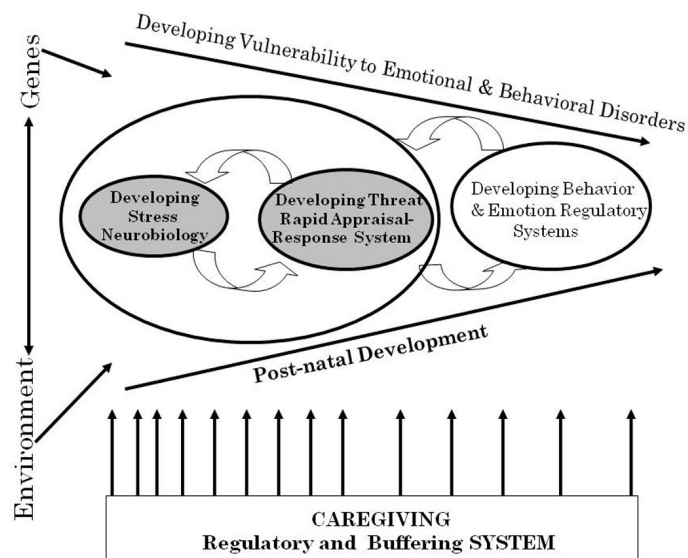
Pesonen and Räikkönen, 2012, 723).⁹ Higher levels of cortisol are correlated with amplified stress response (Loman and Gunnar, 2010, 871), and the decreased quantity of hippocampal GRs is correlated with reduced stress regulation, including prolonged stress reactions (Pesonen and Räikkönen, 2012, 723-4). Lower levels of GRs are also associated with increased risk for suicide, as one study indicates (Arabadzisz et al, 2010, 1106).

Studies on this topic have concluded that ELS has short-term and long-term effects on cortisol levels and GR expression in the hippocampus. These effects suggest that structures involved in stress- and threat-response systems are highly plastic early in life (Cirulli, 2001, 324), and that damaging them can have serious consequences for socioemotional development (Loman and Gunnar, 2010, 871). One of the notable effects ELS has is that it removes the main potential buffer for stress-response, which can lead to the damage noted.

Attachment security--the presence of a responsive caregiver-- is an important buffer in threat and stress responses. Loman and Gunnar notes, "Multiple studies have demonstrated that it is difficult to elevate cortisol in young children when a parent with whom they have a secure attachment relationship is present" (871). For children who suffer ELS, the opposite is true: they have difficulty using the presence of care-giver as a buffer to stress. They are more likely than their counterparts to have prolonged cortisol elevations in their stress responses. Importantly, this finding has been reached in studies conducted on neglect and abuse (871-2). This can lead short-term and long-term¹⁰ to hypersensitivity of the "developing threat appraisal" and response systems, increased risk for behavioral and emotional problems, and increased vulnerability to stress (871).

⁹ Though Loman and Gunnar note that chronic neglect can result in *hypocortisolism*: "low early morning levels of cortisol and blunted ACTH and cortisol responses to stressors" (Loman and Gunnar, 2010, 869).

¹⁰ In other words, these effects can develop during childhood or adolescence and persist into adulthood.

FIGURE 1¹¹

In Figure 1, we can see the trajectory of the development of the stress- and threat-response systems. While some of the interactions of the elements in this image are a bit unclear, what is well-communicated is that caregiving that serves its role as a buffer is vital to normal development of these systems. Where there is not such care giving the development of vulnerability to disorders can increase, meaning there is dysfunction in the threat- and stress-response systems.

¹¹ "This represents the Center's [the Neurobehavioral Development Center (University of Minnesota)] working model. It is purposely general enough to apply to the various model systems (rodent, non-human primate, human) studied by Center faculty. The model assumes that both genes and environment will influence developing vulnerability to emotional and behavioral disorders. The partially converging arrows running from left to right are meant to suggest diminishing (but continuing) plasticity of the neurobiological systems underlying risks for emotional and behavioral disorders. The facets of neurobiology depicted in the model are the neurobiology of stress and the neurobiology of rapid threat appraisal and response, along with developing behavioral and emotional regulatory systems. Stress- and Threat-response systems are depicted to the left within a larger circle to indicate their earlier emergence in development. The circular arrows connecting these two systems are meant to reflect their mutual influence on one another. Emotional and behavioral regulatory systems reflect cortico-limbic systems whose development is depicted to the right to indicate that it is somewhat later developing. The circular arrows drawn between emotional and behavioral regulatory systems and the circle containing the stress- and threat-response systems indicate mutual influence across development. Finally, the aspect of the environment most notable to our Center's model is shown along the bottom of the figure. This is the caregiving regulatory system. The spacing of the arrows is designed to indicate that this system has more influence earlier in development, but depending on the species, may continue to play a role well for prolonged periods of the organism's life" (Loman and Gunnar, 2010, 869).

One last neurobiological effect of ELS is on the central nervous system's (CNS) corticotropin-releasing factor (CRF) systems (Heim et al, 2000; Sanchez, 2006). These systems play an important role in some main body functions. Importantly these functions include elevated startle responses to sound and threat responses, such as freezing and fighting, and an increased vulnerability to fear conditioning (Heim et al, 2000, 593). ELS can cause hyperactivity to stress in the CRF; if the hyperactivity remains, it may contribute to an individual's developing mood or anxiety disorders (Heim et al, 2000, 593; Sanchez, 2006, 627).

The neuroscience described in this subsection is correlated with psychological disorders or the potential to develop disorders. That is, on the more macroscopic level, individuals with ELS are more likely to develop attachment and anxiety disorders, aggression, and other medical and psychopathological problems mentioned previously (Silverman et al., 1996, 710). Attachment is key to building healthy social relationships, and is important for our discussion, so I shall start with that.

Attachment, as with any learned behavior, results from adaptation. We are born with billions of neurons, but the connections between them develop over time.¹² These can form in two ways, according to Sandra Twardosz and John Lutzker. They describe *experience-expectant development* as the overproduction, and then trimming and organizing, of synapses. This process of *synaptogenesis* followed by "pruning" is found in the development of visual, auditory, and motor systems, and also in language centers of the brain and the prefrontal cortex (Twardosz and Lutzker, 2010, 62-3). The second process is *experience-dependent development*: the creation of new synapses or

¹² I recognize this is still on the microscopic level, but this should be discussed before talking about attachment on the macroscopic level. My thanks to David Faraci for noting this.

modification of old synapses in *response to* experience. It is thought that experience-expectant development is complete by early adulthood, so most of our life-learning is through experience-dependent development.

ELS can affect development by depriving children of sensory experience. In cases of neglect, there is simply a lack of social experience for the child to react to; in cases of abuse, a child reacts to negative and threatening experience. Twardosz and Lutzker explain, "There is a sensitive period during early development when exposure to high levels of stress hormones produces a cascade of neurobiological effects that guide the brain to develop in ways that make it more responsive to threat" (Twardosz and Lutzker, 2010, 64). In response to the stressful environment, the brain adapts to obtain greater chance of survival in circumstances of "deprivation and danger" (64). This can affect both the creation and pruning of synapses between neurons.

This adaptation can affect a child's attachment bond to her caretaker, who is a source of stress. Attachment can be strong even in physically abusive relationships: indeed, these attachment bonds are likely to be *stronger*, as terrorized individuals are more likely to need comfort (Cloitre et al., 2006, 14). Marylene Cloitre et al. describe this cycle, "Children assume that their parents have their best interests in mind, despite the abuse. Frightened and confused, they are seeking safety and security; the provision of any shred of care, affection, or attention will draw them closer to their parents" (15). Yet this is also a relationship fraught with fear, suspicion, and confusion, so while the bond is strong, the child has not learned the dynamics of healthy relationships (20). One possible way in which adaptation can manifest in such a situation, for example, is identification with the source of neglect or threat (Finzi et al, 2003, 388). The child is in a situation in

which he may either feel abandoned or betrayed by his caretaker, or he may come to agree with them that he is a "bad child", for example (388).

Further damage can be done in the development of emotions, particularly in the case of abuse. In a situation with healthy parenting, children learn to identify and explain their emotions and perceptions. For example, if a child lets go of a balloon and starts to cry, a parent may indicate that the child is sad that she lost her balloon, and then seek to mollify the situation somehow. Yet in abusive situations, the parent may react by acting derisively and suspiciously toward the child's emotions. This could even be the case after the child experiences physical abuse, such as claiming the child was not hurt or is lying about her emotions. In other words, in cases of healthy parenting, a child may learn to identify, explain, and then cope with her emotions: in cases of ELS, this process is absent or confused (Cloitre et al, 2006, 22-4).¹³ This extends not only to the child's own emotions, but also those of others. An abused child may react to the crying, for example, of others with suspicion and lack of empathy (Scannapieco and Connell-Carrick, 2005, 106-108). Kathryn Hildyard and David Wolfe explain, "Compared to... nonmaltreated children, the stories of the maltreated children suggested that they had a much more negative view of the social world, one that ... tended to be preoccupied with negative relational dynamics, to the exclusion of positive relational dynamics such as empathy,

¹³ Here a question may arise as to whether this process is *emotional* or *cognitive*. The definition of these terms and the distinction between them are unclear and have been debated at least since William James presented his account of emotional experience (Friedman, 2010). Yet increasingly it seems that this distinction is not stark, but fuzzy: emotion and cognition are considered interdependent and are taken to be jointly necessary for emotional experience (Friedman, 2010; Lazarus, 1991; Duncan and Barrett, 2007). Indeed, we can see even a biological overlap, as some systems serve both types of brain processes (Gray, 2008, 271-272). While we distinguish *cognitive* and *emotional* in language, empirical evidence seems to hold that this distinction does not hold ontologically in the brain itself (Gray, 2008, 270). As this is a side issue, there is not sufficient time to recapture the discussion about this topic, but it is worth noting that it is becoming more common to view these processes as in interplay: both are substantive parts to our emotional experience. See also Valerie Hardcastle's discussion about pain as an example of emotional experience that combines both aspects, a feature of emotion that is frequently overlooked by philosophers (Hardcastle, 2001). My thanks to Joe Pitt for noting this objection.

peer acceptance, and problem-solving" (Hildyard and Wolfe, 2002, 687). In other words, perception and identification of emotions, both important in attachment and relationships, can become skewed for an individual with ELS.¹⁴

A possible framework to understand such an atypical worldview could be provided by understanding the differences of *guilt* and *shame*, as moral emotions. This avenue of inquiry into the effects of ELS is recent, but there is evidence to suggest that individuals with ELS are more likely to develop shame complexes, which can actually be undermining to moral behavior. This understanding represents a possibility to relate aberrant behaviors-- such as increased aggression, suspicion, and other negative social relations and dynamics-- to an individual's self-image and behavior. I will describe some of the research on this topic in the next subsection.

iii. Shame and Guilt

Laypeople and psychologists alike frequently treat guilt and shame synonymously, yet many psychologists argue this is mistaken (Tangney et al, 1992, 669-670; Tangney et al, 2007; Webb et al, 2007; Dutton et al, 1995; Hoglund and Nicholas, 1995; Orth et al, 2006). Guilt is seen as a productive moral emotion, where shame is maladaptive. Guilt is negative affect toward action (or omission thereof): there is remorse for the specific occurrence. For example, an expression of guilt might be: "It was wrong to bump into that person because I was in a rush; she could have been hurt" or "I should

¹⁴ One may note I do not discuss precise numbers for the degree to which ELS is correlated with certain effects. Some of the literature provide statistics in discussing results of studies done with a select group of participants, but the number and type of participants vary across studies, making it difficult for me to discuss numbers. Further, some literature (namely review papers) do not discuss numbers much, if at all. Hence, the correlations are statistically significant, but data will vary from study to study. In future projects, I hope to resolve these issues more precisely, but they should be noted here. My thanks to Jim Klage for this point.

have helped my mother with the groceries; I regret not helping." Shame, on the other hand, is negative affect toward the entire self. The self is "painfully scrutinized and negatively evaluated" (Tangney et al, 1992, 670). Example expressions of shame might be: "I bumped into that person and she could have been hurt. I am thoughtless and mean" or "I should have helped my mother with the groceries; I'm just lazy and selfish." We can see in the examples the stark distinction between these emotions: guilt is recognition and remorse over an action, not remorse over being one's self. Guilt may be characterized by "My behavior is bad", whereas shame may be characterized by "I myself am bad" (Webb et al, 2007, 1144).

It should be emphasized that shame is an extremely painful emotion. The entire self is negatively evaluated, which may be accompanied by feelings of worthlessness, powerlessness, exposure, lack of control, and feeling small (Tangney et al, 1996, 743; Tangney et al, 1992, 670). While a person who feels guilt over an action may try to remedy the situation, one who feels shame may feel out of control, unable to make amends for some reason or other (Hoglund and Nicholas, 1995, 142). Shame has been positively linked with depression, both in childhood and adulthood, and is amplified through rumination (Orth et al., 2006, 1615).

More importantly, shame impairs one's relations to others in a variety of ways. First, shame can result in reduced empathy, when one reacts instead by avoidance or hostility (Orth et al, 2006, 1609; Hoglund and Nicholas, 1995, 142; Dutton et al, 1995, 127-9; Tangney et al, 1992, 670-1; Stuewig and McCloskey, 2005, 327). Shame can be accompanied by feelings of wanting to hide, or to disappear. June Price Tangney describes this vividly: "Shame motivates a desire to hide-- to sink into the floor and

disappear" (Tangney et al, 1992, 670). Importantly, this feeling can occur even when the individual is alone: shame need not be a public emotion. But wanting to disappear is only one measure of avoidance that shame can elicit. Another common measure is blaming and suspiciousness of others (Hoglund and Nicholas, 1995, 142), and outright denial of wrongdoing (Dutton et al, 1995, 129).

Shame also is linked to increased rates of hostility and aggression. Tangney hypothesizes that this could be the result of anger episodes motivated by shame. Shame motivates hostility toward the self, but this hostility can be projected onto others, who are seen suspiciously, or as rejecting the person feeling the shame (Tangney et al, 1992, 670; Hoglund and Nicholas, 1995, 142). According to Tangney, the accompanying feelings of lack of control and powerlessness become relevant, because the anger episodes may be irrational attempts to regain control of oneself. Tangney states, "From the initial passive and disabling experience of shame, the individual attempts to mobilize the self and gain control through active anger and aggression" (670). This move can in turn cause further shame, particularly if the individual recognizes the unjustness of her actions. Yet studies indicate that shame is "acutely painful", and it can provoke "an irrational and generally counterproductive rage reaction" (670).

Empirical evidence suggests shame is linked to anger, hostility, or aggression, and guilt is inversely related to these types of negative affect. One such study, conducted by Jeffrey Stuewig and Laura McCloskey, was a longitudinal study of children over eight years. The study showed shame-proneness was correlated with higher degrees of hostility and aggression among children, adolescents, and adults; guilt-proneness, on the other

hand, was correlated with lower degrees of hostility and aggression (Stuewig and McCloskey, 2005, 327).

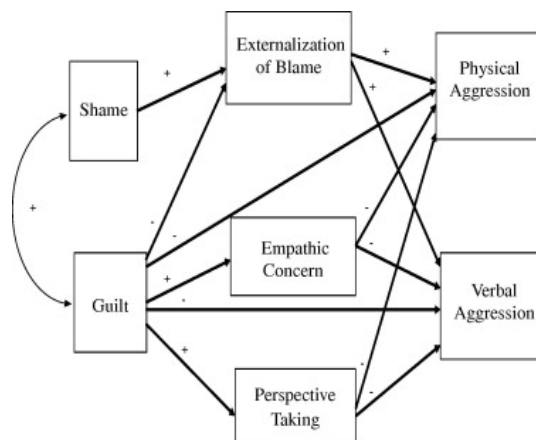


FIGURE 2¹⁵

In Figure 2, we have an image depicting the positive associations of shame with externalization of blame and aggression, negative with empathy, and the positive associations of guilt with empathy and negative associations with externalization of blame and aggression. Other studies indicate the same relationships demonstrated in the figure above (Stuewig et al., 2010; Tangney et al, 1992, 670-4; Tangney et al, 2007; Webb et al, 2007, 1146-1151; Dutton et al, 1995, 123-129; Hogg and Nicholas, 1995, 144-155). Tangney, Hogg and Nicholas, and Webb performed studies with college students; Dutton performed studies with males with a history of domestic violence, as he was studying the role of shame in transmission of abuse.

Another important positive association is that between ELS, particularly abuse, and shame-proneness. A number of studies suggest a positive association of abuse (of any type) with shame (Tangney et al, 2007, 13). Some have suggested that maltreatment in

¹⁵ Jeffrey Stuewig, June P. Tangney, Caron Heigel, Laura Hart, Laura McCloskey. "Shaming, blaming, and maiming: Functional links among the moral emotions, externalization of blame, and aggression." *Journal of Research in Personality*. 44:1. February 2010, Pages 91-102.

early life is the source for adult psychopathology associated with shame, and some studies indicate a positive correlation between maladaptive parenting styles and shame-proneness (Webb et al, 2007, 1144). In addition, some of the physiological correlates of shame are increased levels of cortisol and biological measures of increased stress, such as heightened cardiovascular activity (though this is not the *only* measure of stress, nor is heightened cardiovascular activity restricted to indicating *only* stress) (Tangney et al, 2007, 12).

June Tagney, Jeff Stuewig, and Debra Mashek suggest that empathy is not an emotion but an emotional process. It requires being able to take another's perspective, to understand accurately their emotional experiences, and to have a range of emotional experiences oneself (Tangney et al, 2007, 18).¹⁶ If we accept this explanation of empathy, we can see how shame can interfere with this process at each point. If a person is prone to shame, then he may not empathize with others, particularly if he is the wrong-doer. A shame-prone individual is more likely to want to avoid others and to blame them, even be angry at and hostile toward them. This makes it difficult to understand the experiences of others. Further, given the positive correlation between shame and severe depression, it is not clear that shame-prone individuals possess a full range of emotions themselves (Orth et al, 2006). On the other hand, guilt facilitates this process: it is positively correlated with empathy and a desire to make amends for wrong-doing (Tangney et al, 2007, 18-19). An abused child may come to see himself as deserving his treatment and may believe he is simply bad, and, without intervention, this self-image can persist into adult-hood, to the detriment of the individual's health and relations with others.

¹⁶ We will see in the next chapter that David Shoemaker has a similar definition.

Conclusion

This chapter aimed to review the psychological and neuroscientific literature on the effects of ELS, spanning across multiple levels of scope. ELS has profound physiological and psychological effects on the development of an individual. Individuals with ELS are more likely to develop sensitive HPA-axes and CRF systems, indicating vulnerability to increased stress- and threat-responses, depression, mood and anxiety disorders, and other health effects. In psychological studies, ELS has positive associations with depression, aggression, and a lack of emotional understanding. It further has a positive correlation with the development of shame-proneness, which can lead to hostility and anger toward oneself and others, and a lack of empathy with oneself and with others.

What I showed in this chapter is a brief sketch of the effects of ELS on an individual, both short-term and long-term, as described in psychological and neuroscientific literature; more importantly, I described that these individuals can have atypical perceptions and worldview. Individuals with severe ELS may have experienced childhoods fraught with violence, rejection, shaming, suspicion, and mistrust. We can imagine, or even read about, cases in which children would feel anxious nearly constantly because they never knew when a caretaker might hit them, or they would be lured close when frightened and then hit once they were close enough. They may experience pain and then be told that they are not in pain, or that they are making it up. Such individuals may see the world as hostile toward them, or may see themselves as somehow fundamentally flawed because they are treated in such ways. While the literature is recent

and not all results agree, it does seem to be the case that children in these environments attempt to adapt and make sense of their world, as every human does. Yet adapting to extreme circumstances of abuse or neglect, and trying to make sense of one's deprivation and maltreatment, can very well lead to behavior and emotions that are maladaptive in more passive environments (Twardosz and Lutzker, 2010, 63-4). It is this picture to which I will return in Chapter 3. Now I turn to reviewing the philosophical literature on moral responsibility.

II. MORAL RESPONSIBILITY

Introduction

There are multiple accounts of moral responsibility. Some have chosen to focus on why a person acts in the way she does; to this end they may address, for example, issues of the reasons a person has for her actions, whether these actions reflect the agent's character, etc. Others focus on the issue of *choice*: was an agent able to choose a different path, what conditions prevent one from being free to choose a certain path, and so on. I will categorize these as the 'reasons strategy' and the 'alternative possibilities' strategy, respectively. The former holds, roughly, that an agent's moral responsibility is constituted by her ability to respond appropriately to moral reasons (Talbert, 2008; Fischer, 1999, 127-128).¹⁷ The latter holds one is morally responsible for a given action so long as one freely chooses that action among different possible alternatives one might take. I will discuss these as strategies one might take in addressing the topic of moral responsibility. I will sketch these briefly in the first section.

The discussion is also not just about moral responsibility broadly construed, but also about the issue of demarcating who are morally responsible parties and who are not. Even differing accounts of moral responsibility hold similar criteria for determining when someone is morally responsible, though with varying justifications.

¹⁷ The way in which "reasons responsiveness" (as Fischer calls it) is discussed may seem to depart a bit from ordinary usage of 'reason'. But the idea, as will be explained in Subsection B, is that for any given act that we may judge as morally wrong/right, there are moral reasons (not) to do it. If an individual understands those reasons, then she may be held morally responsible for her actions (Fischer, 1999, 127-8). So, 'reasons' here indicates reasons typically held by a moral community, and individuals may respond (understand) those reasons or not. In this way, 'reasons' as a term becomes a bit more reified (similar to 'principle') than we might otherwise be used to. My thanks to Joe Pitt for noting this incongruity.

A strategy for determining who are morally responsible agents is to focus on why some parties are exempt or excused¹⁸ from moral responsibility; indeed, David Shoemaker cites this as a "typical strategy" (Shoemaker, 2007, 71). Favorite examples of exempt parties include psychopaths, autistic individuals, children, mentally handicapped individuals, and others, though it is controversial as to whether such parties are indeed not morally responsible agents. Yet the categories of people mentioned here are just for ease of example: they exemplify certain features that we consider exculpatory of blame or praise.

In this chapter, I will briefly explain some of the different views in the discussion on moral responsibility. I will then explain how the issue of psychological abnormality has been treated in the discussion of moral responsibility in Strawsonian accounts.¹⁹ This chapter will be divided into two sections. The first section will have three subsections; in the first two will be a brief overview of the reasons and alternative possibilities strategies in addressing moral responsibility respectively. The third subsection will then focus mainly on Strawsonian strategy. The second main section will describe when someone is not held to be morally responsible.

i. Different Accounts of Moral Responsibility

a. Alternative Possibilities

The alternative possibilities strategy generally holds that a person's being morally responsible consists in her ability to choose, free from ignorance or force, her path of action from among alternatives (Fischer, 1999, 98-99; Frankfurt, 1969, 829; Brown,

¹⁸ These terms will be explained in a later section. For now it will suffice to say that these are different ways in which a person may be found not to be morally responsible.

¹⁹ These are accounts that roughly follow P. F. Strawson's descriptive account of moral responsibility, to be discussed shortly.

2006, 265-66). Harry Frankfurt characterizes this principle, as he calls it, as: "A person is morally responsible for what he has done only if he could have done otherwise" (Frankfurt, 1969, 829). In other words, we cannot be morally responsible for options unavailable to us or out of our control, nor can we be held morally responsible for actions for which there is no available alternative. For example, if it is physically impossible for me to rescue an individual in distress, then the option of rescuing him is unavailable to me and I cannot be held morally responsible for not trying to rescue him.

We can construe this principle as claiming that in order to count as viable possibilities open to an individual, at least two conditions must be met. First, there must be more than one "genuinely" available path (Fischer, 1999, 98-99). In other words, the agent must be able to make a choice between alternatives that he could reasonably be expected to take. We theoretically have the available possibility to kill ourselves, for example, in most decisions that we make in order to avoid making the decision, but we would not likely consider this a genuinely viable option.²⁰ Rather, the emphasis is on relevant, available possibilities.

The second condition is that the agent is in control of his choice of possibility: he is not forced to choose, nor is his choice made by chance (99). For example, walking into a coffee shop, I can choose to buy a coffee, or rob the coffee shop, or mug all of the individuals in the coffee shop. These constitute alternative possibilities that I can freely choose among, so that if I choose to rob the coffee shop, I am morally responsible for that act, given that I had the available option of simply buying a coffee.

²⁰ We may claim, in some cases, that people should have died instead of acting; for example, we may think Nazis who had a choice between dying and perpetrating atrocities should have chosen the former. Yet in other cases, if a person is forced to choose between death and some given action (e.g., driving a violent criminal away from the scene of the crime), he is coerced. Hence, his action is excused, though being killed is a possibility open to him (it just may not be a genuine option). My thanks to Jim Klagge for noting this objection.

There are several difficulties for this view worth noting, though there is not sufficient space to examine them thoroughly here. The first is that this account, with its emphasis on choice, is vulnerable to the discussion on free-will and determinism. That is, it appears that if causal determinism (roughly, the arrangement of the world at T_1 entails the arrangement of the world at T_2 , and so on indefinitely) is true, then agents do not have alternative possibilities. John Fischer explains, "If my choices and actions are consequences of the past and laws of nature, then I do not have genuinely open alternative possibilities (of the sort required for moral responsibility)" (100). In other words, if the physical world is determined, then we and the choices we make, as part of that physical world, are also determined.

This is further complicated because the brain *is*, as Patricia Churchland calls it, a "causal machine" (Churchland, 2002, 204). If we want to hold that choice is *completely* free, or completely undetermined, then "choice" may be random, and so not what we would intuitively consider to be choice (as intentional or purposeful). If something is intentionally chosen, then it would seem to be determined at least in part (by beliefs, input stimuli, earlier brain states, or something else, depending on whether we have a belief-desire account of action, or some other type of account). Then the issue becomes how much an event or possibility could be determined while also being held to be "freely chosen": this difficulty is the one addressed by compatibilist pictures, which hold that free-will can be accommodated even in a deterministic world (Fischer, 1999, 100-103; Churchland, 2002, 203-236; Gazzaniga, 2005, 90-102).

A second important difficulty also concerns the issue of choice, in that in some Frankfurt cases,²¹ the distinction between what we are forced or coerced to do and what

²¹ Named after the philosopher who came up with them, Harry Frankfurt.

we choose to do can be blurred. For example, consider the case of Jones, who is forced to do something he independently decided to do. Say, for example, that Jones had chosen to kill Smith, his rival for a promotion. However, prior to his enacting this plan, another villain comes to Jones and forces him to kill Smith (say, for example, that he puts a knife in Jones's hand and threatens him at gunpoint with the claim that he will kill Jones *and* Smith if Jones does not cooperate).²² Because Jones is forced, he does not have genuine alternative possibilities open to him to perform otherwise, so it would seem that he is not morally responsible for killing Smith. Yet he had decided to kill Smith! Frankfurt asks, "Can we say of Jones that he was coerced to do something, when he had already decided on his own to do it...? Or would it be more correct to say that Jones was not coerced to do what he did, even though he himself recognized that....he had to do it?" (Frankfurt, 1969, 833). The problematic question here is whether we can be forced to do that which we want or have decided to do anyway, and the alternative possibilities principle does not recognize this distinction. However, others contend that there is a "flicker of freedom" in such cases where so-called coerced agents have freedom to change their plans (Fischer, 1999, 109-112; Brown, 2006, 268-272).

In this subsection, I have given a brief overview of the alternative possibilities strategy in approaching the question of who is morally responsible, and have also described two key difficulties for the view. I turn now to doing the same for the strategy of focusing on the reasons behind one's action for determining moral responsibility.

b. Reasons

The reasons strategy is to hold an agent's moral responsibility as constituted by her ability to respond appropriately to moral reasons for doing something (Fischer, 1999;

²² The details of the example are mine

Dancy, 2007; Vargas, forthcoming; Talbert, 2008; Levy, 2011). Common-sensically, we tend to think that we have reasons behind our actions, moral or otherwise. That I am out of milk is a reason for me to buy milk, or that I want to avoid being in trouble with the IRS is a reason for me to pay my taxes.

Likewise, there are reasons for moral actions, and holding a person morally responsible is to expect that they can respond appropriately to moral reasons (Fischer, 1999, 127; Levy, 2011, 243; Talbert, 2008, 516-525; Vargas, forthcoming, 2). There are some complications that arise in determining who can respond appropriately to moral reasons (e.g., it is controversial as to whether psychopaths can) and whether one must be conscious of the reasons. First, however, I will flesh out a bit more the basic view.

To be morally responsible, on the reasons view, is to be able to understand moral reasons.²³ Individual moral agents will act in accordance with those reasons and be redressed by others when they act counter to them. That is, if someone acts against a moral principle, the moral community can rightly claim that he or she had a reason *not* to act in such a way: namely, a moral reason (Vargas, forthcoming, 2). That murder is wrong, for example, is a moral reason; this reason should have an influence on my behavior (Fischer, 1999, 127). In the course of reasoning about whether I should murder someone or not, this should be an (overwhelming) consideration in my deliberations. Likewise, in judging someone for her actions, we will consider her reasons for acting as well as the reasons for refraining from acting (Talbert, 2008, 516-525).²⁴

There is some disagreement over whether one must be aware of her reasons or not in order to be held morally responsible for her actions (Levy, 2011). Some philosophers,

²³ We will see in the next section that this overlaps somewhat with Strawsonian views

²⁴ There can be a prioritizing of reasons, and even conflicting reasons. We may, and likely will, have some reasons outweigh others; Jonathan Dancy puts this in terms of favouring relations among reasons: we will favor one reason more than another (Dancy, 2007).

such as Susan Wolf and Gary Watson,²⁵ hold that one's actions should reflect one's character-- the reasons for action should be representative of one's self, but may not need to be consciously held at the time of the action (Smith, 2008, 368-371). Others, such as Angela Smith, hold that the actions need not be voluntary and conscious, but still must be in line with one's judgments and attitudes (383). Neil Levy, on the other hand, holds that we do need to be aware of these reasons at the time of our action. He states, "For them [his opponents mentioned above], as we have seen, an agent may be morally responsible for actions the moral significance of which goes unrecognized by him or her" (Levy, 2011, 247). He argues otherwise: that if we are unaware of the moral reasons involved in a given case, we may be excused (247).

There is also some disagreement among scholars as to what sorts of people have access to moral reasons; psychopaths are a favorite example. David Shoemaker and Matthew Talbert are among those who argue that psychopaths can understand moral reasons. Shoemaker argues that at least some psychopaths are able to carry on discussions about moral reasons, which would indicate that they have understanding of them (Shoemaker, 2007, 80-83). He states, "They can understand the consequences of their actions, engage in moral conversations, and grasp general moral principles" (81). What is lacking in their moral agency is their lack of *motivation* to act in accordance with moral reasons (81-83).²⁶ Talbert more or less agrees: he argues that psychopaths do not seem to

²⁵ These philosophers are not reasons theorists, but some aspects of their view may help clarify reasons strategy. My thanks to David Faraci for noting this.

²⁶ Yet, this aspect of his argument is unconvincing. At least some of the information he gives about psychopaths is inconsistent. For example, he claims "The psychopath is psychologically just like the rest of us except in virtue of his incapacity for complex emotions" (80). Yet first, if this were the case, then the psychopath is indeed not "just like the rest of us," as complex emotions are a key part to our psychology, and as processes in the brain are generally interconnected, it is unlikely that one can have a missing "part" with all else remaining the same. Second, there is empirical evidence to suggest psychopaths are deficient in negative affective emotions, which leads them to treat moral terms as equivalent to legal ones, so the argument here risks equivocation (see Prinz, 2006, 32). Further, Shoemaker's citations are to philosophers, and none to scientists of any sort on the topic. Even further, he supports his arguments with examples of

act compulsively, as do kleptomaniacs, and seem capable of prudential reasoning and understanding moral reasons derivatively (Talbert, 2008, 520-525). Neil Levy, on the other hand, argues psychopaths do not have access to moral reasons and cannot control their actions to be in accordance with moral reasons (Levy, 2008, 129-132).

A major objection to moral responsibility accounts that rest on the reason strategy comes from *situationist ethics*, which investigates moral judgments using social psychology (Vargas, forthcoming; Upton, 2009; Prinz, 2006). Psychological research on making moral judgments has indicated that our moral judgments may be swayed by our emotions (not relevant to the judgment we are making) and our physical environment. Jesse Prinz notes empirical findings that indicate individuals make harsher moral judgments when they have negative emotions or when they are in a disgusting environment (Prinz, 2006, 30-32). Manuel Vargas similarly points out that studies have shown, for example, that people are more likely to be good samaritans when they have benefited from a stroke of good luck (Vargas, forthcoming, 3-5). These social psychological studies offer some reason to doubt the claim that our moral judgments and decisions are primarily reason-based, given the influence emotions and environment have on those judgments.²⁷

In this subsection, I have sketched the reasons strategy, which at first seems very close to our common sense intuitions about how we make our decisions and judgments. Yet empirical findings from social psychology casts doubt on whether we fully

fictional psychopaths such as Hannibal Lecter and Eric Cartman (*South Park*). For these reasons, we should at least be skeptical of the claims Shoemaker makes about psychopathy.

²⁷ Note that this does not address the question of whether we are responsive to reasons, only that our moral judgments or decisions may not be the result of a reasoned process. Yet the point here is that reasons accounts hold that we are responsible for actions when we have a *reason* to act otherwise, and situationist ethics is denying this sort of rational agency (Vargas, forthcoming, 7-8). My thanks to David Faraci for noting this objection.

understand our reasoning process from the armchair. In the next section, I will begin detailing Strawsonian accounts of moral responsibility.

c. Strawsonian

It is important to note P. F. Strawson's motivations for his account of moral responsibility that he presents in "Freedom and Resentment".²⁸ In this essay, Strawson is concerned with the metaphysical debate over free will and determinism, and he argues for an "optimistic" way out of the debate. If we are determinists, we may be "pessimists" or "optimists" about our moral practices. Pessimists will claim that we should retain moral practices (to be described shortly) for practical purposes: our practices lead to efficient cooperation in society, for example.²⁹ However, Strawson argues for an optimistic stance, which is that our moral practices are part of our human nature; hence, even if determinism is true, we will still retain our practices surrounding moral responsibility. He states, "The human commitment to participation in ordinary interpersonal relationships is, I think, too ... deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships" (Fischer and Ravizza, 1993, 54). By this he means that metaphysical hypotheses on free-will and determinism do not affect our lives in a substantial way; even if determinism were true, we would still have the same expectations of goodwill from those in relationships with us.

Strawson distinguishes his position from the pessimist's by emphasizing that moral practices are in our very nature. Strawson states, "It is far from wrong to emphasize the efficacy of all those practices [as the pessimist does]...what *is* wrong is to forget that

²⁸ My thanks to Ted Parent for noting this.

²⁹ This position is similar to what Patricia Churchland suggests. See Churchland, 2002.

these practices... really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes" (Fischer and Ravizza, 1993, 66). This passage captures the distinction between pessimists and optimists: pessimists see moral practices as useful devices for effective social cooperation and optimists see them not just as devices, but also as genuine expressions of our attitudes toward others. In the following section, I will describe further what those practices are.

Strawsonian accounts generally hold, roughly speaking, that to be morally responsible for an action or event is to be held morally responsible for it by the moral community.³⁰ In other words, there is no status of being "morally responsible" beyond what your community holds you morally responsible for. Strawson argues that whether we hold a person morally responsible in practice leads to our understanding of our judgments about whether a person is morally responsible. Others have modified his view slightly to add that individuals should be *appropriately* held morally responsible by the community (Fischer and Ravizza, 1993, 15-19). R. Jay Wallace, for example, puts this in terms of holding someone morally responsible as defining that person's being morally responsible, so long as the individual is fairly held to be morally responsible (Fischer, 1999, 95). Fischer suggests using terms of "appropriate recipients" of others' reactive attitudes (Fischer and Ravizza, 1993, 18).³¹

Strawson importantly likens the moral relationships between people to other sorts of relationships: friendships, familial relationships, romances (49-50). Each person in the moral community is in a moral relationship with each other member. As with any relationship, there are expectations about how one should be treated and how one should

³⁰ This captures Strawsonian accounts generally; Strawson's account is slightly different.

³¹ This term will be explained shortly.

treat others. Gary Watson captures this nicely: "The idea... that we are responsible is to be understood by the practice...of expressing our concerns and demands about our treatment of one another" (Watson, 2012, 4). We demand and expect good will from others, and holding each other morally responsible is our *practice* that expresses and enforces those expectations.

Strawson cashes out the notion of holding an individual responsible with reactive attitudes: those attitudes that arise from social interactions, such as praise, blame, anger, etc. (Fischer and Ravizza, 1993, 19, 48-50, 56-60). Reactive attitudes are the expressions we have about how we have been treated: they express whether our expectations of good will have been met. What this means, essentially, is that we hold that someone acted wrongly because we, as a community or as individuals, *feel* wronged by that person.

Reactive attitudes may be expressed not just on behalf of oneself, but on behalf of any member of the moral community or of the community itself as *moral indignation* (Watson, 2012, 4). Strawson states, "The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves [or others]" (Fischer and Ravizza, 1993, 57). In other words, there is a default assumption about how others should treat us: if a person expresses *ill will* toward us, we will react with *resentment*, a negative reactive attitude (56).

It is important to note that these reactive attitudes are not "mere effusions of feeling, unaffected by facts," as Watson expresses it; yet neither do reactive attitudes require "rational justification" (Watson, 2012, 4). Rather, reactive attitudes are based on our assessments of the intentions of others and their actions toward us, as well as

conventions that we are taught and hold as a community (Watson, 2012, 5; Fischer and Ravizza, 1993, 53-4; Hume, 1739, B3:P2:S2). Strawson states, “What I have called the participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions” (53). By this he means that we react both to actions and to attitudes, which will become important when we discuss excusing and exempting conditions. If a person bumps into me, I might be annoyed at first, until I realize that he had tripped on an uneven sidewalk. Because he did not intend to bump into me and did not show ill will toward me, and because there was a clear cause beyond his control for the event, my negative reactive feelings would dissipate.

One toothsome aspect about Strawson's account is that it allows him to skirt the debate surrounding free-will and determinism. He holds that our practice of holding others morally responsible expresses our *nature* as human beings (Fischer and Ravizza, 1993, 66). Our behavior and practices will not change depending on whether or not we have free-will: we will continue to have friendships, emotions, etc., and we will continue to maintain moral communities (45-48, 64-66). According to Strawson, we may modify our account of reactive attitudes and moral practices, but as they are part of our nature, we should not abandon them, even in light of "panicky metaphysics" (66).³²

One objection to Strawson's account, as Fischer and Ravizza note, is that we tend to think there is a substantive difference between *holding* someone responsible and that person actually *being* responsible (18). It seems at least possible that an individual could be held by his community to be morally responsible for some action that he does not

³² We may think that the same might be said of reasons accounts. However, situationist ethics may dispute the claim that the moral practices described by reasons accounts are *actually* in our nature, as there may be empirical evidence (from social psychology) that would suggest otherwise. My thanks to David Faraci for noting this possibility.

seem to be responsible for.³³ For example, we could imagine a moral community that felt resentment toward eccentric individuals who do not abide by communal norms. We might object that resentment is reserved for objecting to injury or offense, but we could in turn imagine that the community is greatly offended by these eccentric individuals (e.g., many in the United States are offended by flag-burning and may claim some sort of injury is committed by the action). What actions we are blameworthy or praiseworthy for seems somewhat at the whims of the moral community of which we are a part. This may be worrisome as it may seem arbitrary, or if it is the result of historical contingency,³⁴ then it may lead to concerns of relativism.

We also might worry about the justification of reactive attitudes, given empirical findings from social psychology that present a problem for the reasons strategy. Angela Smith notes similarly that our intuitions about whether or not it is fair to hold an individual morally responsible can be influenced by irrelevant factors (Smith, 2007, 466). Smith's notion can be supported by social psychological studies, which have shown these factors can be as arbitrary as the mood or physical environment of the person making the moral judgment. Thalia Wheatley and Jonathan Haidt, for example, conducted studies in which participants were given posthypnotic suggestions to feel disgust when encountering a particular word, e.g., 'take'. This feeling of disgust sharply increased the degree of severity individuals had in their moral judgments. In one study, they were presented with a scenario involving a high school student council member, Dan, who attempts to schedule topics of discussion interesting to all members of his academic

³³ Of course, this objection would simply be incoherent to a Strawsonian, as being held morally responsible and being morally responsible are the same thing. If we put that to the side for now, and think of a different definition of moral responsibility, then someone being *held* morally responsible and actually *being* morally responsible come apart. My thanks to Joe Pitt for noting this.

³⁴ My thanks to Joe Pitt for this point.

community. Despite the innocuousness of the scenario, individuals who were disgusted pronounced harsh moral judgments on Dan. Some claimed he was "up to something", others that he was a "popularity-seeking snob" (Wheatley and Haidt, 2005, 780-783).

Nevertheless, Strawson's account maintains its appeal, especially as it is malleable in light of these concerns. In particular, as we will detail in the next section, he makes room for cases in which the suspension of reactive attitudes is appropriate. Further, he indicates that our beliefs and practices are modifiable, should the community feel there is sufficient reason to do so (66). So, for example, in light of objections from situationist ethics (drawing upon social psychological research), we may be more rigorous in justifying reactive attitudes and moral judgments. In Chapter V, I will do something similar.

In this subsection, I have provided a brief overview of Strawson's account of moral responsibility. One aspect of this account that I have not yet addressed are those cases in which we suspend our reactive attitudes because the individual is not an appropriate target for those attitudes. In cases of excusing behavior that leads to negative reactive attitudes, one takes an objective standpoint from which one can determine whether the agent in question is either a) responsible for a given action or b) capable of engaging in moral relationships. In other words, if a person is able to participate socially with the community, then she is judged as a moral agent by others' reactive attitudes. Yet even moral agents may be excused for some behavior, if they are held not to be at fault (50-55). However, this overlaps considerably with the discussion of excusing and exempting conditions, so I will discuss this issue more thoroughly in the next section.

ii. Excusing and Exempting Conditions

Examining why an individual is not morally responsible can help shed light on why individuals *are* morally responsible. As David Shoemaker notes, one way to parse conditions for membership in the moral community is to examine why the outliers are outliers (Shoemaker, 71). Similarly, we can try to understand what makes an action inexcusable by considering what we hold to be excusable.

On many accounts of moral responsibility, a person may be held not to be morally responsible or blameworthy for a variety of reasons. Speaking generally of different account of moral responsibility, these possible reasons fall into one of two categories: 1) exemptions—e.g., for those who are severely psychologically disabled, or who are children, etc.—and 2) excuses—e.g., for cases in which an individual is acting against her will, or in ignorance, or by accident. Excusing conditions are those that show the individual is not to blame: e.g., cases of accidents, or being forced by another person or circumstances. Exempting conditions are those that apply when a person is exempted from the moral community, e.g., for reasons of mental deficiency (Fischer and Ravizza, 1993, 5-8; Watson, 2012, 4-10).³⁵ The Strawsonian literature, specifically, is not much different than the general approach. An individual is free from moral responsibility if she falls under an exempting condition or if her action is excused (Fischer and Ravizza, 1993, 6, 20).

³⁵ Another means by which we may consider an individual not to be morally blameworthy is if he has a justifying reason for his action. If a person has sufficient reason for performing some action—e.g., lying to Nazis at the door to protect others—then we may say this reason justifies the action taken (lying, in this case). Here we will hold a person responsible for his action and yet not blame him for it, even if we think that lying is generally wrong (Buss, 1997). However, this could probably be construed as being somewhat *coerced* to lie (hence an excused action). It seems we could probably capture justifying reasons under either exempting or excusing conditions, so I will stay with this method of explaining the suspension of judging someone to be morally responsible.

Members of the moral community will hold an individual morally responsible unless he warrants “special considerations” (Fischer and Ravizza, 1993, 50). These special considerations are to be assessed from an objective point of view and serve to “inhibit” our negative reactive attitudes toward others (Watson, 2012, 5-6). In other words, we may excuse the actions of the person because the person himself is not to blame: rather, there are circumstances that lead to the wrongdoing. Or, we may come to realize that the person was not an appropriate agent to place moral demands on in the first place: that is, we may exempt a person from moral responsibility (Fischer and Ravizza, 1993, 50-55; Watson, 2012, 5-6). Let us look at these conditions further.

Excusing conditions are those that show the individual is not to blame: e.g., cases of accidents, or being forced by another person or circumstances. Strawson refers to excusing conditions as a kind of plea. He characterizes this kind of plea as follows:

None of them invites us to suspend towards the agent, either at the time of his action or in general, our ordinary reactive attitudes. They do not invite us to view the *agent* as one in respect of whom these attitudes are in any way inappropriate. They invite us to view the *injury* as one in respect of which a particular one of these attitudes is inappropriate. They do not invite us to see the *agent* as other than a fully responsible agent. They invite us to see the *injury* as one for which he was not fully, or at all, responsible. They do not suggest that the agent is in any way an inappropriate object of that kind of demand for goodwill or regard which is reflected in our ordinary reactive attitudes. They suggest instead that the fact of injury was not in this case incompatible with that demand’s being fulfilled, that the fact of injury was quite consistent with the agent’s attitude and intentions being just what we demand they should be. (51, underlined emphasis added)

In this passage, Strawson picks out several general features that he takes to characterize excusing conditions. Importantly, the focus of the excuse is on the result of an action and not the agent in question. Our reactive attitudes toward the agent are not suspended: rather, the negativity of the attitudes is mitigated by recognizing that the result should not

be attributed to the agent. The character or personality of the agent in question does not seem to matter in determining whether his action is excused in many cases where these conditions apply. We can imagine interchanging any individual into the agent's spot with the same result: any agent in that specific situation would not be found morally responsible. The exemption is independent of the agent in this way. This is what Watson means when he states that excuses may be characterized as saying the individual is not to blame (Watson, 2012, 5).

Exempting conditions, on the other hand, are quite focused on the agent. Exempting conditions are those in which an agent cannot be considered to be part of the moral community, and this can manifest as permanent or temporary. Children, for example, are "permanently" exempt (at least, until they reach adulthood), whereas temporary insanity, being temporary, is not a case for permanent global exemption; rather, an individual with temporary insanity is temporarily exempt. We may think of temporary exemptions, following Strawson, as those that apply when the agent is not "himself" (Fischer and Ravizza, 1993, 51). Strawson states that cases where exempting conditions apply (whether permanently or temporarily) are cases in which the agent is "incapacitated in some or all respects for ordinary inter-personal relationships" (55).³⁶ Watson cashes this point out quite nicely: "[exempting conditions] bear upon the question of whether the agent is...seen as a responsible agent, as a potential term in moral relationships, as a member (albeit, perhaps, in less than good standing) of the moral community" (Watson, 2012, 5-6). Put another way, the question about whether someone is exempt from moral responsibility is really a question about whether that person can be

³⁶ One might object here that certain excusing conditions--e.g., having a gun to one's head-- are captured by this description also. But Strawson has in mind here psychological incapacitation (as we shall see shortly); presumably, an individual who is *forced* to act a certain way under threat still has the capacity to recognize what she is doing. My thanks to David Faraci for noting this.

in moral relationships with others: that is, whether we can actually make a moral demand of that person in the first place and, more importantly, whether that individual is open to moral address by the community (5-9).

Importantly, Strawson emphasizes that individuals permanently exempt from the moral community *are themselves*.³⁷ In these cases, the circumstances are "normal" but the agent is "psychologically abnormal" (Fischer and Ravizza, 1993, 52). Strawson states, "The agent was himself; but he is warped or deranged, neurotic or just a child" (52). Hence, it is not the case that an individual is exempt because of a condition (psychological or otherwise); rather, the individual *is* psychologically abnormal and so cannot be a member of the moral community (hence, he is exempt from moral responsibility and the moral community). He *himself* is unable to participate in moral relationships. This is what is meant by exemptions from moral responsibility for Strawson.³⁸

The example of psychopaths may help shed some light on exempting conditions. Much focus in the literature has been on psychopaths, who are considered by some scholars to be globally exempted: many scholars debate whether psychopaths are free from moral responsibility in all cases and whether they are not able to participate in the moral community (Buss, 1997, 337-338; Watson, 2012, 12-17; Fischer and Ravizza, 1993, 20-25; Levy, 2008, 129-131; Shoemaker, 2007, 77-85). Watson explains that while we may react negatively to such individuals, we cannot consider such individuals as able to participate in "moral relationships" (Watson, 2012, 5). We can think of this

³⁷ This will be important for Ch. IV, in which I address the common intuition that individuals are exonerated when 'their brain makes them do it'.

³⁸ This contrasts with the point I will discuss in Ch. IV. Briefly put, some will hold that a person cannot be responsible for his biological conditions that make him do certain actions. Hence, an *individual* should be exempt because what caused the action was his biology, not his 'self'. Unfortunately, Strawson's point here gets lost in the discussion on exempting that comes after him (e.g., Watson, 2012).

pragmatically as well: those who cannot understand what it is to be a participant in a moral community cannot really be held responsible in failing to fulfill that role. Rather, they are simply outside of the moral community (Watson, 2012, 12-17; Levy, 2008, 129-135).³⁹

It should be noted that we need not take these two categories—exempting and excusing conditions—as exhaustive of the possibilities. Strawson himself states, “It needs no saying now how multifarious these [special] considerations are” when it comes to incidents in which we must inhibit our reactive attitudes (Fischer and Ravizza, 1993, 50).⁴⁰

Conclusion

In this chapter, I have provided an overview of two important topics in moral responsibility. First, I covered some of the different approaches one might take to answering the question of why we find someone morally responsible, looking in turn at strategies from alternative possibilities, reasons, and Strawson. I also noted some objections for each strategy.

Second, I gave an overview of the reasons we might find someone *not* to be morally responsible. I have provided a brief overview of the exempting and excusing conditions that allow us to suspend moral judgments on an individual. While these may be fairly in line with common intuitions about moral responsibility, there is some

³⁹ This is controversial. Some, such as Shoemaker and Talbert, will hold that psychopaths are not exempt individuals. See subsection *b* of this chapter.

⁴⁰ It should be noted also that there will be varying degrees of moral responsibility; moral responsibility is not always an all-or-nothing affair. We may be *partly* responsible for actions. Yet Strawson himself (as well as others) does not address the variations of reactive attitudes or the variety of different cases that could be addressed, and admits he must ignore them and deal in "crude dichotomies" (Fischer and Ravizza, 1993, 50, 52). My thanks to Ted Parent for noting this point.

question as to whether that is a *good* thing. I will elaborate on this claim in ensuing chapters.

In the next two chapters, I will elaborate on some reasons to be skeptical of excusing and exempting. First, this distinction is already fuzzy, but furthermore, some cases can blur the distinction as provided. I will argue that an individual with ELS could provide such a case. Second, both excusing and exempting an individual on the grounds of psychological reasons leads us into a strange dualistic picture of responsibility. In short, we will need to modify the distinction and modify what or whom we hold responsible for events and actions in light of psychological evidence. I will propose the first modification at the end of Ch. III, and the second modification in Chapter V.

III. EXCUSING/EXEMPTING CONDITIONS

Psychological and neuroscientific studies about the long-term effects of Early Life Stress (ELS) provide an interesting case to examine under a Strawsonian framework of excusing and exempting individuals from moral responsibility. The reason for this peculiarity is that we can imagine cases of wrong-doing in which an individual who endured ELS acted wrongly but understandably (i.e., she would not seem to be responsible). Yet, her action would not be *excused*, strictly speaking, because the action itself is *wrong* given the circumstances. But neither would we consider her *exempt*, because she can be a term in moral relationships. Hence, it seems something interesting is going on in this potential case. I will devote the rest of this chapter to this discussion.

First, I must clarify at least some of the effects of ELS on which I will focus.⁴¹ Studies have shown that individuals who endured some form of ELS are statistically more likely to have both atypical behavior and, correspondingly, atypical neurology. ELS can result in neurological effects on individuals, even as adults, that can diminish their ability to engage healthily with others, affecting many, if not all, of their social relations (Cloitre, Cohen, and Koenen, 2006). More specifically, and related to the topic at hand, ELS statistically produces heightened fear and stress responses in many survivors, even into adulthood (Loman and Gunnar, 2010; Twardosz and Lutzker, 2010; Gould et al., 2012; and others). A threat response can be triggered in an individual who had ELS in atypical conditions as the stress and threat systems have irregularities on a structural

⁴¹ Most of this is simply drawing attention to information covered in Chapter I.

level. By this I mean that ELS has been shown to increase the level of glucocorticoids in an individual, which directly affects her amygdala system functions, the system which plays a large role in the body's 'threat system': in effect, the higher level of glucocorticoids leads to a heightened responsiveness in the threat system.

In addition, studies of abused children have indicated special attenuation among subjects to anger cues. Pollak and Sinha (2002) found abused children to be more likely than non-abused children (the controls) to identify facial expressions (in a continuum of neutral to angry expressions) as angry. Similarly, Pollak and Tolly-Schnell (2003) found that abused children had more difficulty disengaging from angry, but not happy, faces than did controls (measured by increases in brain electrical activity). One last example is that Pollak and Kistler (2002) saw results that indicated abused children had "atypical perceptual preferences" in gauging the intensity of angry faces, but not other facial emotions (Pollak and Sanchez, 2009, 505). In other words, the abused children were more attenuated to the differing intensities of displays of anger than controls, and less so for other emotions. M. Mar Sanchez and Seth D. Pollak state, "These findings are consistent with the view that infants need to adjust or tune their preexisting perceptual mechanisms to process salient aspects of their environments" (505). Abused children also are more likely to remember and assimilate angry expressions and cues than non-abused children (502). Briefly put, we can understand this data as indicating that abused children seem more likely to develop a focus on anger: they develop in volatile and unpredictable environments, so they become especially attenuated to cues of impending threat (namely, anger cues).

This narrative can carry through to adulthood. Those who had ELS are more likely to have difficulty with emotional awareness and regulation, which can make adult interpersonal relationships difficult. Some of the symptoms that survivors of ELS may exhibit are anxiety, distrust of positive emotions or interactions, confusions about feelings, and aggression. When abused children, for example, react in pain, their abusers may deny that there *is* pain or that the children are wrong to feel pain. As such, in adulthood, these individuals may have difficulty recognizing their emotions, and may believe the lessons they learned about vulnerability applies to others (so, in cases where other people are upset, some individuals with damaging childhoods may not fully empathize) (Cloitre, Cohen, and Koenen, 2006, 129-155).

In broader and less precise terms, we might think that ELS can result long term in hyper-defensiveness in an individual, even into adulthood (Loman and Gunnar, 2010). If we think about the nature of child abuse, this result does not seem very surprising: the individuals in early life live generally in a state of fear and uncertainty, so being “on guard” against possible threats is understandable (we might think). Roughly, that such individuals have heightened threat and stress response systems and are attuned to cues of anger supports this narrative. This is not to say that all individuals who suffered ELS could be so characterized, but we can imagine that some individuals may exhibit these symptoms.

Before I continue along this route, I should address a possible objection that these neuropsychological considerations could support the claim that individuals who had ELS are permanently *exempt* from moral responsibility. I will here address the possibility of permanent exemption, and will leave temporary exemption to the side for now. It would

be excessively strong to exempt such individuals *permanently*, as it does not seem as though they are incapable of forming moral relationships with others. There would not be an explanation for morally wrong action, for example, in cases where nothing occurs or is present that could possibly be taken as a threat by a given individual. And there is not sufficient empirical evidence to indicate that individuals who suffered ELS are incapable of understanding moral reasoning or devoid of empathy, though as previously noted, some individuals display less empathy in certain situations than non-abused individuals. Rather, they seem to meet the typical requirements for engaging in moral relationships: they can have empathy and can understand moral reasons (Shoemaker, 2007; Tangney et al, 2007, 18-19).

For the remainder of the paper, I will consider a hypothetical individual with a history of ELS—Ralph—who is hyper-defensive: as is the case with some individuals who experienced ELS, Ralph is attenuated to expressions of anger and has an atypically heightened threat and stress response system. We can imagine a case in which Ralph is “provoked” to “defensive” action that we would consider to be morally wrong. Say, an acquaintance of Ralph’s says a certain insult that Ralph’s caretaker used to say before she abused him. The acquaintance is blocking the only exit to the room. Ralph strikes the acquaintance and bowls over him so that he may get out of the room, feeling that he is threatened. Is he morally responsible for harming the acquaintance? Let us consider this question in terms of trying to determine whether Ralph is exempted or excused from moral responsibility in this case.

We have already discussed permanent exemptions, which do not seem to apply to Ralph or individuals like him, so let us turn to temporary exemptions. There is some

hesitation to considering Ralph temporarily exempt. For one, many examples of temporary exemption make it seem as though temporary exemption is a “one time deal,” so to speak. Such examples include cases of hypnosis, temporary insanity, being slipped a mind-controlling drug, etc. Not only does Ralph’s case seem repeatable, but the source is not *random* as we may consider hypnosis and mind-controlling drugs to be. Moreover, while Ralph’s reaction makes him seem unreasonable, it is not clear that even in the moment of his reaction that he is incapable of being in a moral relationship with others. What makes Ralph’s reaction atypical is that he has a *skewed perception* of the circumstances and environment he is in, which leads him to act *wrongly*. However, it does not follow from this that he lacks empathy or understanding of moral reasoning, *even temporarily*.

To push this point a bit further, imagine that what Ralph perceives is actually the case: his acquaintance, Bob, really *is* threatening him. Bob is insulting and menacing Ralph, and displays intent to act violently. Ralph, feeling threatened, strikes Bob and bowls him over in order to get out of the room, as Bob is blocking the only exit. We would likely feel here that Ralph’s actions are largely *justified*: what he displayed was not exactly intent to *harm Bob*, rather intent to escape a dangerous situation. Hence, Ralph’s ability to reason morally and to have empathy with others does not seem at issue in the case where he *perceives* the situation to be this way (despite that it actually is not). So, Ralph does not seem exempt, permanently or temporarily, on that basis.⁴²

Let us consider the possibility that Ralph is *excused* from moral responsibility.

Recall from earlier that according to Strawson, excuses “invite us to see the *injury* as one

⁴² In other words, Ralph meets the requirement for being considered a member in the moral community. It is not the case that he is “morally undeveloped” (Fischer and Ravizza, 1993, 52). What is being pointed out here is that when Strawson discusses psychological abnormalities as reasons for exempting individuals, he does not seem to have in mind individuals like Ralph.

for which he was not fully, or at all, responsible” (Fischer and Ravizza, 1993, 51). An individual is excused on the basis of not being in control of the action: the individual could not have acted otherwise, and so the action itself cannot be attributed to the agent when it comes to assessing moral responsibility. For example, if a person is physically moving my arm so that I stab someone, I am excused for the stabbing—I am not morally responsible.⁴³ If I am ignorant of the consequences of my action—e.g., a child is lying on the road disguised as a speed bump and I run over him—then I would not be held to be morally responsible. As mentioned earlier, for excuses, the focus is heavily on the circumstances and the action, not as much on the agent.

For this reason, Ralph’s case does not seem to be one in which we would consider him excused for the actions he took. The focus is on *Ralph* (as it would be with considering him to be exempt or not) and not on the actions. Consider if we exchange Ralph for a neurotypical individual, Jane, in our thought experiment. If Jane were to act in the same way in the same situation with Bob, we would consider her to have unnecessarily attacked Bob in response to an insult. We would find it morally wrong that she escalated the situation to one of violence and that she injured Bob. What makes the situation difficult to assess is precisely the presence of Ralph, who is doing the acting. To consider Ralph excused, in other words, seems to confuse the causal story behind his actions: what is at issue is that Ralph suffers from skewed perception, leading him to act in abnormal ways. Excused actions, on the other hand, seem to be characterized by a lack of *control* of one’s actions because of external circumstances—and Ralph is not out of control, in this scenario, rather he is acting mistakenly in self-defense.

⁴³ Indeed, we may not even regard it as *my* action. My thanks to Jim Klagge for noting this.

At least in this case, then, exempting and excusing conditions do not fully and separately capture our moral intuitions about whether Ralph is blameworthy or not. Considerations of exemption seem to have some relevancy because Ralph's psychology is the key ingredient in why we may find him not to be morally responsible for his actions. Yet, as is captured in considerations of his actions being excused, we would not consider Ralph to be ineligible as a term in moral relationships (so he is not exempt from the moral community): rather, we think that the circumstances played a heavy role in his action because they *triggered* a learned and emotional response from him. Had the situation been different, Ralph may not have acted in the same way.

In this chapter, I have presented an example case that illustrates the vagueness of the exempting and excusing distinction. However, it is possible that this lack of clarity is unimportant. Whether or not we consider exonerating individuals on the basis of psychological reasons to be excusing or exempting does not seem worrisome in itself.

Yet if we change the distinction to be that between external/internal explanations for actions, then we can avoid the vagueness illustrated in this chapter. Instead of discussing whether someone is excused or exempted from moral responsibility for a given action, we can discuss instead whether she is exonerated for external or internal reasons. That is, we can posit that if external circumstances forced her to behave in such a way (as when she is forced at gunpoint to act), then there is an external reason for her to be exonerated. If a feature of her biology caused her to make a given action, we may determine there is an internal reason for her to be exonerated. For example, in the case of an individual having a seizure that hits another person, there is an internal explanation to

be given that would exonerate him. Subsequently I will refer to individuals being 'exonerated'.

But there is still a worry left in this discussion. In discussing reasons for exoneration based on psychological abnormalities, we tend to treat the "individual" in question as distinct from her brain. That is, our treatment of psychological defects in the conversation on moral responsibility leads us to a dualistic picture of people. Hence, even if we characterize the new distinction as between external/internal explanations, we need further clarification on what we mean by exonerating an individual because of internal features. The next chapter will clarify the concern about dualism and I will propose the solution in Ch. V.

IV. RESPONSIBILITY AND DUALISM

Concerns about one's psychological development as related to issues of moral responsibility seem to manifest frequently in one of two forms.⁴⁴ First, there is a common intuition that we are not responsible for our biology. Second, there is a common intuition that individuals have control over their life-path. The thought is that our fate is not decided by our genetics and events that happen to us; rather, we decide our actions based on our beliefs and desires.

Both of these intuitions, once fully explored, lead us to dualism regarding the mind/brain. I will assume, for the purposes of this thesis, that a dualistic picture of the self is not desirable. While I recognize that this is a somewhat controversial assumption, a defense of a reductive materialist (sometimes called 'physicalist') account of the mind/brain is a book, or series of books, in itself. While I recognize this may limit my audience, it should be obvious by this point that my discussion heavily relies on neuroscience as providing knowledge about the brain and how individuals function.⁴⁵ For other defenses of materialism and criticisms of dualisms, I would refer readers to philosophers who have written volumes on the subject: Paul and Patricia Churchland, Jaegwon Kim, Dan Dennett, and others.

⁴⁴ I do not mean that these are exhaustive, rather that these seem to be the more common types of intuitions on this topic.

⁴⁵ In other words, I have probably already limited my audience in such a way.

Let me be clear about this assumption: we *are* our brain.⁴⁶ My account is a reductive materialist one (so, mental states reduce to brain states), but that is not all I mean by saying "we are our brain". There is an intuition that certain parts of our brain activity make up our 'self' or 'personality' and that others do not. Typically, the former are associated with conscious, mental states. The notion seems to be this: the regulation of my blood-pressure is a brain activity, but it is not a mental (conscious) one and is not part of my personality, character, or self (however one wishes to characterize it). In discussing the autonomic nervous system, this is a fairly clear notion. Yet it becomes unclear when we start to talk about perceptual systems. *How* somebody perceives the world seems to be very much part of "them" and their "personality".⁴⁷ For example, depressed individuals see the world as grayer than it is (Prinz, 2012, 16). As mentioned previously, abused children are more likely than non-abused counter-parts to have special attenuation to anger expressions (Pollak and Sanchez, 2009, 505). We might think of these commonsensically as vicious circles: one might see the world as grayer because one is depressed, and the absence of bright, cheery colors may in turn reinforce depressive symptoms.

This will be more thoroughly discussed in the third subsection to this chapter, but it seems disingenuous to claim 'lower-level' brain operations (as on the perceptual systems) are not 'part of us'. It should be clear from preceding and subsequent discussion that we could not claim even that they are not part of 'the morally relevant' parts of us, as they are quite relevant in moral action and decisions (they provide the input upon which

⁴⁶ Jim Klage has pointed out that the grammar of "I am my brain" suggests there is an 'I' in possession of 'brain' (given the possessive 'my'). The grammar here is misleading, as Jim has noted. It would be better, really, to say 'the brain is me', or something along those lines.

⁴⁷ Many thanks to David Faraci for helpful discussion on this point.

we act). Hence, my position is slightly stronger than claiming just that mental states are brain states. We are our brain, not just the mental states.⁴⁸

Now that my assumptions are clear, I shall launch into the discussion of why this picture of moral responsibility may not be appealing to us who are materialists (of my type). This chapter will be divided into three sections. The first two sections are intended to lay out the intuitions in question. The first section will deal with the first intuition: we are not responsible for our biology. The second will address the second: we can overcome our backgrounds, and are responsible for doing so. The third section is to explain the relevant neuroscience and the implications it has for these intuitions.

i. "My Brain Made Me Do It"

We commonly think that we are not responsible for our biology. It is out of my control, for example, that I have brown eyes. Yet this can be complicated because some aspects ourselves *can* be changed or controlled, including our psychology. As noted in Chapter I, our brains are remarkably plastic. Yet there is empirical evidence to suggest that some features of ourselves may be unalterable. For example, it seems that psychopaths are incapable of empathy, which may be an untreatable condition.⁴⁹ But there are multiple questions at hand here: if we can control our biology (e.g., impulses), how much do we control? Further, how do we explain that controlling agent in neuroscientific terms?

I will progress through this section by starting with some illustrative cases to further explain the intuition at play here. First, I will consider a hypothetical case of a

⁴⁸ Though I would not extend this to the brain-stem and autonomic nervous system generally, at least in a robust way.

⁴⁹ Psychopaths are identified by behavior (see Shoemaker, 2007 for the criteria), but it is a biological condition. See, e.g., Weber et al., 2008.

color-blind individual, then the case of Phineas Gage, then the case of Ralph. We will see through that progression that the intuition becomes weaker. Second, I will demonstrate the concerns in these examples, particularly as they relate to dualism.

Imagine a color-blind individual, Kim, who is in an intractable situation appropriate to a philosophical thought-experiment. This situation is that she is before two differently colored buttons and is given instructions on what they do. The red button will diffuse missiles that are heading toward a city, and the blue button will divide the missiles so that both the original and an additional city will be destroyed by missiles. Kim has to push a button, and so she ends up pushing the blue button. Since she is color-blind, she could not differentiate between the buttons.

What we should explain here is not whether Kim is responsible, but *why* she is not. It is clear that this would be a case where Kim is exonerated of moral responsibility; she *could not* possibly know what color button she is pushing, and as she must push a button, it is random guessing on her part. In other words, her biology is such that the information required to make the morally right decision is inaccessible to her. Her perceptive abilities being atypical exonerate her from responsibility in this case.

Now consider the case of Phineas Gage, a man who suffered brain damage in 1848. After suffering damage to his frontal brain regions, Gage had profound psychological changes. His impulse control and social abilities were impaired. People close to him reported he was "no longer Gage" (Gazzaniga, 2005, 96). This raises interesting questions. Should we understand Gage's inappropriate actions as those his brain, affected by his brain damage, did, but that do not reflect 'who Gage is'? Or should

we hold that Gage changed (i.e., his brain changed, and he *is* his brain) and so he performed those actions?

Lastly, let us reconsider the case of Ralph, who struck his coworker, Bob. *Prima facie*, this case may seem remarkably similar to Kim's. Ralph's inability to perceive accurately his environment leads him to perform a morally wrong action. Yet the details of the two cases are different, and where Kim's case seems clearly exonerated, Ralph's is not. Even if we find Ralph not to be morally responsible, it requires far more investigation and explanation (even in spite of the fact that Kim's actions led to more injury).

Let us assume, for now, that Ralph is not morally responsible: the way that his brain functions is such that he interpreted Bob as a threat and acted in self-defense. Notice the way I have explained this to make it understandable and exonerated. Had I written, "Ralph is not morally responsible: he interpreted Bob as a threat and acted in self-defense," it may have given us pause. Without reference to the brain, the sentence suggests Ralph *actively* or *purposefully* interpreted Bob in such a way, not that it was automatic and instinctual. We might say that Ralph's actions do not reflect Ralph as an individual, similar to what is said in some accounts of moral responsibility (see Ch. 2b). Yet if we hold that Ralph perceiving something and his brain perceiving something are the same thing, then these sentences should have the same meaning.

If we are materialists, then we must accept that our brain makes us do *everything*. There is nothing that we do that our brain does not play a causal role in. But, one might say, we are unaware of many of the inferences and processes our brain undertakes. We are not aware of the coordinate-transformations our brain makes when we reach for the

cup of coffee that we see, for example (Churchland, 2002, 59-81). Yet I can *consciously* make the decision to have a cup of coffee. Here it seems that 'my brain' is aware of processes that 'I' am not, but both are aware of the intended result (i.e., grabbing the coffee).⁵⁰

However, this does not get us out of our dilemma yet. Here is our dilemma: we have an intuition that our brains perform actions without "us" being in control or, sometimes, even conscious of them. But how do we explain what we mean by "us"-- the controlling agent-- without holding a dualistic picture of the brain? For what we consciously decide to do is just another series of brain processes: just as with *unconscious* activities, conscious activities rely on information received from our environment that gets interpreted and transferred throughout our brain (Churchland, 2012, 35-90).⁵¹ In all three of the cases above-- Kim's, Gage's, and Ralph's-- there is some "flaw" in the system (i.e., the brain) that either prevents or skews the reception, interpretation, or transference of information. For any activity we perform (consciously or unconsciously) there is a causal story to be told by way of the "inputs" of information the brain receives, the structure of the brain and the processes therein, and the "outputs" resulting from those processes. Indeed, Patricia Churchland states, "The brain does indeed appear to be a causal machine. So far, there is no evidence at all that some neuronal events happen without any cause" (Churchland, 2002, 204).⁵² Now, what parts

⁵⁰ I will return to this line also in the third subsection of this chapter.

⁵¹ These brain processes will be further discussed in the third subsection.

⁵² One might object here that there is some indeterminacy at some level in the brain, either from quantum indeterminacy or from neuronal noise. (My thanks to Jim Klagge, Ted Parent, and David Faraci for noting this possible objection.) Let me briefly address quantum indeterminacy first. This hypothesis has not been shown to be wrong, but there are some reasons to doubt it as it stands. First, Max Tegmark argues that the brain is better considered a classical, rather than quantum, system due to the way the brain makes calculations (see Tegmark, 1999). Second, quantum indeterminacy could not do the work we would need in order for it to account for free will: invoking quantum indeterminacy for interrupting the causal chain would yield *random* actions, not *intended* ones (see Churchland, 2002, 205-210).

of this causal machine is the person responsible for, particularly since the person simply *is* that causal machine?

Let us look again at the claim "My brain made me do it," as we started with. We tend to think that there are aspects of our biology that we are not responsible for, and if those aspects cause some event or some series of events, we are not responsible for those. Yet, our brain is responsible for *everything* we do and are. What seems to be behind the intuitions involved in the example cases given is the notion that there is a controlling agent. But, as materialists, we must recognize that whatever that "controlling agent" is, it is part of the causal machine that is the brain. Hence, if we accept this notion as a legitimate exonerating condition, we seem to have to embrace a dualistic picture. The intuition requires we endorse the notion that in some cases, what *I* am responsible for and what *my brain* is responsible for diverge. But this means we have a dualistic picture of the agent.

The notion that there is a "controlling agent" leads to the intuition to be discussed in the next section, and then will be examined given the neuroscience in the third. There are some cases in which we will claim that an agent "should have known better": we are, as self-aware beings, able to rise above our biology. We have control over ourselves. In the cases mentioned in this section, Kim clearly does not have control over whether she is color-blind. In Gage's case, given the brain damage he suffered, he seems to have lost his

There is also a hypothesis that there exists indeterminacy from neural noise. Michael Shadlen and Adina Roskies explain that neurons emit variable action potential per unit time even under identical conditions. That is, a neuron may emit ten spikes (as the action potential per unit time) in a given controlled condition and six in an exactly replicated controlled condition. This is what is meant by "noise": the neurons may fire irregularly (Shadlen and Roskies, 2012, 6-7).

Yet this hypothesis is perhaps even more contested. What one counts as neuronal noise and what one counts as neuronal signal is going to be different for the test subject and the experimenter: this is the first way to contest Shadlen and Roskies' conclusions (Knoblauch, 1-4). Neuronal noise also is expected in the brain (as it is a dynamical system) (Seth, 2008, 60). Most importantly, however, the variability in neuronal spikes is explainable as effects of temporally modulated inputs (for more on this, see Knoblauch and Palm, 2005, 83-90).

ability to control some of his social impulses. Yet in Ralph's case, we may think that he had the ability to control his reaction but did not. That is, it might seem he had a choice about how to act. Perhaps he should have known better and should not have struck Bob. This question will be further explored in the next section.

ii. "Should Have Known Better"

The other dualistic intuition is that genetics and early life experience can be seen as difficulties to be transcended. In these cases, a person may not be exonerated from moral responsibility if she does not seem *forced* to behave in a certain way by her biology. Even if a person, for example, has the genetic predisposition to psychopathy and faces severe childhood trauma, these two ingredients for psychopathy do not *predict* that the person will become a psychopath. In addition, there is the notion that we normally exhibit impulse control: even if we *are* psychopaths, we can "know better"; that is, we can know that certain acts are morally wrong and so we can control ourselves. Yet this notion is hard to reconcile with neuroscience: if the mechanisms required for inhibitions are dysfunctional, then what is it that is supposed to control our actions? Or, in other cases (as in Ralph's), if the information that our brain is working with is not accurately reflective of the situation, can we be expected to overrule our perceptions?

I will address these questions in this section. First, I will start by questioning our intuitive notion of the 'controlling agent', bringing in a non-moral case (that of appetite) that we normally take to be in our control. Next I will question what sorts of expectations we can hold of someone, looking at their biology and experience.

Common intuition, rightly, is that we have control over whether we act on our impulses. It may be the case that a person would like to be drunk every night, but he has

control over whether he drinks or not (unless he is an alcoholic). But what exactly is doing the controlling? We tend to think that actions we have control over are those that are voluntarily or intentionally done: theoretically, we *could* have not done them. But this is a definitional claim, not an explanatory one (otherwise it would be circular), and so it needs further parsing (Churchland, 2002, 207-211).⁵³

Consider, for example, the case of appetite. We tend to think individuals have control over the amount they eat, and that overeating is an issue of will-power. But there is some empirical evidence that may lead one to question this notion. Leptin, a hormone released by fat cells, basically controls feelings of hunger and the satiation of it. But leptin receptors can be mutated, and we can predict how overweight a subject will be based on the kinds of mutation of the leptin receptors (shown in mice studies). Churchland states, "If a person is born with the db mutation ... and if, in consequence, he feels as ravenous at the end of dinner as at the beginning, it seems inevitable that he will overeat ... such a person will [likely] have less control over his eating behavior than a person with the standard version of the leptin receptor" (218). In other words, at least some cases of overeating will be explained by leptin receptor mutations, and it is unclear to what degree control fits into this picture.⁵⁴ We might think that if a person is aware of this problem, then he can stop eating at an arbitrary point (hence exhibiting some control over his impulses); yet this response is problematic in regards to other inputs.

Reconsider now Ralph's case of dysfunctional perception that leads him to strike Bob. Let us say Ralph is aware that he is hyper-responsive to threat signals and can

⁵³ In other words, those actions over which we have control *just are* ones that we chose voluntarily when we could have done something else. Having control over actions is not *explained by* doing those actions instead of others.

⁵⁴ In other words, the mechanisms by which we normally prevent ourselves from over-eating are dysfunctional in at least some individuals: we stop eating when we feel satiated, but these individuals do not have that sensation.

misinterpret some situations as threatening when they actually are not. We may think that, as in the case above, he might just have an arbitrary controlling point: he can actively try to maintain calmness in cases of stress. In some cases, this may work; yet we can see already that this will be difficult in other cases, including the situation with Bob. If Ralph's perception is of a threat, how is he supposed to distinguish cases of actual threat and cases of misperceived threat (cases where he should "know better" than to react uninhibited)? This becomes more complex also when we consider that some fear responses (such as startle responses) are involuntary.

The other difficulty is that trying to understand the notion of a 'controlling agent' in the context of the brain being a causal mechanism leads to a weak notion of control and voluntariness (Churchland, 2002, 210-11; Gazzaniga, 2005, 101). The brain, as mentioned earlier, is a causal machine in which inputs determine outputs mechanistically (recall, for example, the description of the function of the HPA axis from Chapter I). Michael Gazzaniga states, "The neuroscientists cannot talk about the brain's culpability any more than the watchmaker can blame the clock. Responsibility has not been denied; it is simply absent from the neuroscientific description of human behavior" (101). In other words, we think that the brain's activity is determined by inputs that causally lead to a series of events and activities in the brain, which ends when it dies. There is one long causal story to each brain.

To say that something different could have occurred-- i.e., that we could have chosen not to do something or to do something else-- is just to say that there could have been different antecedent conditions within the brain (Churchland, 2002, 210-11). But if there are different inputs, then of course there will be different outputs: this claim is

simply weak and uninteresting. Yet if all antecedent conditions remain the same, then there is only one outcome. It is unclear, then, what we mean when we say we "could have done otherwise", or "should have known better" (as I have stated it in this section). We know what we know, and to claim we *should* have known something else is just to claim antecedent conditions *should* have been different.

But depending on how far back we go in the brain's causal story, the "control" and "choices" that are such antecedent conditions are not clearly voluntary or involuntary. That is, let us say that we claim Ralph should have known better than to interpret Bob as a threat and so should not have struck him. What is it that we mean when we make that claim? The elements at play in the situation are Bob's actions (that seem threatening) and Ralph's perception of the event. Ralph obviously has no choice in or control over Bob's actions, so that leaves us with the latter element. But if we claim that Ralph somehow should have had more control over his perception of events, it is not clear what we mean. It would seem unfair that we should say Ralph should have had control over events in his childhood that contributed to, if not caused, the hyperactivity in his threat responses and attenuation to anger. Yet it is also unclear to what degree, if any, we have control over the way in which our experiences form us, particularly experiences we have as children. Churchland states, "Increasingly, it seems unlikely that there is a sharp distinction... between being in control and being out of control, either in terms of behavioral conditions or ... underlying neurobiology. This implies not that there is no distinction, but only that whatever the distinction, it is not sharp" (211).

Given the difficulty in differentiating between what we would consider voluntary, we should at least be cautious in making claims about what people *should* have known or

had control over. In some cases, this will seem a clear-cut issue: Kim does not have control over her color-blindness. Yet in others, such as Ralph's, it becomes more difficult to determine what actions we should hold him responsible for. If it is not clear the extent to which the brain can regulate whatever process is relevant-- e.g., evaluating perceptual input as accurate or inaccurate-- then it appears we are holding the individual responsible for something for which his brain does not have access to. This should be no more justified to us than holding Kim responsible for her color-blindness.

iii. Who, or what, is the agent?

Now that I have described the intuitions and mentioned briefly some of the problematic points, I want to delve more thoroughly into what neuroscience has to tell us about how it is that we make decisions. This is quite relevant to the discussion at hand; we tend to think that conscious decisions are distinct from those our brain makes that 'we' are unaware of: e.g., the regulation of vital functions by the autonomic nervous system. This intuition is not incorrect: obviously these two processes are going to be different and will involve different parts of the brain. Yet, as we will see, what neuroscience indicates about decision-making poses serious problems for the notions of 'the controlling agent' and 'voluntary action' at play in the intuitions in the previous subsections.

Recall from Ch. 1 that synaptic connections between neurons develop over time. These can form in two ways: *experience-expectant development* and *experience-dependent development*. The first is the overproduction, and then "pruning" --trimming and organizing-- of synapses (Twardosz and Lutzker, 2010, 62-3). The second process is

the creation of new synapses or modification of old synapses in *response to* experience. Most of our life-learning is through experience-dependent development.

These synaptic connections have a role of transforming information as it moves through the brain. The idea is this: incoming information (from, say, the retina) goes through a series of transformations in neural processing. Paul Churchland aptly describes this as a ladder-like structure: information comes from the bottom rung (the retina) and goes up the subsequent rungs (the LGN, the primary visual cortex, etc.). The synaptic connections are what make the information transformations: to continue the ladder metaphor, they are the sides of the ladder connecting one rung to the next. So, we have populations of neurons connecting (through synapses) onto other populations of neurons, where the connections are transforming the representations in the process (somewhat like the process of translating from one language to another) (Churchland, 2012, 35-37).⁵⁵

What the transformation is depends on synaptic weights. Churchland states, "A set of synaptic connections functions as a transformer, and the specific *nature* of the transformation effected is determined by the specific *configuration* of connection strengths or weights across that transforming population of synapses" (37). These weights are determined by the "training" of our synaptic connections: they are not static or fixed (37-39). What he means here, roughly speaking, is that if our synaptic weights were not significant (i.e., if we had not trained our synapses), then we receive just noise from our perceptual organs. Yet because of synaptic weightings, we "train" ourselves to recognize features that fall into *perceptual categories*, e.g., the texture of fur on an animal (38-45).

We train ourselves to have these perceptual "skills"-- they are *acquired*, much as riding a

⁵⁵ Churchland offers a neural network account of modeling brain function that I cannot fully capture here. For brevity's sake, I leave out details about information processing and representing done by populations of neurons in "activation spaces". Yet I want to note here that this is the somewhat simplified version of Churchland's discussion.

bike or playing the piano (49). Churchland states, "What happens, as sensory information ascends such a ladder, is its progressive transformation into a succession of distinct representational formats, formats that embody the brain's background 'expectations' concerning the possible ways-in-which the world can be" (35). In other words, through the process of transforming representations as we move along the ladder of the relevant brain system, the representation ends up embodying not *just* the sensory input initially received, but also our expectations of the world given the training of synaptic connections.

So how does this training work? A large part comes from habit, or the practicing of perceptual skills, and conditioning (49). What this means is that rather than trying to compress all the input we receive (which is not possible), we instead seek out commonly encountered inputs. That is, at the second rung of our ladder, we sort out *preferred input stimuli* as the information that will continue up the ladder (60-63). As we move up the ladder, there will be similar preferential weightings: e.g., in using certain features to identify a face as male or female, we may treat a square jaw line as having higher weight for categorizing the face as male (60-62). We can then use recurrent feedback to determine the weights of synapses to minimize the processing of information that is not useful.

What happens also in training synaptic connections is that at some level of our ladder, synaptic connections do not only project upward but also downward as well. These are '*recurrent*' *axonal projections* that "*control*" lower populations of neurons to a significant degree (148). The recurrent network can train its synaptic weights such that input can move more efficiently to *attractor points*. These function similarly to the

preference weightings described previously: they provide "a further dimension of recognitional focus" in coming to a representational prototype (149). For example, say we know that we are at a cow farm and notice a four-legged bulky creature in the distance. We can use that background information to assist in recognizing the creature as a cow, as opposed to a horse, say. The background knowledge that we have will vary in location depending on how evolutionarily recent the know-how is (150).⁵⁶

Our decision-making processes are just another system in the brain, whose outcomes are determined by inputs and synaptic weightings. When we make a decision, we tend to form a preference, act upon that preference, and then receive and evaluate the outcome (Ernst and Paulus, 2005, 1). The neurological processes involved in decision-making encompass various brain areas and types of neurotransmitters, yet we can identify these and even manipulate them experimentally (Schall, 2001, 40).

The key part to the process of formulating preferences and acting upon them is the difference between expected outcomes and actual outcomes. Our preference for the expected outcome functions similar to attractor points, mentioned before. The preferences form as a result of rewards. Jeffrey Schall explains, "A key insight into the brain mechanisms of utility was the discovery that animals and humans will work to obtain electrical stimulation delivered to certain parts of the brain. Brain-stimulation reward shares many properties with natural rewards" (Schall, 2001, 38). The rewards we *do* receive in the outcome of an action will attribute a weight to certain preferences: the larger the reward, the stronger the preference.

⁵⁶ Primitive know-how --e.g., how to breathe-- will be in the autonomic nervous system. But more sophisticated motor know-how (e.g., how to dodge) would be in the motor and premotor areas of the cerebral cortex (150).

When the expected value of the reward differs from the actual value we receive, this difference becomes a "learning signal" (Ernst and Paulus, 2005, 3). Monique Ernst and Martin Paulus explain, "This value difference is critical to learning processes. Electrophysiologic work in monkeys has demonstrated that dopamine neurons code the value difference between the expected and actual value of outcomes, and this value difference serves as a learning signal that permits behavior to become adaptive" (3). We form associations with outcomes based on these values: this type of association is suggested to occur in the medial prefrontal cortex. If these associations are positive, a certain preference will have more weight; if they are negative (as with regret), we will likely modify future behavior (so, we may prune the weight the preference had) (Ernst and Paulus, 2005, 3-4; Schall, 2001, 39-40).

What is important to emphasize is that the brain is a self-regulating system. In so far as there is a controlling agent for a given action, it is to be found in a given level of a recurrent network, and depending on the action, the controlling agent will be different depending on context. Decision-making processes rely on input from other systems (what we are making a decision *about*). This input has already been transformed from one representation to another depending on synaptic weights, and options presented for making a decision are similarly weighted. The brain can edit these options automatically or after deliberation, but the option 'chosen' will be based on the weighted value of that option as determined by its association with positive rewards, as coded by dopamine neurons.

This should shed light on our discussion of involuntary versus voluntary actions and the intuition of a controlling agent mentioned in the previous subsection. We see in

the description of the neurology of decision-making that making 'voluntary' choices is simply another process in the causal mechanism that is the brain, and our choices are determined by inputs and synaptic weightings. Our higher-level decisions, while vastly more complex, function in very similar ways to lower-level perceptual skills. Our preference weightings in perceptual systems and in decision-making processes are formed and trained through habit and regulation by supervisory systems.⁵⁷

Given this discussion, what does it mean to say "My brain made me do it" or "He should have known better"? In addressing the first point, the descriptions of how we make decisions indicates that the brain *always* determines the actions we make, on some level or other. The second point seems to suggest that an individual should have had background knowledge informing the regulation of the weights of his synaptic connections that he lacked. However, this second point seems confused even conceptually: how is one supposed to know something he does not know?⁵⁸

Yet the first point warrants more attention. One might reply that there is a difference between atypical neurology affecting our actions (e.g., as in Ralph's case) and acting intentionally. One seems to occur below the level of consciousness and the other is conscious action.⁵⁹ Yet, ultimately it seems to be the case that our decisions-- conscious or not-- are made based on input and the weightings of values of options determined by a feedback loop. Our learning process that results from the decision-making process is sensitive: it can be damaged such that certain options are given higher weights even if they are maladaptive. Areas associated with emotional control-- the amygdala, in particular-- are especially important in this regard as outcome stimuli can be associated

⁵⁷ An interesting note is that the HPA-axis is such a supervisory system.

⁵⁸ Keep in mind that even our perceptual skill is know-how that forms background knowledge for the brain (Churchland, 2012, 48-50).

⁵⁹ My thanks to David Faraci for noting this objection.

with positive or negative affect. Ernst and Paulus state, "Early dysfunction in these regions and associated networks could compromise significantly the development of adaptive decision making" (Ernst and Paulus, 2005, 4). Substance-abuse disorders, schizophrenia, PTSD, and anxiety disorders (including anxiety associated with attentional bias toward threat) all have been associated with dysfunction in decision-making processes (4-6).

In other words, while conscious action may be on a higher level of processing than just our perceptual systems, the connectionist (neural network) structure⁶⁰ applies in both cases. Outputs of populations of neurons are determined by inputs and synaptic weightings, and the weightings are trained and adjusted through feedback loops accounting for outcomes and background knowledge and expectations. So, whether on the level of conscious action or on the level of perceptual processing, "our brain makes us do it."

One might respond further that there are morally relevant parts of the brain (such as parts governing intentionality) versus non-morally relevant parts (such as the autonomic nervous system), and in talking about whether our brain "makes us do" something, we refer to the latter. Take two cases: A) John hits Lou because he is having a seizure; B) John hits Lou because he enjoys it.⁶¹ These seem remarkably different, and B) certainly seems morally wrong. We could give a story about B) in that John is acting from the selection of a certain option based on input and preference weightings, but we might think that it is simply wrong that he has those preference weightings. Yet this seems to get back to the issue of what the controlling agent in a given case is. In A), there

⁶⁰ The 'ladder' structure discussed earlier.

⁶¹ My thanks to David Faraci for noting this kind of objection.

is no *conscious* control over a seizure, in B) there seems to be *conscious* control over what option John picks (either hit or do not hit Lou). But the story about B) here is oversimplified. There is control in B) if the recurrent network is functioning properly: decision-making processes rely on the functioning of the networks involved. If the relevant systems are dysfunctional in case B), then in essence we have the same story for A) and B): John has maladaptive functioning that ends up causing a harmful action.

Conclusion

In this chapter, I have argued that employing psychological excuses for moral responsibility leads us into a dualistic picture of the individual. In some cases, we hold that a person's biological dysfunction is responsible for a given action, but not the person. In others, we hold that despite some biological malfunctioning, an individual ought to have "known better": he should be able to evaluate and control his impulses. Without further explication of the notion of 'control', these kinds of excuses seem to smack of dualism. Yet when we dig into a neurobiological account of 'control', we encounter that the brain is a causal mechanism, in which events are determined by antecedent conditions. If we hold that an individual makes a decision or other action that is not part of this causal mechanism (i.e., is not determined by antecedent conditions), then that is to uphold dualism.

The purpose of this chapter is not to claim that moral responsibility does not exist or to claim that we have no 'control' over our actions. Rather, the purpose was to push the issue of how we actually understand what we mean by those claims. In particular, we should be cautious in saying that someone is exonerated from moral responsibility

because of a psychological defect. There seems no *prima facie* clear justification for privileging some psychological defects over others; after all, it is difficult even to explain fully what we mean by claiming some people be aware of and be able to control their psychological defects, but not hold others to the same standard. In examining moral responsibility, we should be cautious about *who* or *what* we take to be the agent being evaluated.

IN THE NEXT CHAPTER, I WILL PROPOSE ANOTHER MODIFICATION TO A STRAWSONIAN ACCOUNT OF MORAL RESPONSIBILITY THAT WILL HOPEFULLY AVOID THIS CONCERN ABOUT DUALISM BROUGHT ABOUT BY USING PSYCHOLOGICAL DEFECTS AS EXONERATING CONDITIONS BASED ON INTERNAL FEATURES. RECALL THAT IN CHAPTER III, I SUGGESTED REDRAWING THE EXCUSING/EXEMPTING DISTINCTION AS AN EXTERNAL/INTERNAL DISTINCTION. IN THIS CHAPTER, I HIGHLIGHTED A CONCERN FOR THE DISCUSSION OF EXONERATION BASED ON INTERNAL FEATURES, WHICH IS THAT SUCH A DISCUSSION CAN LEAD TO A DUALISTIC PICTURE OF THE AGENTS IN QUESTION. TO AVOID THIS CONCERN, I WILL PROPOSE IN THE NEXT CHAPTER THAT WE SIMPLY STICK TO TALKING ABOUT BRAINS.

V. TOWARD A NEURO-RESPONSIBILITY ACCOUNT

In this chapter, I propose a slight modification to Strawson's account of moral responsibility that should get us out of concerns arising in the previous chapter. What I propose is that we do away with exonerating individuals for psychological reasons,⁶² and instead begin a discussion that centers around determining *what* is responsible for any given event or action. Once we determine what is responsible for an event, we can begin a process of intervention. One might possibly see this modification as changing the topic from *moral* responsibility to just 'responsibility' as we construe it broadly. While we may have to reconceive 'moral responsibility' slightly, it is still 'moral responsibility' in the Strawsonian sense: we, as a community, are still holding something morally responsible for an action based on our reactive attitudes. This modified notion of 'moral responsibility' is more tailored to causal responsibility. What we hold morally responsible for a given action is what is causally responsible for that action.⁶³

First, I shall briefly recapture Strawson's account. Recall from Chapter 2 that on a Strawsonian account, our practices, as a moral community, of holding a person to be morally responsible reflects an understanding of our judgments of whether that person is morally responsible. Strawson argues that whether we *hold* a person morally responsible

⁶² This will not completely do away with the external/internal distinction; rather, internal features will be significantly limited (to cases such as an individual having a seizure or color-blindness, where there is a clear exonerating factor).

⁶³ This is not to say that causal responsibility and moral responsibility are the same thing. *X* can be causally responsible for *Y* and assessments of moral responsibility will not apply. E.g., the sun is causally responsible for the growth of plants. The constraints on my modification of 'moral responsibility' will become clearer throughout the chapter.

in practice determines whether we have judged that person *to be* morally responsible. We make these judgments through *reactive attitudes*: those attitudes that arise from social interactions, such as resentment, blame, or praise, to name a few. In cases of excusing behavior that leads to negative reactive attitudes, one takes an objective standpoint from which one can determine whether the agent in question is capable of engaging in social interactions. In other words, if a person is able to participate socially with the community, then he is judged as a moral agent by others' reactive attitudes (Fischer and Ravizza, 1993, 50-55).

I propose a slight reconfiguring and narrowing of this account. Imagine some morally wrong event happens, such as Ralph striking Bob. If we were to witness this occasion knowing nothing of Ralph or Bob, we would react with moral indignation on Bob's behalf. This resentment would be directed at Ralph, initially, as the perceived cause of the injury to Bob. Yet as we learn more about the situation, that resentment may change in severity or may redirect to another source. For example, if we discover that Bob actually told Ralph to strike him to wake him up, our resentment may dissipate; if we discover that Bob burned down Ralph's house the previous night, our resentment may lessen from sympathy. Lastly, if we find that Jane was actually moving Ralph's arm to strike Bob, our resentment redirects to be toward Jane.

Let us now consider the original scenario: that Ralph is hyper-responsive to threat. In Strawson's account, we may see this as an occasion to suspend our reactive attitudes and take an objective viewpoint to determine whether Ralph's actions are excusable (Fischer and Ravizza, 1993, 50; Watson, 2012, 5-6).⁶⁴ Yet, as Watson notes

⁶⁴ Recall that the 'objective viewpoint' is the stance we undertake when we suspend our reactive attitudes (Fischer and Ravizza, 1993, 50-55).

with regards to the Robert Harris case,⁶⁵ the reactive attitudes themselves do not dissipate. We are still appalled by Harris's actions, even if we come to an understanding that he may not be fully responsible (Watson, 2012, 12-17).

Instead, we should redirect those reactive attitudes to their appropriate target, namely, whatever specifically led to the action in question. After all, when reactive attitudes arise even in Strawson's account, they are in reaction to the action or event in question, not the *agent*, though they are directed toward the agent (e.g., blaming the agent for some action). What I propose is essentially no different, except that the agent of the action is to be reconceived to be what is responsible for the action, narrowly construed. In Ralph's case, what seems to be responsible (mainly) is a malfunctioning HPA-axis. The malfunctioning of the HPA-axis was the cause for the event that gave rise to our reactive attitudes; hence, it is the target of those attitudes. Once we locate and hold responsible whatever malfunction or process *is* responsible for the action, we can better assess productive actions to take.

That is, once we have located the responsible process or malfunction, we can intervene. In Ralph's case, the assignment of responsibility to his malfunctioning HPA-axis indicates a structure that could lead to reoccurring instances of a similar problem. The hyper-reactive stress response could be provoked in another stressful situation, leading to further injury of other people. Yet if we locate the cause of the action, we can proceed to treat Ralph for his malfunctioning HPA-axis, either through therapy or medication, or through a combination of both.

⁶⁵ Harris coldly murdered two youths, and yet suffered his entire life from extreme abuse from both parents.

While this approach may seem counter-intuitive at first, there are multiple benefits to it over the traditional treatment of morally responsible agents. First, we avoid the dualistic view described in the previous chapter. Instead of debating whether an individual's biology is distinct from her moral character, we are holding the biological units morally responsible for those actions for which they are *causally* responsible. In other words, we will not have cases where we are treating the individual as distinct from her brain.⁶⁶

One might object that we have now introduced, however, a new problem of determining what that biological unit is. This seems an empirical question: it might be the case that a single receptor type is the reason for a certain malfunctioning, as is the case with leptin receptor mutations being responsible for some cases of over-eating. Yet in other cases, it could be the case that a neurological system is malfunctioning, as may be the case with oversensitivity of one's threat-response system.

A second benefit to this approach is that it is productive and forward-looking. The focus is on locating the source of maladaptive behavior and then finding a treatment for it. This will better protect the community against recurring instances of harmful behavior from the same sources. One might say that current accounts of moral (and legal) responsibility accomplish the same end: if we determine a person is not morally responsible for his actions because some part of his brain "made him do it," we can then treat him for whatever impairment he has. Yet we do not *always* engage in inquiry into the neurology of individuals with morally wrong behavior-- these individuals may go

⁶⁶ This does mean that if we were to talk about holding *individuals* morally responsible, on my account, people would be morally responsible for whatever their brains are morally responsible for. But it is more precise to talk just about brains, at least in discussing internal explanations. My thanks to David Faraci for noting this.

untreated. Further, this means only that the main difference between the traditional account and my own is that I am eliminating the middle step of blaming the individual.

One might respond to this that it is not clear whether treatment should be the preferred method. Can we punish or blame someone without treating them? Will this depend on the severity of the crime? E.g., should we treat mass-murderers?⁶⁷ These are difficult questions. At least some neurological conditions (e.g., psychopathy) currently seem either untreatable or strongly resistant to treatment. Society will have to determine a way to deal with cases where individuals who pose a threat to the community are untreatable. For example, we may send them to prison.⁶⁸

The answer to the second-question is essentially bullet-biting: despite severity, if we can determine what is causally responsible for a given action and it is treatable, then we should treat the condition. This is not a comfortable position to hold. There are-- rightly-- strong, angry reactions to criminals who coldly murder, rape, or torture others. But if the perpetrator of a violent crime has some neurological condition that is treatable, then why should we *not* treat him? It does not seem that any party involved (society, the criminal, the victims' loved ones) gains something from withholding treatment of the criminal; rather, it seems at least society and the criminal *would* gain something: the criminal would be better able to function in a moral community and would likely experience better mental health. Thus, I am willing to bite this bullet. We should not collapse notions of moral responsibility and vengeance.

A third benefit to my approach is that it appears to skirt underdetermination by social psychological evidence, mentioned in Chapter II. Situationist ethics, armed with

⁶⁷ My thanks to Jim Klagge for pointing out these objections.

⁶⁸ This broader scope of implementing this account of moral responsibility on a societal level is a future project.

social psychological studies, have suggested that our moral judgments are influenced by our own emotions, physical environments, and other unexpected sources (Vargas, forthcoming; Upton, 2009; Prinz, 2006; Hume, 1739). Prinz, for example, notes that our judgments can be harsher if we are in disgusting environments (Prinz, 2006, 30-32). While these empirical findings may suggest that our reactive attitudes could be unduly influenced by irrelevant sources, this worry is eliminated in my approach. It is an empirical question as to what activity in one's brain is responsible for a given action: whatever the emotions of someone judging the situation, the cause of the action will remain the same. There is a fact of the matter to appeal to, rather than relying solely on the reactive attitudes of the community making the judgment.

Of course, one may object on exactly those grounds: that I am discussing *empirical* questions, and the issue of moral responsibility is *normative*. It may appear that I am committing a naturalistic fallacy of sorts. Further, *people* are morally responsible, not objects. If a tree falls on my car, we do not hold it morally responsible for vandalism. It is not clear whether we can hold, for example, hormonal receptor mutations to be *moral agents*, capable of being considered *morally responsible* for an action.

These are strong objections that highlight that my approach may not be in line with moral intuitions about the nature of normativity. Yet again, this returns to the issue of dualism in our considerations of moral responsibility. To hold a person morally responsible for an action *just is* to hold his brain responsible for the action. *All* of our actions are caused by our brain. What my account proposes is simply to narrow the scope so that we may determine what exactly the brain was doing when the grievous action occurred: what *exactly* the cause was. To say the brain as a whole was the cause is a

broadly sweeping statement, and as our neuroscience continues to be refined, we are better able to pinpoint the functions of mechanisms in the brain and offer narrower explanations for behavior and phenomena in the brain.

There is little doubt that this may be an unpopular view. It may seem to make humans mechanistic. Well, we are. As described in the previous chapter, there are difficulties in explaining concepts integral to the traditional conversation of moral responsibility-- e.g., 'free will', 'choice', 'control'-- in neuroscientific terms. If we reject dualism and if we accept that neuroscience provides the best explanations of human behavior and brain activity that we have access to, then the difficulty in reconciling neuroscientific understanding of the brain and our concepts involved in moral judgments should make us reconsider those concepts.

It is worth noting here that Strawson's account is nearly a direct replica of David Hume's. I will elaborate on this briefly. In his *Treatise of Human Nature*, Hume argued that our emotional reactions constitute moral judgments. According to Hume, what is morally right or wrong is simply what yields pleasure or discomfort respectively. Hume states, "In all enquiries concerning these moral distinctions [between right and wrong], it will be sufficient to shew the principles, which make us feel a satisfaction or uneasiness from the survey of any character, in order to satisfy us why the character is laudable or blameable" (Hume, 1739, B3:P1:S2). By this he means that when we discuss what is morally right and wrong, we need to inquire into the reactions "we" (the community) have to those claims or agents, who have committed some act. If we are repulsed at the claim, 'Skinning dogs alive is a fun hobby for a rainy afternoon,' then this constitutes that

activity being morally *wrong* in principle.⁶⁹ If we come across an individual who has committed this repulsive act, then our "very *feeling*" of repulsion constitutes our *blaming* the individual for the action, i.e., for holding him morally responsible for it (ibid, italics are Hume's). It is not that the feeling gives *rise* to the moral judgment or informs it: the feeling *is* the judgment (ibid).

Hume's account should sound familiar: at least on the topic of moral judgments and determining moral responsibility, Hume sounds no different from Strawson, writing roughly two centuries later. This would suggest that the *way* we discuss moral responsibility has not changed substantially for about two centuries. Certainly there are a variety of accounts of moral responsibility (Ch. II) that have arisen in the meantime (the field is productive), but it may be time to update the content.

⁶⁹ It is worth noting, though it takes us a bit astray, that for Hume, many such principles are conventional: they come into being over time in our communities and then are taught to us, for better or worse (Hume, 1739, B3:P2:S2). One may think, for example, of consensual incestuous relations among infertile siblings. This narrow case seems immune to common objections against incest: there are not complications of power-relations (as a parent-child relationship has) or potential genetic deformities to offspring (the individuals are infertile). Yet many have the "gut" reaction that these individuals would be doing something *morally* wrong. (See also Prinz, 2006.) If we take up Hume's view, we can explain this as being the reactions toward the violation of conventions we have come to accept (e.g., siblings do not engage in romance with one another).

CONCLUSION

I have argued that in at least Strawsonian accounts of moral responsibility, the way in which we treat exonerations of moral responsibility resulting from psychological or neurobiological abnormalities can result in significant worries, particularly if we want to maintain a materialist view of the brain. The worry is that the way in which we define *excusing* versus *exempting* conditions does not seem to hold up in all cases: there are some cases in which this distinction is not so distinct. The second worry is that in our debates about who is the agent responsible for a given action in these cases, we tend to treat the "brain" and the "individual" as distinct entities, and they are not. What we treat as the properties of the "individual" or the "agent"-- the personality or character, the ability to make choices, etc.-- is still just the *brain*.

As an attempt to circumvent these concerns, I proposed two modifications to Strawson's account. The first was to redraw the distinction given along the lines of external and internal explanations. The second was to propose a modification to the way we discuss cases of internal explanation. Instead of discussing why an individual may or may not be responsible given her biology, I propose we discuss what occurred in the brain that led to a particular action being performed. Whatever specifically is the cause of an action is what we should hold to be responsible for the action.

The purpose in proposing these modifications is not to claim that discussions of moral responsibility are false or misguided. Rather, my intention is to modify an existing framework with insights from neuroscience. Hume and Strawson's accounts rightly place

emphasis on the role of the community in judging moral behavior. We are morally responsible for injuring those in our community, and so it is appropriate that the reactions of our community to our actions is determinant in moral judgments. If we can make this explanation more reconcilable with neuroscience, this is a possibility that we should consider. Strawson himself noted the revisability of his account in light of changing beliefs in the community: this is the revision I propose in light of current scientific understanding of human behavior (Fischer and Ravizza, 1993, 66). Furthermore, Hume-- who is the apparent father of this sort of moral responsibility account-- might even *agree* with such modifications, given the importance he placed on psychology even in his own time.

Lastly, this thesis offers something further.⁷⁰ There are concerns-- even fears-- about the implications that neuroscience has for our accounts of ethics and morality. On the question of moral responsibility, it might seem to some that neuroscience and accounts of moral responsibility are incompatible. Or it may seem, as Strawson discusses, that we are left with a "pessimistic" view of moral responsibility: our practices of moral responsibility are merely empty, expedient practices to ensure social cooperation. Yet neither of these have to be the case. I have modified Strawson's account so that it is consistent with neuroscience but also consistent with our values. The modifications focus on clarifying what went wrong that led to a harmful action, and then *treating* that problem so that it does not arise again for the individual or for the community. Rather than viewing neuroscience as a threat to the way we do philosophy, we should view it as a science that provides new insights that we can use to strengthen our understanding of moral agency. We can be optimistic *and* scientific.

⁷⁰ My thanks to Dan Linford for this point.

REFERENCES

Arabadzis, Dimitrula et al. (2010). "Primate Early Life Stress Leads to Long-Term Mild Hippocampal Decreases in Corticosteroid Receptor Expression." *Biological Psychiatry* 67: 1106-1109.

ed. Bechtel, William; Mandik, Pete; Mundale, Jennifer, and Stufflebeam, Robert S. *Philosophy and the Neurosciences: A Reader*. Malden, Massachusetts: Blackwell Publishers, 2001.

Brown, Vivienne (2006). "Choice, Moral Responsibility and Alternative Possibilities." *Ethical Theory and Moral Practice* 9 (3): 265-288.

Buss, Sarah (1997). "Justified wrongdoing." *Noûs* 31 (3):337-369.

Churchland, Patricia. *Brain-wise: Studies in Neurophilosophy*. Cambridge: Massachusetts Institute of Technology, 2002.

Churchland, Paul. *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. Cambridge: Massachusetts Institute of Technology, 2012.

Cirulli, Francesca (2001). "Role of environmental factors on brain development and nerve growth factor expression." *Physiology and Behavior* 73: 321-330.

Cloitre, Marylène, Lisa R. Cohen, and Karestan C. Koenen. *Treating Survivors of Childhood Abuse: Psychotherapy for the Interrupted Life*. New York: Guilford Press, 2006. Print.

Craver, Carl. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press, 2007.

Dancy, Jonathan (2007). "Defending the right." *Journal of Moral Philosophy* 4 (1):85-98.

De, Haan M, and Megan R. Gunnar. *Handbook of Developmental Social Neuroscience*. New York: Guilford Press, 2009. Print.

Duncan, Seth, and Barrett, Lisa Feldman (2007). "Affect is a form of cognition: A neurobiological analysis." *Cognition & Emotion* 21 (6): 1184-1211.

Dutton, Donald G.; van Ginkel, Cynthia, and Starzomski, Andrew (1995). "The Role of Shame and Guilt in the Intergenerational Transmission of Abusiveness." *Violence and Victims* 10 (2): 121-130.

Ernst, Monique, and Paulus, Martin P. (2005). "Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective." *Biological Psychiatry* 58.8: 597-604.

Finzi, Ricky; Har-Even, Dov, and Weizman, Abraham (2003). "Comparison of ego defenses among physically abused children, neglected, and non-maltreated children." *Comprehensive Psychiatry* 44 (5): 388-395.

ed. Fischer, John M, and Mark Ravizza. *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press, 1993. Print.

Fischer, John Martin (1999). "Recent work on moral responsibility." *Ethics* 110 (1):93–139.

Frankfurt, Harry G. (1969). "Alternate Responsibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829-839.

Friedman, Bruce H. (2010). "Feelings and the body: The Jamesian perspective on autonomic specificity of emotion." *Biological Psychology* 84 (3): 383-393.

Gould, Felicia; Clark, Jennifer; Heim, Christine; Harvey, Philip D.; Majer, Matthias; Nemeroff, Charles B. 2012. "The effects of child abuse and neglect on cognitive functioning in adulthood." *Journal of Psychiatric Research* 46: 500-506.

Gray, Jeffrey A. (1990). "Brain Systems that Mediate both Emotion and Cognition." *Cognition & Emotion* 4 (3): 269-288.

Heim, Christine et al. (2000). "Pituitary-Adrenal and Autonomic Responses to Stress in Women after Sexual and Physical Abuse in Childhood." *Journal of American Medical Association* 284 (5): 592-597.

Hildyard, Kathryn L. and Wolfe, David A. (2002). "Child neglect: developmental issues and outcomes." *Child Abuse & Neglect* 26: 679-695.

Hoglund, Collete L. and Nicholas, Karen B. (1995). "Shame, Guilt, and Anger in College Students Exposed to Abusive Family Environments." *Journal of Family Violence* 10 (2): 141-156.

Hume, David. *A Treatise of Human Nature*. Oxford: Clarendon Press, 1896. Accessed from <http://oll.libertyfund.org/title/342> on 2013-04-03

Kennett, Jeanette. *Agency and Responsibility: A Common-Sense Moral Psychology*. Oxford: Clarendon Press, 2001. Print.

Knoblauch, Andreas, and Günther Palm (2005). "What is signal and what is noise in the brain?." *Biosystems* 79.1-3: 83-90.

Lazarus, Richard S. (1991). "Cognition and Motivation in Emotion." *American Psychological Association* 46 (4): 352-267.

Levy, Neil (2011). "Expressing who we are: Moral responsibility and awareness of our reasons for action." *Analytic Philosophy* 52 (4):243-261.

_____ (2007). "The responsibility of the psychopath revisited." *Philosophy, Psychiatry, and Psychology* 14 (2):pp. 129-138.

_____ (2006). "Cognitive scientific challenges to morality." *Philosophical Psychology* 19 (5):567 – 587.

Loman, Michelle M. and Gunnar, Megan R (2010). "Early experience and the development of stress reactivity and regulation in children." *Neuroscience and Behavioral Reviews* 34: 867-876

Orth, Ulrich; Berking, Matthias, and Burkhardt, Simone (2006). "Self-Conscious Emotions and Depression: Rumination Explains Why Shame But Not Guilt is Maladaptive." *Personality and Social Psychology Bulletin* 32: 1608-1619.

Pesonen, Anu-Katriina and Räikkönen (2012). "The lifespan consequences of early life stress." *Physiology & Behavior* 106: 722-727.

Prinz, Jesse J (2006). "The Emotional Basis of Moral Judgments." *Philosophical Explorations* 9 (1): 29-43.

_____. *The Conscious Brain: How Attention Engenders Experience*. Oxford: Oxford University Press, 2012.

Sanchez, Mar M. (2006). "The impact of early adverse care on HPA axis development: Nonhuman primate models." *Hormones and Behavior* 50: 623-631.

Scannapieco, Maria, and Kelli Connell-Carrick. *Understanding Child Maltreatment: An Ecological and Developmental Perspective*. New York: Oxford University Press, 2005. Print.

Schall, Jeffrey D. (2001). "Neural basis of deciding, choosing and acting." *Nature Reviews Neuroscience* 2.1: 33-42.

Seth, Anil K. (2008). "Causal networks in simulated neural systems." *Cognitive Neurodynamics* 2.1: 49-64.

Shadlen, Michael N., and Roskies, Adina L. (2012) "The neurobiology of decision-making and responsibility: Reconciling mechanism and mindedness." *Frontiers in Neuroscience* 6: 1-12.

Shoemaker, David (2007). "Moral address, moral responsibility, and the boundaries of the moral community." *Ethics* 118 (1):70-108.

Silverman, Amy B.; Reinherz, Helen Z., and Giaconia, Rose M. (1996). "The long-term sequelae of child and adolescent abuse: a longitudinal community study." *Child Abuse & Neglect* 20 (8): 709-723.

Smith, Angela M. (2007). "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11 (4): 465-484.

_____ (2008). "Control, Responsibility, and moral assessment." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 138 (3): 367-392.

Stein, Abby. *Prologue to Violence: Child Abuse, Dissociation, and Crime*. Mahwah, N.J: Analytic Press, 2007. Print.

Stuewig, Jeffrey, and McCloskey, Laura A (2005). "The Relation of Child Maltreatment to Shame and Guilt Among Adolescents: Psychological Routes to Depression and Delinquency." *Child Maltreatment* 10: 324-336.

Swogger, Marc T.; You, Sungeun; Cashman-Brown, Sarah; Conner, Kenneth R. (2011). "Childhood physical abuse, aggression, and suicide attempts among criminal offenders." *Psychiatry Research* 185: 363-367.

Talbert, Matthew (2008). "Blame and responsiveness to moral reasons: Are psychopaths blameworthy?" *Pacific Philosophical Quarterly* 89 (4):516-535.

Tangney, June Price; Wagner, Patricia; Fletcher, Carey, and Gramzow, Richard (1992). "Shamed Into Anger? The Relation of Shame and Guilt to Anger and Self-Reported Aggression." *Journal of Personality and Social Psychology* 62 (4): 669-675.

Tangney, June Price (1996). "Conceptual and Methodological Issues in the Assessment of Shame and Guilt." *Behavior, Research and Therapy* 34 (9): 741-754.

Tangney, June Price; Stuewig, Jeff, and Mashek, Debra J. (2007). "Moral Emotions and Moral Behavior." *Annual Reviews Psychology* 58: 345-372.

Tegmark, Max (2000). "Importance of quantum decoherence in brain processes." *Physical Review E* 61.4: 1-13.

Twardosz, Sandra and Lutzker, John R. (2010). "Child maltreatment and the developing brain: A review of neuroscience perspectives." *Aggression and Violent Behavior* 15: 59-68.

Upton, Candace L. (2009). "Virtue ethics and moral psychology: The situationism debate." *Journal of Ethics* 13 (2-3):103 - 115.

Vargas, Manuel (forthcoming). "Situationism and Moral Responsibility: Free Will in Fragments." In Tillman Vierkant, Julian Kiverstein & Andy Clark (eds.), *Decomposing the Will*. Oxford UP.

Watson, Gary. (2004). "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press, 2012.

Webb, Marcia et al. (2007). "Shame, guilt, symptoms of depression, and reported history of psychological maltreatment." *Child Abuse & Neglect* 31: 1143-1153.

Weber, Sabrina, et al. (2008). "Structural brain abnormalities in psychopaths—A review." *Behavioral sciences & the law* 26.1: 7-28.

Thalia Wheatley and Jonathan Haidt (2005). "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16: 780-784.