Transformation and Exoneration:
Exploring Responsibility Attribution Over Time

Nicole Ramsoomair
Department of Philosophy, McGill University, Montreal
February, 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the Philosophy Ph.D

TRANSFORMATION AND EXONERATION: TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to start by first thanking my advisor Prof. Natalie Stoljar for her continuous and unfailing support of my Ph.D study, related research and personal growth. Through each academic achievement, challenge and wordy draft to be reviewed, she never wavered in her encouragement. She provided guidance though personal milestones, unfortunate setbacks and academic achievements. I do not even think it is even possible to recount the numerous ways she has helped through this process both professionally and personally. I have been immensely fortunate to have her as a supervisor to guide me through this portion of my life. For everything she has done and all the times she has gone beyond, I would like to extend my sincerest gratitude.

To my co-supervisors Prof. David Davies and Prof. Kristin Voigt. Prof. Davies, I would like to thank for his vigilance in trudging through early drafts of my manuscript and offering some invaluable insights. Prof. Voigt's expertise also proved an immense help in working the thesis into the final product.

I would also be remiss not to mention the excellent administrative staff I was fortunate enough to have in support. I would like to thank: Angela Fotopoulos, Andrew Stoten and Mylissa Faulkner who went out of their way to help when it was truly needed.

I was also fortunate enough to receive support and feedback during a manuscript workshop with a number of fellow students including Frédérick Armstrong, Éliot Litalien, Muhammad Velji, Aberdeen Berry, Celia Edell, Hugo Cossette-Lefebvre and Joey Van Weelden. The insights garnered from this workshop challenged me to really think through my arguments in ways I had not previously considered. I would like to extend my thanks to each for their support and time. In particular I would like to single out Joey; his help in the workshop was one of many, many instances in which he provided support as a colleague and a friend.

A very special thanks goes out to Michel Xhignesse and Éliot Litalien (once again) who, without any hesitation, helped me whip the French version of the abstract into shape. Thanks for being there for me and not laughing at my oftentimes-stilted use of the language.

To the other friends I met along the way: I would like to thank Karina Vold for being a pillar of encouragement and friendship throughout this process. Her strength and tenacity during each challenge was truly inspiring. There have also been a number of other fellow students who in more ways than I can fully mention here have provided support, friendship and reassurance during my time at McGill: thanks to Tim Juvshik, David Gaber, Jess Barnes, Melanie Coughlin and Maiya Jordan.

Another special thanks to Prof. Emily Carson and Prof. Hasana Sharp whose advice and support as graduate advisors really helped me get through some of the challenging times.

That brings me to my family. My mother, Anne Ramsoomair, has been there for me through all the highs and all the lows. When I moved to Montreal she called everyday like clockwork just to see how I was doing. When I moved back to Ontario she was there providing unfailing support and love for our little family. As well, my father, Franklin Ramsoomair has provided me with his support throughout my entire academic career. Whether it was last minute requests to look at my work as an undergrad or the advice given to me as graduate, I could count on him to come, coffee in hand, to help wherever he could. As for my brothers, Craig and Scott, they were always available to provide a laugh and cheer me up as I worked through this thesis. I was quite lucky to have brothers who double as best friends.

Finally, I would like to close by dedicating this thesis to those nearest to my heart and address them specifically: Mike and Sebastian.

Mike, you have supported me in every way possible throughout this Ph.D. from the very start. You encouraged me to apply for the program, you took over the household duties when deadlines were too much, held my hand through illness, and pushed me to continue in the face of what seemed to be insurmountable challenges. Your love, encouragement and support guided me through the toughest times. I, without any exaggeration, would not have ever been able to complete this thesis if it were not for you. You are my best friend, my partner and my greatest supporter. I love you more than I could ever express in words.

Lastly, Sebastian, your arrival into this world made the work challenging at times. Yet, this was because your beautiful smile, laugh and warmth always proved to be an irresistible distraction. If you ever get a chance to read this, I would like you to know that I finished this thesis, not in spite of you, but because of you. I really do mean that. You make be want to do more and be more in order to make you proud to call me your mother. You both amaze and inspire me with every hug, kiss and “love you” yelled at the top of your little lungs.

Mike and Sebastian, you two are the greatest joys in my life and with this thesis finally completed, I cannot wait to start the next chapter our lives together.

Abstract (English)

This thesis explores conditions of responsibility in the face of radical personality change. Whether it is a reformed criminal who disavows gang associations or an alcoholic who repudiates her past actions, such cases exemplify an affective break with the past despite retaining psychological continuity. In these cases, while numerical identity seems to be secured, the radical qualitative differences seemingly throw responsibility attributions into question. By focusing on cases of radical personality change, such as criminal rehabilitation, I question whether such qualitative change alone can in fact undermine responsibility for past crimes. I start by defining the limits of one's qualitative character in terms of the persistence of what I call the *moral self*. With a basis in Lockean theory of identity, I argue that this self is best described as an agent's continued evaluative perspective, constituted by only those psychological aspects in which agents may be appropriately *answerable.* I show that sufficient change to this evaluative self can unsettle conditions of answerability necessary for responsibility attributions. Surprisingly, however, I will also suggest that even though a loss of answerability may undermine whether a criminal remains responsible for past crimes, such a loss does not yet determine whether one is rehabilitated. What matters in rehabilitative change is not necessarily the fact that change occurred, but whether change occurred in a specific manner. In particular, I argue that desisting criminals may be better considered rehabilitated if they take responsibility for their former crimes by narratively appropriating the past. The articulation of a narrative can extend a sense of answerability and offer guided change in ways that offset the force of blame and warrant forgiveness. Thus, change within the self may disconnect one from the past enough to undermine responsibility attributions. To be rehabilitated, however, involves retaining a connection to that past by taking responsibility for it.

Abstrait (Français)

Cette thèse explore les conditions de la responsabilité en introduisant une nouvelle conception de soi. Le but de cette exploration est de répondre aux questions posées par les changements radicaux de personnalité. Qu'il s'agisse d'un criminel réformé qui désavoue ses associations de gangs ou un alcoolique qui répudie ses actions passées, ces cas illustrent une rupture affective avec le passé malgré la continuité psychologique des agents ; sont-ils les mêmes personnes avec les mêmes responsabilités pour leurs passés ? La personne demeure numériquement identique tout en étant qualitativement différente. En mettant l'accent sur les cas de changement radical de personnalité, tels que la réhabilitation des criminels, je me demanderai si un tel changement qualitatif peut, à lui seul, affaiblir la responsabilité des crimes du passé. Je commence par définir les limites du caractère qualitatif d'une personne en termes de persistance de ce que j'appelle le « *soi moral* ». En me basant sur la théorie de l'identité de Locke, je soutiens que ce soi est mieux décrit comme la perspective évaluative continue d'un agent, constituée uniquement des aspects psychologiques desquels l'agent peut être tenu de répondre. Je démontre ensuite que des changements suffisants dans ce soi moral peuvent perturber les conditions qui rendent l'exigence de justification adéquate et qui sont nécessaires aux attributions de responsabilité. De manière surprenante, cependant, je également que, même si le fait qu'exiger d'un agent qu'il fournisse une justification devienne inadéquat peut nuire à la responsabilité d'un criminel pour ses crimes passés, une telle perte ne permet pas encore de déterminer si un individu est réhabilité. Ce qui compte dans la réadaptation n'est pas nécessairement le fait que le changement s'est produit, mais bien le fait que le changement se soit produit de manière spécifique. Je soutiens en particulier que les criminels condamnés pourraient être mieux réadaptés s'ils assument la responsabilité de leurs crimes antérieurs en s'appropriant leur passé sous une forme narrative. L'articulation d'un récit peut élargir le sens de l'exigence de justification et offrir un changement guidé de manière à contrecarrer la force du blâme et à justifier le pardon. Ainsi, le changement de soi peut suffisamment déconnecter l'agent de son passé pour nuire aux attributions de responsabilité. Être réhabilité implique cependant la conservation d'un lien avec ce passé en en assumant la responsabilité. Ainsi, les changements qualitatifs peuvent créer un nouveau soi moral, mais le fait d'être un nouveau soi moral ne suffit pas pour la réhabilitation.

Introduction

"How can a person express contrition if he's not guilty?"

(Stanley Williams, New York Times Interview. Dec 2nd 2005)

In personal identity theory, the prospect of two or more distinct identities within the same life is an intriguing possibility. Identity theorists have tended to conceive of this prospect metaphysically and have focused on interruptions in one's psychological life due to memory loss or other types of discontinuity. They propose that metaphysical sameness, whether through psychological or physical continuity, grounds our concerns about the persistence of the person and anchors notions of moral responsibility. Agents maintain responsibility for wrongdoing insofar as they have the same (or sufficiently continuous) psychology or physiology. However, even if identity is necessary for moral responsibility, it is not clearly sufficient given that metaphysical tests do not take into account the affective and emotional changes a person can experience over a period of time. While change of personality through maturation is inevitable, there are some changes that, on the face of it, seem much more radical. It is these sorts of changes, I will argue, that make the most significant difference when attributing continued moral responsibility.

Consider the complications that arise from the life story of one of the founding members of the notorious street gang, "the Crips", Stanley "Tookie" Williams. On December 1st 2005, Williams was put to death after Gov. Arnold Schwarzenegger rejected an executive appeal for clemency to commute his death penalty sentence to life in prison. There was no new evidence to back any suggestion that it was anyone other

than him who committed the violent murders with which he was charged. The governor detailed his crimes and included a "strong and compelling" list of evidence that left "no reason to second guess the jury's decision of guilt or raise significant doubts or serious reservations…"[1] But this was not the argument presented by his lawyers. When Williams' supporters yelled "The state of California just killed an innocent man!" in the witness media room after execution, they were not referring to a mistake within the judicial process itself.[2] Williams' lawyers and supporters alike suggested that Williams was worthy of mercy because of his claims of redemption. Due to the personality change he underwent in prison, they argued that there was a new man up for execution.

The outrage that followed his death sprang from the notion that the man facing the sentence was recognizably different from the one that committed murder in 1979 in ways that seemed to mitigate moral responsibility. While on death row, Williams renounced his former gang affiliations and actively spoke against them through ongoing community projects and authoring children's literary works. His work against gang violence eventually led to a Nobel Peace Prize nomination. If metaphysical questions of sameness, which Marya Schechtman calls "re-identification" questions, are all that matters in these cases, then Williams remains responsible.[3] Williams no doubt retained many memories of his past actions and is clearly metaphysically connected to the man he was when he (allegedly) committed the crimes. It is the change in how he viewed his

[1] Schwarzenegger, Arnold. "Statement of Decision: Request for Clemency by Stanley
[2] Warren, Jennifer and Mura Dolan. "Tookie Williams Is Executed," *LA Times*, Last modified Dec 13, 2005. http://www.latimes.com/local/la-me-execution13dec13-story.html.
[3] Schechtman, Marya. *The Constitution of Selves*. (Ithaca: Cornell University Press, 1996): 76.

past and his disavowal of it that raises questions for continued identity attribution for the purposes of moral responsibility. His supporters argued that the thorough change in Williams represented more than reform. They suggested that a new man had emerged within the prison cell. What changed was his perspective and attitude towards the past and it was these changes that seemed to justify an appeal for clemency.

In responding to the appeal, Gov. Schwarzenegger was left to answer whether Williams' redemption was "complete and sincere" or "just a hollow promise."[4] My question for this thesis however is not about the details of this case or how to determine the veracity of another's statements. Instead, I want to explore the claim of innocence made by Williams' supporters and ask whether attitudinal shifts could ever warrant absolution from responsibility. Certainly, stories of reform are commonplace in literature and real life, yet while each case may differ in content, they remain linked through the sentiment reformed individuals are likely to express: "That is not me anymore." These individuals seem no longer able to psychologically inhabit the life they once lived. Their attitudes, judgements and ways in which they approach the world have radically, if not fundamentally changed. In these cases, we are no longer looking at the continuation of identity proper – that is, numerical sameness – but the continuation of a vague and almost inexplicable notion of the self or 'qualitative' sameness. Forceful repudiation of and alienation from past actions suggest that in a subtle, but no less important sense, these agents have become different persons. Hence, the traditional focus on the metaphysical conditions of identity do not necessarily accommodate the possibility that metaphysical continuity and the continuity of the personality could be distinct conditions

[4] Schwarzenegger, "Statement of Decision", 5.

of persistence through time and have different implications for questions of moral responsibility. Whether or not Williams satisfies these criteria, however, does not answer whether or not it is appropriate to continue holding him morally responsible and whether the execution was indeed justified. That is, following radical changes of character and personality, does the agent remain apt for attributions of moral responsibility over time? I argue that these cases raise deep questions about both responsibility and the continuation of the self.

Chapter Outline

The question of innocence introduced by Williams and his supporters is more complicated than simply asking if he was the one to commit the crime. Instead, we need to ask whether it is possible to attribute his criminal past to the person he was before he was executed. Could Williams' apparent rehabilitation amount to exoneration on the grounds of being a new self? I see the answer to this question as involving four subsets of questions: "What are the conditions of a self at a time?", "What makes one responsibility-apt?", "What are the conditions of responsibility-aptness over time?" and "What matters for rehabilitation?". If these questions are answered, I believe we would then have a means of clarifying Williams' situation.

I take the question 'What is the self?' to be an important first question as some conception of self or self-loss is assumed in questions of extended responsibility. We speak of 'losing oneself' in an identity crisis, when committing an act we deeply regret or, to borrow Locke's terms, when one can be said to be "not himself" or "beside

himself".[5] A change has occurred, something that we have taken as central to 'who one is' has been lost, and how we treat the individual thereafter will be affected. But what exactly has been lost? I begin to form an answer to this question that I would like the reader to envision as a process analogous to that of a sculptor working on some bronze. In the first chapter, I begin with the lump of bronze and slowly chip away until we get something that has the shape of a self that could be a basis for responsibility attributions. Subsequent chapters will then work to refine this shape until we arrive at something with sharp features that is more clearly recognizable as the kind of moral self I am ultimately after. After the moral self has been defined in this way, I move on to the third and fourth questions that concern when responsibility-aptness may continue and when a person may be properly considered rehabilitated.

I will begin this process with the work of John Locke as the base material, then chisel away by appealing to contemporary theories of responsibility. Starting with Locke may seem bizarre because he is often considered the founder of the discipline I distinguished from my inquiry just a moment ago, namely one that is concerned with numerical identity conditions over time. I assure the reader that this starting point is not a mistake. To begin to answer the question of the self, I want to first uncover a notion of the self buried within a potential misinterpretation of Locke's work. I will push against the tradition of Lockean criticism claiming that he held a deeply problematic memory theory of identity. I argue alongside recent interpretations of Locke that he was actually unconcerned with the question of the continuance of metaphysical identity proper in his

[5] Locke, John. *An Essay Concerning Human Understanding*. (Raleigh, N.C: Alex Catalogue, 1990): II.xxvii.17.

seminal chapter: "On Identity and Diversity". He was not a memory theorist, nor was he particularly interested in offering an account of numerical sameness. I argue that this new interpretation leads us down an intriguing path and towards a new Lockean inspired notion of the self as defined by persisting phenomenology.

Chapter two turns to how we should understand the idea of a persisting Lockean consciousness. Although traditionally interpreted as constituted by memory, persisting consciousness actually involves more than recollection or even the aptitude for reflection. Drawing on a number of contemporary theorists, I argue that consciousness, rather than being a reflective operation of the mind, is instead a pre-reflective and non-egological apprehension of an experience, which in turn provides the basis for the experience of the self and an answer to my first question ("What are the conditions of a self at a time?"). What constitutes consciousness is that which affects and influences experience at an unselfconscious and immediate level. It is this sort of influence and framing of the world that generates a distinctive phenomenological unit that I take to be the basis of *the self*. I argue that the self in general consists in consciousness understood as what influences and inflects present experiences or *the motivational profile* that creates a distinct phenomenological subject.

The notion of a motivational profile is useful to highlight the main features of a self. However, it is too broad to accommodate the notion of being morally responsible. Chapter three will focus on this problem of over-inclusiveness. In order to meet these worries, I refine the concept of the self by taking a cue from Angela Smith's notion of answerability. I argue that the aspects of the motivational profile in which persons are responsible are also those in which they are answerable. I consequently narrow the

phenomenological profile to what I call the *evaluative profile*. What is important for an evaluative profile is not the specific collection of evaluative beliefs and desires themselves, but how these aspects of one's psychology coalesce and contribute to one's general outlook on the world. I will thus answer the second question concerning responsibility-aptness ("What makes one responsibility-apt?") by arguing that actions that can be connected to these sorts of evaluative aspects are in consequence responsibility-apt because they represent and reflect the self or more appropriately, the *moral self*.

The focus on the phenomenological perspective means that the moral self is not necessarily unified along traditional parameters. In particular, I will argue in chapter four that the test of which psychological aspects of the self are attributable to the moral self is one of *evaluative sensitivity*, which leaves open the possibility that selves may be conflicted. This chapter will also begin to address some potential criticisms of my view, in particular whether the self needs to be deeply integrated in order to be responsibility-apt. I will argue that the test of evaluative sensitivity allows the possibility of a fragmented and conflicted moral self. This is important because it begins to demystify the situation of rehabilitated offenders. No longer are the kinds of conflicts that characterize who they are exceptional, but decidedly the norm. Regular law-abiding citizens also experience the kind of psychological conflict that may be experienced by rehabilitated offenders.

However, once moral selves are understood as loose clusters of evaluative attitudes and beliefs in this manner, it is not clear how such selves could possibly persist over time. Chapter five will shift from the previous focus on the self *at a time* and begin to

answer the third question of responsibility-aptness *over time*. I will no longer be working to sculpt the shape of the self, but will be asking how this newly defined moral self persists through time and which specific features need to be retained through fragmentation and conflict. We will see that to be responsibility-apt does not require the exact same constitution of the evaluative profile over time. Just as the notion of answerability (evaluative sensitivity) characterizes moral responsibility at a time, I argue that agents remain responsibility-apt over time only if they remain answerable. Responsibility-aptness may be preserved though alteration - even radical alteration - because of what I call *evaluative access,* namely the ability at time to inhabit former values and beliefs. When we are dealing with a changed self, such as a criminal offender, what matters is not only whether they repudiate the past, but also whether they are deeply connected to it.

Chapter six will focus on setting the parameters on just how connected an agent must be to the past by drawing on arguments from Sorites-type cases. I will suggest that being answerable may be consistent with much phenomenological change as long as the agent is *saliently similar* in some relevant respects. Being absolved of responsibility-aptness for a particular act does not require a wholesale change, but only a change that is relevant to the act in question. This chapter represents an answer to the third question ("What are the conditions of responsibility-aptness over time?"). The chapter also shows that it is possible for a later self to be radically different from an earlier self yet still answerable, and also no longer answerable despite being fairly similar.

In chapter seven, I will argue that the salient similarity highlighted in the previous chapter may be maintained by the articulation of a narrative, which extends

answerability. Narratives can extend evaluative access due to the way they maintain salient similarity of the current moral self to the past through their backward and forward looking-functions. The narrative provides a link to former evaluative beliefs that keeps them at the forefront of the individual's conscious experience and conditions how they experience the world thereafter. As a result, the narrative extends answerability in a personal way.

However, the fact that narratives have this function brings to light a twist in the project. In chapter eight, I will return to the initial question of responsibility of the criminal offender, and contend that the conditions of responsibility-aptness identified in this thesis to not settle the independent questions of blame, forgiveness or even whether the criminal should be considered rehabilitated. This is an unexpected result of the discussion in the thesis, but this result also answers the forth question ("What matters for rehabilitation?"). Once we focus on the notion of narrative articulation, we see that taking responsibility and extending answerability in the way I described in chapter seven may allow responsibility-aptness to persist and at the same time deliver the conditions required to withhold blame, warrant forgiveness and treat the offender as rehabilitated. I adopt recent theories of blame and forgiveness as essentially communicative to make this argument.

Against the claims of the protestors at Williams' execution, we should not ask whether Williams was a 'new self', but focus on how he conceived of himself and his past before his execution. We should be concerned as to whether he has properly taken responsibility for his past through narrative appropriation of it. Change without such narrative appropriation is problematically unguided and contingent in a way that would

not satisfy the communicative conditions necessary for victims to forswear blame, or forgive Williams and consider him rehabilitated. Thus, to ask whether criminal offenders are not responsible for their past is different than asking if they are rehabilitated. The former implies that there is significant and sufficient change to one's evaluative profile, but to be considered rehabilitated requires an extension of the moral self.

What are the conditions of a self at a time?

Chapter 1: Lockean Consciousness and the Forensic Unit

Introduction

Continuing the sculptor analogy, this chapter will start by clarifying the kind of material I will use to define the self. In particular, I will begin with recent interpretations of John Locke's notion of consciousness. In *An Essay Concerning Human Understanding*, Locke argued that identity is determined by a person's "consciousness."[6] What exactly is meant by consciousness and what implications it has for personal identity theory however has been subject to much debate. Galen Strawson and Matthew Stuart in their recent interpretations of Lockean identity theory have suggested that Locke, in contrast to how he has traditionally been interpreted, surprisingly was not concerned with many of the questions considered fundamental to personal identity theory.[7] Locke's work may be better understood as determining the limits of identity insofar as it tracks the extent of an individual's moral responsibility.

The first section will begin to question traditional interpretations of Locke's theory, while §2 will suggest a revised interpretation. Rather than seeking persistence conditions to determine identity conditions, the focus will be narrowed to the requirements for proper attributability and characterization, which (as we will see) will be distinguished as a separate and secondary question from identity determination using insights from Marya

[6] Locke, *Essay,* II.xxvii.9.
[7] See Perry and Olsen for a good overview on the major themes in personal identity Oslen, Eric.T. "Personal Identity" In *The Blackwell Guide to Philosophy of Mind,* edited by Stephen P. Stich and Ted A.Warfield. 352-368, (Oxford: Blackwell Pub, 2003).

Schectman's work in §3. Overall, the purpose of this chapter is thus to lay the boundaries for a more narrowed approach to persistence, responsibility and change over time.

1. **Memories and Consciousness**

Although the term has been highly contested, one thing that is clear of Lockean "persons" is that they persist insofar as the consciousness extends and no further.[8] Locke clearly states:

> … since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity, i.e. the sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.[9]

As a mental capacity that can "repeat the idea of any past action with the same consciousness it had of it first; and with the same consciousness it has of any present action", memory has traditionally and historically been regarded as the fundamental faculty that constitutes Lockean consciousness.[10] He speaks of consciousness as fundamentally connected to memory when he states, "to remember is to perceive anything with memory, or with a consciousness that it was perceived or known before."[11] On this reading, the extension of consciousness consists in the sameness of episodic memory (autobiographical memory that can be recalled) as both necessary and sufficient for sameness of persons. It is sufficient in that past action may correctly be attributed to the person when the present subject remembers having the thought of or having performed the action, and necessary, given that without memory recall, there is no link to

[8] Ibid,. II.xxvii.28.
[9] Ibid,. II.xxvii.9.
[10] Ibid,. II.xxvii.10.
[11] Ibid,. I.iii.22.

the past action. Memory, on this interpretation is how "the consciousness of this present thinking thing can join itself". [12] Once remembered, this connection "makes the same person, and is one self with it, and with nothing else; and so attributes to itself, and owns all the actions of that thing, as its own."[13] In what follows, I would like to question this traditional interpretation of Locke. The heavy criticism this interpretation has received has been so devastating to the theory it is a wonder why Locke did not see it himself. Rather than thinking Locke held a deeply mistaken theory we might instead assume that the theory itself has been misunderstood.

Locke's work, understood as a simple memory theory of identity, has received numerous criticisms almost as well known as the theory itself.[14] Perhaps the most notorious objection to the simple memory theory derives from the question of the permanence of memory relations and forgetfulness. That is, "if all I am is my memories, what happens to me when I forget?" Famously, Thomas Reid takes up this question through recounting the life of a brave officer. He asks us to imagine:

> A brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging[15].

---

[12] Ibid,. II.xxvii.19.
[13] Ibid.
[14] At this point we might also question how to distinguish veridical memory from that which may be falsely remembered. Unfortunately, in this thesis, the answer to this question will not appear in until much later and in the context of narrative self-constitution in chapter seven. There, I argue that there is space for embellished or inaccurate self-conceptions (which may include memories) insofar if those factual inaccuracies enlarge more general truths about the self.
[15] Reid, Thomas. *The Works of Thomas Reid: With an Account of His Life and Writings*. (Charlestown: Printed and Published by Samuel Etheridge, Jun'r, 1813): VI.III.351.

Reid's charge against Locke's simple memory theory is that even though it seems intuitively plausible that the general and the flogged boy are the same person, this cannot be so on Locke's account. Given that the general has no recollection of stealing the apples from the orchard or the ensuing flogging, the lack of sufficient memory connections bars an attribution of identity. Consciousness as composed of episodic memory is too demanding in that it cannot accommodate the possibility of ordinary forgetting. Locke makes clear that memory is intransitive, whereas identity requires a transitive relation.

As a result of such criticisms, many have sought to answer Reid's charge by modifying the simple memory theory and insisting that these refinements would be welcomed by Locke or at least consistent with his broader claims.

1(a). Psychological Continuity and Connectedness

Most Neo-Lockean attempts have focused on connecting the general to the flogged boy by means of a succession of appropriate memory links or *psychological connections*. This approach is concerned with highlighting the vast array of psychological connections - including connections between, persisting beliefs, values, desires or intentions and the corresponding actions - that provide continuity between earlier and past persons when memory connections fail to hold.

Perhaps the most famous Neo-Lockean theorist of this variety is Derek Parfit. He argues that mental activities carry a temporal reach to past psychological relations that connect earlier experiences that, when taken together "overlap like the strands in a rope"

to substantiate a continuity when connectedness fails.[16] As the overlap continues, we are given a clear link to the past through an ancestral relation.

Psychological continuity theories, like that proposed by Parfit, provide a means to remedy the Lockean account. Psychological continuity acts as a bridge to connect the person when memory fails. Hence, the general need not remember all that has happened before if the 'rope', constituted by various psychological connections, remains continuous. As we track the length of the rope, there is a measure of continuity even if particular strands of memory are no longer directly connected.

As noted by J.L Mackie, however, even if this approach saves the Lockean memory theory, he would likely not endorse it. Mackie states of these sorts of continuity theories, to conceive of Locke's memory criterion in this manner "is a revision, not an interpretation, of Locke's account. Not only does he not say this, he commits himself explicitly to a different view."[17] Even if the psychological continuity approach provides a measure of persistence over time, the fragmentary nature of consciousness Locke describes seems to bar such an interpretation. Locke asks:

> Suppose I wholly lose the memory of some parts of my life, beyond a possibility of retrieving them, so that perhaps I shall never be conscious of them again; yet am I not the same person that did those actions, had those thoughts that I once was conscious of, though I have now forgot them? [18]

Locke answers this question by noting that insofar as it is "possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt

[16] Parfit, Derek. *Reasons and Persons*. (Oxford: Clarendon Press, 1987): 222.
[17] Mackie, J. L. *Problems from Locke*. (Oxford: Clarendon Press, 1976): 182.
[18] Locke, *Essay*. II.xxvii.20.

the same man would at different times make different persons."[19] Locke not only acknowledges the possibility of irretrievable memory loss, but also argues that, when such an event occurs, the same person can be said to no longer exist. Several times in his chapter on identity, Locke emphasizes that personal identity extends only "as far as that consciousness reaches, and no further."[20] Thus, if ancestral relations were applied here to ensure continuity, it would be no longer clear as to why Locke would consider such a loss of consciousness a threat to one's identity. The general might be continuous with the flogged boy, but this is not enough of a connection for Locke. His insistence on the consequences of such a loss of consciousness seems to bar any natural progression to a theory of memory continuity as an ancestral relation. Indeed, Locke goes further even to suggest that the same consciousness is required if we are to say that the person remains through sleep as he states, "If the same Socrates waking and sleeping does not partake of the same consciousness, Socrates waking and sleeping is not the same person."[21] Despite the apparent continuity that could connect a person through sleep, without consciousness, we can be sure on Locke's account that the "the selfsame person was no longer in that man."[22]

According to Locke, for any period of time, regardless of whether the duration lasts only a minute, if an item cannot be recalled, it simply is not part of your person. If this were Locke's basis for determining identity, the irretrievable loss would suggest that persons could come into and out of existence during everyday bouts of sleep or

[19] Ibid.
[20] Ibid,. II.xxvii.17.
[21] Ibid,. IIxxvii.19.
[22] Ibid,. II.xxvii.20.

forgetfulness. Given the apparent counter-intuitiveness of such a claim, it would seem that Locke simply did not realize the full implications of his remarks. Matthew Stuart surprisingly suggests however, that we should take Locke's statements at face value and derive a more charitable interpretation. The strangeness of Locke's work is mitigated when considering it in conjunction with Locke's stance on the relativity of identity. To dismiss Locke as having made a serious error not only misses an important aspect of Lockean identity conditions, but also commits him to answering a question that he may not even have asked.

1(b). Relativity of Identity

The idea that a person can cease due to forgetfulness is based on the principle that it is only when a past mental item can be joined to a present one by consciousness that it can be part of your present person. To fully appreciate this claim and its deeper implications, we need to distinguish the kind of persistence condition Locke uses to characterize persons.

It is important to note that he is using the term "person" in a different manner than is usually conceived. Conventionally, the focus of personal identity theory has been on the relation of numerical sameness and the question of re-identification of the person. In this traditional interpretation, when asking if the person is the same, we are asking whether there is a relation of numerical sameness between persons at earlier and later times. This however is not Locke's usage insofar as Lockean persons do not behave in a manner consistent with what is required for numerical identity to obtain. When a past event is forgotten, it may no longer be part of the person, yet the forgotten event may still

constitute a part of a person's numerical identity. This is because Locke uses the term 'person' as a particular sortal term in a way that I will now explain.

Locke is careful to stress at the beginning of his chapter that identity conditions are relative depending on the kind of entity in question. Importantly for Locke, there is a sense in which I could be said to remain despite irretrievable memory loss. Yet, in saying that I remain, "we must here take notice what the word I is applied to; which, in this case, is the man only."[23] When determining the identity of the entity, we must first be careful to distinguish what kind of entity it is. Starting with the simplest constitution, for Locke, the span of existence for masses of matter and other living things is determined by spatial-temporal continuity in different ways. For non-living masses of matter, identity will change any time a particle is removed or added. Masses of matter can be organized in various ways, but they need the same particles to remain the same masses of matter. Living bodies - including plants, animals and even man considered as a human being- are also mereological sums of these atoms, yet Locke argues that identity can be preserved through both large-scale and minute changes in these sums of particles. The different bits of matter participate in that life at different times, thus allowing the identity of an acorn to remain connected to the oak tree insofar as it is "partaking in the common life" of its species determined by a "vital union and shape."[24]

Much the same can be said of Lockean "man" (human) and "person". The man is constituted by various bits of matter, unified in an overarching life. In turn, the man may constitute the person. The man does not require any specific collections of matter nor

[23] Ibid.
[24] Ibid., II.xxvi.28

does the person require the particular man. Although the man and person usually occur together and we often speak this way of persons, Locke is clear that these are separable and need not coincide. He illustrates this through several body swapping scenarios, including a potential exchange of bodies between a price and cobbler. The result is the consciousness (person) in the body (man) of the cobbler and vice versa. To be the same substance or man does not necessarily mean one is the same person. Locke continues:

> It is not therefore unity of substance that comprehends all sorts of identity, or will determine it in every case; but to conceive and judge of it aright, we must consider what idea the word it is applied to stands for: it being one thing to be the same SUBSTANCE, another the same MAN, and a third the same PERSON, if PERSON, MAN, and SUBSTANCE, are three names standing for three different ideas … [25]

Importantly for Locke, man, persons, and substances have identity conditions that differ in rather dramatic ways. The identity of man provides a means to determine where a human being's spatial-temporal limits lie. However, Lockean persons, under the current interpretation, cannot be fully understood in the same spatio-temporal manner because identity conditions differ according to the kind of thing or entity we are considering. Locke argues for a *relative identity theory* where "man" and "person" are not the same and will have differing identity conditions relative to the kinds of things they actually are. When considering the entire temporal span of a particular person's life, the relative identity theory tells us that although there are many actions that belong to the whole person in terms of his numerical identity, not all of these events, thoughts, or actions adhering to the man can equally be said to constitute the person. To conflate the person and man is to make the mistake of assuming that what you are is essentially a man where continued numerical existence does not extend once the man does not.

[25] Ibid., II.xxvii.8.

Rather, 'person' functions on Locke's account as a particular, but nevertheless peculiar, sortal term or "abstract idea" as Locke defines it.[26]

Locke distinguishes persons considered specifically as persons, from the question of the persistence of the human being in general. Forgetfulness may lead to radical disconnects in temporal continuity of persons, but there remains the "man" to instantiate a persisting existence between these apparent lapses. Locke distinguishes between what he terms "real" and "nominal" essences and we can only have knowledge of the latter.[27] A real essence refers to the physical constitution in the sub-microscopic physical structure that causes the observable qualities in substances. A nominal essence by contrast, signifies the abstract ideas of those entities whose identity is under consideration. To define a particular nominal essence requires piecing together a collection of particular qualities (primary, secondary, etc.) that can produce ideas within the mind. It is a naming process that employs the qualities associated with a certain idea in order to create specific taxonomical categories. Once applied to the discussion of persons, Locke can be read as attempting not to define the real essence of a person but one nominal description of the many that persons can fall under as a matter of convention. Persons are considered under a name that denotes a specific idea and "for whatever makes the specific idea to which the name is applied, if that idea be steadily kept to, the distinction of anything into the same and divers will easily be conceived, and there can arise no doubt about it."[28] The nominal essence used to define persons in this

[26] Ibid., III.iii.15.
[27] Ibid.
[28] Ibid.

case refers specifically to persons in regard to responsibility attributions.[29] It is a "forensick term" (*sic*), which "appropriates action and merit" in carrying the weight of determinations of moral responsibility.[30]

If Locke's concern for persistence of the person is defined specifically in forensic, nominal terms, then we can agree that Reid's general may not be the same person as the young boy, as this does not force us to assume that the boy and the general are not numerically the same; rather the extent of responsibility for stealing the apples from the orchard may have simply run its limit. The malleability of the continuance of persons supplies a point of expiry when responsibility no longer holds. He is the same man, but it is questionable as to whether he is the same person responsible for the theft. Consequently Lockean identity conditions are not absolute, but are relative to the kinds of 'forensic' concerns we might have when attributing responsibility.

2. Uncovering the Moral Identity

The restricted focus on forensic concerns has been criticized on a number of grounds. In particular, J.L. Mackie argues that if Locke's theory of personal identity is interpreted in this manner, then it no longer qualifies as a theory of personal identity at all. It is not a theory of when a person is thought to begin and end, but what is "better described as a theory of action appropriation."[31] The theory on this interpretation merely provides grounds as to whether an action is attributable to a person, but is silent on the numerical identity of that person. Stuart suggests that Locke could simply respond:

[29] See Udo Theil's work "The Early Modern Subject", specifically §3.2 for a discussion of these themes.
[30] Ibid., II.xxvii.26.
[31] Mackie, *Problems,* 184.

"That's not the objection, that's the theory!"[32] If we are to read Locke with charity, we might instead think of his theory as determining how actions and past experiences are to be considered as part of the present person for the purposes of attributing moral responsibility.

The problem and apparent absurdity generally attributed to Locke's account, is thus from "names ill-used, than from any obscurity in things themselves."[33] Indeed, as noted by Stuart, because Locke "thinks that the span of a person's existence can decrease or increase abruptly as time passes, his account of persons is at once weirder, more interesting, and less easily refuted than it is frequently taken to be…"[34] Persons permit a kind of "gappiness" in their span of existence as they can increase or decrease and even move in and out of existence over time.[35] Stuart continues:

> Ordinarily we presume that our temporal extent can increase only by the addition of segments at the later end and the accumulation of thoughts and actions in the present. On Locke's account, a person's temporal extent can increase by the addition of thoughts and actions that occurred in the past. Ordinarily we presume that our existence is, and must be, continuous. On Locke's account, the temporal extent of persons can be very *gappy*.[36]

Given the distinct way the Lockean person behaves, we may question whether Locke's account was meant to answer questions of numerical persistence. As Strawson argues, "We have, then, a conflict. We have a respect in which personal identity over time may be said to be a *gappy* thing, and a sense in which personal identity over time

---

[32] Stuart, Matthew. *Locke's Metaphysics*. (First ed. Oxford: Clarendon Press, 2013): 379.
[33] Locke, *Essay*, II.xxvii.28.
[34] Stuart, *Metaphysic*s, 341.
[35] Strawson. Galen. *Locke on Personal Identity: Consciousness and Concernment*. (Princeton: Princeton University Press, 2014.): 8.
[36] Stuart, *Metaphysics*, 374. Emphasis mine

must be a matter of genuine continuity."[37] The concept of a Lockean person thus behaves in way quite unlike other sortal concepts (like "human being" or "thinking thing") as it is compressed and made wholly forensic.[38] It tracks not the person's spatio-temporal continuity (and perhaps even presupposes it), but the nominal essence of a person *qua* unit of moral responsibility. The lack of transitivity may indeed be a problem if we are concerned with providing the numerical persistence conditions of persons, but not if we remain concerned with responsibility attributions in particular. To take Locke's work on identity as internally consistent and allow that the existence of persons ends so far as consciousness reaches, we may have to admit that the often-heralded forefather of personal identity theory was rather uninterested in its canonical question of numerical continuity.

Strawson embraces this interpretation of Locke (although he denies that Locke's theory – even considered as of a forensic term – can be fully understood as a simple memory theory). In particular, he interprets Locke as being quite indifferent about the numerical identity of the subject of experience that constitutes the person. He argues instead that the temporal continuity underpinning numerical identity is taken as a given in Locke's text. Strawson writes:

> For one way to characterize the central error in the interpretation of Locke's discussion of personal identity is to say that it rests on a failure to recognize this point, a failure to realize that Locke simply *assumes* the continuity of the [subject of experience] component of [person]. He takes it as given, even as he engages in dramatic thought-experiments, involving soul-jumping and body-hopping, when considering what might underwrite this continuity (what might "carry" it, what it might "reside" in).[39]

[37] Strawson, *Personal Identity*, 8. Emphasis mine.
[38] Stuart argues that Locke holds a simple memory theory. Yet, I will contest this presumption in the next chapter.
[39] Strawson, *Personal Identity*, 10.

Strawson claims that the central error that arises when interpreting Locke is the tendency to characterize his work as attempting to give an account of the numerical identity of the subject of experience that realizes the person. Lockean persons require some substantial realization yet, on Strawson's reading, we are told that Locke simply defines the subject of experience who may qualify as a person and asks about the identity of this subject considered specifically in moral terms. That is, what realizes the person is left an open question as he concentrates on the forensic concerns.

The person as a nominal term tracks aspects of who we are that are important for attributions of moral responsibility. In one of Locke's most re-iterated passages, he states:

> This being premised, to find wherein personal identity consists, we must consider what person stands for which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it.[40]

Stuart breaks down this passage in order to see that a person under Locke's definition must first have (i) reason: as the ability to discover immediate ideas from reflection and (ii) reflection: as the ability to perceive the operations of the mind including ideas of mental activities involving thinking, believing, willing etc. Included in this capacity for reflection is the ability to take one's own thoughts as the object of thought.[41] Together (i) and (ii) form the necessary conditions for persons as the base capabilities required to support more complex operations. They are necessary, but not sufficient for personhood. Next, (iii) the ability to consider itself as itself, which moves

[40] Locke, *Essay,* II.xxvii.9.
[41] Stuart, *Metaphysics*, 341.

beyond mere reflection in that this ability requires employing the idea of a person and applying it to oneself. This condition picks out the higher order capabilities of human persons such as self-consciousness. That is, self-consciousness requires being able to consider oneself the same thinking thing at different times and places. These components of the Lockean definition of the person supply the criteria necessary to make the person a proper object of reward and punishment. Hence, Locke first gives the cognitive definition to delineate the sort of properties required to be persons of a particular kind. As noted by Strawson, he starts with "the fact of enduring and continuously existing, personalities, human subjects of experience who are born, live, and die, who are—he assumes—eventually resurrected, who can feel pleasure and pain, happiness and misery, and who are, quite crucially, 'capable of a law.'[42] This move defines the necessary conditions to qualify as a person considered specifically in forensic terms.

The practice of setting rewards and punishments necessitates an entity that can understand himself as persisting and is able to appreciate the consequences of actions. Then, only after the subject of the inquiry is defined, does Locke move to the central condition of persistence as the extension of consciousness, which need not necessarily involve the aforementioned traits. To fully grasp this structure, Strawson asks us to picture Locke's project as first defining the subject of our inquiry, then defining what is important for responsibility attributions as a matter of picking candidates from a pool of properties associated with the person or human being. The task is to determine what aspects of the pool should be considered as most important for our forensic concerns.

[42] Strawson, *Personal Identity*, 23.

Locke's structure for defining the moral identity of persons will be similar to my own and, here, is where the sculptor analogy framing my project may be especially helpful. The process will by first forming the shape of the self considered specifically in forensic terms. Afterward, rather than picking candidates from a pool of possibilities, the process will involve gradually chipping away at various aspects of person's psychological life in order to render the final product an accurate representation of the self considered specifically in terms of being an appropriate target for responsibility attributions. For reasons I will develop in the next chapter, I will not necessarily settle on the same candidates as Locke even if I take my choice to be Lockean-inspired. First, however, I will clarify the kinds of questions that will concern my inquiry by exploring some important distinctions made by Marya Schechtman in order to not only better situate my project, but also draw further parallels to Locke's work.

3. Characterizing the Inquiry

Marya Schechtman's terminology concerning selves brings some needed clarity to both Locke's work and my general aims in this thesis. The interpretation of Locke so far suggests that his concept of persons is better suited to answering, not a question of "re-identification" that is traditionally attributed to him, but rather one closer to what Schechtman calls "the characterization question" – a question that concerns the "set of characteristics that make a person who she truly is."[43] The characterization question is an important one and I will close this chapter by further elaborating what is involved in answering it.

[43] Schechtman, *The Constitution of Selves*, 76.

Schechtman argues that, buried in the debates within personal identity theory, there are two strands of inquiry that have not been adequately distinguished. The first is the "re-identification question," that asks, "what makes a characteristic part of a person's history" in the sense required for assessing whether the "basic" attribution of the characteristic to that person is justified. [44] The second is the "characterization question," that asks what makes that characteristic "*truly* hers."[45] We can see this distinction in the kind of redemptive change Williams undergoes. As Schechtman argues, "We might say that [he] has become a different person, but there is some sense in which we clearly do not mean it. The change is only remarkable because [he] also remains the same person."[46] If we were to ask the question of re-identification, it is obvious that Williams is the same numerically speaking. He survived in that prison cell in a *basic* sense. He is the same man as the one who (allegedly) committed the crimes so he satisfies the conditions of *basic attributability*. Yet, there is a further sense of survival to be considered: there is an alleged failure in "*subtle"* survival and in that sense it is not as clear that he should be characterized as the same.[47] The affective discontinuity between the earlier and the later reformed Williams challenges the conclusion that he has the same character. I will call the sort of attributability that concerns continued character - *deep attributability*. Deep attributability determines the sphere for which persons are apt for attributions of moral responsibility or *responsibility-apt*. That is, it may or may not be

[44] Ibid.
[45] Ibid.
[46] Schectman, Marya, "Empathic Access: The Missing Ingredient of Personal Identity." In *Personal Identity*, edited by Raymond Martin and John Barresi. (Oxford: Blackwell, 2003): 241.
[47] Ibid., 256. Emphasis mine.

legitimate to hold the agent responsible for a previous crime or transgression (because there might be some legitimate excusing conditions), but they are at least 'open to' or 'apt for' attributions of moral responsibility because they satisfy the conditions of deep attributability.[48]

The distinction between re-identification and characterization is important as it moves the line of inquiry away from identity as numerical sameness. Recently however, Schechtman has noted that the notion of characterization may be more complicated than she initially supposed.[49] She reframes Locke's project as concerned with what she calls the "forensic unit."[50] The "forensic unit" determines "the limits within which questions about responsibility and self-interest are appropriately raised", but not necessarily settled.[51] Importantly, articulating what is involved in the constitution of the forensic unit represents a significant initial step but does not quite get us to what I have called

[48] I should note that here I am referring to justifying conditions as opposed to excusing ones. A justification does not necessarily speak to whether or not the action is deeply attributable. Instead, when an action is justified it might be understandable or reasonable given certain circumstances and thus a reason to not hold the agent morally responsible. It justifies the act without denying that it was indeed attributable to the agent.

[49] In her recent work, Schechtman has now turned her focus to broadening the concept of person to mean more than what is important for forensic concerns. She argues that because "people in our lives are practically significant to us in all kinds of ways", any definition of the general identity of a person needs "to allow for the personhood of infants and the cognitively impaired, and so to explain how a Lockean person can be the same person as her earlier infant self and later demented self."(Schechtman, *Staying Alive*, 68,103). Schechtman is right that Locke's focus on the forensic does not encompass all that might be practically important in being a person and constitute proper attributability in these general terms. Nevertheless, I see my concerns in this thesis as less ambitious than Schechtman's and importantly narrower. It may be the case that Schechtman's account better answers a broader question of what it means to be a person. The Lockean person is a limited term, but one that I see as usefully limited given my line of inquiry. See *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. (New York, NY: Oxford University Press): 2014.

[50] Schechtman, *Staying Alive*, 14.

[51] Ibid.

"deep" attributability. For instance, on Schechtman's analysis, the potential transfer of consciousness in prince and cobbler body-swapping scenarios highlights the Lockean forensic unit (but not necessarily the moral self that is required for attributions of moral responsibility). She suggests that Locke's argument concludes by advising that "contrary to what we might think before reflecting on the matter, it is not sameness of body, or even soul, that sets the limits of a forensic unit; it is rather sameness of consciousness."[52] For the purposes of attributing responsibility, we "would look not for the right man, but for the right person."[53] Locke's inquiry thus shows where best to look in order to initiate the characterization question. In other words, Lockean consciousness describes the necessary but not sufficient conditions for responsibility-attributions. It is one step narrower than what might re-identify 'the man', but the person as a forensic unit is still not narrow enough to define the moral self. A further step is needed.

To spell out Locke's aim more clearly, we might reframe the difference between the forensic unit and the moral self in a different way. Here, we could say that if an item belongs to the forensic unit rather than the moral self, that item is attributable to the person, but not deeply attributable in the sense required for attributions of moral responsibility. For instance, let us use Schechtman's example of Jane knocking over a lamp. Jane, while rough housing with her brother, Dick, and her dog, Spot, accidently knocks over a lamp while her mother was out of the room. If Jane's mother were serious, and perhaps a little melodramatic about determining "whose body actually impacted the lamp", she might, just as would be appropriate for a criminal investigation,

[52] Ibid.,19.
[53] Ibid., 16.

"first check to see if they have the right [person] using tools like fingerprints and DNA analysis" before fully appraising Jane's responsibility.[54] Basic attribution is satisfied once the mother is able to determine it was Jane who was the proximate cause of the broken lamp. However, Locke's argument is that even though Jane's mother arrived at the appropriate target, being the same human animal to have broken the lamp is not what matter's most for our forensic concerns. She should track consciousness instead. Yet, even once it is determined that Jane has the same consciousness as the earlier person who broke the lamp, in order for the act of breaking the lamp to be *deeply attributable*, more evidence is needed. This is the question of whether Jane is responsibility-apt for breaking the lamp. Facing her accusatory mother, Jane could note that "she had one of her seizures or fainting spells and flailed into the lamp, or perhaps Spot ran in and knocked her over, or perhaps she is only eighteen months old and has no idea of the consequences of putting her hands on the lamp."[55] In each case, Jane's consciousness persists and is identified as the proper target of our forensic concerns, but we have not yet determined whether Jane is suitably responsibility-apt. Hence, identifying the proper forensic unit is analogous to when "mother chooses *Jane* as the child to question" whereas determinations concerning the moral self involve asking whether the action was indeed deeply attributable to Jane.[56]

In this thesis, questions about the "moral self" concern the "actions and experiences for which [an agent] is in fact held rightly accountable."[57] Defining the

[54] Ibid.
[55] Ibid.
[56] Ibid., 15.
[57] Ibid.

forensic unit (the person, or consciousness) does not quite answer the characterization question and what it means for an action or experience to be deeply attributable. Thus, I see my project – namely, the focus on the persistence conditions of the moral self – as speaking to this absence in Locke. However, this focus does not mean that I am fully unconcerned with the broader notion of the forensic unit. In fact, it is through developing an account of the constitution of the forensic unit that my project gets off the ground. Much of the thesis will be aimed at understanding moral responsibility after radical character change by using Schechtman's proposed distinctions between the forensic unit and moral self. I will take Locke's lead and start with persons specifically as forensic units and offer a new interpretation of Lockean consciousness in order to arrive at a more complicated question of deep attributability and proper characterization. These questions will then be further refined as distinct from conditions of *blameworthiness*. Although persistence of the moral self make blaming activity appropriate, whether one is open to blame also involve ruptures to social relationships that determinations of responsibility-aptness is unable to fully address. I will also, like Locke, presuppose that the persistence conditions required for re-identification are a conceptually prior, yet distinct, issue from my proposed line of inquiry. Like a mallet removing swathes of material from the initial lump of bronze, his insights will set the basic shape of what we might call the self as a forensic unit. I will later work this initial form into a final finish and something resembling a moral self. Once this is completed, I will discuss the conditions of persistence for this newly formed moral self, responsibility-aptness and blameworthiness.

Conclusion

With the initial distinctions in place, I hope that this chapter has opened the door to reconceiving Locke's work and shifting the direction of inquiry to the one that frames the approach in this thesis. It has been my contention, alongside Stuart and Strawson, that Locke's conception of consciousness does not answer or seek to answer a question of numerical re-identification, but rather one of characterization. Following Schechtman, Locke's concern may be even further narrowed. She argues that his account can be seen as setting the limits of the forensic unit, which in turn are necessary to define the limits of the moral self. Yet, regardless of how suggestive this interpretation may be, Locke's forensic unit does not provide an answer to the limits of the moral self. We know that characterization requires the extension of consciousness, but what are the sufficient conditions for the conclusion that the same character persists? I will argue in later chapters that the persistence of a moral self – the persistence of the character required for agents to be responsibility-apt – rests on the notion of answerability. For now and in the following chapter, I would like to explore the Lockean notion of consciousness in order to set the limits of the forensic unit.

Chapter 2: Reconceiving Locke's Theory and the Limits of Consciousness

Introduction

In the last chapter, I suggested that we follow Marya Schechtman and interpret Locke as delineating the conditions of the forensic unit, rather than on questions of identity proper. This chapter will start by better clarifying what the forensic unit could mean on Locke's terms. The 'could' here should be emphasized as I do not intend to provide a definitive analysis of Locke's account, but merely offer one suggestive interpretation. This definition will serve as the most inclusive definition of the self that will be narrowed in the coming chapters. I propose beginning this inquiry by asking what it means to be conscious of something on Lockean terms in order to define consciousness in general. By the end we should have the preliminary shape of the forensic unit as derived from Lockean consciousness that will be further refined to resemble a specifically moral self as the thesis progresses.

This chapter will proceed as follows: As the forensic unit is defined by the extension of consciousness, I will start by attempting to analyze what consciousness might mean. In §1 I will first question the traditional interpretation of persisting consciousness as "memory". In §2 I offer my alternative interpretation. I suggest that Lockean consciousness could be read as distinctly phenomenological by drawing on some arguments made by a number of Lockean theorists including Udo Thiel, Matthew Stuart and Galen Strawson. Consciousness may be usefully understood as a kind of self-presence implicit in and inflecting our experience of the world. This sort of self-presence

is not unlike the perspectival self found in Dan Zahavi's work.[58] Section 3 continues this line of argument and connects this extended interpretation of Locke to one of his critics, Edmund Husserl and the phenomenological experience of the self. [59] In §4 I will outline how the consciousness, framed as this phenomenological perspective, might provide a basis on which to begin to define the self as uniquely attributable once supplemented with considerations concerning the complexity of consciousness as given by Marya Schechtman. *The self*, defined as a unique and somewhat enduring perspective on the world, is then the basis of the forensic unit. Finally, § 5 will show how a collection of psychological aspects that I call the *motivational profile*, comes to constitute this self.

Overall, I argue that our forensic concerns track this perspectival sense of self because not only is this perspective uniquely attributable to the specific agent, but can also provide a fitting pragmatic ground to hold one responsibility-apt as well. Knowing something about the general outlook of offenders can give insight to how they will specifically act, thus providing an ideal base in which to make determinations concerning responsibility-aptness in the future.

1. Lockean Consciousness, Reflexivity and Reflection

In the previous chapter I suggested in passing that Locke might not be offering a memory theory. Yet I have not said much about why this might be so. In order to better frame what could replace memory as defining consciousness, I would first like to explore why we might reject a memory criterion. In the literature we see that equating

[58] Zahavi, Dan. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. (Cambridge, Mass: MIT Press, 2005).
[59] Husserl, Edmund. *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*. (Edmund Husserl Collected Works, V. 4. Dordrecht: Kluwer Academic, 1991).

consciousness with memory has recently been subject to debate with a number of theorists questioning the legitimacy of such an interpretation. Each argue that memory connections should not be considered the sole feature that underlies the extension of the Lockean person. For instance, Marya Schechtman notes that, if we think Locke a memory theorist, "[t]rying to understand [his work] this way leaves us … with the nagging question of why he never says that memory connections constitute personal identity."[60] Locke speaks of memory in other passages of his work, yet when it comes to identity, "he *always* talks about extension of consciousness and *never* about memory connections."[61]

However, it is not just Locke's neglect to specifically frame consciousness as memory that urges a reinterpretation of his work. Locke's reported embrace of a memory criterion is curious in light of an immediate, experiential quality he sees in consciousness and this quality need not be associated with memory alone. In fact it seems to be present in a variety of ways that does not seem to presuppose explicit recollection or reflection. In this section, I wish to reframe Lockean consciousness as something more fundamental than memory even if this interpretation stretches what was actually meant by Locke. The point nevertheless is to move away from a reflective interpretation of what it means to be conscious to an interpretation that emphasizes the *reflexivity* of conscious experience.

Traditionally, the kind of awareness that accompanies thought and "all the actions of that thing as its own" is taken to imply a capacity for reflection.[62] Many have understood Locke as equating being conscious of something as a higher order monitoring

[60] Schechtman, *Constitution*, 107.
[61] Ibid.
[62] Locke, *Essay*, II.xxvii.17.

process that takes mental operations as objects of thought. As Angela Coventry and Uriah Kriegel suggest:

> On this model of consciousness, then, consciousness is an extrinsic property of S_1: there is nothing in S_1 itself that makes it conscious. What makes it conscious is an altogether different state, S_2. So S_1 is conscious in virtue of its relational property of being appropriately represented by a separate state.[63]

Through reflection one can acquire ideas of the operations of the mind and much of what Locke says about consciousness seems to support this sort of position. Indeed as Udo Theil notes:

> Locke… states that reflection is 'the Perception of the Operations of our own Minds within us'… and a few sections later that consciousness is 'the perception of what passes in a Man's own mind'… The two statements are not identical, but they would seem to be sufficiently similar to suggest that consciousness and reflection are the same thing for Locke.[64]

As noted by Thiel, Locke seems to treat reflection as analogous to sensory perception as a kind of "inner sense."[65] Indeed, in Locke's own words, consciousness is "that notice which the mind takes of its own operations, and the manner of them, by reason whereof there come to be ideas of these operations in the understanding."[66] Under this interpretation, the notice we take of the operations of the mind obtains through an inward perception where such operations are made objects for observation. Theil continues, "If there is a difference between consciousness and reflection at all, it would seem to be a difference that relates to the object, not to the nature of the activity

---

[63] Coventry, Angela, and Uriah Kriegel. "Locke on Consciousness." *History of Philosophy Quarterly*. 25.3 (2008): 222.
[64] Thiel. *Early Modern*, 111 quoting Locke, *Essay*, II.i.4, II.i.19
[65] Locke, *Essay*, II.i.4.
[66] Ibid.

involved."[67] Consciousness, on this interpretation, requires a higher order state to render other mental activity conscious.

Memory might involve this sort of reflective, higher-order awareness as it extends and retrieves these past experienced objects for further examination. They are brought to the mind to be perceived once more. Locke writes of memory:

> In remembering, the mind is often active. In this secondary perception, as I may so call it, or viewing again the ideas that are lodged in the memory, the mind is oftentimes more than barely passive; the appearance of those dormant pictures depending sometimes on the will.[68]

Memory, like the reflective reading of consciousness, requires the mind to "turn as it were the eye of the soul upon" past impressions and "take notice of them" and "renew its acquaintance with them."[69] The recall of memory is but an inward means of reflection to examine what was once experienced.

However, the fact that Locke describes consciousness as "inseparable from thought" raises a number of questions as to whether consciousness can truly be considered the sort of higher-order process this interpretation suggests.[70] When paired with this idea that we are conscious of all thought, the reflective interpretation seems to involve a taxing double awareness, which in turn threatens the possibility of an infinite regress. When the agent thinks or feels a particular sensation, she must be capable of reflecting on and directing her intentional aim to her own mental states as a kind of higher order self-monitoring. Each act and each sensation would involve an awareness of

---

[67] Theil, *Early Modern*, 110.
[68] Ibid., II.x.7.
[69] Ibid.
[70] Reid, *The Works of Thomas Reid*, VI.iii.346.

both the object and a secondary awareness of the process itself, thus creating a problem of regress. As described by Shelly Weinberg:

> All thinking, for Locke, is conscious, and consciousness is thought to be some sort of perception of a perception. It follows that if consciousness bears a relation to ideas that produces more ideas, then any perception by consciousness results in a mental state that must itself be perceived by consciousness. This results in another mental state of which we must be conscious, and so on.[71]

Quite simply, the interpretation that Locke endorses a higher order theory cannot be squared with the notion that we are necessarily conscious of all mental states. Indeed, as noted by Coventry and Kriegel we cannot equate consciousness with higher order monitoring precisely because it is "untenable in conjunction with Locke's view that all mental states are conscious."[72] They argue that we should understand consciousness as something other than reflection. Otherwise, Locke's theory would be internally incoherent.

One could try to save reading consciousness as reflection by arguing that not all mental states are conscious, but this would not be true to Locke's theory. Locke famously argues that we cannot have knowledge in the mind without perceiving that we do.[73] All thought or experience is accompanied by some measure of awareness of the operations of the mind. It is not only impossible to "perceive without perceiving that he does perceive" when one is thinking, but also "when we see, hear, smell, taste, feel, meditate or will anything, we know that we do so. Thus it is always as to our present

---

[71] Weinberg, Shelly. "The Coherence of Consciousness in Locke's 'Essay.'" *History of Philosophy Quarterly*, vol. 25, no. 1(2008): 25.
[72] Coventry & Kriegel, "Locke on Consciousness", 227.
[73] Locke, *Essay*, II.x.2.

sensations and perceptions: and by this everyone is to himself that which he calls self." [74] For Locke, "consciousness...is inseparable from thinking" and that "thinking consists in being conscious that one thinks" denoting an immediate awareness contained within and not directed at acts of thought.[75] Sometimes we do take a reflective stance when reporting a pain or explaining a sensation, yet the majority of our bodily sensations already carry an immediate given-ness that allows them to be experienced as my own without reflection. It is not just that something has been experienced, but it is I who is having these experiences without an explicit representation of this fact.

In what follows, I will suggest that Locke did not necessarily have a higher-order process in mind when defining consciousness, but something more like the immediate and minimal self-presence not unlike what is seen in the work of Edmund Husserl and Dan Zahavi. Although undoubtedly controversial, I will try to show that psychological aspects may be considered part of conscious experience through an immediate affective awareness, without the need for introspection or higher order processes. I do not claim that this interpretation of Locke is historically accurate, but I would like to elaborate on these suggestions to see if we can arrive at a more useful interpretation of consciousness for my purposes of defining the forensic unit.

2. Immediate Awareness and Ipseity

Although he does not make it explicit, Locke seems to allow for an immediate affective connection to both bodily sensation and thought without a higher-order monitoring process. There is a sense of ownership or 'mineness', as such experience is

[74] Ibid., II.xxvii.9.
[75] Ibid.

felt to be an intrinsic property of action and thought. Unlike reflection this sort of reflexivity "is not a relation which may hold sometimes but not other times."[76] As noted by Schechtman there is a sense in which Locke extends consciousness, not on the basis of reflection, but reflexivity. She states that on Locke's account, "present actions are made part of a persons consciousness by affecting his well-being or causing pleasure or pain." [77] Like the experience of physical pleasures and pains, items are part of consciousness "by *feeling* its effects." [78] Locke writes of bodily sensations, a "little finger is as much a part of [the agent] as what is most so." [79] The agent is affected by the sensations garnered from the finger "and so attributes to [himself] and owns all the actions of that [finger] as its own." [80] He even comes to entertain the possibility that if the finger were to become detached, it too would have "its own peculiar consciousness" now separate from the subject, yet also notably without the higher order capacities traditionally attributed to consciousness.[81] But once detached the now fingerless body can no longer feel these affects and would not "at all be concerned for it, as a part of itself, or could own any of its actions, or have any of them imputed to him." [82] Thus, I would argue that we are concerned for our bodies, not through a cognitive acknowledgment that it is a part of us, but because what happens to the body affects us. There is an immediate self-presence in these sensations that could arguably accompany thought as well.

---

[76] Theil, *Early Modern*, 116.
[77] Schechtman, *Constitution*, 109.
[78] Ibid.
[79] Locke, *Essay*, II.xxvii.17.
[80] Ibid.
[81] Ibid., II.xxvii.18.
[82] Ibid.

There is a sense in which Locke's theory allows for reflexivity without reflection for mental items. Udo Theil argues that Lockean consciousness should be understood as something more fundamental than reflection. More specifically, conscious experience provides the material necessary for higher order reflection. Theil states, "Without consciousness, reflection would not have any objects upon which to reflect. Both sensation and reflection are conscious acts; but for Locke they are not necessarily accompanied by an act of reflection".[83] Consciousness, Thiel argues, is the source of our ideas. It is inherent in thought or experience just as the awareness of pain or hunger consists in the feeling itself. As Theil writes, "In the same way [as one experiences hunger], consciousness is not something that needs to be added to thinking externally; rather, it is an aspect of thinking itself".[84] In short, the idea is that "consciousness is not a mental act additional to the original perceiving of an idea, but rather something internal to it. This requires that we see the perception of an idea as a complex mental act that includes also being conscious that we are perceiving the idea." [85] So, rather than positing a higher order perception to perceive the lower states, we may think of consciousness rather as a "one-level account of consciousness" or "same-order perception."[86]

Consciousness is thus not grounded in a separate state additional to perceiving, but something internal to the very act of perceiving the object. This means that certain experiences and psychological effects can be conscious in two ways. They can be drawn up to consciousness by reflection, but also inform and continually update current

[83] Thiel, *Early Modern*, 115.
[84] Ibid., 116.
[85] Weinberg, "Coherence", 26.
[86] Conventry & Kriegel, "Locke on Consciousness", 226.

experience pre-reflectively. It is the latter interpretation that captures the immediacy inherent in Lockean consciousness. We can reflect on perceptions, but to be aware or conscious may also be understood as "a reflexive and proprietary constituent of ordinary perception."[87] The "inner perception" suggested by Locke may instead refer to the kind of perception Galen Strawson refers to as its "*field of from-the-inside giveness."*[88] We are given a "non-thetic", "pre-reflective" and "non egological" condition to qualify as part of consciousness.[89] As Strawson continues, the sense of ownership given in consciousness is "devoid of any expressed, spelled-out, distinct representation of the subject of experience or self or 'ego' considered specifically as such."[90] Indeed, there is an immediate non-observational access to myself involving a basic sense of self-presence or "auto-affection" or what Jean-Paul Sartre referred to as "ipseity." [91,92] There is a basic sort of self-presence implicit in our experience of the world that gives a different sort of explanation as to what it means to be self-conscious.

Of course, self-consciousness may take on many meanings from this sort of reflective turn to the feeling one gets in an awkward situation, but neither of these are the kind of self-consciousness I am attempting to highlight here. For instance, it may be the case that when those like Hume turn the mind's eye onto its own operations; they may

[87] Ibid., 26.
[88] Strawson, *Personal Identity*, 50.
[89] Ibid.*,* 42.
[90] Ibid.
[91] Parnas, Josef, and Louis A. Sass. "The Structure of Self Consciousness in Schizophrenia." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher (Oxford University Press, 2011): 428.
[92] Sartre, Jean-Paul. *Being and Nothingness: An Essay in Phenomenological Ontology*. Translated by Sarah Richmond. (Abingdon: Routledge, 2018): 126.

find no more than the inconceivable "rapidity" and "perpetual flux" of consciousness. [93] But self-consciousness that I am attempting to attribute to Locke does not require the kind of reflection Hume describes. Instead, I see it as similar to consciousness as characterized by Dan Zahavi. He argues that the self is "not a quality or datum of experience on a par with, say, the scent of crushed mint leaves or the taste of chocolate ... It refers to the first-personal presence of all my experiential content; it refers to the experiential *perspectivalness* of phenomenal consciousness."[94] Zahavi continues:

> The self currently under consideration—and let us simply call it the experiential self—is not a separately existing entity—it is not something that exists independently of, in separation from, or in opposition to the stream of consciousness—but neither is it simply reducible to a specific experience or (sub)set of experiences; nor is it, for that matter, a mere social construct that evolves through time. Rather, it is taken to be an integral part of our conscious life.[95]

As Zahavi argues, the (minimal or core) self "possesses [an] experiential reality and that it can be identified with the ubiquitous first-personal character of the experiential phenomena."[96] It is a sense of self that structures how we experience the world, not something that can be perceived and understood separately from our experiences. If we think the mind "a kind of theatre", as does Hume, "where several perceptions successively make their appearance; pass, re-pass, glide away and mingle in an infinite variety of postures and situations", the kind of self-consciousness that I am concerned with here means that at no time would we ever see this self glide past the

---

[93] Hume, David. *An Inquiry Concerning Human Understanding: With a Supplement, an Abstract of a Treatise of Human Nature*. (Indianapolis: Bobbs-Merrill Educational Pub, 1955): 1.4.6.
[94] Zahavi, *Subjectivity,* 22. Emphasis mine.
[95] Ibid.,18.
[96] Ibid.

theatre stage because the self is inherent in the act of perceiving that stage.[97] The sense of self on these accounts is thus a fundamental feature of consciousness. It is the non-inferential, non-reflective awareness of our own occurring thoughts, pain, perceptions and feelings. Each experience appears in a "first-person mode of presentation that immediately reveals them as one's own."[98] The first-personal mode here is important, because it not only distinguishes what is my own through an immediate self-presence, but it may also distinguish which aspects of my past and memories continue to inform who I am at the moment. Pre-reflective processes can update current experience and create a distinct phenomenological unit.

It is this self as primarily understood in this immediate, perspectival sense that will constitute the basis for my first and most inclusive definition of the self as the forensic unit later in this chapter. First however, I would like to explore why focusing on this perspectival self might be fitting to constitute the forensic unit.

3. Pre-reflective Apprehension Over Time

To continue the interpretation of consciousness at a time and see how it extends over time, we might want to say that consciousness extends only if the current perspective extends. The way one views and experiences the world remains the same or at least sufficiently similar. Thus, in what follows, I will suggest that consciousness extends when persons can be characterized as having a similar enough perspective on the world. Past values, experiences and other influences may come to frame how the world is experienced and shapes who the agent is in the present in the same way it did in the

[97] Hume, *Inquiry*, 1.4.6.
[98] Parnas and Sass, "Self Consciousness in Schizophrenia.", 430.

past. This is important because knowing something about the agent's perspective on the world provides a measure of predictability fitting for determining the extent of responsibility attributions.

The kind of extended perspectival self I have in mind is similar to the sense of self in current experience described in Edmund Husserl's *The Phenomenology of the Consciousness of Internal Time*. Husserl depicts an immediate sense of ownership through what he calls the *retentional* and *protentional* track of consciousness.[99] There is a structure of conscious experience given by two factors at the level of a pre-reflective awareness that allows for the kind of auto-affection, as the self affecting the self, that results in an experience as being specifically and immediately my own. To use Husserl's famous example, when listening to a melody, we do not simply hear the sounding of each note in isolation, but experience it as a unity with the impact of the highs, lows, tempo and the various chords as partially derived from what was heard previously and what is expected to follow.[100] If each moment occurred separately from the others, the unity of the melody would be lost. For Husserl, our experience of a melody as unified displays the flow structure of consciousness. The experience of a musical note occupies consciousness for only a moment, but once the moment passes the note does not disappear from consciousness altogether. It survives as a form of retention. The retentional track keeps the intentional sense of the note available after it has been sounded. Complementing this retentional track is a sense of anticipation created by the

[99] See *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*. (Edmund Husserl Collected Works, V. 4. Dordrecht: Kluwer Academic, 1991): §8 and §24.
[100] Husserl*, Phenomenology,* 2. §21.

protentional track. When listening to a melody, it allows for moments of surprise when the music turns from the expected or a sense of satisfaction when the anticipation is confirmed. Indeed, when speaking I also have an anticipatory sense of where the sentence is going even if it is not completely definite. Working anticipation keeps thoughts on track and provides a sense of where my spoken thought is headed. Protentioning provides conscious experience with an intentional sense that something will happen next. Thus, conscious experience is not given by the perception of isolated moments. Rather our experience of the world at a time is connected to both the past and present. At each moment, there is a pre-reflective sense of what I was just thinking (retention) or perceiving coupled with a notion that such experiences will continue (protention).[101]

Shaun Gallagher argues that retention and protention together provide a sense of ownership and agency within immediate cognitive thought. The continuity given by the retentional track allows a thought to be experienced as one's own, as belonging to the same stream of consciousness. When we utter a sentence, retention functions much like working memory as it retains the sense of the earlier words I have just spoken. This function is part of the longitudinal aspect that delivers the sense that the spoken words are indeed mine. Ownership and agency generates a basic sense of *ipseity* that shapes experience. As Gallagher explains, "The words do not become part of a free-floating anonymity, nor do they seem to belong to someone else; they remain, for me, part of the

[101] The phenomenological perspective gestured to here is similar to the kind of phenomenological self envisioned by Marya Schechtman who likens one's experience of the world to particular flavours of soup (See Schechtman, *Constitution*, 143). This account will be explored further in chapter five.

sentence that *I* am in the process of uttering, because they remain part of my stream of consciousness."[102] The retentional-protentional structure of consciousness allows for self-identity in the "changing flow of consciousness."[103] The disparate moments of experience are perceived as a unity even if the unity is primarily subjective in character. The subject persists and is connected to each fleeting moment, which in turn gives rise to the phenomena of a continuing sense of self embedded within immediate experience. Thus, experiences are inherently and tacitly connected to my past, present and anticipated future in a way that generates the kind of immediacy that can be drawn (even if tangentially) from Locke.[104]

This discussion of the continuing sense of self and the kind of perspective it implies will be important in providing the foundation for my first attempt at defining the self as the forensic unit. The self is not necessarily the collection of memories or experiences that are retained over time, but the phenomenological perspective the retention of those memories and experiences constitute. The retentional track is deeply implicated in how we interact with the world by structuring how the world is experienced, which includes everyday occurrences such as every time I see a recognizable face or even when I step onto the floor expecting it to be solid. Both my body and thought process shape themselves, according to what is expected, informed by

---

[102] Gallagher, Shaun. *How the Body Shapes the Mind*. (Oxford: Oxford University Press, 2005): 192.
[103] Ibid., 192.
[104] This Husserlian interpretation of time-consciousness is sometimes viewed as a more modern version of Saint Augustine's concept of *distentio animi*. The idea is that the past and the future extend the mind through memory on the one hand and expectation on the other. For a brief overview see Malan G.J. "Ricœur on Time: From Husserl to Augustine." *Hts Teologiese Studies / Theological Studies* 73, no. 1 (2017).

a past of what was once experienced. The world and even the kinaesthetic possibilities my body presents remain familiar due to this influence. Due to the protentional track we then see the processes renewed, as the current experience (structured by the retentional track) is brought to shape future possibilities and expectations.

Of course, this structuring of experience does not require explicit recollection. In normal phenomenology, the world in which I experience is familiar to me, but this is not because I have stored memories available for access to draw on and apply to the world. Instead, it is because my past continues to inform the present at a pre-reflective and tacit level. Without the need to recall, the past informs and unifies my experience of the world. What is important in understanding the sense of "mineness" given by the protentional and retentional track is the perspective that is produced by being affected in this manner. The forensic unit may then first be understood as the continuing phenomenological perspective and defined by *how one approaches the world*.[105]

4. A Unique Perspective

The notion of the perspectival or phenomenological self that is endorsed by these theorists provides an intriguing way to interpret the sense of self-presence found in Locke's work. Yet, if my aim is to provide an interpretation of the forensic unit, it is not yet clear why this sort of self-presence, which is not clearly moral in nature, would provide the proper basis for such a unit. For instance, there is also undoubtedly a specific way it is like to be a dog or infant. Yet, the bare fact that there is such a perspective

[105] What constitutes the protentional and retentional track at one moment might not be the same over time. Given that I am concerned with this perspective over time, the kind of flux seen here is a problem, but one that I will not fully address until chapter six. Briefly looking ahead, the constitution of the motivational profile need not be the exact same as long as it is similar in the ways that matter for the attribution in question.

seems to be too minimal to support responsibility attributions. That is, there may be something that it is like to be a dog, but simply because a dog may experience the world with a particular perspective, it is hard to see how this gets us to the forensic unit.

There may indeed be a minimal sense of self, a perspective on the world that all conscious creatures possess, but there is something uniquely complex about that perspective when possessed by creatures with a more multifaceted psychology. As noted by Marya Schechtman (reflecting on Jeff McMahan's work), "The difference between that creature continuing and its being replaced by another with a similarly pleasant life becomes very thin."[106] Although I would not deny this sort of self-experience in animals and the very young, there still may not be much in those self-experiences that makes them deeply personal or moral.[107] How a newborn would experience the world is not too different from the one born down the hall. Aside from common desires for food, comfort and closeness, there is little complexity in the way of life experience, attitudes or beliefs to make the perspective of a newborn deeply personal. It is complexity that makes the experiential or perspectival self deeply personal and, hence, a fitting basis for determining persistence over time for the purposes of determining the forensic unit.

Like a melody inflected by the note that preceded it, one's previous experience provides a sense of individuality to one's perspective. Consider Schechtman's example of a childhood spent in economic insecurity. Once in adulthood, the individual may still

[106] Schechtman Marya. "The Size of Self", in *Narrative, Philosophy and Life*, edited by Speight, Allen (Dordrecht Netherlands: Springer, 2015): 42.
[107] Consciousness interpreted in this way could easily apply to other complex creatures like humans. The upshot of this means that creatures like dolphins, cetaceans or other kinds of intellectually advanced animals could very well be responsibility-apt and this is a consequence I would fully accept.

experience a sense of frugality long after the adult has come into wealth. The way she sees the world is deeply inflected by these earlier episodes. She may scoff at the price of a luxury apartment, feel a sense of joy when finding an item on sale or experience insecurity when hobnobbing around other wealthy folk. It is not that the wealthy adult now remembers the times when money was unavailable with every purchase made, or explicitly juxtaposes those experiences when at an exclusive party. Instead, those experiences from the past are able to affect one's present experience in a more global manner. The past conditions the present without mediation of any specific memory or perhaps even conscious awareness. Instead, the past inflects one's experience like notes previously heard. As Schechtman describes it, one's past "provides a backdrop that affects the quality of almost all of one's day-to-day experience."[108] The past informs experience, unselfconsciously and from the inside even if we are not specifically aware of this ability.

So, what do life episodes have to do with this core, minimal self that is suggested by Husserl and Zahavi. The answer is, as I have suggested, because they add complexity to our experiences. The self can be made more complex given the sorts of experiences one undergoes and the complexity of the psychology of those possessing that minimal self. In his sense, the minimal self will be " intertwined with, shaped, and contextualized by memories, expressive behavior, and social interaction, by passively acquired habits,

[108] Schechtman, *Constitution*, 111.

inclinations, associations, etc."[109] This intertwining I suggest contributes to the uniqueness of one's particular perspective as well.

The collection of experiences we accrue in a life makes the experience of that life distinctive. Persons are not just affected by financial woes, but the myriad of past experiences that renders one's current perspective deeply complex and personal. Consider, as Schechtman does, the melody analogy once again. She asks us to consider Mozart's *Ah Je Vous Dirai Maman*. This work:

> …starts with the simple folk theme known, among other things, as *Twinkle, Twinkle Little Star*, and goes on to present twelve variations, some of which are quite complicated. Although each variation is a *version* of the simple theme, none is generated by the simple addition of other notes, and there is no note-for-note reproduction of the original in the sophisticated variations. The musical sophistication of the variations is not achieved by placing something else on top of the original melody, but rather by transforming and complexifying it.[110]

Using Schechtman's analogy to music, we can see how the phenomenological, perspectival self may be made more or less complex. The addition of sophisticated psychological capacities and experiences is not "like plunking out the simple theme with the right hand and then adding some sophisticated left hand pyrotechnics on top of it"[111]. Instead, as Schechtman argues, "it is more like replacing the simple plunking with one of the variations."[112] Likewise, how I approach the world, I would argue, is not only made more complex by the various experiences I undergo, but it makes that subjective experience of what it is like to be me, distinctively my own. It is a variation on a theme,

---

[109] Zahavi, Dan, "The Unity of Consciousness." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher. (Oxford: Oxford University Press, 2011): 332-33.
[110] Schechtman,"The Size of Self", 43.
[111] Ibid.
[112] Ibid.

but a variation that may be considered truly my own. Perhaps, like 'complexifying' a melody we might call my particular version *Ah Je Vous Dirai Maman, the Nicole Remix*.

To be clear, the variation that constitutes the Nicole remix need not have the same stylistic elements as a kind of personal signature written on each experience. For instance, I am not saying that there is a distinct 'Nicole' way of listening to music or reading the paper. Instead, one's perspective is uniquely personal due to the way each element varies to be collectively unique.[113] So there is not some way of seeing the world that is distinctly my own, rather it is a unique collection of different variations on this simple theme and not a particular style I bring to each rendition.

So while dogs and infants possess a self, that self may be more minimal insofar it is not inflected by various lived experiences, as we would see in an adult. Perhaps this is why Locke requires of persons to be a "thinking intelligent being that has reason and reflection, and can consider itself as itself, the same thinking thing."[114] As necessary conditions for personhood, it is not as if these capacities make up consciousness, but could potentially imply the kind of complexity we are after. The person as a forensic unit is one that can reason, reflect and think of itself as a self, "which it does only *by that* consciousness which is inseparable from thinking." Like a simple melody or baseline, it is "by that consciousness" that these capacities are preformed and thereby able to

[113] Theoretically, this perspective could be duplicated as theorists like Parfit have forcefully shown in teletransportation scenarios. I do not, however, think it deeply puzzling that both duplicates could be characterized as the same. After all, it would not be unreasonable to treat both duplicates as responsibility-apt as both would approach the world in the same way. Both arguably have the same need for rehabilitative methods given that criminal outlook has been duplicated. See Parfit, *Reasons and Persons*, Chapter 10.

[114] Locke, *Essay,* II.xxvii.9.

'complexify' conscious experience. Creatures with the necessary capacities Locke describes, would thus meet a minimal sort of threshold of complexity of consciousness (not possessed by animals and babies) to be the appropriate target of our forensic concerns. It is not that all consciousness must possess such capacities, but all that can be included as part of the forensic unit that requires a sufficiently complex consciousness. One's past experiences, beliefs, attitudes and even reflective capacities can add a degree of complexity to one's perspective that, when taken together, is not only distinctive of the particular individual, but is also is appropriately complex enough to support forensic concerns.

There are even some benefits to framing consciousness in the perspectival sense suggested here. In particular, we are given a variety of implicit ways a person can remain connected to the past by focusing on how the affect from the past shapes the present. A person may not explicitly remember certain episodes in the past, but can still remain connected to it in this deep and personal way as long as those past psychological features still affect the present perspective.[115] Certain memories, judgements, desires and the like affect and shape a discrete phenomenological experience of the world (whether implicitly or explicitly) and it is this phenomenological unit we are tracking when we say that an individual remains the same self as before. Consciousness comes to constitute the person in Lockean terms, who is then the appropriate unit for responsibility attributions. Thus, to be same forensic unit over time will be determined by the maintenance and

[115] At this point, there may be a number a questions concerning what sorts of influences should be considered connected to the person and those that are not. The focus of chapter three will concern making these finer distinctions to determine what may be considered 'truly' part of the person and what influences bar this kind of characterization.

persistence of the individual's unique and sufficiently complex phenomenological perspective.

5. The Motivational Profile

To avoid confusion and any conflation with the traditional way Locke has been interpreted, I will rename the forensic unit highlighted in this chapter, *the self*. I would also like to simplify things further and use a more general term to identify all these implicit and explicit psychological features that come to constitute the self: *the motivational profile*. The self is one's phenomenological perspective that corresponds to how one approaches the world, while the motivational profile constitutes this self.

The Self (Constituted by the Motivational Profile): The continuing first-personal perspective constituted by the collection of psychological features of the motivational profile.

Motivational profile (Constitutes the self): The collection of psychological features that come to inflect the agent's current perspective or outlook.

We are concerned with the individual's outlook specifically at a time and to ask questions of moral responsibility and ask whether past action or experience continues to frame or influence that outlook. It is here then that we receive a fuller explanation for the "gappiness" seen in Lockean persons.[116] One's outlook is not determined by spatio-temporal continuity, but can reach back to elements of the distant past just like it does for that which is currently experienced. When we ask if it is indeed the same self, we are asking whether the individual has a similar (though often not identical) motivational profile, which is to say that the past and present self are sufficiently similar through

[116] Strawson, *Personal Identity*, 8.

comparison on their phenomenological perspective or outlook.[117] The motivational profile constitutes the self as any and all influences that come of inflect and influence one's current perspective. When this motivational profile is generally retained, we might say that the individual's outlook is similar enough to the past individual as comparable affect and influences press upon the actual subject in their actions and understanding of the world.

Past desire, intentions, memories and even implicit features on this reading have the ability to continue to inform one's present experiences, define one's present outlook and colour one's experience of the world. For instance, Reid's general, who can no longer remember stealing apples from the orchard, would not be considered as the same self as the young boy if aspects of the boy's motivational profile no longer constitutes and informs who the general is now. The relation that determines this outlook is not transitive nor need it be to underpin notions of responsibility. If aspects of one's past have no influence on who one is now, we might just say that the self simply has not extended or has hit an expiration.

Conclusion

The concept of the self that is outlined in this chapter leaves many questions central to personal identity theory unanswered. Arguably, however, answering questions of re-identification was not Locke's purpose, nor is it mine. Identifying and tracking the motivational profile is a distinct pursuit from what is traditionally engaged in by Lockean theorists. Using Schechtman's terms, the notion of the forensic unit, I have renamed as

[117] The fact that this motivational profile constantly changes due to one's experiences is a problem for my account. In chapter six I will try to address this concern by more clearly spelling out what I mean by "sufficiently similar" here.

the self, may in turn be used to sort out some of these questions by delineating a persisting phenomenological subject as a viable basis for practical concerns such as responsibility. I would like to close this chapter by briefly considering why this perspectival self is important for framing the forensic unit (although this theme will be revisited as the thesis progresses).

The self considered as a perspective on the world I see as being deeply personal in an important sense. It represents, not just what the person wishes or intends to do, but characterizes the very way they see the world and operate within it. It characterizes them as individuals within the world in a way that also matters for responsibility attributions and forensic concerns. Knowing how an individual sees the world in some sense can also offer a measure of predictability in that it gives us a sense of what she will do or will not do given the circumstances. Knowing one has the same perspective, gives us ground to say whether a questionable act will be committed again, (if this is indeed what we care about when determining responsibility-aptness). So overall, the perspectival or phenomenological self not only offers something deeply personal, but it provides a fitting pragmatic ground to guide our forensic concerns in determining responsibility as a means of predictability.

Of course, as it stands now the concept of the self I have provided is subject to a number of criticisms with the most problematic being that the definition is both over-inclusive and vague. Importantly, the perspectival self might characterize what matters for the forensic unit, it does not yet answer the characterization question as determining

whether an action was "*truly* hers" in terms of the moral self.[118] In chapter three and four, I will refine the boundaries of the self as a forensic unit in order to come to an answer of what it means to be the same moral self over time in chapters five and six. For now and, hopefully through this chapter, the general structure is clear. The self (defined by one's phenomenological experience) is constituted by the motivational profile (that which inflects and influences this experience) that in turn will be generally tracked to see if we are dealing will the same self over time.

[118] Schechtman, *The Constitution of Selves*, 76.

What makes one responsibility-apt?

Chapter 3: Incorporation of the Rational Relations View

Introduction

So far I have only begun to mould the preliminary shape of what may constitute the self as a Lockean forensic unit. I argued that the self might be characterized as the individual's ongoing phenomenological perspective, constituted and made unique through the constituents of a motivational profile.[119] Yet, by delineating a phenomenological perspective we have something that might form a sufficient basis for the forensic unit, but have not yet defined the moral self. As it stands, the definition of a self and the motivational profile that constitutes it is over-inclusive and missing the finer features that mark certain psychological aspects as specifically moral.

This chapter explores several ways to chisel away at the multitude of psychological aspects that shape one's perspective in order to spell out the conditions under which individual's attitudes and actions can be deeply attributed and, hence, be responsibility-apt. One way of understanding 'deep attributability' in this sense is to consider only that which is the product of one's will, with the will being understood as a deliberative capacity. In §1 and §2 I briefly analyze two volitional accounts that focus on the will in this manner: identification and control. I argue that each (for different reasons) is too under-inclusive. In §3, I explore an alternative account due to George Sher and argue that

[119] In what follows, I will be taking a rather naïve and limited view concerning psychological motivations. I will be primarily concerned with beliefs, attitudes, desires and emotions as the basic psychologically motivating attitudes. I am sure there is room for debate on whether actions can be fully traced to such psychological entities, but I will leave that aside for another discussion. Also, I will put aside some more complex motivations including moods and intentions for the sake of clarity as others of done. See Arpaly, Nomy, and Timothy Schroeder. "Praise, Blame and the Whole self." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 93, no. 2 (1999).

it is over-inclusive. In §4, I explore Smith's rational relations view and her notion of 'answerability' as a solution to identifying the fine line between over and under-inclusivity. This approach shifts the analysis from the deliberative *exercise* of the will to the *quality* of that will. In §5, I will review some criticisms of her account due to David Shoemaker and argue that he may be half-right. As a result, I argue for an account that will sit between Smith and Shoemaker by offering different interpretations of both 'answerability' and 'judgment sensitivity.' Using these interpretations, in §6 I will propose a test to delineate psychological aspects that bear on whether the agent is responsibility-apt and in §7 define the parameters of what is constitutive of the moral self or what I will call the "evaluative profile."

Overall, while this chapter is mainly expository, the key concepts introduced here will have important implications for the discussion of responsibility-aptness over time in the following chapters. In particular, the notion of answerability introduced in this chapter will serve as a necessary condition to determine responsibility-aptness over time. Crimes may be deeply attributed and continue to be deeply attributed to the criminal as long as she remains answerable in the ways elaborated in this chapter.

1. Making the Delineation

Given the range of influences that inflect one's experience of the world, it is clear that although these may help to shape one's perspective, not all influences are specifically moral in nature. There may be aspects of the motivational profile that are "boring", to use Galen Strawson's terms, as they are uninteresting for determinations of

responsibility.[120] These aspects may include colour or food preferences that influence the agent in many ways, but generally do not contribute to anything of moral significance.[121] Yet there are also attitudes or perspectives of the agent that are not morally boring in this sense, but are of questionable attributability.[122] For instance, if an individual were to viciously insult a friend due to suffering from Tourette's syndrome, attributing responsibility for the scathing remark due to a tic would seem to not just be inapt, but wrong. However, the motivational profile, as I have framed it, does not exclude these kinds of pathological influences. In what follows, I would like to address this issue by exploring ways to narrow the range of influences that may be used to determine responsibility-aptness.

1. First Suggestion: Frankfurt and Identification

One means of discriminating between these influences is to use the "internal-external distinction" introduced by Harry Frankfurt.[123] On his terms we may distinguish 'external' influences like nervous tics and other like phenomena that "assail the agent from without" from aspects of the self that are 'internal' and otherwise representative of the *real self*.[124] As kind a of general responsibility theory, *real self views* aim to pick out a class of influences that are more representative of the self than others. Frankfurt may

[120] Strawson, *Personal Identity,* 75.
[121] In chapter six I will consider how to exclude 'boring' influences that may come to shape who you are but are of no significant moral import. Whether an aspect is boring is relative to the particular line of inquiry. It is possible that colour or food preferences could be relevant if they somehow contributed to a morally culpable act or attitude.
[122] Ibid., 75.
[123] Velleman , David. "Identification and Identity." In *Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton. (Cambridge, Mass.: MIT Press, 2002): 92.
[124] Frankfurt, Harry G. *The Importance of What We Care About: Philosophical Essays*. (Cambridge: Cambridge University Press, 1988).

be classified as such due to the way his tests of 'endorsement' and 'identification' delineate what counts as representative of the self into a smaller subset of psychological activity. His work may be considered to be a volitional account as well because some volitional activity (such as endorsement) is required to delineate the real self. Volitional accounts are "a cluster of distinct views which share a common assumption: that choice, decision, or susceptibility to voluntary control is a necessary condition of responsibility (for attitudes as well as actions)."[125]

For Frankfurt, agents are morally responsible for attitudes and actions if they identify with or endorse what he calls "first-order desires" that bear on those actions and attitudes.[126] Take the example of quitting smoking. Anyone who has quit smoking knows how difficult it is to achieve this end. Some may repeatedly quit, but often begrudgingly come back to the habit. Deep in the throws of one's craving the former smoker might feel an undeniable and insatiable pull to smoke, which is felt as *external* insofar it does not reflect the kind of self the agent wants to have. Frankfurt argues however that 'second-order volitions' represent the real self. These are second-order desires for a particular first-order desire to be effective, or "move a person all the way to action."[127] The unwilling smoker has a second-order volition not to smoke and therefore experiences the craving as external, an assault on her will. The desire to smoke is not deeply attributable to the unwilling smoker because it does not represent her real self. On the other hand, if the smoker identifies with the first-order craving by way of a second-order desire to be 'cool' like the classic film stars of the past, this willing smoker

[125] Smith. "Activity and Passivity", 238.
[126] Frankfurt. *Importance*, 19.
[127] Ibid., 4.

would not feel resistance as she is acting on the desire she endorses. For Frankfurt, the willing smoker has "the will [she] wants", and hence the act of smoking is deeply attributable to her [128]

We may ask why second-order volitions represent the real self more than first-order desires? Frankfurt clarifies his account by noting "the mere fact that one desire occupies a higher level than another in the hierarchy system seems plainly insufficient to endow it with greater authority or with any constitutive legitimacy."[129] Arguing that the higher-order desire is more reflective of the true self than the lower due only to their hierarchal ranking threatens a regress. Frankfurt responds to the problem using an analogy. He asks us to imagine a math student devising multiple formulas to check an answer because "what leads people to form desires of higher orders is similar to what leads them to go over arithmetic."[130] Once we suppose the student is fully confident that he would obtain the same answer if he were to continue, he could "without arbitrariness" cease further inquiry.[131] The future is "transparent to him, and his decision that a certain answer is correct resounds endlessly in just this sense: it enables him to anticipate the outcomes of an indefinite number of possible further calculations."[132] He commits to this formula as a necessary means to end his inquiry. Similarly, we make these sorts of commitments with respect to our second-order desires. We are committed to them unless there is some reason for doubt. Like the math student committing to a particular formula, what we care about determines how our decisions are made thereafter and is a kind of

[128] Ibid.,20.
[129] Ibid., 166.
[130] Ibid.
[131] Ibid.,168
[132] Ibid.

self-legislation. Those decisions are then our own in a fundamental way and embody who we are as agents.[133]

1(a). Problems with Identification as a Basis for Responsibility

On Frankfurt's account, a charge of over-inclusivity can be avoided through the device of 'identification' with lower-order desires. For example, one is not responsible for a desire to steal, even if that desire is part of the motivational profile, unless that desire is endorsed by second-order volitions. However, if we use Frankfurt to solve one problem, we would unfortunately land on another. The theory may be unable to explain examples where individuals are responsibility-apt for aspects of their characters with which they do not identify. In other words, Frankfurt's position is under-inclusive.

David Velleman asks us to consider Freud's "Rat Man" who was psychologically divided between love and hate for his father.[134] Because of this inner conflict, the Rat

[133] In a later work, Frankfurt complicates this picture with the introduction of what he terms, "volitional necessities." Volitional necessities are volitional incapacities where a person can do no other but A or are unable to refrain from doing *A* due to the kinds of cares and commitments the agent holds. To do otherwise would be "unthinkable."(Frankfurt, *Necessity, Volition, and Love*, 142.) Unity of the self may be threatened in time of conflict or ambivalence. Frankfurt provides us with the image of Agamemnon at Aulis to demonstrate this. Asked to sacrifice his daughter Iphigenia to Artemis in order to gain the necessary winds to set sail**,** Agamemnon finds himself torn within an "inescapable conflict between two equally defining elements of his own nature": Either sacrifice his beloved daughter or forgo the glory of the war that awaits him. (Ibid.,139) Regardless of what he chooses in this moment (even though he chooses the former in the end), Agamemnon damages himself to the extent he can no longer be considered the same self. The Agamemnon that began the journey to Troy was fundamentally different than the one later murdered by Clytemnestra. We know this because, if the agent is able to act in a way contrary to his volitional necessities and sacrifice his daughter, this shows the action to have become thinkable, which tells us something fundamental about his deepest commitments. See Frankfurt, Harry. *Necessity, Volition, and Love*. Cambridge, U.K: Cambridge University Press, 1999.

[134] Velleman, James D. "Identification and Identity" In *Self to Self: Selected Essays*. (Cambridge: Cambridge University Press, 2006): 342.

Man's agency was often undermined by "repeatedly doing and undoing an action, or thinking and contradicting a thought."[135] Rather than acknowledging the conflicted state of love and hatred, he would only acknowledge the love and deny the hateful thoughts, which he experienced as psychologically external to himself. But does he act in a manner that makes us think the hate for his father is not deeply attributable to his character? Is he really overtaken by hate or could it just be that he does not love his father in the way he thought?

The Rat Man does not endorse the aspects of his character/motivational profile that express hate for his father, but we can imagine many cases in which most persons would be responsible for acting on such lower-order desires. If a lack of identification excuses one from moral responsibility, it would do so too widely. Rather than saying one aspect of the Rat Man is more attributable than the other, I would argue that we should simply allow for the possibility of deep internal conflict. Both his love and resentment can attract attributions of moral responsibility because both can be seen as part of the self and play a role in constituting the Rat Man's experience of the world. Because of such limitations, we need to look elsewhere and away from Frankfurt's notion of identification to refine the motivational profile.

The case of the Rat man at least moves us closer to a view concerning the quality of one's will due to the fact that persons may be responsibility-apt for more than is the product of an exercise of the will, but what may overtake it and cause them to act in ways they would rather not.

[135] Ibid. 343.

2. Second Suggestion: Choice and Control

We saw that a former smoker may be overtaken by a desire to smoke, but for some theorists, the fact that she is overcome or does not identify with such a desire does not readily excuse. Prior-choice accounts argue that the causal history of the act or attitude can reveal a choice somewhere in the lineage of the action, which is the basis of the attribution of responsibility in the present. So even if the former smoker does not identify with her choice to smoke and gives into an insatiable craving, these actions may be responsibility–apt given her initial choice to smoke in the first place.

Choice may offer a necessary condition of attributability, but as some theorists argue, it is not sufficient. After all, it is possible that the smoker did not know she would become so dependent as to make quitting tremendously arduous. If it was reasonable for her to have not foreseen the potential outcomes of her actions, then according to some prior-choice theories, she may be excused even if the action could be appropriately traced to a prior-choice. Take for example Manuel Vargas' epistemic condition for responsibility. He states, "For an agent to be responsible for some outcome (whether an action or consequence) the outcome must be reasonably foreseeable for that agent at some suitable prior time."[136] On this view, to be responsible requires that not only is the action causally traceable to the agent's choice, but the agent also needs to act in light of potential knowledge of the wrong-making features of the present action.

[136] Vargas, Manuel. "The Trouble with Tracing." *Midwest Studies in Philosophy*. 29.1 (2005): 274.

For our purposes then, if we take choice as a means of delineating what is external, then we would only be responsible for what is either under my control or sensitive to the sorts of choices I made somewhere down the line if it was possible that I could foresee the consequences.

2(a). Problems with Prior Choice and Control

Like the criterion of identification, the test of prior-choice seems to be under-inclusive. Consider George Sher's story of the "Hot Dog" for example. In this scenario:

> Alessandra, a soccer mom, has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family's border collie, Bathsheba, who rides in the back of the van. Although it is very hot, the pick-up has never taken long, so Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, Alessandra is greeted by a tangled tale of misbehaviour, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba unconscious from heat prostration.[137]

According to Sher, we correctly attribute blame to Alessandra. Yet, Sheba's condition was the result of a lapse in judgement and not the result of any voluntary or intentional choice. Alessandra did not choose to forget. Even if we appeal to the epistemic condition, it is not clear that Sheba's condition would have been reasonably foreseeable to Alessandra because the arbitrary circumstances that distracted her were not foreseeable. The difficulty for volitional accounts like this one is that the basis of responsibility "appears to lie not in the agent's conscious will, but something that overtakes it."[138] Like the Rat Man, it seems that she is responsible due to an aspect of her moral character or some 'quality of her will' that allowed her to forget the dog. One

[137] Sher, George. *Who Knew?: Responsibility Without Awareness*. (Oxford: Oxford University Press, 2009): 25.
[138] Ibid., 26.

might claim that Alessandra does not care enough about her dog and it is this character disposition that grounds her responsibility-aptness despite the actions not clearly being traceable to some prior choice.

There are two ways the prior-choice views might be saved from objections like Sher's. First, we might argue outcomes of lapses such as forgetfulness are not suitably connected to choice to be a proper counterexample. The prior-choice view suggests that it is only if such aspects of character were the outcomes of some past voluntary decision or choice that agents could be responsible for them. Thus, leaving Sheba may be the outcome of a choice, but it is hard to connect this choice to the act of forgetting.

However, just as was the case for the condition of identification, once a direct connection to choice is applied to conditions of responsibility-aptness, we see that such a stringent condition would excuse many other actions that seem to attract attributions of responsibility and be under-inclusive as a result. In particular, I would point to spontaneous emotional reactions, outbursts, or even spur of the moment actions that may speak to one's character and are intuitively responsibility-apt in ways forgetting may not. Consider Peter Railton's example of Christine speeding down a country road.[139] Driving at such speeds requires her full attention as she scans for potholes and other cars. In turning a corner, she spots an elderly driver. Though she could easily speed past him, she instantly slows down enough to wave and make eye contact. She acts without a moment's hesitation, and notably without reviewing the reasons to act as she did. Christine slowed down because she noticed the anxious look on the elderly driver's face

[139] Railton, Peter, "Practical Competence and Fluent Agency." In *Reasons for Action*, edited by Sobel, David and Steven Wall. (Cambridge, UK: Cambridge University Press, 2009): 104-105.

that triggered both empathy and a negative affect at the prospect of startling the other driver. According to Railton, even the most basic deliberative decisions are partly constituted by unconscious and intuitive impulses and affects, which motivate and favour some reasons over others.[140] Moreover, Christine seems praiseworthy for her actions despite a lack of a voluntary or rational choice to look out for the elderly driver. This suggests that the prior choice view is too narrow: we are responsibility-apt for many aspects of the self (of our moral characters) that cannot be traced to prior voluntary decisions including spontaneous and emotional reactions. Indeed, borrowing from Thomas Nagel, we could argue for some sort of constitutive moral luck as luck in "the kind of person you are, where this is not just a question of what you deliberately do, but of your inclinations, capacities and temperament."[141] If we insist on a direct causal connection, we would not only exclude Alessandra's forgetfulness, but Christine's spontaneous kindness as well.

[140] David Shoemaker argues that these cases of emotional and spontaneous reactions may belong to a separate category of responsibility he calls "responsibility without answerability" as a type of aretaic appraisal. (Shoemaker, *Attributability, Answerability and Accountability,* 609*)* Here, it could be argued that such an account could provide the middle way between choice and the seeming lack of responsibility exhibited in these cases. Perhaps the actions of Christine and Alessandra are not subject to claims of full responsibility, but perhaps a breed of aretaic appraisal. This is a tempting means of solving the issue, yet I do not want to throw my weight behind this account due to further reasons of cogency. The reasons for introducing this third way have been recently criticized by Angela Smith. I find these criticisms convincing, although I will not rehearse them here. Instead, a couple of specific issues I take with the account will be discussed later in this chapter. So while Shoemaker's claims may provide an alternative approach, there are further reasons we might want to reject it, even if it solves the issue in this instance. See Shoemaker, David. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics*. 121.3 (2011).
[141] Nagel, Thomas. *Mortal Questions*. (London: Canto, 1991).

There may, nevertheless be a second response that could save the prior-choice view from the charge of over-inclusiveness by limiting the causal story through foreseeability. John Martin Fischer and Neal A. Tognazzini argue that responsibility-aptness depends on how broadly we understand the epistemic condition concerning foreseen outcomes. They give the example of Jeff, a current jerk who once adopted a "jerk" persona in order to attract the opposite sex.[142] They argue that he is responsible for being a jerk because "he freely decided to become a jerk at some point in the past, and it is reasonable to expect Jeff's younger self to have known that becoming a jerk would in all probability lead him to perform jerk-like actions."[143] So one can be responsible for character traits in the present when they result from decisions in the past that end up cultivating these traits. If so, perhaps the same could be true of Christine. Maybe it could be said that somewhere in her history, Christine decided to cultivate kindness and foresaw that this decision would lead to her acting as a kind person in the future. So even if there were no direct link from choice to action, as long as the outcome was foreseeable and causally related, Christine would be responsibility-apt on the prior-choice view.

The problem of under-inclusivity however is not necessarily resolved by lessening the conditions on what it means for an act to be traceable to one's choices. In fact, attempting to include Christine in this manner would render the condition over-inclusive instead. Everything we do might be connected to a choice somewhere in the generative history of the act in question. Consider Nagel once again, and specifically his concept of

[142] Fisher, Martin and Neal A. Tognazzini. "The Triumph of Tracing" In *Deep Control: Essays on Free Will and Value*, edited by John M. Fischer (Oxford: Oxford University Press, 2012): 539.
[143] Ibid., 539.

moral luck, we might also argue that when an action is traced to prior choice, it is always possible that we may find the preceding action not subject to the same control in a way that undermines whether the act was actually chosen. Given this possibility, it is not clear where the tracing condition might reach a limit given that it is always possible to tell some sort of causal story in how the act came about.

Moreover, adding the epistemic condition shifts the account from being under-inclusive to over-inclusive in ways that are difficult to justify. The charge of over-inclusivity arguably holds given that if the epistemic condition includes Christine, then it includes nearly everything we do. For many, if not all actions, it is possible to know that it will have some general, non-specific effects in the future. Fisher and Tognazzi acknowledge the generality of the foreseeable outcomes condition as they state, "After all, everyone ought to be able to foresee that they might inadvertently offend someone at some point in their lives!"[144] Yet, they do not see it as a problem because we can be responsible for actions that we might not be blameworthy for.[145] Even if responsibility and blame are distinct, and I suspect they are, the proposed distinction is only suggested without any clear reason to see why blame and being responsibility-apt may come apart. We are then left to speculate, without clear answers, as to why being responsibility-apt for nearly every action is acceptable.

[144] Ibid., 225 n16.
[145] In chapter eight I will argue for a distinction like this. Blame and being responsibility-apt are distinct and may help support Fisher and Tognazzi's claims to some extent. I will nevertheless maintain an account of responsibility that does not rely on choice. The reason is that I see responsibility as being more deeply connected to the agent and her valuations than can be provided by a choice/control account.

The largest problem for the prior-choice view, however, is that it requires an over complicated and tenuous story as to why persons are responsibility-apt for certain acts. In particular, prior-choice attributes responsibility to choices that may potentially be far removed from the current questionable action as long as a (reasonably foreseeable) causal link may be established. Yet, I do not usually praise your act of good will because twenty years ago you thought to adopt a mantra of repeatedly paying it forward that only now has become second nature. Construed more broadly, maybe it was not even a mantra, but a spur of the moment choice to be more positive or something seemingly unrelated like choosing to be kind to whomever wears blue or sequins (or better yet blue sequins). There is a level of detachment from the act that determines responsibility-aptness and the individual's current state. We might be able to trace Jeff's current jerkiness to a decision to put himself in that jerk-like state, yet it is still at least conceptually possible that the jerk-like actions function now with only a tenuous ancestral connection to who he is in the present. Jeff may no longer be a jerk, even if a foreseeable choice caused him to act like one. Prior-choice may give us a causal story about how the act came about without a necessary connection to the individual's current beliefs and attitudes.

3. Third Suggestion: Causal Structures

In the last section I argued that Christine should be responsible for her actions whereas prior choice views could only tell a convoluted story as to why this was so. As for Alessandra, I will leave a fuller discussion of where I stand in regards to her responsibility-aptness for a little later in this chapter. For now, I will turn to Sher's own solution to the question of Alessandra's moral responsibility. He suggests that she is

responsible for forgetting about Sheba because of the enduring "causal structure whose elements interact in ways that give rise to these responsibility related activities."[146] That is, the responsible self on Sher's account is to be identified with "whatever psychological and physical structures sustain [her] normal patterns of functioning."[147] The self is a composite of not only the "conscious center of will", but also the "enduring causal structure whose elements interact in ways that give rise to these responsibility-related activities."[148] To answer the question of "what makes someone the particular responsible agent he is … we must look beyond the agent's consciousness and reason-responsiveness to the causal structures that sustain them."[149] Thus, Sher's conception allows for attributions of responsibility for unwitting, unintentional, and emotional actions as long these arise from processes of the normal functioning agent.

3(a). Problems with Causal Structures

It is questionable whether Sher's suggestion really refines the motivational profile enough, at least to be useful for our purposes. Indeed, on Sher's terms it is possible that my normal functioning includes my poor hearing as it readily and consistently influences what I hear or do not hear. When I miss something someone has said, although this failure of hearing is traceable to the underlying psychophysical structures that sustain what counts as my normal functioning, am I responsible for this aural failure?

Citing the over-inclusive nature of Sher's causally defined self, Angela Smith argues that Sher's criterion is not limited enough to worries like these. She states:

---

[146] Sher, *Who Knew,* 121.
[147] Ibid.
[148] Ibid., 123.
[149] Ibid., 134.

The basic problem here, as I see it, is that merely citing a causal connection between some failure of awareness and the workings of the vast "psychophysical structure" that generally sustains our intellectual activities cannot establish the right *kind* of connection between an agent and her wrongdoing to justify us in regarding her as responsible for it.[150]

Indeed, in trying to avoid the under-inclusiveness charge, it seems as if Sher has lost sight of why exactly we would consider persons responsible. If we attribute an action to an agent on the basis of a causal link, and also recognize that the action fails to meet a standard of goodness, this tells us little in the way of whether blame or other reactive attitudes are prima facie appropriate. As Smith further explains:

When we blame someone for a cruel action or attitude, for example, we do not seem to be saying merely that she has a quality that fails to meet a certain objective standard of moral goodness (as my hearing failed to meet an objective standard of aural goodness); we seem to be saying that she has failed in some way, and that she is open to serious moral criticism for this failure.[151]

Responsibility and blame require more than judging the quality of an individual's actions against predetermined criteria as this does not get to the heart of why we consider persons responsible.[152] We are looking for what Thomas Nagel once referred to as "not what happens to a person, but *of him*."[153] In other words, when we say that a person is morally responsible we are not saying that:

---

[150] Smith, Angela. 2010. "Book Review: Who Knew? Responsibility Without Awareness." *Social Theory and Practice* 36(3)(2010): 523.
[151] Smith, Angela. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138(3), (2008): 374.
[152] Smith's argument is similar to that prominently given by Susan Wolf against real-self views in her work, "Freedom Within Reason", Wolf questions whether the conditions for attributability in real self views are sufficient to ground genuine attribution of responsibility. Just as we can attribute bad qualities to an earthquake or one's ability to hear properly, Wolf argues that responsibility, taken as a kind of attributability, consists of the same sort of trivial grading without the characteristic force of responsibility. See Wolf, Susan. *Freedom Within Reason*. (New York: Oxford University Press, 1994).
[153] Nagel, *Mortal Questions*. 36.

…a certain event or state of affairs is fortunate or unfortunate or even terrible. It is not an evaluation of the world, or of an individual as part of the world… We are judging him, rather than his existence or characteristics.[154]

Pointing to the psychophysical structure that is responsible for a bad outcome does not offer more than a descriptive appraisal of the person. For instance, the forgotten care of Sheba may just have been a similar "glitch in her psychophysical system" not unlike the "glitch" that caused me not to hear what was said.[155] Sher thus would need to explain why some psychophysical structures give rise to outcomes for which we are responsibility-apt (such as forgetting) and others (such as poor hearing) do not. As Smith argues with regard to vision:

My vision, of course, regularly 'determine[s] the contents of [my] conscious thoughts and deliberative activities,' and would thus count as one of my 'constitutive features' on Sher's definition. But given my tiny blind spot, there are going to be occasions when I am not aware of features of my surroundings that may be morally or prudentially relevant. Would the fact that my failure in such a case is caused by one of my 'constitutive features' show that I am therefore responsible and to blame for it?[156]

Insofar as we would not consider an individual responsible for psychical limitations such as a visual or aural failure, Sher's causal structure account is not sufficiently fine-grained to delineate the moral self.

4. Fourth Suggestion: Rational Relations View

This section will defend Smith's "rational relations" account as the most promising answer to our problem. Unlike the previous volitional accounts that focus on the exercise of the will, Smith attributes responsibility on the basis of something like P.F. Strawson's notion of the "quality of will" introduced in his influential work, "Freedom and

[154]Ibid.,36.
[155] Smith, "Review", 523.
[156] Smith, "Review", 524. quoting George Sher, *Who Knew?*, 121.

Resentment."[157] This sort of view does not attend just to the exercise of the will in the agent's actions or choices themselves. Instead the focus is on the representative connection of the questionable act or attitude has with the agent's 'real self.' This corresponds to Frankfurt's approach insofar it distinguishes some actions and psychological activity as more deeply attributable to the self. Yet unlike Frankfurt and as usefully summarized by Michael J. Zimmerman, the "will" here does not necessarily refer to some chosen course of action among various alternatives.[158] Instead:

> Strawson's use of the term is much broader. He is primarily concerned with whether we show good will or ill will (or indifference) toward others, and in this context 'will' encompasses a wide variety of attitudes that we take toward others through the choices we make or, indeed, the choices we do not make.[159]

Smith's view is similar because she proposes that we assess the moral worth of something like the agent's moral character through by what is revealed through certain attitudes, acts and even telling omissions. On this view, as long as the action or attitude is rationally connected with the agent's judgmental activity, they may be said to be responsibility-apt.

Smith's account does not merely provide a means to assess the individual according to some proposed standard, as might be the result of Sher's account. She argues that "moral praise or blame, unlike assessments of native intelligence or [hearing], seem to go beyond a mere unwelcome description…."[160] This is because if

---

[157] Strawson, Peter F. "Freedom and Resentment." Proceedings of the British Academy, Edited by Gary Watson. Volume 48 (Oxford: 1962):15.
[158] Zimmerman, Michael. J. "Moral Responsibility and the Quality of the Will", in *Responsibility : The Epistemic Condition*. Edited by Philip Robichaud and Jan Willem Wieland, (Oxford: Oxford University Press, 2017): 220.
[159] Ibid.,220.
[160] Smith, *Book Review*, 380.

you miss something important I said due to bad hearing, it would not be intelligible ask you to justify this lapse. Bad hearing "bears no relation to [your] own judgmental activity."[161] The key here is the connection that deeper moral criticisms have with the agent's judgmental activity and Smith argues that this is best captured when it is appropriate to regard the agent as *answerable*. She states, "[i]n order for an agent to be answerable for something, it seems that thing must be connected to her in such a way that it makes sense to ask her to rationally defend or justify it. This, in turn, suggests that the thing in question must in some way reflect her own judgment or assessment of reasons."[162]

"To reflect" one's judgments or assessment of reasons can be understood in a manner similar to T.M. Scanlon's older accounts of judgment sensitivity.[163] This sensitivity is understood as a kind of conditional of the form: "…if one sincerely holds a particular evaluative judgment, then the mental state in question should (or should not) occur." [164,165] For instance, in one scenario it might be that if the individual judges a vase to be of worth, we would expect him not to intentionally smash it. If a couple were

[161] Ibid., 380.
[162] Smith, Angela. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." (Ethics 122, 2012): 579.
[163] I say older because as Smith notes, he has more recently backed away from notions of judgment sensitivity to emphasize other aspects of responsibility.
[164] Smith, "Activity and Passivity", 253.
[165] Scanlon, defines such sensitivity as including judgements an "ideally rational person would come to have whenever that person judged there to be sufficient reasons for them … [or]… 'extinguish' when that person judged them not to be supported by reasons of the appropriate kind" (Scanlon, *What We Owe,* 20). Smith uses the notion in a less stringent way and does not presuppose an ideal rational agent. The rational connection to one's evaluative attitudes and judgments on her account simply means that the judgment, attitude, behavior, commitment, etc. bears a rational connection to the evaluative attitudes the agent holds. See Scanlon, T. M., *What We Owe to Each Other*. (Cambridge, Mass: Belknap Press of Harvard University Press, 1998).

parents we would expect them to care for their child. Likewise, if Alessandra cares about her dog, then we can assume she would not leave it in a hot car. Overall, persons may be considered answerable (and hence responsibility-apt) if it is at least in principle legitimate to ask them to respond, whereas judgment sensitivity marks the class of actions that may be open to this demand.

Under the rational relations view, Alessandra may only be seen as not answerable if her forgetting Sheba was the product of a "glitch in her psycho-physical system."[166] Consider "Forgotten Baby Syndrome" as it has been dubbed in the media.[167] In the United States, about thirty-seven young children die every year due to vehicular heatstroke with over half of those incidents due to a parent or caregiver forgetting the child in the car.[168] What is important in these cases is not that they involve children rather than dogs, but the general inability to infer anything about the parent's evaluative profile in most cases. In a NBC interview, David Diamond explains that often the phenomena is not necessarily the result of a lack of caring on the parent's part, but due to competing processes in the brain. He describes the conflict in terms of a tennis match in which, "[t]he basal ganglia allows a tennis player to hit the ball in an almost reflexive way, while the hippocampus and the frontal cortex allow the player to devise a strategy."[169] These two systems can compete, especially when there is a change in the routine. The parent may think that the child, while quietly in the back, is at daycare and

[166] Smith, "Book Review", 523.
[167] Williams, C. A. and A. J. Grundstein. "Children Forgotten in Hot Cars: a Mental Models Approach for Improving Public Health Messaging." *Journal of the International Society for Child and Adolescent Injury Prevention*. (2017): 279.
[168] Ibid.
[169] Rosenblatt, Kalhan. "Hot Car Deaths: Scientists Detail Why Parents Forget Their Children." NBC news. Jun. 27.2017 / 11:18.

may even have a false memory blurred with all the other times the morning went as planned. Coupled with the sleep deprivation and often-insufficient parental leave that forces the parents back to work early in the child's development, these circumstances are primed for such fatal mistakes. In this sense, forgetting Sheba may too be excused if it functioned like a glitch of this sort. She would be excused because she is not answerable in this scenario. To ask her to justify why she left Sheba in the car would be as intelligible as asking her to justify why she sneezed as both are not judgment sensitive.

5. Clarifications through Shoemaker

By using judgement sensitivity as a means to delineate the scope of the motivational profile, Smith's view would help to avoid the charge of over-inclusivity by involving only those aspects of the self that represent the "basic evaluative framework through which we view the world."[170] There is a clear justificatory connection that grounds deep attributability, not just basic attributability that is given by a causal connection. It is another question, however, as to whether it can avoid a charge of under-inclusivity.

David Shoemaker might argue that the rational relations view is still too narrow. He denies that moral responsibility should be understood only in terms of answerability because it is possible for actions to be attributable and represent one's evaluative commitments without the individual necessarily being answerable for them. The focus on judgment sensitivity in the rational relations view, he argues, excludes emotional commitments, non-rational attitudes and behaviour for which it does not seem appropriate to demand an answer or justification. He gives the example of a mother's

[170] Ibid.,251.

"groundless emotional commitment" to her murderous son.[171] The emotional mother might say:

> After my child has become a serial killer, for instance, I may arrive at the consciously held propositional belief that he's a worthless human being, that he's dead to me. And yet when I read of his upcoming execution, I may well up with tears or fall into a depression. "I still care about him," I may say. "There are no reasons to do so—he's an awful man—but it still matters to me what happens to him."[172]

Shoemaker argues that even though the mother's care is "simply devoid of resources necessary to engage with [a] communicative attempt", she may nevertheless be morally responsible for it because the emotional plea is attributable to her moral character.[173] Responsibility-aptness may hold in spite of a lack of answerability.

Shoemaker is right that the focus on answerability might appear to make the rational relations view overly rationalistic and under-inclusive as a result. However, I think it is possible that Smith's account could include the kinds of examples he gives, although not necessarily for his reasons. In the following sections, I offer two general responses to Shoemaker's interpretation of answerability and judgement sensitivity as to counter his claim to under-inclusivity. Each of these sections will contain a number of smaller objections that question Shoemaker's criticism of the answerability test; they also offer a means to clarify my alternative interpretation of Smith's claims that amalgamates her view with Shoemaker's. I hope to show that the emotional mother is responsibility-apt, not as a counterexample (as Shoemaker argues), but *because* she satisfies the answerability test.

---

[171] Shoemaker, "Attributability, Answerability, and Accountability", 611.
[172] Ibid.,610.
[173] Ibid., 611.

5(a). First Response to Shoemaker: What it means to be Answerable

First, there are a number of smaller objections we could make against Shoemaker's claims starting by asking about what it means for an attitude or action to be attributable without being answerable for it. It is not clear to me that conditions of attributability would even hold if the emotional mother could not answer for the love she feels. If the mother's emotional state could not be modulated by her evaluative beliefs at all – if they are completely judgment insensitive – Why would we think it expressive of her moral character? If the love were entirely disconnected from her evaluative judgments, it would be more akin to a state like hunger that an aspect of moral character. The love she feels would then be functionally equivalent to psychological disorders, glitches or physical responses that are not responsibility-apt.

Secondly, it is also not clear in the example given by Shoemaker that no answer, in principle could be provided even if it is difficult to do so. The mother may have some inchoate, unpersuasive, or hard to articulate reason to ground the love she has for her murderous son. Generally, I do not think that we should take it as a sign of a lack of answerability when an answer is not fully formed by the individual. Despite the conflicted sense in which the mother may love her son, it is not against all reason, but perhaps only against good, articulated reason.

Smith argues that with considerations of answerability, "we must show that the agent is connected to that thing in a way that makes these answerability demands intelligible."[174] An agent is not answerable only if they have ready answers in the face of

[174] Smith A.M. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122, no. 3 (2012): 578.

challenges, but that the demand for an answer is intelligible or fitting. Take for example Andrea Westlund's case of "Betty" who "confounds her doctors by refusing potentially life-saving skin-graft surgery."[175] Her refusal was not because she was incapable of answering, but that she "reject[ed] as unreasonable the very demand that she give reasons" at all.[176] Due to not valuing justificatory dialogue, she "simply shut down when pressed to give reasons."[177] It is possible that Betty did not have any justificatory reasons in mind, but this does not necessarily undermine her answerability. As Westlund argues those like Betty may be answerable because they "manifest responsiveness to justificatory challenges … even while disvaluing and refusing to engage in certain practices, including practices in which they are pressed to cite their reasons in the face of direct questioning."[178] Indeed, this apparent lack of readiness to engage in dialogue is not uncommon and might even characterize someone with a stubborn mindset. Stubbornness might not only cause individuals to refuse engagement with justificatory challenges, but could even blind the individual (even if temporarily) to other possibilities. Here too, the lack of readiness to provide a response does not undermine the fittingness of asking for a response in the same way asking someone to justify a sneeze might. Answerability holds as long as the demand is at least intelligible.

The potential intelligibility of requesting a response as determining answerability might nevertheless seem too thin a requirement for answerability and might not be enough to allay Shoemaker's worries entirely. He argues that in order to be answerable

[175] Westlund, Andrea C. "Rethinking Relational Autonomy." *Hypatia* (24, no. 4, 2009): 37.
[176] Ibid.
[177] Ibid.
[178] Ibid.,40.

the agent could reasonably be asked not just for the considerations that "she judged to count in favor of F-ing but also, 'Why did you F *instead of not-F*?''[179] Answerability on Shoemaker's terms depends on whether one has the ability to govern oneself in light of "instead of reasons." [180] Of course there would not be any 'instead of' reasons when one sneezes and it is not clear that the emotional mother or Betty even had such counter reasons on hand, but again I do not think not being introspective or refusing to be introspective enough as to devise contrastive reasons gets to the heart of what it means to be answerable.[181] Again, it remains intelligible to request a response even if the agent did not have such contrastive reasons in mind. This is because, while the action itself may not be committed with many these reasons in mind, the fact there is a lack of contrastive reasons could be due to the agent's evaluative beliefs, which renders her answerable to some extent.

Consider Betty once again. Arguably, because she simply "shut down" in the face of questioning and did "not engage in an internal give and take of reasons of the sort her doctors hoped for …", she had no contrastive reasons in mind when she refused treatment. [182] Yet, she lacked these contrastive reasons because she did not value justificatory dialogue and this says something important about her and her evaluative

[179] Shoemaker, David. *Responsibility from the Margins*. (Oxford: Oxford University Press, 2015): 75.
[180] Shoemaker, *Responsibility from the Margins*, 76.
[181] Shoemaker uses the example of a psychopath who is "unable to perceive any facts about others' normative perspectives as (even appearing to be) reasons" and thus, "cannot be answerable for his judgment that others' interests are worthless, in virtue of the fact that he has no access to the relevant "'instead of' regard-based reasons" (Shoemaker, *Margins*, 218).
[182] Westlund, "Rethinking Relational Autonomy", 37.

system. The same can be said of stubbornness that blinds one to 'instead of' reasons. After all, stubbornness is generally considered a vice, not an excuse and its expression may be dependent on 'who the person is' and the kinds of valuations they hold. On my account, the emotional mother, Betty, or the stubbornly blind would only not be answerable if each lacked the capacity to assess the action in light of "instead of" reasons *and* that the lack was not due to something within the agent's evaluative constitution. Against Shoemaker, I would argue that only then would the influence be properly external as consistent with an answerability test.[183]

Of course arguing in this manner against Shoemaker might seem a little hypocritical. After all, I seem to be using tracing conditions that I rejected for prior choice theories. It is true that the sort of indirect answerability advocated in Betty's case might be one step removed and trace in some sense, but what it traces to is importantly different and I think it's an advantage for this interpretation. Attributability based on the

[183] In light of these responses to Shoemaker I hope to have shown where he and I diverge, while also defending evaluative sensitivity as distinguishing what is and is not responsibility-apt. To avoid confusion, I want to briefly note one further difference between our respective projects. Indeed, Shoemaker outlines a comprehensive theory of responsibility responses that separates the kinds of objects taken by our reactive attitudes and what needs to be the case in order to render these specific responses appropriate. He generally adheres to what he calls the "H-tradition" or "holding-responsible", which "maintains that our responses most fundamentally help constitute what it is to be responsible" (Shoemaker, *Margins*, 19). Regarding persons as responsible consists in the sorts of reactive attitudes we have and whether or not they fit their circumstances. I however belong to what he terms the "B-Tradition" or "being responsible" (Ibid.). This tradition claims "our responses are most fundamentally our best epistemic trackers of the attitude-independent facts about responsibility" (Ibid.). Being responsible amounts some antecedent fact, in our case it is whether the evaluative profile is reflected in the action or attitude that would "count as evidence about when to hold people responsible" (Ibid.). In later chapters, I will defend this position further and show the difference between being responsible and holding responsible (as in being to blame) as something like a Venn diagram with moral responsibility inhabiting the overlapping subsections.

extent of answerability differs from prior-choice theories insofar as Smith's view tracks the sorts of valuations the agent currently holds rather than the genesis of the act or attitude. She states:

> The genetic strategy forces us to view our own attitudes as the mere products of our own actions, like the bodily conditions we produce through training, exercise, or excessive alcohol consumption. Both of these accounts fail to capture the special fact about attitudes, which is that they are judgment-sensitive responses to the world around us.[184]

Here, Smith is pointing to the kind of disconnect that undermined prior-choice theories as appropriate to distinguish between internal from external aspects at the beginning of the chapter. In regards to Jeff's responsibility for acting like a Jerk, I argued that establishing a link between the initial choice that generated the act in question says little on who the person is and the sorts of evaluations they hold (as they currently stand). The story as to why this connection makes one responsibility-apt is tenuous at best.

Answerability by contrast offers a deeply personal connection that prior-choice theory is unable to provide. Part of the requirement of intelligibility of requesting a response requires an answer to be personal in the sense that it involves reasons that the agent currently holds. Persons are only responsible for what reflects who they are and not who they might have been when a generative choice was made. What matters instead for answerability is what I will call *personal answerability* as offering a necessary connection between responsibility-aptness and the agent's current evaluative system.[185] The term personal answerability emphasizes the necessary condition of the agent holding

[184] Smith, Angela M. "Attitudes, Tracing, and Control." *Journal of Applied Philosophy* 32, no. 2 (2015): 127.

[185] This means that if an action or attitude was the product of choice or even appropriately traced to a choice somewhere in its lineage, the individual may not be appropriately responsibility-apt if it does not reflect who they are now.

the evaluative aspects in which a response is requested. For the rest of this thesis, the notion of answerability will be understood with this condition implied. For now it is important to note how it provides a justificatory relation, not an explanatory and causal account about how the act came about.[186]

5(b). Second Response to Shoemaker: What it means to be Judgement Sensitive

The answerability test targets only those aspects of the self that reflect the agent's valuations in the world, which arguably includes the kinds of cares that characterize Shoemaker's emotional mother. I have argued that she is answerable because despite not having a ready answer, it still is intelligible to request a response. Yet, we could also argue that she is answerable because it is intelligible to answer for more than Shoemaker presumes is implied by answerability. Despite Smith's use of the term 'judgment' in the notions 'judgment sensitivity' and 'evaluative judgment', these terms should not be understood through the rationalistic lens that Shoemaker attributes to the account. Smith argues for a much laxer sense. She states, "The reason for this looseness is that I want to make clear that the judgments I am concerned with are not necessarily consciously held propositional beliefs, but rather tendencies to regard certain things as having evaluative significance."[187] This may involve a wide array of psychological activity including the kinds of cares and commitments Shoemaker identifies. Smith notes that she uses the term

[186] I would even go as far to posit that assessing the content of the attitudes the very reason we find theorists preoccupied with choice and voluntary control. It is not that responsibility relies on something having been chosen, but that a choice allows for the inference to one's character in a clear and non-tangential way. Clarity of inference does not mean that choice renders persons more or less responsible if we are looking to assess who they are. Choice and control conditions are not peripheral to responsibility assessments. They are just not central to them and can be used as a means to support the kind of assessment as advocated by Smith.

[187] Smith. "Activity and Passivity", 251.

'judgment' despite the potential confusions in order to denote the stability of the kinds of dispositions she is concerned with. They are "standing commitments" and not "merely one time assessments."[188,189]

For the sake of clarity, I would like to rename Smith's ' judgement sensitivity' as 'evaluative sensitivity'. These are the aspects of the self that are evaluatively sensitive, reveal the individual's valuing activity and the agent is responsibility-apt for only these aspects of the self. Summarized here:

Evaluative sensitivity: Actions or attitudes that are reflective of the individual's evaluative stance.[190]

I would argue that if an attitude, belief or state is evalutively sensitive then the agent is answerable for acts reflective of those evaluative aspects and are, hence, responsibility-apt. This means that judgment sensitivity construed as evaluative sensitivity includes the kinds of cares, commitments, choices, endorsements or whatever evaluative activity that Shoemaker includes as responsibility-apt. Yet, against Shoemaker, they are included due to being open to answerability demands. In the mother's case, what may appear to be a "groundless emotional commitment" might be better characterized as what I call an intermediate state for which persons are indirectly answerable, as I will now explain.[191]

---

[188] Ibid., 251, n.27

[189] I would not, however, emphasize them as standing commitments. As I will argue in chapter four, although commitments, cares and judgments might be fleeting, they are nevertheless deeply attributable.

[190] I would like to thank Joey Van Weeldon for draw my attention to how overly rationalistic the term 'judgment sensitivity' was and for offering the term 'evaluative sensitivity' in its place.

[191] Shoemaker. *Attributability*, 611.

5(c). Evaluative Sensitivity in Action: Intermediate States

Evaluative sensitivity sheds further light on why persons may be answerable for certain cares and commitments even if it seems that the agent could not help but feel these emotions and attitudes. Arguably the love experienced by the mother is not *sui generis*; it is likely connected to her other beliefs, attitudes and valuations. She may not be able to give a direct answer as to why she continues to love her son, or a defense that employs a specific belief about love, but she may be answerable on the basis of these connected states.[192] I call states that involve indirect answerability *intermediate states*

[192] The notion of these intermediate states, also help shed light in similar problems concerning responsibility attributions of what Nomy Arpaly and Timothy Schroeder have called "inverse akrasia."(Arpaly and Schroeder, "Whole self", 164.). One example includes Mark Twain's Huckleberry Finn's praiseworthiness for his resistance to racism. Generally we tend to think these actions like Finn's are praiseworthy, but on what basis is not always clear. Finn acted contrary to his best judgment and against his values despite laudably not turning in his friend Jim who was a runaway slave. I would argue that both regular and inverse akratics are responsible for their actions because their motivations are evaluatively sensitive in either case. For instance, Finn's actions are not the result of blind mechanisms working on him. Arpaly and Schroeder explicitly deny that he can be characterized as such as they argue, "It would be wrong to think that Finn is squeamish, unable to see a man in chains, or blindly attached to every adult that is nice to him. Finn is praiseworthy because he is averse to turning Jim in for morally significant reasons" (Ibid.,164.). The emotional reasons are morally significant and attributable to him. But they are just as attributable as his explicit beliefs and attitudes about racism. The wrong and right making features could lie instead in whether Finn was right to follow gut feelings on the matter. Perhaps we could turn to Tappolet and argue that like an insightful hard-boiled detective relying on his gut, Finn "has well-tuned self-monitoring habits, such that [he] would not have relied on [his gut feelings] if there had been reason for [him] to believe that [his] anger was misleading [him] (Tappolet, "Emotions, Reasons and Autonomy",173). He may be in the wrong to act in such a way if these skills are not "well-tuned" in a way that would be haphazard at best (Tappolet. *Emotions,*173). He could be critiqued regardless of whether his emotional intuitions were right. In this case if the reason he was right was merely contingent. In either case, whether or not we attribute blame or praise does not make much of the difference for responsibility, as both regular and inverse akrasia are attributable. See Arpaly, Nomy and Timothy Schroeder. "Praise, Blame and the Whole self." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 93, no. 2 (1999) and

because they are felt experiences that result from the agent's particular perspective on the world. The extension of answerability to intermediate states will broaden their application to the mother's love and also clarify the conditions of answerability. It will also shed light on why inadvertent acts and spur of the moment decisions, as we saw with Christine, might be responsibility-apt despite the apparent disconnect to the agent's otherwise explicit judgements and beliefs. [193]

I consider cares such as love and other similar states like jealousy as evaluatively sensitive in a way physical responses and the like are not because the manifestation of such depends on one's evaluative profile being constituted in a certain way. Such states are not a fully physical response (like e.g. hunger) because the experience of such requires both the holding of and the perception of certain evaluative attitudes.

Consider Christine Tappolet's account of emotions. She argues that certain emotional states can be understood as perceptual experiences of evaluative properties. Emotions, she argues, are the result of perceiving a certain evaluative property in an object or situation. Tappolet gives the following analogy:

> Like sensory perceptions, emotions appear to have a characteristic phenomenology; both emotions and sensory perceptions are in general caused by things in our environment; both fail to be directly subject to the will; both appear to have correctness conditions, in the sense that they can be assessed in terms of how they fit the world; and finally both can conflict with judgment.[194]

Tappolet, Christine. "Emotions, Reasons and Autonomy." In *Autonomy, Oppression, and Gender*, edited by Andrea Veltman and Mark Piper, (Oxford: Oxford University Press, 2014).

[193] It is then my contention that responsibility is unitary because answerability, whether indirect or with regards to emotional pleas, holds in all of these cases in which the agent is intuitively responsibility-apt.

[194] Tappolet, *Emotions,* 170

Emotions pick out information in the environment in ways that inform and motivate us to act. For instance, fear informs the agent of danger independently of a judgment that the object should elicit fear. The object is simply represented as fearsome. Thus, "emotions can key us to real and important considerations that speak in favor of acting in certain ways, without always presenting that information to us in a way susceptible to conscious articulation."[195] It may or may not be the case that the murderous son possesses the property of loveableness objectively speaking, but because the mother is who she is, the world is represented to her in a particular manner due to her evaluative constitution.

Smith makes a similar distinction to mark out what is evaluatively sensitive and asks us to consider perception to make her point clear. In perception whether or not you see a straw as bent within water is not dependent on any evaluative judgements. How I understand the perception may differ, but the actual content of it does not change depending on what I believe. Smith acknowledges that there may be some sense in which perception may be said to be sensitive to our valuations. She states:

> In the case of attitudes like shame, jealousy, fear, [love,] and admiration, the evaluative judgments are themselves partially constitutive of the attitudes in question. Our attitudes are not merely the causal effects of our judgments (in the way that some of our physical reactions may be causal effects of our judgments). They are, rather, active states, in the sense that they essentially involve our judgmental activity.[196]

States like those enumerated by Smith are arguably moral indicators that suggest a deeper evaluative judgement due to their constitution.

[195] Ibid., 170
[196] Smith, "Activity and Passivity", 258.

If the feeling of certain emotions like love is constituted by the discernment of the love-salient properties in particular situations, then it is both an evaluatively sensitive and visceral response. For instance, consider a situation described by Lawrence Blum. He asks us to picture John and Joan on a packed subway train with no empty seats. There is one woman standing while trying to carry some heavy packages. In this case, "John is not particularly paying attention to the woman, but he is cognizant of her. Joan, by contrast, is distinctly aware that the woman is uncomfortable."[197] There is a difference in how the situation is perceived that is not necessarily confined to the facts of the situation, but John and Joan's particular perspectives. Blum continues:

> [T]he difference between John's and Joan's perceptions of the situation lies not only in the relation between that perception and the taking of beneficent action. It lies in the fact of perception itself. We can see this more clearly if we imagine John's and Joan's perceptions to be fairly typical of each of them. John, let us say, often fails to take in people's discomfort, whereas Joan is characteristically sensitive to such discomfort.[198]

It would seem that in virtue of what is salient to John, it is possible to infer a character defect of some sort. I argue that John's non-action is responsibility-apt as his perception or lack-there-of is related to his evaluative aspects given that these influence how he views the world. Thus if John is responsibility-apt in this case, he may also be responsibility-apt for a whole host of emotions that do not necessarily reflect his explicit judgements or commitments. Yet, the fact that these emotions have a basis in his evaluative constitution renders them the kinds of things that are intelligible to request a response.

---

[197] Blum, Lawrence A. *Moral Perception and Particularity*. (Cambridge, England: Cambridge University Press, 1994). 31.
[198] Ibid., 33.

The love of the irrational mother may be of the same sort of answerability on the basis of evaluative sensitivity. She is answerable for her love even if she does not hold any explicit judgments of the sort. She has reason to feel the way she does and the fact that she feels and perceives love for her son at all is a reflection of that.

6. A Counterfactual Test

In cases where we might be unsure as to whether the agent is answerable - as with the emotional mother - we need to ask: if the mother's evaluative activity that pertained to the issue at hand changed, would she still love her son? Through questions like these, I propose that we can determine whether evaluative sensitivity holds. What is important in this counterfactual is whether the mother's love (or any other questionable state including those made without 'instead of reasons') would have occurred regardless of his or her evaluative constitution. If so, then the love is not responsibility-apt. If not, then it says something about the mother in a way that maintains deep attributability. If she would still love her son despite having a radically different (and perhaps even opposite) evaluative stance then the love would be against all her evaluative activity and not be answerable as a result. What matters is not necessarily whether the mother has certain valuations in mind as explicit propositional beliefs, but that her attitudes and actions are at least evaluatively sensitive in a way that opens her for answerability demands. Against Shoemaker, evaluative sensitivity does not require the agent to have acted in light of or even have access to "instead of reasons."[199] Rather, what is deeply attributable to the agent depends on whether the actions, behaviour or traits could have been otherwise due to the agent's evaluative constitution.

[199] Shoemaker, *Margins*, 75.

This counterfactual may also shed light on Christine's actions that were introduced earlier in the chapter. She would be praiseworthy due to what this spontaneous emotional response reveals about her evaluative activity and concern for others. The act of slowing down was deeply revealing of Christine's evaluative perspective not because it arose out what seems to be an unthinking habit, but it is act that is deeply connected to Christine's evaluative constitution. Arguably, had she not had the kinds of evaluative beliefs and attitudes she had, she might not have been able to see the concerned look on the elderly driver's face in the same way the mother might not have seen the loveableness of her son had she been constituted differently. On a counterfactual, had Christine's evaluative stance been different, she might not have slowed down.[200]

7. The Motivational and Evaluative Profile

Smith's notion of answerability combined with the wider scope introduced by Shoemaker, gives us the tools needed to refine the motivational profile into something

[200] Of course, persons may not be responsibility-apt for just anything they fail to do as a result of this counterfactual test. Love, good pet-ownership or morality calls for our actions to be limited or influenced in certain ways that display our evaluative commitment to them. We might even want to say that some demands should necessarily influence our actions if my identity properly locates me in a particular normative domain. It makes sense to demand an answer for an action or inaction because the performance or non-performance is governed by some normative standard based on the role occupied. If I properly stand in this role, I can be questioned for actions that pertain to that role. For instance, as a mother, I may be blameworthy if I shrug some parenting duties or forget something important I should have known otherwise. The caveat I am adding here links to Smith's distinction between depth and significance. Depth refers to assessments of the individual as an agent, while significance refers to the normative domain the action falls under (e.g. This can be a moral one). We might want to say then that this proposed counterfactual only applies if the action falls in the right normative domain. The general idea is that if an agent values something, judges it to be of worth, and clearly occupies a particular role that presupposes those values and judgements, the agent's patterns of thought, feeling and motivation to act should be affected in consequence. So even during a lapse, as long as one belongs to the normative category, they may be said to be responsibility-apt due to this counterfactual test.

that is more recognizable as a moral self. Here I call the general subset of responsibility–apt influences (those that are evaluatively sensitive) the *evaluative profile.* This grounds attributions of moral responsibility and is distinguished from the *motivational profile* that comprises the whole of the motivations that affect one's perspective.

The distinction between the evaluative and the motivational profiles parallel Gary Watson's distinction between evaluative and motivational systems although I use the terms somewhat differently.[201] The motivational system, for Watson, consists of all psychological aspects, including desires and cravings that move the agent to action. The evaluative system by contrast is a "set of considerations which, when combined with his factual beliefs (and probability estimates), yields judgements of the form: The thing for me to do in these circumstances, all things considered is, A."[202] It is the normative system connected to what the agent identifies as good and worthwhile, which may "be said to constitute one's standpoint, the point of view from which one judges the world."[203] So, in a manner similar to Frankfurt, Watson's distinction shows that the motivational profile can be likened to *desiring*, while the evaluative concerns thinking worthwhile or good*, ceteris peribus*. The difference between my use of the evaluative profile and Watson's is that I would include *mutatis mutandis* as well as *ceteris peribus* valuations because I see both as being deeply attributable for the purposes of responsibility attribution.[204] As I argued earlier and against Frankfurt, desires and

[201] Watson, Gary. *Free Will*. (Oxford: Oxford University Press, 2003): 215.
[202] Ibid.
[203] Ibid., 216.
[204] The reason for this difference will be revisited in the next chapter. I argue that aspects of the evaluative profile that are not well integrated either semantically or evaluative are

attitudes in which we experience alienation and estrangement might nevertheless be deeply attributable even if we wished they were not.

Taking on board Smith's rational relations view, Shoemaker's objections, and Watson's terminology with the distinctions introduced in the last chapter, I will define the motivational and evaluative profiles as follows:

Motivational profile (Constitutes the forensic unit): Includes all psychological features that inflect and influence one's experience of the world (including physical motivations such as desires, but also impulses, nervous tics and the like.) A psychological feature X is generally *attributable* to an agent A's general self iff X is part of the causal structure that explains A's behaviour

Evaluative Profile (constitutes the moral self): Includes all psychological features that are evaluatively sensitive in Smith's sense (i.e. satisfy the answerability test). A psychological feature X is deeply attributable to A's ("real") moral self iff it is appropriate that an agent A is answerable for X (which can include *ceteris peribus* and *mutatis mutandis* valuations)

We are responsibility-apt when the actions or attitudes are sensitive to the wide array of valuations contained in our evaluative profile as composed of care, commitments, values or simply "tendencies to regard certain things as having evaluative significance."[205] When taken together the evaluative profile "make[s] up the basic evaluative framework through which we view the world."[206] With the inclusion of insights from Smith, Shoemaker and Watson, I have now arrived at what I see to be a plausible answer to our initial question of the moral self that initiated this chapter.

8. Conclusion

---

nevertheless deeply reflective even if they do not represent one's *ceteris paribus* valuations.

[205] Smith, "Activity and Passivity", 251.

[206] Ibid., 251.

Overall, while basic attributability locates the proximate cause of an event or action, the kind of attributability here allows for a deeper assessment of the individual. The bounds of what is responsibility-apt are not drawn by identification, of choice or psychophysical structures, but the parameters set out by what is and what is not evaluatively sensitive. So not only does the rational relations view provide a compelling rationale for our everyday assessments of others, but also it further shapes the boundaries of the self. Limiting selfhood to only what is intentional or deliberate would provide us with an impoverished picture of all that we can be legitimately held responsible for. Part of what informs my present outlook may be evaluative beliefs and attitudes that I may have never explicitly considered, but are nevertheless attributable to me. I may also act on cares that I may not be explicitly aware of, but reflect who I am as a valuing agent. We do not simply grade the agent against some normative standard, but evaluate certain aspects of them as a reflection of their agency. We should not take an inchoate or emotional response as lacking possible answers. Many emotions function like intermediate states and the reason for the saliency of what is perceived is due to the agent's constitution. Because the evaluative profile contains only what is evaluatively sensitive, it represents the person as an agent and this includes all attitudes, beliefs and behaviors that would be otherwise if one's evaluative beliefs changed. Attribution is deep once it concerns the evaluative activity of the agent and, hence, an appropriate target for a demand of justification. The self (forensic unit) encompasses the whole motivational profile with all the influences that come to inflect our experience, but the moral self extend only insofar as actions stem from this agential subset of attitudes. This

subset of evaluative beliefs and attitudes then warrants the claim to being representative of the 'real self' and avoids the charge of over-inclusiveness.

Chapter 4: Conflicting Clusters of Selfhood

Introduction

The last chapter defined the shape of the moral self in terms of whatever is reflective of an agent's evaluative activity on the basis of an answerability test. This inclusion provided a means to refine the defining features of the self as something particularly moral in nature by including only those aspects of the self that are evalutively sensitive. This chapter in turn will be concerned with polishing any residual rough edges by addressing a couple of potential criticisms that would question whether evaluative sensitivity alone provides a stringent enough test for responsibility-aptness.

I will divide the chapter into two general parts corresponding to two objections to my account thus far. Section 1 will discuss the conditions under which we can be responsible for actions and attitudes that escape conscious awareness. I call this the *awareness objection*. I discuss the challenging case of *implicit biases* that will form a case study for the rest of the chapter. I argue that we can be responsible for actions influenced by implicit biases when the latter function as something like the intermediate states discussed in the last chapter. Implicit biases are also relevant to the second objection that I call the *integration objection*. Not only might implicit attitudes conflict with the agent's explicit and endorsed beliefs, they often do not represent the agent's all things considered beliefs and attitudes. This section explores "whole self" theories such as those of Nomy Arpaly, Timothy Schroder and Neil Levy who argue that integration of

the self is necessary for responsibility-aptness. [207] I argue that there is little reason to think that the self is unified in the manner suggested: even in everyday circumstances, fragmentation and internal conflict do not undermine responsibility-aptness.

The exploration of these two objections serves a larger purpose for the thesis. If we start with the idea that the self is generally and usually in such disarray, then the kind of conflict that characterizes rehabilitative change would no longer be an exception to the standard cases of continuation. Criminal rehabilitation may represent a more extreme case of discontinuity, yet, if this chapter is correct, then the difference in these cases may only be of degree and not kind.

1. The Awareness Objection

As we saw in the last chapter, one's evaluative outlook could be inferred from cases of cognitive failures such as lapses in judgment and forgetfulness. Sher's example of Alessandra forgetting Sheba in a hot car served as a poignant example of possible responsibility due to forgetfulness. Evaluative sensitivity opens the door to a wide variety of actions. For instance, responsibility may be attributed despite a lack of direct control or control in the causal history of the action. As Smith argues, we may even be responsible for "involuntary reactions" and "morally objectionable desires." [208] The former may include laughing at a malicious joke, while the latter may include simply having "inherently objectionable" desires such as wanting the suffering of animals.[209]

[207] Levy, Neil. "Implicit Bias and Moral Responsibility: Probing the Data." *Philosophy and Phenomenological Research* 94, no. 1 (2017): 3-26. and Levy, Neil. "Neither Fish nor Fowl: Implicit Attitudes As Patchy Endorsements." *Noûs* 49, no. 4 (2015): 800-23.
[208] Smith, Angela M. "Attitudes, Tracing, and Control." *Journal of Applied Philosophy*. 32.2 (2015): 119.
[209] Ibid.,119.

Whether or not there was a prior choice to laugh or act on a morally repugnant desire does not absolve one of responsibility for finding the joke funny or having the desire in the first place because we are assessing the agent's quality of the will. Yet, it could be argued that by focusing on the quality of the will we are missing something important, namely the exercise of the will. As noted by Thomas Nagel, "A person may be greedy, envious, cowardly, cold, ungenerous, unkind, vain or conceited, but behave perfectly by a monumental effort of the will."[210] That is, one could argue that success in the active suppression of the attitude and not merely holding it is what should be assessed for purposes of responsibility attribution.

I would like to explore the phenomenon of implicit biases in order to counter this sort of objection. Implicit biases exert influence despite the agent's inability to exercise their will in the acquisition or manifestation of these biases. Intuitively, it might be thought that they are not responsibility-apt in a way an individual might be for inappropriate laughter or repugnant desires. Implicit biases force us to ask how it is that we can be responsible for an attitude that the agent did not know she had and would disavow if she was made aware of it. We need not question whether there was a "monumental effort of the will" because the will was never engaged in the manifestation of the bias.[211] I argue that only some attitudes that escape conscious awareness are responsibility-apt. A lack of awareness (or integration as we will see) *tout court* does not excuse responsibility-aptness even if it may provide a justifying factor in whether we

[210] Nagel. *Mortal Questions*, 32.
[211] Ibid., 34.

hold the agent culpable. Rather, responsibility for unconscious or unintegrated aspects of the self depends on the degree of evaluative sensitivity.

1(a). Implicit Biases

As the name implies, "implicit biases" are acculturated biases that influence the agent's judgments in unconscious ways. These biases have been used to explain a number of phenomena including race or gender preferences in the selection of CVs and discriminatory physical responses including eye blinking or sitting further away from people according to their race, gender or other identity. Even more strikingly, it has been suggested that they affect interpretations of empirical observations such as an innocuous object in the hand of a black man being perceived as a gun.[212] The term has been used as somewhat of a catchall for a number of disparate phenomena and as more and more research has come to light, the definition has shifted.[213] That being said, I wish to start with an initial definition of implicit biases as associative mechanisms.

Implicit biases, when framed as associative mechanisms, operate independently of awareness or rational control. Consider Jennifer Saul's definition: implicit biases are "unconscious biases that affect the way we perceive, evaluate or interact with people

[212] See Payne, B. Keith, Alan J. Lambert, and Larry L. Jacoby. "Best Laid Plans: Effects of Goals on Accessibility Bias and Cognitive Control in Race-Based Misperceptions of Weapons." *Journal of Experimental Social Psychology* 38, no. 4 (2002): 384-96, Ashby Plant, E. and B. Michelle Peruche. "The Consequences of Race for Police Officers' Responses to Criminal Suspects." *Psychological Science* 16, no. 3 (2005): 180-83 and Correll, Joshua, Sean M. Hudson, Steffanie Guillermo, and Debbie S. Ma. "The Police Officer's Dilemma: A Decade of Research on Racial Bias in the Decision to Shoot." *Social and Personality Psychology Compass* 8, no. 5 (2014): 201-13.

[213] For an overview, See Brownstein, Michael, and Jennifer Mather Saul, eds. *Implicit Bias and Philosophy*. Volume 1, Metaphysics and Epistemology. Oxford, UK: Oxford University Press, 2016.

from groups that our biases 'target'." [214] In a study by Ulman and Cohen (and presented by Neil Levy), subjects were asked to rate the suitability of various candidates for the position of police chief. As Levy elaborates:

> One candidate was presented as 'streetwise' but lacking in formal education while the other had the opposite profile. Ulman and Cohen varied the sex of the candidates across conditions, so that some subjects got a male streetwise candidate and female well-educated candidate while others got the reverse. Subjects were also required to indicate the importance of the criteria listed for suitability for the job of police chief, as well as indicate their degree of confidence that their decision making process was objective. [215]

The results of the study were perhaps not particularly surprising. The subjects considered the male candidate significantly better qualified under all conditions. The importance of being either "streetwise" or well educated shifted when they were assessing the male candidate, thereby showing a preference for him regardless of the criteria given. [216] But in what way could they be responsibility-apt when the subject did not know that the bias was operative?

It seems that there are a couple of reasons to think persons may not be responsibility-apt for such biases being operative in their actions and decision-making. Saul, for instance, cautions against assigning blame for implicit biases due to the way they may elide conscious control. She states:

> I think it is also important to abandon the view that all biases against stigmatized groups are *blameworthy*… A person should not be blamed for an implicit bias of which they are completely *unaware* that results solely from the fact that they live in a sexist culture. Even once they become aware that they are likely to have implicit biases, they do not

---

[214] Saul, J. "Implicit Bias, Stereotype Threat, and Women in Philosophy". *Women in Philosophy What Needs to Change?*, edited by Katrina Hutchinson and Fiona Jenkins. (New York: Oxford University Press, 2013): 40.
[215] Levy, Neil. *Consciousness and Moral Responsibility*. (New York: Oxford University Press, 2014): 94.
[216] Ibid.

instantly become able to control their biases, and so they should not be blamed for them.[217]

She notes that persons may be blameworthy for such biases only if "… they fail to act properly on the knowledge that they are likely to be biased—e.g., by investigating and implementing remedies to deal with their biases." [218] Saul's emphasis is on the precondition of control for responsibility-aptness, but there is another way to excuse persons that is implicit in her statements. In particular, the participants in the streetwise study would not be responsibility-apt in Saul's sense not only because they could not control the acquisition or expression of the bias, but also they needed to become aware of the potential for the bias as a precondition to such control. Even if persons are not able to directly control the bias after acquisition, they may be able to implement certain countermeasures (such as annonymizing resumes in the streetwise case) to alleviate the effects of the uncontrollable bias. In either case awareness is a precondition to the kind of control suggested by Saul. In general, it could be argued that persons may be excused for having certain biases operative in their actions, especially if there is little reason to think they could not have become aware enough of this possibility in order to curb or control the influence of such biases. This is the *awareness objection* that will be the focus of the following section.

As the objection goes, persons are responsibility-apt only if they could control the manifestation of the bias and they could only control it if they had been aware of the potential influence. This seems reasonable given how implicit biases (as associative mechanisms) operate. It is hard to see how the participants in the streetwise study could

[217] Saul, "Implicit Bias", 55. Emphasis mine
[218] Ibid.

have known, let alone should have known, their decision was potentially biased enough to have taken appropriate countermeasures. They may have made a biased decision, but they also had no necessary reason to think their choice was biased. Perhaps we could argue that they should have known due to the vast amounts of recent literature on the phenomena. If not, perhaps some sort of self-monitoring should have picked up the influence in their behaviour. However, in both cases, if these sorts of epistemic conditions were made general conditions for determining responsibility-aptness, then these conditions would be implausibly hard to satisfy. The average person cannot be expected to keep abreast of recent psychological findings and implement these findings in her daily life. Likewise, to notice one's behaviour in the manner suggested would require tremendous inward focus that does not apply to people generally. Further, it is not clear whether the participants in the study should have even thought to be this diligent in the first place. As Levy argues, if one is ignorant that their actions are wrong, the agent lacks any "(internal) reason to manage them differently."[219] Responsibility attribution could be called into question given that it is unlikely that the participants had any reason to initially think their choice could be biased and modify their decision making processes accordingly.

Yet, despite the inability to control the expression of the attitude due to a reasonable lack of awareness, I would hold that the participants in the streetwise case are in fact responsibility-apt for their biased decisions. In the next section, I will show that being unaware does not necessarily undermine the aptness of responsibility attribution. I

[219] Levy, Neil. "Culpable Ignorance and Moral Responsibility: A Reply to Fitzpatrick." *Ethics* 119, no. 4 (2009): 737.

will defend the view that evaluative sensitivity alone provides a prima facie reason to think that agents are responsibility-apt.

1(b). Grounding Attitudes

It seems to me that the awareness objection is one of deep attributability. That is, even if the biased decision could be acknowledged as sexist, it would be unfair to attribute that sexism to the individual due to insufficient awareness. Yet I do not see why we should resist these sorts of attributions if the sexist attitudes arise from the agent's own evaluative aspects. In this section, I will argue that an accidental sexist *could* still be culpably sexist. Yet, this 'could' is dependent on what implicit biases turn out to be. If implicit biases are like associative mechanisms, they would be more like hunger or thirst than aspects of the self that reflect who we are and hence would not be ruled responsibility-apt through the counterfactual test (to determine answerability). However, even if the agent were not given a reasonable opportunity to curb these biases by being made aware of them, the fact that they have these evaluatively sensitive biases at all shows them to be reflective of his or her evaluative profile and, hence, responsibility-apt. Evaluative sensitivity, not control or awareness, determines whether or not an agent is responsibility-apt for the expression of bias in one's actions.

New research may now support the possibility that such biases might be evaluatively sensitive enough to open the door to a claim to responsibility-aptness. This research has initially indicated that that implicit biases are not just associative states but may be indirectly connected to evaluative attitudes in the same manner as intermediate states (like love or jealousy) discussed in the last chapter. One experimental finding has suggested that the extent to which biases are manifested is influenced by the degree to

which non-prejudiced behaviour is taken as important. [220] It suggests that there are notable differences in the manifestation of implicit biases between persons who see treating others in non-prejudiced ways as a good in itself, and those who see it as important because of possible social sanction. Devine et al. hypothesize that the manifestation of biases and self-reported preferences diverge according to a greater or lesser concern for self-presentation. As Jules Holroyd further explains:

> Individuals who endorsed nonprejudiced behavior for its own sake (e.g., "I attempt to act in nonprejudiced ways toward Black people because it is personally important to me") rather than for instrumental reasons (e.g., "If I acted prejudiced toward Black people, I would be concerned that others would be angry with me") manifested less bias in experimental conditions. [221]

The manifestation of the bias may even be modulated by one's goals. In particular, having the goal of treating people non-prejudicially is important for inhibiting the effects of bias. Holroyd adds that others, such as Gordon B. Moskowitz and Peizhong Li, have argued that the activation of egalitarian goals shields interference from the associations contrary to such a goal, making it less likely that a person's actions would be affected by the bias. [222] Hence, the gathering evidence is starting to show that commitment to egalitarian goals - for its own sake in particular - can modulate the manifestation of bias. Thus, if these studies are correct, there is reason to think that the manifestation of implicit bias may indeed indicate something about the agent's values and who they are in

---

[220] Devine, Patricia G, E. Ashby Plant, David M Amodio, Eddie Harmon-Jones, and Stephanie L Vance. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice." *Journal of Personality and Social Psychology* 82, no. 5 (2002): 835-48.

[221] Holroyd, Jules. "Implicit Bias, Awareness and Imperfect Cognitions." *Consciousness and Cognition*. 33 (2015): 289.

[222] Moskowitz, Gordon B, and Peizhong Li. "Egalitarian Goals Trigger Stereotype Inhibition: A Proactive Form of Stereotype Control." *Journal of Experimental Social Psychology* 47, no. 1 (2011): 103-16.

the world.[223] We might then draw some preliminary conclusions about answerability: it might be intelligible to require an agent who expresses biases to justify her position on egalitarianism as an intrinsic good or ask her to answer for the importance she places on self-presentation. We can attribute responsibility for the bias in the same way I argued that we can do for intermediate states despite the fact that the acquisition and persistence may be non-volitional in the same relevant ways.

If we remember from the last chapter, intermediate states are the kinds of states whose expression is dependent on the evaluative profile being constituted in a specific manner. Likened to perceptions, these states are not necessarily subject to control, but are attributable because they require some sort of evaluative attitude to be experienced at all. Thus, by way of analogy, we might be able to frame such biases as more like a perception of certain evaluative properties over being simply a dopamine-regulated associative mechanism due to some relevant similarities.

For one, as implicit biases are neither chosen nor controlled by sheer will, much of the same can be said of some intermediate states. I often cannot help acquiring jealous feelings. I can put myself in situations to mitigate the influence and choose other indirect means to alter my jealousy, but again this too can be said of implicit biases. Likewise, if

[223] Devine et al. hypothesize that the reason for the variance among people is due to "a subpersonal or automatic inhibitory system … prevent[ing] the influence of negative associations on behavior and judgment" (Devine et. al, "Regulation of Explicit and Implicit", 281). Having the goal of treating people non-prejudicially is important for inhibiting the effects of bias. Holroyd adds that others, such as Gordon B. Moskowitz and Peizhong Li, have argued that activating a goal that is inconsistent with the negative association can inhibit a stereotype though a processes of "goal shielding": the process of inhibiting distractions to the goal. In this case, it may be that the more readied activation of egalitarian goals shields interference from the associations contrary to goals one has. (Moskowitz and Peizhong, "Stereotype Inhibition", 105)

the cited studies are correct and implicit biases are rationally connected to further general beliefs such as the presence or absence of one in egalitarianism, then the counterfactual could show they could be responsibility-apt on the basis of indirect answerability.[224] That is, had the agent had the relevant beliefs in egalitarianism, the bias would not have manifested. The bias would satisfy the counterfactual test in the same way intermediate states do.

If this rough analogy between biases and intermediate states holds, then persons could be answerable, not for holding the bias, but for not giving sufficient weight to egalitarian beliefs, the lack of which the biases stem from. In these cases, collections of different attitudes and beliefs may be understood as racist, sexist, homophobic or any other systemic form of discrimination when taken together, which means that the agent may not explicitly hold discriminatory beliefs in order to be culpable. These may be passively absorbed from living in a system of discrimination and the agent might not even know that the description applies to her collective attitudes and beliefs. The agent may even explicitly disavow the discrimination in most circumstances and would otherwise take steps to mitigate these affects. Yet, this lack of awareness does not

[224] The argument from the last chapter extends to other objections as well. Some may point to the fact because implicit biases exclude the option to do otherwise, even if they are evaluatively sensitive, they still would not be deeply attributable. Consider again Shoemaker's claim from the last chapter that answerability does not hold if there are no clear 'instead of' reasons. As I argued, the fact that we do not have 'instead of reasons' at our disposal is not what is important. What is important is *why* we don't have them. If the fact that there are no 'instead of reasons' is due to one's evaluative constitution, then this grounds a claim to answerability. If implicit biases operate connected to other attitudes (as has been suggested) there is room to consider them evaluatively sensitive.

exculpate due to the fact that these attitudes are "genuinely her judgments" as long as they are evaluatively sensitive. [225]

We can think of a prejudicial reaction, such as clutching a purse when seeing an African American, as indicative of the values the agent holds or does not hold even if the clutch was inadvertent and seemingly innocuous to the agent at the time. This is because the clutching of the purse could signify either a low valuation of treating a person according to merit over stereotype, which licenses a claim about 'who the person is'. Even if the agent did not know their action could be understood under this discriminatory description, the fact that the biases are expressed at all says something about the agent's perspective in the world.

2. The Integration Objection

As shown in the last section, I accept the potentially controversial point that actions influenced by implicit biases may be responsibility-apt on the basis of evaluative sensitivity alone. At least, they are not excluded from being responsibility-apt just in virtue of being implicit or due to the agent's lack of awareness of them. The second objection I will explore suggests that certain evaluative aspects of the self are not deeply attributable (or less attributable) because these biases are not well integrated with the agent's more settled, all things considered, valuations. This is the *integration objection*.

Studies such as the one cited in the last section have led Levy and others to change their hypothesis about the nature of implicit biases.[226] Levy suggests that, although there

[225] Smith, "Responsibility for Attitudes", 125.
[226] Eric Mandelbaum has further questioned the assumption of implicit biases as insensitive to judgment. His canvass of the literature shows these biases to have some sort of propositional structure we would not expect if the biases were entirely

are some propositional content to these biases, they still do not operate like every day beliefs. According to more recent interpretations, implicit biases may no longer be mere associative mechanisms, insensitive to judgments made by the agent, but are now "patchy endorsements" that straddle the distinction between belief and association.[227] In Levy's words:

The evidence seems to indicate that they are *sui generis* states for which we lack any term in our folk psychological vocabulary. I dub them "patchy endorsements". They are endorsements, because they have some propositional structure, which entails that they

---

associative. For instance, if the bias was purely associative, alteration of them should not involve some level of inference and yet, some associations seem to change through methods other than counterconditioning and extinction. Mandelbaum gives examples of shifts in apparent associations due to further information that would involve an inferential story to explain the change. In one case, the individual's associative mechanisms seemed to reason "the enemy of my enemy is my friend" and shift negative associations depending on this rather complex inference of a positive association due to a relationship with another. As Mandelbaum explains, "…just as you might consciously reason that you should probably like those that Hitler hates and dislike those that Hitler likes, so too it appears that we unconsciously reason this way" (Mandelbaum, "Attitude, Inference, Association", 640). Another interesting case Mandelbaum explores involves the notion of celebrity 'contagion'. Subjects noted a lesser value to an item owned by a celebrity (George Clooney in this case) if that item had been laundered. Mendelbaum continues, "The point to keep your eye on is that a merely associative account cannot explain these types of effects. For example, it's not as if people have strong negative associations with hygiene which could swamp the positive association with Clooneyness. Rather, what's transpiring is that subjects appear to have some propositional state that expresses that the article of clothing contains Clooneyish material. This state is then inferentially promiscuous—it can interact with other knowledge stores in inferential ways. In particular, in this case the subjects' knowledge of what washing entails (e.g., it disinfects clothes) interacts with this propositional state to cause the subject to infer that the Clooney essence will be eliminated if the sweater is washed." (Mandelbaum, "Against Alief", 206). By no means do these studies show that implicit biases are indeed sensitive in this very manner. It is merely suggestive of the idea that these sorts of attitudes may not be as semantically insular as initially supposed. As a result, there may be more at play than the passive absorption of the prevalent sexist and racist attitudes in the formation of implicit biases. See Mandelbaum, Eric. "Attitude, Inference, Association: on the Propositional Structure of Implicit Bias." *Noûs*. 50.3 (2016) and Mandelbaum, Eric. "Against Alief." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*165, no. 1 (2013).

[227] Levy, "Fish nor Fowl", 816.

have satisfaction conditions, so that by tokening them agents are committed to the world being a certain way. But they are patchy: they feature in only some of the kinds of inferences and respond to only some of the kinds of evidence we expect from bona fide beliefs with the same kinds of contents (and they also are sensitive to and respond to representations in ways that beliefs do not). [228]

Levy argues that "existing moral concepts apply relatively poorly to people who harbor such attitudes and to the actions that they cause."[229] Yet he is also cautious not to fully excuse persons for them either. Levy argues that because implicit biases function unlike any state in our folk psychology "…we should hesitate before we blame, or feel shame, or guilt".[230] Implicit attitudes are not clearly propositional in a manner that usually attracts attributions of responsibility but neither should we rule out moral responsibility for them.

I think Levy is right in his assessment of the consequences of patchy endorsements, but only if attributing moral responsibility requires the self to be integrated in a way that patchiness fails to satisfy. In the following, I will argue that the sort of patchiness described by Levy in the case of implicit biases is also true of many other aspects of the moral self. In the case of patchy endorsements, we could, as Levy argues, revise traditional moral concepts to make sense of the patchy nature of these aspects of the moral self. Alternatively, we could keep the traditional concepts and argue that implicit biases fall outside normal responsibility attributions. I will choose a third option. I will start with the notion of a conflicted and patchy self and explain how our traditional moral concepts apply to this revised concept of the self. Patchy endorsements are not atypical enough to require revised moral concepts. For the remainder of this

[228] Ibid., 816.
[229] Ibid.
[230] Ibid., 817.

chapter, I will argue that the force of Levy's objection derives from a very specific (and unconvincing) picture of the moral self as semantically and evaluatively unified.

2(a). From Awareness to Integration

On Levy's terms, the participants in the study could not be responsible for a biased decision if they did not know that decision was biased because it does not represent their general normative stance. Levy argues that for an action to truly express one's normative stance, it needs to express her "global perspective on what matters morally, and not a single attitude, or a set of attitudes that falls short of constituting the agent's evaluative stance." [231] He argues that having a single controversial attitude inadvertently manifest itself in one's actions does not express the agent's evaluative stance 'in the right way' because it does not constitute attitudes that have been properly filtered through the agent's larger moral beliefs. This would mean that evaluative sensitivity alone does not determine deep attributability because the attitude may be "crucially at odds with the states with which we can most securely identify the agent." [232] These biases are as a result "*too alien* to the self to ground moral responsibility." [233] Because the expression of the bias is not tempered by the agent's larger evaluative stance and this lack of awareness implies a lack of semantic integration, Levy calls into question responsibility attribution for such biases.[234]

[231] Levy, Neil. "Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action." *Analytic Philosophy*. 52.4 (2011): 256.
[232] Levy, "Probing the Data", 14.
[233] Ibid., 14. Emphasis mine
[234] Although Levy does not explicitly define what he means by such semantic integration, I take it to be a kind of necessary connectedness to the whole, such as how a word in a sentence is connected to the rest of the sentence.

Levy acknowledges that divergence from one's explicit beliefs cannot be more than a heuristic because such conflicting attitudes may be deeply embedded in other "semantic relations to many other attitudes which themselves are plainly attributable to the agent" and might even be "inferentially linked to many of the *same* attitudes." [235] A conflicting attitude can be deeply integrated in the agent's belief system in ways patchy endorsements are not. So although conflict is telling, it is not necessarily the driving force behind denying responsibility-aptness in these cases. What matters for Levy instead is not whether an attitude conflicts with one's explicit beliefs, but whether it is deeply embedded within one's mental economy.

Awareness generates greater interaction and embeddedness even if the attitude remains in conflict with one's larger evaluative stance. A patchy endorsement is not one's own because it functions like an accidental mental typo that would have otherwise been edited out upon a more thorough review. It may occur multiple times, but sporadically enough that it can only be described as "patchy" at best, exerting itself without reason and without being embedded in a system of inferential relations. In Levy's sense, a bias *could* become legitimately attributable if it was eventually integrated and embedded.[236] A consistent attitude on the other hand may more clearly represent the agent's evaluative stance. Indeed, as Levy states:

> A set of attitudes must be relatively consistent to constitute a stance: a stance consists of a set of mutually supporting attitudes. Real agents do not have perfectly coherent sets of attitudes; rather, they possess a number of attitudes, some of which are in conflict with their stance. Since a stance must be relatively consistent, some of an agent's attitudes do not belong to his or her evaluative stance.[237]

[235] Ibid., 11.
[236] Levy,"Fish nor Fowl", 816.
[237] Ibid., 257.

On Levy's account, even if an attitude were evalutively sensitive, persons would not be responsibility-apt due to the attitude's lack of integration. In what follows, I will argue, contra Levy, that less integrated aspects are nevertheless important for responsibility attributions on the grounds that there is little reason to think such traits are not deeply attributable to the individual. More specifically, I take issue with denying attributability on the basis of a lack of consistency or integration. Levy suggests that the onus is on the attribution theorist to provide "a story to explain why we ought to identify [less integrated aspects] with the self sufficiently strongly to ground moral responsibility in these cases" of conflict and fragmentation.[238] I will argue the opposite. The onus, rather than being on the attribution theorist, falls on those who would deny attributability in such cases as there is little reason to think that such partial evaluative beliefs are any less the agent's own than other beliefs. The wide scope of responsibility attributions mirrors the inconsistency of the evaluative profile in general. It is not just the endorsements that are patchy, but also the general constitution of the self.

2(b). The Whole Self

In this section, I will explore the possibility of a "patchy" self by analyzing an argument in favour of integration.[239] In particular, I will look at Nomy Arpaly and Timothy Schroeder's concept of the *whole self view* to see if there is sufficient reason to require evaluative aspects such as attitudes, commitments, or cares to be either semantically or evaluatively unified with other states to be responsibility-apt. Following this, I will try to provide a reasonable alternative to the whole self view by providing

[238] Levy, "Probing the Data",14.
[239] Levy, "Fish nor Fowl", 816.

another explanation for our intuitions regarding integration, blame and the semantic constitution of the self. In the end, while the arguments do not refute the whole self theory, I hope they will at least establish some reasonable doubt as whether we should deny responsibility-aptness on the basis of lack of integration.

The whole self view defended by Arpaly and Schroder amounts to the claim that:

> …other things being equal, agents are praiseworthy(or blameworthy) for the good (or ill) they do to the extent that the morally relevant beliefs and desires which led them to act were well-integrated (assuming that the act met some very minimal standard of integration).[240]

Despite my disagreements with Shoemaker in the last chapter, we would converge in our opposition to such whole self theories. He writes that as long as one's attitudes "express at least one care or some aspect of one's evaluative stance, are attributable to one, even if they are 'shallow' and even if they conflict with the rest of one's cares or commitments."[241] A begrudging misanthrope, not unlike the one seen in Kant's *Groundwork*, may even deserve acclaim for acting on a morally praiseworthy whim despite not having clear motivations to do so.[242] Further still, he notes that integration may even result from habit or laziness on the part of the agent and consequently it is not a clear basis for praise or blame. He states:

> I have already suggested that sometimes a psychic element's being *less* integrated with one's screwed-up character is cause for greater praise when it moves one. But it also seems that some attitudes could just be well integrated out of long-standing habit or laziness, and it is difficult to see their integration in that case as the source of any aretaic praise or blame we might muster.[243]

---

[240] Ibid., 175
[241] Shoemaker, *Margins*, 137.
[242] Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary J Gregor and Jens Timmermann. (Cambridge: Cambridge University Press, 2012).
[243] Shomaker. *Margins*, 137.

I think that Shoemaker is right in saying that shallow aspects of the self are nevertheless attributable. However, I believe the problem with the view that integration is necessary for attributability goes deeper than the potential issues concerning *how* those aspects have been integrated. I would argue that it does not necessarily matter whether an aspect of the self was integrated due to habit or laziness as long as it was evaluatively sensitive. So although Shoemaker may be right in rejecting whole self theories, I do not think they should be rejected for his reasons. In the following, I would like to consider the cogency of Shoemaker's objection to such theories before turning to my own. I hope that the process of analyzing his claim, rather than bolstering the whole self view, will actually highlight the reasons I think it is mistaken.

Shoemaker is correct that *how* aspects of the self are integrated may be more or less a product of one's agency. This fact, however, is not necessarily a problem for whole self theories. For Arpaly and Schroder, how the attitude gets integrated is not necessarily important. Consider their example of two kleptomaniacs, Lana and Greg. Lana, despite deeply opposing stealing, often finds herself unable to control larcenous desires and shoplifts periodically.[244] Arpaly and Schroeder argue that:

While her thefts may stem from an underlying psychological distress involving doubts of self-worth or related issues, and hence be more integrated than they seem at first blush,

[244] In both cases, my view would first ask of evaluative sensitivity of the experienced urges and ask if Lana or Greg's evaluative beliefs had been different, then would they still steal? If so, then the theft is not attributable but a mania that functions like a phobia as discussed in the last chapter. If it is evaluatively sensitive, then it is attributable and they are responsible for their larcenous urges. On my account, evaluative sensitivity determines the sphere of what we can be responsible for, but importantly it does not determine the level of blameworthiness. Instead, (as I will suggest in chapter eight) blameworthiness and praiseworthiness has more to do with the sorts of social relationships we find ourselves in rather than the constitution of the person.

even so the discord between her urge and her other beliefs and desires is pervasive enough to substantially reduce integration. [245]

Lana's thefts are excused, not because they conflicted with her explicit desires not to steal, but because "the action appears poorly-integrated" with her overall personality.[246] This stands in contrast to Greg who steals "in spite of himself" but still experiences some sort of joyful excitement when he does, due to his general love of thrill seeking.[247] As the desire to steal is integrated in his love for risk taking, Greg is more blameworthy than Lana on the whole self view. So whether or not the desire was born out of habit, laziness, mania or thoughtful acquisition is beside the point.

Shoemaker also argues that, "integration, which is just about the relationship between psychic elements, has no obvious connection to mattering." [248] While I believe he is right again in that there is no *clear* connection between integration and mattering or responsibility, this does not mean one cannot be made. Indeed, I claim with Levy, Arpaly and Schroeder that integration may be usefully connected to notions of praise and blame, but I offer different reasons for the connection. Integration helps determine whether *continuing* blame is appropriate and not whether blame in general is appropriate. Whether an aspect of the self is integrated may matter to moral responsibility by way of what it says about the fittingness of continuing blame or ultimately forgiving.

In what follows I show how integration may be useful in determining the temporal extent of blaming activity, not responsibility-aptness. To show this I would like to first reposition where the conditions of integration and awareness might be most relevant to

[245] Arpaly & Schroeder, "Whole Self", 176.
[246] Ibid.
[247] Ibid.
[248] Shoemaker, *Margins*, 136.

determining responsibility. It is not that they have no connection to mattering as Shoemaker proposes, but they matter in a different way than suggested in these two objections. [249]

2(c). The Moral Responsibility Exchange

Even though I agree with Shoemaker that integration is not clearly connected to mattering, my rejection of conditions like integration and awareness (and control and choice by extension for that matter), does not necessarily mean they are not important in determinations of responsibility more generally.[250] Before I discuss how each is connected to blaming, I want to be clear on the kind of work such conditions are actually doing when attributing responsibility. I will argue that these sorts of conditions do not offer a means for an excuse when determining whether one is responsibility-apt. Integration and awareness conditions may be sufficient for attributability, but the presence or absence of them is most relevant to responsibility assessments in their capacity to justify an agent's actions. The following section will then connect this refined placement of these conditions to blaming attitudes and set us back on track with responses to Shoemaker.

First, I would like to suggest that we understand the terms 'excuse' and 'justification' with reference to Michael McKenna's notion of the "moral responsibility

[249] Ibid.
[250] I see this argument for the placement of awareness and integration conditions as justifying condition as equally applying to control and identification even though I will not be giving a treatment of those conditions separately.

exchange" as it is similar to the legal use of these terms.[251] McKenna sees the process of attributing responsibility as a kind of conversation that determines how the offender will be treated thereafter. Like any conversation there is a give and take that is analogous to the stages that occur before holding a person responsible. The way he defines each stage differs from my account but serves the same function, so I will borrow his titles for each.

I see the first stage – that he calls "Moral Contribution" – as defining whether the attitude or action is responsibility-apt.[252] We can see the agent as "introducing, or risking the possibility of introducing, a meaningful contribution that is a candidate" in a moral responsibility exchange.[253] When an excuse is offered the attributability of the act is denied. Legal excuses can range from diminished capacity including insanity, drunkenness, or automatism. In each case, the psychological impairment halts deep attribution in the first stage of the moral exchange. They are unable to answer for these due to an external impediment (located in the motivational profile) on their behavior. The request for a response can only be properly given at this stage if the contribution was evaluatively sensitive and, hence, open to answerability demands.

In contrast to an excuse, when a justification obtains, it does not entirely exculpate, but provides merely a mitigating factor in how we treat the person if responsible. Justifications do not deny attributability, but are instead the potential answers that can be given when one is properly answerable. This corresponds to the second stage or the

[251] Mckenna, Michael "Directed Blame and Conversation." In *Blame: Its Nature and Norms*. Edited by Justin D. Coates and Neal A Tognazzini. (Oxford University Press, 2012)
[252] Mckenna, *"Directed Blame"*, 128.
[253] Ibid.

stage concerning "Moral Address." [254] The agent may address the charge "by means of offering an excuse, a justification, an apology, and so on." [255] Whether or not this answer is accepted in turn determines the force and fairness of the third more reactionary stage of the exchange – or "Moral Account" - concerning possible responses "say by forgiving, or punishing, or simply ending the exchange and moving on and so on." [256]

An appeal to reasonableness may be one such justification offered in the second stage of this exchange. We might excuse if the agent could not do otherwise, whereas an action could be justified if the agent could *not be expected* to do otherwise. In acting otherwise, the agent would have to assume an unreasonable and substantial burden in complying with the norms set out. Similar to the legal fiction of the *reasonable person*, whether it is fair to punish is responsive to considering how a reasonable person would act given the same limitations and under the same circumstances.

Drawing from the example from chapter three, we see that Alessandra's forgetting about Sheba could be justified in this manner. Even if it turned out that her forgetting Sheba was evaluatively sensitive in some manner and not a glitch in her psychophysical system, this does not mean that she would automatically be held responsible for her forgetfulness, it may be the case that her situation is analogous to the perfect storm that seemingly describes the situation of many new parents when forgetting a child in the car. One's circumstances thus may render the mistake tragically reasonable even if not excusable. As experts are now warning parents, "Any person is capable of forgetting a child in a car under circumstances where a parent is going through a routine and the child

[254] Ibid.
[255] Ibid.
[256] Ibid.

is in the back."[257] As the ubiquity of the situation and the circumstances in which it occurred can attest, it may not be fair, pragmatically or morally to hold some parents responsible for these lapses even if they were deeply attributable to them. Being responsibility-apt does not mean being justifiably blamed or punished. The circumstances that led to Alessandra forgetting Sheba may be generally understandable insofar as most people would do the same if they were put in her shoes. If, however, Alessandra was held up for reasons that were not as understandable (such as chit-chatting with a friend or something else that did not take as large a cognitive load) we may not be as keen to accept the justification and she would be reasonably open to the reactive attitudes and social sanctions that may follow.

In cases like Alessandra, the degree to which the situation is reasonable offers one possible answer in the second stage of the moral responsibility exchange with implications on the third, but one that does not undermine the appropriateness of the answerability demand in the first. It offers a justification and not necessarily an excuse. The same can be said of the phenomena of implicit biases and like intermediate states that may be evaluatively sensitive, but operate without awareness. In order to justify one's biased actions, we need to look outwards to consider the circumstances and not just the agent and her constitution. For instance, a justification consisting of a lack of awareness could justify an act if it is reasonable that the agent did not know a decision was biased. It could be argued that a reasonable person in similar circumstances would not have known their actions were possibly biased. Obscurity and difficulty in obtaining

[257] Ibid.

this knowledge justifies insofar as it would be unreasonable *to expect them to have known* by keeping abreast of scientific findings or through introspection.

Likewise, the fact that a bias is poorly integrated with one's usual evaluative stance may justify as well. Given the situation and the psychological state of the agent, we might think it understandable that this lesser part of the person was expressed. But, just because it is understandable does not make it any less open to responsibility attributions. The biases may be reflective of genuine attitudes held by the agent, but ones that are responsibility-apt even if holding these attitudes could potentially be justified.[258] Being unable to exert something like Nagel's "monumental effort of the will" in order to suppress a held attitude does not necessarily excuse, but a monumental effort as opposed to an apathetic attempt may justify one's culpable actions.[259]

So the effort of the will does not mitigate responsibility aptness, but can justify one's actions in a way that should undermine the application of the third stage of the moral responsibility exchange. I would argue that the idea that such conditions excuse is primarily due to the fact that when they are satisfied, not only do they offer justifications for the offender, but may also lessen some epistemological worries in the process, which makes them intuitively important in attributing responsibility. For instance, if it were the case that an agent chose to do something (with full awareness) in accordance with their second order volitions (with no impediments) and that this was overall well integrated with the agent's usual personality, then this scenario would be to attributability as highly controlled conditions would be to a science experiment. Making the inference to what is

[258] Smith, "Responsibility for Attitudes", 125.
[259] Nagel, *Mortal Questions,* 34.

contained within one's evaluative profile is clear cut in such instances and unhampered by contingencies of circumstance. All of the variables are removed and we have clear sight of "who the person is' without much speculation. However, even if the action is more clearly attributable when awareness and integration conditions are satisfied, I maintain that the initial attribution is nevertheless satisfied by evaluative sensitivity alone.

2(d). Integration and Blame

Positioning conditions like integration as having application in the second stage of the moral responsibility exchange helps elucidate the importance connecting such conditions have with blame. Blame does have a "connection to mattering" that Shoemaker denied, but not within the first stage of the moral responsibility exchange as it is better associated with the third. [260] In this section, I will show that integration has application (as a potential answer in the second stage of the exchange) by potentially mitigating the force of blame (undermining the application of this third stage), but not necessarily excusing the initial attribution (as in the first stage). To show this I would like to briefly discuss the phenomena of blame that will receive a more thorough treatment in chapter eight.

I find T. M Scanlon's account of blame instructive.[261] For Scanlon, blame is deeply interpersonal and responsive to the kinds of relationships persons find themselves in. Relationships, he argues, can be understood as "a set of intentions and expectations about

[260] Shoemaker, *Margins*, 136.
[261] In chapter eight, I will elaborate on this account, but at this point I only want to make a brief mention to substantiate the claim that responsibility and blame are distinct. This bare bones description will receive a more substantial treatment as we continue.

our actions and attitudes toward one another that are justified by certain facts about us."[262] Judgements of ill-will cause the blamer to alter the relationship with the one blamed. Likewise, according to Miranda Fricker's notion of communicative blame, we blame in order to communicate the feeling of being slighted to the wrongdoer and jolt them into engaging in repair of the altered relationship.[263] If the blame achieves this purpose, then it is withdrawn. So, the question of whether blaming responses are appropriate is different from that of responsibility-aptness because they track different conditions: the moral self for responsibility-aptness and social transgression for blame. Blame may be retracted even if responsibility as answerability still holds. This is an argument I will revisit more fully later on. For now, I argue that integration does not call into question attributability, but may instead help indicate whether continued blame is warranted.[264] Once we see blame as a social matter the connection to integration is clearer.

---

[262] Scanlon, Thomas. *What We Owe to Each Other*. (Cambridge, Mass: Belknap Press of Harvard University Press, 1998): 87.

[263] See Miranda Fricker "What's the Point of Blame? A Paradigm Based Explanation" *Nous*. 50.1 (2016).

[264] Matthew Talbert argues for something relatively similar. He separates being blameworthy and the aptness of blaming responses. He argues, "being blameworthy, being an appropriate target of blaming responses—doesn't depend on whether an agent is causally responsible for his faults. This is because (on my view) blaming responses like resentment are largely means of marking and protesting a wrongdoer's objectionable evaluative judgments. So, if facts about how a wrongdoer came to be the way he is do not call into question the moral status of the judgments that inform his behavior, or the attributability of these judgments to him, then they do not call into question the aptness of blaming responses" (Talbert, "'Moral Competence" 55) Actual praise and blame may be usefully distinct from being blameworthy. See Talbert, Matthew. "Akrasia, Awareness and Blameworthiness" In *Responsibility: The Epistemic Condition,* edited by Philip Robichaud, Jan Willem Wieland (Oxford: Oxford University Press, 2017): 47-63.

Returning to the example of the kleptomaniacs, if it is true that the urge to steal is not a pervasive motive, this offers justification that could potentially forestall application of the third stage of the moral responsibility exchange. If a motivation is not well integrated it is unlikely that the agent will commit the action again or at least they are likely to take steps to avoid the same actions. There would be little reason to alter or suspend the relationship that was damaged by the transgression due to its isolated nature. That is, not only might we assume that a well-integrated evaluative aspect is unlikely to fade into indifference, but when it is retained the agent carries a number of other evaluative aspects that would seem to indicate that they approach the world in a manner similar to before and thus give reasons to continue the relationship as it once was. Consequently, integration seems better understood as a heuristic to help predict whether the action will be committed again, which in turn determines whether to continue to blame as the third stage of McKenna's moral responsibility exchange.

Taken together, we see that evaluative sensitivity determines whether an attitude is responsibility-apt (Moral Contribution), while integration or awareness may function as something like a justification (Moral Address) that will help determine whether or not we blame (Moral Account). Integration in particular may even be responsive to judgments of blame or praise by helping to determine responsibility over time due to its predictive function. Not only might we assume that a well-integrated evaluative aspect is unlikely to fade into indifference, but also that the agent's moral self contains related evaluative aspects that would seem to indicate that they will approach the world in a manner similar

to before.[265] We then have the connection to praise and blame Shoemaker argued was absent, but it is satisfied in a manner that does not accept integration as determining what is and what is not responsibility-apt.

2(e). The Conflicted Self

How well integrated the self is does not ground responsibility claims. Here one might object, citing rationally sensitive whims and lapses as accidental, transient and independent of the normal functioning of the moral self. I could suddenly desire on a whim to jettison my thesis draft. This would not only be imprudent, but also uncharacteristic. I may also inadvertently laugh at an off-colour joke for which, as Smith argues, "it seems not only morally objectionable to *express* amusement; it seems morally objectionable, and blameworthy, to *be* amused."[266] I may argue that it is generally not like me to laugh at such jokes and I usually do not. The inadvertent laugh does not cohere with my more settled and integrated disposition and, surely, I would not want to be identified with a lesser part of myself. Likewise, as with implicit biases, how can it be that we are responsible for what we would otherwise disavow or prevent if we were able to? This response depends on how we understand the normal functioning of the moral self. We can either see this self as essentially unified with less-integrated aspects at the margins or we can start with the notion of a 'fragmented' self and see unification as an ideal to strive for. I will choose the latter and ask: "If the moral self is conflicted and non-integrated why should lesser integrated aspects be *less* attributable?"

[265] In chapter six, I will address the importance of being similar as grounding continued responsibility attributions. For now, it is simply important to see how integration might allow us to assume that similarity.
[266] Smith, "Attitudes, Tracing, and Control", 118.

Whole self views assume that normal functioning, as a standard in which to compare what is and what is not attributable, corresponds to an ideally rational, coherent, smoothly functioning agent. I would like to use John Perry's description of Spock as an example. He says of this regular from Star Trek:

> When Mr. Spock is faced with a decision, he deliberates, taking into account all of the goals he has and all that he believes. His desires are ordered by their importance; his beliefs by his degree of confidence in them, and that degree of confidence corresponds to the evidence he has for them. He rationally computes what the best thing to do is, that is, the thing which has the optimal chance of promoting his most important goals, given the beliefs in which he is most confident. [267]

Dr. Spock, with all actions filtering through his more considered normative beliefs, might represent an ideal of integration. Yet, this ideal is not descriptively accurate as to apply to everyday circumstances.

Perhaps the example of Spock is an exaggeration, and whole self theories are not advocating integration as an ideal, but simply noting that when something is integrated it is more attributable. Yet, if being responsibility-apt would only follow if the action were connected to our evaluative aspects due to such integration, I would contend that a great many of our actions would resist responsibility attributions because, descriptively speaking, the self is never fully integrated as a coherent whole. As Perry writes, "My goals and beliefs combine into clusters, often with many common elements, that vie for control of my various systems of effectors. Victory is seldom complete."[268] I would like to concentrate on Perry's notion of "clusters of elements that vie for control" in that I see

[267] Perry, John. "Selves and Self-Concepts." In *Time and Identity*, edited by Joseph K. Campbell, Michael O'Rourke, and Harry Silverstein. (Cambridge, Mass: MIT Press, 2010): 245-246.
[268] Ibid,. 246.

it as entirely possible that, contra Levy, not only are we unlikely to be evaluatively unified, we also may not be semantically unified either.

Christopher Cherniak argues that the kind of idealized rationality exemplified by Spock is decidedly not the norm. He questions whether it should even be considered a normative ideal. He writes, "[I]deal rationality conditions abstract from a fundamental fact of human existence: we are in the finite predicament of having fixed limits on out cognitive resources, in particular, on memory capacity and computing time." [269] As finite beings, it would be too much to think that agents could eliminate all the inconsistencies in their beliefs, let alone believe all the consequences of their beliefs. He continues, "the web of belief is not merely tangled; its fabric of sentences is 'quilted' into a patchwork of relatively independent subsystems. Connections are less likely to be made between these subsets." [270] Considering the non-ideal ways humans interact in the world, it is not a stretch to think that inconsistencies are abundant and perhaps the norm. Why then should the standard of rationality pertain to a world and circumstances persons to not even inhabit? [271]

Indeed, Eric Mendlebaum and Andy Egan also acknowledge the possibility that there may be more than one web of beliefs.[272] The beliefs that characterize the self may

[269] Cherniak, Christopher. *Minimal Rationality*. (Cambridge: MIT Press, 1986): 77.
[270] Ibid., 51.
[271] The evaluative profile involves more evaluative psychological aspects other than beliefs, yet I will concentrate on beliefs here. Beliefs not only form an integral aspect of the evaluative profile, but of all the psychological elements, they best illustrate what I mean by having various webs of evaluative aspects (as analogous to the webs of beliefs here).
[272] The metaphor using 'webs of beliefs' is derived from W.V Quine in the context of epistemic justification. (See Quine, W. V, and J. S Ullian. *The Web of Belief*. 2nd ed. New York: McGraw-Hill, 1978).

be "fragmented" in the sense that some beliefs may be "causally isolated from other beliefs." [273] Mendlebaum continues: "the picture that emerges is one where we cannot, strictly speaking, talk of a person's single stock of beliefs. Rather, each believer will have multiple, synchronously encapsulated webs of belief, but no single overriding web…" [274] Egan tells us, beliefs may be "fragmented" or "compartmentalized" as a means of explaining the sorts of inconsistent beliefs and goals a person may have.[275] So, rather than having a single system or Quinean 'web' of interconnected beliefs to guide "all of our behavior all of the time, we have a number of distinct, compartmentalized systems of belief; different ones of which drive different aspects of our behavior in different contexts." [276]

---

[273] Mandelbaum, Eric. "Attitude, Inference, Association: on the Propositional Structure of Implicit Bias." *Noûs*. 50.3 (2016): 650.

[274] Others have even questioned consistency in the kinds of beliefs one holds and have allowed for the possibility of believing both A and not A in some circumstances. Andrew Huddleston has argued in favour of the possibility of "naughty beliefs" that show these divergent attitudes are not necessarily a different category of psychic elements, but regular old beliefs that are simply behaving badly. (Huddleston, "Naughty Beliefs", 209) Rather than a well mannered adult as beliefs are often pictured to be, they behave more like " a petulant toddler throwing a tantrum, these naughty beliefs, so to speak, put their fingers in their ears and chant, "La La La. I can't hear you…" (Ibid.,218). The "naughty belief" is more recalcitrant than beliefs ought to ideally be (Ibid., 209). He questions creating a separate category for these rebellious beliefs because they may be just as semantically embedded and part of the agent's mental economy as any other belief would be. These naughty psychological states are in the game of belief so to speak, but break the rules to some extent. In any case, whether or not this is a true belief or whether it can be rightly called one is not the point here. Rather, the contradictory nature of such beliefs highlights the fact of diversity in one's evaluative profile. See Huddleston, Andrew. "Naughty Beliefs." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*160, no. 2 (2012).

[275] Egan, Andy. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*140, no. 1 (2008): 48.

[276] Ibid.

Jennifer Radden argues that conflict and fragmentation, far from being a problematic state, is natural to commonplace experience. It is not only a conflict between webs of beliefs, but between the various psychological states as well that contribute to such fragmentation. She states:

> Some assumptions, beliefs, values, and desires are in conflict, and to recognize this is normal. Perhaps it is human too: a degree of harmony so great as to preclude all ambivalence, if such a state is imaginable, suggests a different style of intelligence perhaps of divine or artificial origin.[277]

Aside from being a god or at least a rational alien like Spock, ambivalence, fragmentation and semantic disarray are common to human experience. If it is both human and common, then it is not clear why the presence of these features should necessarily undermine deep attributability. She continues, the "heterogeneities described here are neither puzzling nor strange. Indeed, some of them seem so central to our notion of a human self that we would question whether a more unified self than they allow were truly human." [278]

So if one's beliefs, webs of beliefs and psychological states are not necessarily unified, is it not possible that when these are in conflict and less integrated that they still deserve the classification as aspects of the moral self? Moreover, as the evaluative profile contains cares, commitments and other valuations, might the moral self *itself* be as patchy as these endorsements? As Perry elaborates:

> Suppose I'm at a department meeting. Someone says something, which I interpret as a put down. I become angry. But part of me, an inner voice urges restraint…It has no control of an important part of my agency, that part of it that controls speech. Another cognitive cluster is in charge of that. Some goals that are quite important to me, like not

[277] Radden, Jennifer. *Divided Minds and Successive Selves: Ethical Issues in Disorders of Identity and Personality*. (Cambridge, Mass.: MIT Press, 1996): 16.
[278] Ibid, 23.

appearing foolish, not alienating colleagues, not saying things that will be counterproductive to the deliberations of the department, play no role in this second cluster, which has seized control of my mouth. They are present; they are me, or so I would have thought, things like making sure people know how I feel, defending myself from criticism, and the like.[279]

Assuming Perry's picture, would we offer different attributions of responsibility if the cluster that controls the mouth were less integrated than the one urging restraint? Is he less responsibility-apt than if he had been quiet?

One response that could be made on behalf of whole self theories would be to say that there is a minimum threshold of integration that must be met in order for the action to be part of the moral self. For instance, cases of duress or addiction offer excusing conditions on the grounds that the resulting acts are "extremely poorly-integrated" and are, as a result, "better attributed to circumstances … and so be an act for which the actor is not responsible." [280] I would agree that these are not attributable, but if we stop speaking of aspects being more or less one's own and speak of thresholds instead, it is not clear what more a whole self theory can say above what is already shown by evaluative sensitivity. This is because these aspects are poorly integrated precisely because they are not evalutively sensitive. My account and a whole self theory involving thresholds would land on the same conclusion, but only one relies on a tenuous connection between integration and mattering.

The main problem I see with requiring integration for attribution: why should integration be connected to deep attributability? Moreover, why should it matter for responsibility? That is, if we start with the idea that the self is usually in semantic or

[279] Perry, "Selves and Self Concepts", 246.
[280] Arpaly & Scheoder, "Whole self", 176.

evaluative disarray, it seems as if the onus has shifted. Human agents are not like Dr. Spock and we will need a further explanation as to why non-integrated aspects do not belong to the self other than their apparent 'alienness.' Just because an action does not represent who we would most like to be or express itself often in our lives does not by itself mean that it does not reflect who we are in some manner. The self may be composed of clusters of different beliefs, cares, commitments and desires, none of which is in particular control. After throwing my thesis out the window on a whim, it still would nevertheless be appropriate to question me on why I did that. It is intelligible to ask me why I acted on a whim. Barring any explicit mind control, mind-altering substances or extreme exhaustion, the whim is still evaluatively sensitive and, hence, attributable to me, even if it was not well integrated.

3. Conclusion

It may be the case that one may more consistently act on a deeply held evaluative belief or having such a belief may help one favour one course of action over another, but this does not mean what is less consistent, overridden or unchosen is not attributable as a result. An implicitly racist act is still reflective of racist attitudes. I suspect that we generally think actions resulting from less integrated aspects of our constitution intuitively attract less responsibility due to the fact that we generally think ourselves more unified in not just our aims, but in our psychological constitution, than we really are. This is especially true if we take a cue from Perry, Cherniak, Mendlebaum, Egan, and Radden to entertain the idea of a conflicted, fragmented and patchy moral self. This self, as I see it, is a collection of conflicted desires, attitudes and judgements, none of which is more or less attributable. Considerations of responsibility are essentially

concerned with ‘who the person is’, not who they are *mostly*. The onus is on those who would argue otherwise.

What are the conditions of responsibility-aptness over time?

Chapter 5: Change, Alteration and Replacement

Introduction

At this point, the shape and features of the self should be clear enough that I will now shift from exploring what makes persons responsibility-apt *at a time* to what makes them responsibility-apt *over time*. Just as we may question whether a sculpture is the same sculpture when it has gone through a continual process of erosion and restoration, I will address the ways in which the self may degrade and change so as to call responsibility attributions into question. One immediate problem is that the sort of deep attributability I have proposed thus far seems incapable of being applied to diachronic extension. As I argued in chapter two, the self is a "gappy" forensic term that does not necessarily satisfy the persistence conditions normally required for identity over time.[281, 282] Worse, in the previous chapter, I argued that the moral self can be "patchy" in the sense that the evaluative profile that constitutes the self is not always unified either semantically (in terms of connections between psychological criteria) or evaluatively (in terms of patterns of endorsement). [283] As a result, does it even make sense to speak of persistence over time of something so *patchy* in its extension and *gappy* in its constitution?

I will organize this chapter around these two problematic features of the moral self. I will explore whether the features described as gappy and patchy undermine continued answerability. In §1, I will illustrate a couple of ways change may occur through drawing

[281] Strawson. *Consciousness and Concernment*, 8.
[282] I argued that the self as a forensic unit is gappy, but I think the arguments can equally apply to the moral self as well even if I have not explicitly done so.
[283] Levy, "Fish nor Fowl", 816.

on and elaborating from an example from Anthony Burgess's novel *A Clockwork Orange*. [284] In §2 I will look to L.A. Paul's account of transformative experiences as a means to demonstrate what I call *the problem of gappiness*.[285] I will argue that responsibility-aptness holds despite the apparent gaps that occur when one undergoes a transformative experience. The exploration of Paul's work also reveals a preliminary distinction between alteration and replacement within the evaluative profile that will narrow the kinds of changes a person can undergo and still be answerable. This distinction will nevertheless be too broad to be plausible. I will further refine this distinction to argue that alteration signals the persistence of what I call "evaluative access" to preserve answerability over time as a means of qualifying what is meant by the alteration/replacement distinction in the following chapter. In §3 and §4, I explore Marya Schechtman's account of narrative identity and sympathetic identification to illustrate *the problem of patchiness*. For Schechtman, the self may be evaluatively patchy only to a certain extent. She argues that a loss of the self may occur in instances of deep repudiation and shame, yet in §5, I will reinterpret these examples to show them as primary cases of the expansion of the evaluative profile despite repudiation and conflict through the alteration/replacement distinction. As I argue in §6, answerability holds through deep repudiation and shame despite how patchy these aspects may be. Overall, whether the agent is transformed in ways suggested by Paul or fiercely repudiates her past as Schechtman contends, answerability holds through these changes as long as that change represents a modification rather than a destruction of one's former valuations.

[284] Burgess, Anthony. *A Clockwork Orange*. (New York: W.W. Norton, 1986).
[285] Paul, L. A. *Transformative Experience*. (Oxford: Oxford University Press, 2014).

Answerability can extend through both radical transformation and affective breaks with one's past.

1. The Changing "Raskazz"[286]

To begin, I would like to consider the depictions of transformation in the novel *A Clockwork Orange*. The story focuses on the violent and sadistic protagonist, Alex. A teenager engaged in a criminal life, Alex and his comrades make a habit of committing and reveling in violence. This life is brought to a halt after Alex is sent to prison for the murder of an elderly woman. As a way of minimizing his sentence, he is offered an experimental procedure, "Ludovico's technique," that will rehabilitate him in a matter of weeks.[287] It is a form of associative learning in which, while being subjected to movies depicting excessive violence, he is given a drug that causes debilitating nausea. Afterward, his murderous impulses are followed by an intense sickness, which subdues the once aggressive criminal. Any behavioural change in Alex was not accompanied by a change in his evaluative profile. As Alex was humiliated by being forced to grovel at an attendant's feet, he thinks to himself:

> Now I knew that I'd have to be real skorry and get my cut-throat brivita out before this horrible sickness whooshed up… But, O'brothers, as my rooker reached for the brivita in my inside carman I got this like picture in my mind's glazzy of this insulting chelloveck howling for mercy with red red krovy all streaming out of his rot, and hot overtake, and I viddied that I'd have to change the way I thought about this rotten veck…[288]

Indeed, he retains sadistic thoughts, but quells them in fear of the overwhelming sickness. At most, if we use the term 'rehabilitation' loosely, we may be able to say that

[286] This term can be found in the fictional language, "Nadsat", used by Alex and his fellow gang members. The word "raskazz" may be translated to mean "story." See Burgess, *Clockwork*, 164.
[287] Ibid., 30.
[288] Ibid., 93.

Alex is rehabilitated insofar as he cannot engage in the violent acts he once enjoyed. Yet, while Ludovico's technique is operative, at most we see a criminal with his hands metaphorically tied, whereas who he is and who he conceives himself to be intuitively remains the same. After all, it seems that if it were not for the extreme bouts of sickness, Alex would not desist from criminal behaviour. This lack of change is vividly portrayed in Stanley Kubrick's film adaptation.[289] After the technique is reversed, the closing scene depicts Alex resuming his former fantasies and menacingly telling the audience, "I was cured alright."[290] Let us refer to this version of Alex as *Movie Alex*.

At the end of the novel in the original United Kingdom edition, there seems to be an internally motivated change in Alex once the technique was reversed. Alex notices that, not only is there a change in his aesthetic tastes, but he finds himself yearning for a domestic life. He begins to notice slight, but no less profound changes in himself. He is bored of his violent, criminal life and finds himself rather envying an old friend who has made a new life with a wife and child. Alex no longer appreciates the booming concertos that were once a soundtrack to his violent acts, develops a taste for beer over milk, and finds himself strangely annoyed at the lack of sentimentality of his former companions. He thinks to himself:

> Perhaps I was getting too old for the sort of jeezny I had been leading, brothers. I was eighteen now, just gone. Eighteen was not a young age. At eighteen, Wolfgang Amadeus had written concertos and symphonies…Eighteen was not all that young an age, then. But what was I going to do? ... And now I felt this bolshy big hollow inside my plot, feeling very surprised too at myself. I knew what was happening, O my brothers I was like growing up.[291]

---

[289] Kubrick, Stanley, director. *A Clockwork Orange*. Warner Bros, 1971.
[290] Ibid.
[291] Burgess, *Clockwork*, 140.

Is Alex reformed now? Intuitively, it seems as if he is beginning to show signs of reformation that were lacking after the externally imposed technique. By all indications, 'who he is' seems to be changing as he approaches the world in a different manner than before. I will call this Alex at the end of the novel, *Novel Alex*.

Despite the differences between Movie and Novel Alex, both are arguably responsibility-apt. At least, if we think he is responsibility-apt for crimes committed in one scenario, we should also think him responsibility-apt in the other. To see this, consider what would happen if the Ludovico's technique were never reversed and it remained operative for Movie Alex. Let's call him *Cured Alex* to avoid confusion. This Alex represents the result of the successful application of Ludovico's technique. At first the process may limit his sadistic desires as an external imposition restricting him from acting in accordance with his evaluative stance. Yet, the debilitating nausea might encourage him to form something like an adaptive preference for domesticity. Eventually Cured Alex's change might resemble the kind of change Novel Alex undergoes. He may alter his preferences in such a way that he would no longer gravitate toward violence once the possibility is taken off the table.[292]

Adapting one's preferences to the situation does not require anything as extreme as Ludovico's technique in order to achieve this adjustment. The same could be said of a maturing partygoer declining a drink. She may decide to forgo a night out not only due to maturing tastes, but those maturing tastes may be partially formed due to the fact that she is unable to physically recover from a hangover like she once did. This physical and

[292] I initially conceived of the versions of Alex as being different in terms of responsibility attribution. However, many thanks to Muhammad Velji for pointing out just how similar each are.

evaluatively insensitive aspect of herself can change her preference to drink, but not in a way such that we would consider her no longer responsibility-apt for that decision. Indeed, the world is riddled with affordances and limitations on our preferences that may end up altering our perspective on the world. The reality of our situation and the world around us may make an impression on our agency in a manner not unlike what is seen when Alex undergoes Ludovico's technique once we are given time for the technique to work. If one is responsibility-apt, then we should think the other is too.

Despite initially invoking varied intuitions, the difference between the three versions of Alex may only be of degree and not kind. In what follows, I will defend the view that responsibility-aptness is compatible with a wide variety of changes a person can undergo. Whether it is the application of an experimental procedure, maturation, personal transformation or even deep alienation, responsibility-aptness holds in each due to an extension of answerability.

2. The Gappy Self

In everyday instances there are few times where one's evaluative profile operative in one moment is the exact same profile operative in the next. Mood, context or even what one is currently thinking about can alter which evaluative aspects are operative at time. I call this sort of day-to-day variation, *the problem of gappiness*, in reference to Strawson's description of the variable persistence conditions of one's phenomenological perspective (as opposed to numerical identity conditions). Just as it is not common to act and speak differently around family then if one were among colleagues, the sorts of evaluative aspects operative at one time, may recede and barely influence in other similar situations. This gappiness stems from the fact that the evaluative profile is not one set of

valuations that shapes each experience in the same way. As I argued in chapter two, the evaluative profile is unique to the particular agent, not because there is one fully constituted evaluative profile brought to each experience in the same manner every time, but instead it is a unique collection of evaluative aspects that may impress on experiences in different ways at different times that is specifically attributable to the individual (like a particular musical composition as seen in chapter two). Taken together the ways these aspects impress are unique, even if quite variable in different circumstances. The problem for our purposes that concern identifying responsibility-aptness over time is that the constant and mundane variation makes it difficult to determine the conditions under which responsibility-aptness is lost. How much gappiness is too much gappiness?

We require a means to limit when a person is responsibility-apt, but also need to be careful in defining this limit. If we require the exact same constitution of the evaluative profile in order to support responsibility-aptness, then we may be forced to say that the loss of self occurs more regularly than would be intuitive. Yet, if we are to accommodate variation, this raises a question of how much is too much and when is the self gappy enough to no longer be responsibility-apt? In this section I will explore one potential answer from the work of L.A Paul on transformative experiences. I will start by showing that responsibility-aptness is compatible with the sorts of life-altering experiences described by Paul. Yet, even if her examples do not provide the needed distinction, I will pull from a couple her insights to define a more appropriate boundary to define a limit of responsibility-aptness.

2(a). Raising the Bar: Epistemic and Personal Transformation

Paul mentions two connected ways persons may transform that seem to make sense of the changes Alex undergoes. First, an experience is *epistemically transformative* if it gives you new 'what it's like' information that you could not previously access.[293] Different experiences contain for us different phenomenological and cognitive values and it is through experiences such as "hearing beautiful music", "tasting a ripe peach", or "experiencing a major life event" that we are able to access these values for ourselves.[294] Through the addition of these "subjective values", subsequent value formation and decision-making is then enriched and partly informed by this new experience. Past values might also be reevaluated in light of the new evaluative bar that has now been set.[295]

Epistemic transformation seen here can lead to what Paul considers "personal transformation" as a significant alteration in your priorities, preferences, and self-conception.[296] Persons in this sense might be personally transformed after:

> …gaining a new sensory ability, having a traumatic accident, undergoing major surgery, winning an Olympic gold medal, participating in a revolution, having a religious conversion, having a child, experiencing the death of a parent… and so on.[297]

In each case, the experience changes "what it is like to be you" by changing "your point of view, and by extension your personal preferences, and perhaps even change the kind of person that you are and take yourself to be."[298] The changes are so significant

[293] Paul, *Transformative*, 10.
[294] Ibid.,13.
[295] Ibid.
[296] Ibid.,16.
[297] Ibid.
[298] Ibid.

that we can say it altered the self and one's "core preferences." [299] Such experiences represent a gap in one's evaluative constitution that would, on Paul's terms, warrant calling the agent a different self.

Paul's account claims that personally transformative experiences change how one experiences oneself and the world, but we are told very little about *how much* the person needs to change, *what* needs to change, if there are any *thresholds*, or even what is meant by the "*core values*." [300] Might there even be a degree of change that would warrant the label of a new self?[301] Of course, these are not Paul's questions, but the examples of personal transformation she uses are problematic when combined with my account due to the way such changes would potentially mitigate moral responsibility far too readily. Having a child may change one's outlook by dramatically showing what it is like to give birth, to operate on a torturous lack of sleep or to experience deep and unconditional love in ways never thought possible before. Likewise, other sorts of experience, like undergoing surgery, may allow the agent to physically experience the world in from a new physical vantage point. A death of a family member and similar trauma may shatter one's sense of self to the point that they are unrecognizable.[302] I may be transformed

[299] Ibid.
[300] I will begin to answer these sorts of questions in the next chapter. I will outline the extent in which the evaluative profile would need to be replaced in or to limit continued attributions of responsibility. For now, I wish to focus on a distinction between alteration and replacement.
[301] I should note that Paul does not make the distinction between the self and moral self I made previously. I take this nevertheless to be an account of the moral self as it concerns changes of the evaluative self (as it pertains to decision theory, not an account of strict identity).
[302] I would like to thank Muhammad Velji again for noting that the example of trauma may be more complicated than I have mentioned here. Susan Brison, for instance, argues that trauma victims provide a poignant example of a thorough disintegration of the self.

after these experiences, but personal transformation does not seem to be incompatible with responsibility-aptness over time. Surely I remain responsibility-apt for my past actions whether or not I choose to have a child, win a gold medal at the Olympics, or experience the death of a loved one. Otherwise, rehabilitation and redemption would just require engaging in any one of these activities. We would no longer require prisons but parental leave instead.

Compounding the problem is the fact that the sorts of experiences mentioned by Paul change persons in radically different ways. Each of these experiences may affect the individual differently due to their previous contexts, experiences and personal preferences, but when are such experiences transformative enough for a change of self and for whom? If we think that one is responsibility-apt after having undergone these experiences, then responsibility-mitigating personal transformation needs to mean more than a change in one's core preferences and experiencing the world in a different manner.

They are made to be "helpless in the face of a force that is perceived to be life-threatening" (Brison, "Trauma, Memory and Personal Identity", 13.) Brison describes the effect of trauma as the "obliteration of one's former emotional repertoire" leaving behind only "counterfactual, propositional knowledge of emotions"(Brison, *Aftermath,* 21). As she explains, "Not only is one's memories of an earlier life lost, along with the ability to envision a future, one's basic cognitive and emotional capacities are gone, or radically altered, as well." (Ibid.) The dissolution of the self in traumatic events becomes an "epistemological crisis" from which the survivor lacks "all bearings" in their ability to "navigate the world and understand oneself" (Ibid). How might we characterize responsibility in the face of such dissolution is a difficult question indeed and one that might be larger than I am able to treat here. For now I can only gesture to a possible line of inquiry that could be explored as a potential wrinkle in the account I have provided. See Brison, Susan, "Trauma, Memory and Personal Identity." In *Feminists Rethink the Self*, edited by Diana Tietjens Meyers. (Boulder: Westview Press, 1997).
and Brison, Susan. *Aftermath : Violence and the Remaking of a Self*. (Princeton: Princeton University Press, 2002).

2(b). Brainwashing, Replacement and Answerability

I would first note that epistemic and personal transformations on Paul's terms are neither mutually inclusive nor exclusive. Of course, it is possible to have experienced something epistemically transformative without it being personally transformative. Conversely, a personally transformative change might occur without any new phenomenological information. Perhaps, for example, it might be possible to be personally transformed through methods other than epistemic transformation. [303] So personal transformation need not involve epistemic transformation, and each may be distinguished not by the kinds of experiences one undergoes as the mechanism of change, but what the experience changes in the agent's evaluative profile.

We know that an epistemically transformative experience gives you "new information in terms of your experience" whereas personally transformative experiences by contrast are those in which a person's preferences, desires and self-conception are

[303] One example of this might be the act of naming workplace sexual harassment, in which the creation of a community helped give voice to the wrong suffered by women in general. Take the example of Camita Wood as given by Miranda Fricker. She describes Wood as suffering from repeated harassment at work by her employer, enough so to cause her physical as well as emotional stress. After a transfer was denied, she promptly quit. However, unable to articulate her experiences and being wrought with embarrassment, her decision to quit appeared to be pursued for purely personal reasons and any unemployment insurance was denied as a result. Eventually, she joined a seminar on unwanted sexual advances, where it became clear that each woman suffered this kind of experience in some form. The collective experience was finally termed "sexual harassment." Arguably, there was not a new experience that initiated this alteration, but a reframing of her experience that led to a personal transformation. Perhaps we could argue that the discussion of sexual harassment was epistemologically transformative, but this does not seem to be a case of undergoing a new experience in which the additive information changes one's values. The same experiences are understood in a new light due to the conceptual, rather than epistemological change. See Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. (Oxford: Oxford University Press, 2007).

dramatically altered.[304] So, if epistemic transformation changes the ordering of one's preferences through new epistemic information and does not necessarily lead to personal transformation, we see that alteration is possible without a loss of self. As a result, personal transformation needs to mean more than alteration of one's values and I would suggest that it be understood as a replacement or loss of them. On these terms, pregnancy and having a child would only be personally transformative if the experience did more than shift one's priorities, but thoroughly replace the kinds of valuations one had previously. Thus a personally transformative experience would change "how you experience who you are" because one now operates with a new, rather than altered, evaluative perspective. [305] Loss or replacement, I argue, determines whether the same moral self persists.

It might be illustrative to imagine a third way Alex could change. This Alex, rather than being subject to forced associative learning, has his preferences, desires and evaluative judgements wiped clean and replaced with a new set through some sort of extensive brainwashing. Alex's former evaluative profile is extinguished and replaced. But this is not a kind of hypnosis that would act on him like some alien force disconnected from his will. It represents the implantation of an entirely new evaluative profile — a comprehensive attitudinal overhaul.[306] This Alex is altered in a way that

[304] Paul, *Transformative*, 16.
[305] Ibid.,17
[306] In fact procedures like this are no longer solely the fodder of science fiction scenarios, but have been proposed in treatment of offenders. Recently, procedures using psychotropic drugs to directly alter the brains neurotransmissions and manipulate memory retention have been proposed alongside the possibility of implanted devices that produce "repetitive transcranial magnetic and direct current stimulation, and deep brain stimulation" in order to alter and "affect a vast array of complex, interrelated brain

would render him a new self and without responsibility for any past exploits. He is now only responsible for actions committed after the brainwashing technique. Let us call this Alex, *Brainwashed Alex* to separate him from the previous ones found in the movie and novel. Alex would no longer be responsibility-apt for actions prior to the brainwashing because his case presents us with a replacement that undermines continued answerability.

Persons readily change throughout a life, but it is only when their former beliefs and attitudes are significantly replaced that the grounds for being responsibility-apt are undermined. The fact that it was a loss rather than an alteration is important because it

---

processes associated with memory, cognition, motivation, emotional regulation, empathy, and moral judgment" (Craig, "Incarceration, Direct Brain Intervention, and the Right to Mental Integrity",114.) Indeed, Nicole A Vincent argues that persons may not even be responsible even if the procedure aims at returning the offender's capacities in order to stand trial. She argues there is "mounting empirical evidence [that] substantiates the worry that direct brain interventions might have adverse effects on such things as authenticity and personal identity by significantly altering character and personality" (Vincent, "Restoring Responsibility", 30). Indeed, her reasons for thinking responsibility are diminished would be similar to mine. In particular, it is not clear that the restored offender is indeed answerable for his or her past actions. She states, "it might not be legitimate to expect a treated person to answer for their earlier self's actions because the content of their testimony might be unrepresentative due to their altered character, or to attribute responsibility and to hold responsible the treated person for what their earlier self did, because we might now be dealing with someone who is in no position to answer (usually to society, but perhaps even to themselves) for the actions of the person that originally committed the crime" (Vincent, "Restoring Responsibility", 31). I accept that they are no longer responsible if the evaluative profile has been replaced, but even if the replacement is the aim, I suggest that aside from concerns stemming from bodily integrity and authenticity, that we would not want such a procedure because the change does not address the criminal wrongdoing. It merely eliminates it, so the criminal before us is no longer responsible, but not in a way that would justify forgiveness or considering the criminal rehabilitated for reasons that will be made clear in chapter eight. See: Craig, Jared N. "Incarceration, Direct Brain Intervention, and the Right to Mental Integrity - a Reply to Thomas Douglas." *Neuroethics* 9, no. 2 (2016) and Vincent, Nicole A. "Restoring Responsibility: Promoting Justice, Therapy and Reform through Direct Brain Interventions." *Criminal Law and Philosophy: An International Journal for Philosophy of Crime, Criminal Law and Punishment*, no. 1 (2014).

signals not just a change in Brainwashed Alex's phenomenological perspective, but a limit on answerability. While an explanatory answer as to why he committed past crimes (perhaps one that was even rich in detail) could be given after a replacement of former evaluative aspects, it would be in the third person. Smith argues, "In order for an agent to be answerable for something, it seems that thing must be connected to her in such a way that it makes sense to ask her to rationally defend or justify it." [307] Brainwashed Alex has no connection to the evaluative profile of his past and perhaps not even an ancestral one as the Brainwashing was inflicted on him. He represents, not a gap in the evaluative profile, but a severance.

In more ordinary cases of course, a past action also remains connected to the agent in the present due to being part of one's history, but the answerability test requires more than an explanation of events that occurred; it requires the sense of what I called *personal answerability* as relating to the agent's current evaluative stance that I introduced in chapter three. When the act is disconnected from one's current evaluative stance, we would be demanding a justification of evaluative activity that is as if it belonged to another. The three Alexes differ from the Brainwashed Alex in that they speak in the first person of the desire/values they once had, rather than in the third-person about another who once had those desires/values. I am suggesting therefore that answerability over time withstands variation and alteration, but not replacement on the basis of this sense of personal answerability.

Movie, Novel and Cured Alex all remain responsibility-apt because at no point does it remain unintelligible to consider them personally answerable. There is an

[307] Smith, "Control, Responsibility, and Moral Assessment", 369.

alteration within their subjective preferences and the change they each undergo represents a supplementation rather than a full transformation. Their change could be likened to a change in one's subjective values not unlike what is seen in J.S Mill's work. Mill argued that after having experienced higher and lower pleasure, agents would strongly prefer the former rather than the latter as the result of an expansion of subjective values.[308] It is not a loss within one's evaluative profile but an addition. For instance, the subjective value gained from listening to classical music for the first time may cause persons to rearrange what they formerly thought most worthwhile. Perhaps, like Alex, after listening to Beethoven, their whole world is infused and inflected by his 9th Symphony as a soundtrack to their exploits. Mill's sophisticated hedonists now prefer Beethoven to burgers, but they are not different selves as a result because at no point do they experience a loss of evaluative aspects. When pressed for an answer, the explanation will still involve evaluative aspects these agents currently hold because any change is primarily additive on their evaluative profile. These agents would still be answerable even if their evaluative profiles were now infused with a love for Beethoven.

It is the fact that during such alterations, former evaluative aspects are retained that explains why persons are responsibility-apt through small-scale day-to-day change as well as more monumental change like having a child. The new love is life altering as it changes nearly every aspect of who you are, but it does not necessarily replace these aspects. It also explains why Alex is responsibility-apt in all scenarios except brainwashing.

[308] Mill, John Stuart. *Ultiliatrianism.* Edited by Andrew Bailey. (Peterborough, Ontario, Canada: Broadview Press): 32-34.

Movie Alex is clearly answerable for his past. This Alex looks fondly at past violent crimes with a sense of yearning and anger at his inability to continue his criminal life. He recognizes himself to be the same criminal as he retains a good deal of his former judgments despite the inability to act on them when the technique is operative. He remains the same because any transformation due to Ludovico's technique is epistemic in that he gains new phenomenological information (that criminal impulses will make him violently ill), but this does not lead to a personal transformation. There was no replacement of his values even if new ones were inserted. These experiences are not personally transformative even if they are epistemically so and changed how Alex approaches the world.

The change in Novel Alex is not as clear. We begin to see a subtle alteration of his character that puts into question to what extent his former evaluative profile persists. I would argue that Alex is still responsibility-apt in this case because the alteration can be likened to ordinary maturation, or the gradual ebb and flow of altering and rearranging one's values.[309] The slowly maturing Alex may still remember the excitement he once "viddied" when engaging in violence and he may even feel a certain sense of nostalgia when passing the "Korova Milkbar."[310] Alex may not be fully sympathetic to his past, but at this point, I doubt Alex has undergone a loss of self even if this indicates an eventual personal transformation. This is also true of Cured Alex, whose possible success with Ludovico's technique mirrors Novel Alex in ways Burgess did not intend. In any of these cases Alex is arguably still the same despite the different ways the

[309] Schechtman, "Empathic Access", 246.
[310] Burgess, *Clockwork*, 3.

changes occurred because the change amounted to an alteration that preserved answerability. The difference between the kinds of change concerns the likelihood of complete personal transformation down the line. The connection to the past evaluative stance may be gappy and perhaps strained, but not so much as to rend apart the connective thread.

3. The Patchy Self

The distinction between additive change and replacement shows that the agent can remain answerable and therefore responsibility-apt despite significant epistemological transformation and alteration. I argued that responsibility-undermining personal transformation, rather than being defined as a change in one's core preferences, should be understood as requiring replacement of those preferences. Yet, it could be argued that this condition risks an implausibly high bar that could allow responsibility-aptness to hold even when agents deeply and fiercely repudiate who they once were.

One might argue that if these agents are still responsibility-apt, then it seems that we have lost sight of what really counts in determinations of responsibility-aptness. Why should we consider persons responsibility-apt if they approach the world in a radically different manner than before? I call this this *the problem of patchiness* in reference to the fragmented moral self I introduced in the last chapter. I suggested that there is little reason to think evaluative aspects that are not semantically or unified are not attributable to the agent if we start with the idea that the self is generally patchy. In this section I what to explore just how patchy that self can be by exploring the possibility of answerability in cases of extreme evaluative fragmentation in order to counter claims of an implausibly high bar.

3(a). Lowering the Bar: Narratives and Sympathetic Transformation

Marya Schechtman's notion of empathic access may be seen to challenge the possibility of answerability in the face of radical evaluative fragmentation and patchiness. Although it does not do so explicitly, the conditions under which a loss of the moral self occurs calls into question the changes in Alex (due to Ludovico's technique and maturation) as generally preserving the self, answerability and responsibility-aptness by extension. On her account, if the alteration leads to a moral self that is thoroughly unsympathetic to their former beliefs and desires the self is not preserved. This idea is housed within her theory of narrative identity, so in order to understand it we must first take a step back to look at her narrative identity thesis as a whole. I will argue in the end that answerability holds despite the kind of 'patchiness' as alienation implied by Schechtman's account. That is, persons may be answerable despite radical evaluative fragmentation as seen in cases of alienation and repudiation of the past.

According to Schechtman's narrative view, agents create an identity by forming an autobiographical narrative of their lives. We perceive our lives not as a series of incidents, but part of an ongoing narrative where our experienced past and anticipated future is woven into our present context. To see this, consider the perspectival self first discussed in chapter two. There, I argued alongside Schechtman that given the personal experiences one undergoes, this continuing perspective could be made more or less complex. Narratives do not add to the minimal self by adding a certain capacity on top of another. Narratives instead change that perspective and how one approaches the world

by "transforming and complexifying it."[311] In this way self-narratives not only function as organizing principles to integrate experience and make sense of one's past, but also go some way to shaping how one experiences the world.

On this account, once a consistent narrative is in place and informing one's experiences, we are now able to speak in terms of sameness rather than simple similarity. Rather than a disparate collection of evaluative aspects and life events influencing the present self, the narrative unifies and makes sense of such aspects as part of the same unfolding story. Hence, the articulation of an "identity constituting narrative alters the nature of the individual's experience in a way that extends consciousness over time, producing a persisting experiencer who is the primary experiencing subject." [312] From the first-person perspective, the kind of organization the narrative provides to one's current phenomenology produces more than qualitative similarity, but a kind of persisting subject through the way it maintains consistency in the phenomenological self.

As a helpful analogy, Schechtman compares the composition of a person as a "stew or soup" as "each ingredient contributes to the flavour of the whole and is itself altered by being simmered together with others." [313] Narratives ensure that this ingredient list is generally followed for an extended duration, thus reducing both 'gappiness' and 'patchiness' of the self. As a result, the self in the past and the self now are the same in virtue of being the same phenomenological unit whose psychological ingredients (including attitudes, evaluative judgements and beliefs) are generally held together and

[311] Schechtman, Marya. " 'The Size of the Self": Minimalist Selves and Narrative Self-Constitution." In *Narrative, Philosophy and Life*, edited by Allen Speight. (Dordrecht: Springer, 2015): 43.
[312] Schechtman, *Constitution of Selves*, 149.
[313] Ibid., 143.

made to persist through the continuation of an identity-constituting narrative. Thus, a narrative for Schechtman is not just a story we occasionally tell to make our past intelligible, it is rather an ongoing and active orientation that retains a particular flavour and inflects and characterizes one's unique perspective.

3(b). Empathic Access

According to Schechtman, central to maintaining the particular 'flavour' of one's 'phenomenological soup' is a specific kind of affective connection; the loss of which can signal a narrowing of the experiencing subject: "empathic access." [314] She states:

> What is needed for empathic access is … not an exact recreation of past emotions, thoughts and feelings, nor just an ability to call them up from a first person perspective. What is needed is this ability plus a fundamental sympathy for these states, which are recalled in this way."[315]

Empathic access, as a particular type of memory connection, is not dry and descriptive, but deeply embedded within our emotional states through experiencing the underlying emotions and thoughts felt during the remembered event. This access is also coupled with "an additional requirement of a generally sympathetic (or at least non-hostile) attitude towards those emotions, thoughts and feelings." [316] Sympathy is thus a kind of endorsement and is bound to obtain if the agent's current phenomenological and affective repertoire in the present is generally the same as it once was. If not, we can say that the agent may have lost "access to one's past point of view." [317] Without empathic access, "they cannot feel as they felt before or look at the world through the same eyes."

---

[314] Schechtman, "Empathic Access", 254.
[315] Ibid.
[316] Ibid.
[317] Schectman, Marya. "A Mess Indeed." In *Art, Mind, and Narrative: Themes from the Work of Peter Goldie*, edited by Julian Dod. Oxford: Oxford University Press, 2016): 19.

[318] The idea is that when sympathy no longer extends, the values and beliefs one once held no longer factor into current decisions.

If we consider the notion of empathic access in context with the narrative view, we see the importance of maintaining such empathic connections. Narratives act like a list of ingredients that need to be in place in order for the self to remain the same self and retain the particular flavour of the phenomenological "soup." [319] Importantly, like a soup that requires each ingredient to give it a particular taste, when there are dramatic shifts in the narrative thread (as we would get with a loss of empathic access) this signals a dramatically different way of arranging and understanding psychological ingredients that now come to inform current experience. There would be, on Schechtman's view, a new self.

To show how this empathic disconnect works, Schechtman describes a trio of matrons that showcase, to varying degrees, the distinctions between identity threatening change and ordinary maturation. Schechtman asks us to imagine a former party girl turned sober matron not unlike the one briefly mentioned at the beginning of this chapter and even earlier in chapter one. This now sober and serious matron may have changed some of her preferences due to some adaptive mechanisms, but regardless of how these preferences were formed, she no longer has the desire to act in the same manner as she once did. She can remember what it was like to be that young and carefree, but she now "knows how empty, tedious and ultimately disappointing those parties became and how

[318] Ibid, 20.
[319] Schechtman, *Constitution*, 143.

pleasant it is now to get some rest…" [320] Has the wild child continued in her present embodiment as a toned-down, but well rested mother? Schechtman would answer 'yes'. Even if this matron does not act in a way that she once did and chooses sleep rather than all-nighters, she has not completely lost access to her "past phenomenology, she has only placed it in a broader context which causes her to make different life choices." [321] To have empathic access to one's former self is to retain the "phenomenological and behavioural connection we desire in the present", yet this does not necessarily need to materialize in the same behaviour as before.[322] As Schechtman explains, her "old impulses act as a check on the new, making [her] consider carefully [her] motives or drawing [her] to a more balanced picture…." [323] In this sense, changes in behaviour may simply reflect a more considered decision as seen in ordinary maturation. The former constitution of the evaluative profile is present to a certain extent, but its influence on her behavior is tempered by further considerations.

We could also picture this matron in a more intermediate position as being "somewhat less-serious." [324] This matron remembers her past with fondness and perhaps a "friendly embarrassment", like that of "remembering the naïve passion of a first love…"[325] So long as the past sympathies are "given some weight" the "less-serious matron" may feel a need to justify actions to herself when she deviates from past

[320] Schechtman, "Empathic Access", 245-246.
[321] Ibid.,245.
[322] Ibid.
[323] Ibid.,252.
[324] Ibid.
[325] Ibid.

ideals.[326] This changed matron still remembers the excitement of getting dressed for a night out. She feels a great deal of sympathy for her former perspective, which in turn permeates her experience of the world.

Finally, compare this "sober" and "less serious" matron to Schechtman's account of the "mortified matron." [327] Like the previous two, the mortified matron has access to both the memory and emotion of her past. However, unlike them, these memories are now connected to "shame and disgust." [328] The antics of the former party girl are now seen as "sinful impulses" with immense hostility.[329] Like a religious convert trying to purge herself of sinful desires, she lacks empathic access and would not be the same self on Schechtman's account.[330] Unlike the sober and less-serious matrons who portray pictures of growth and expansion of various beliefs, desires and attitudes, Schechtman argues that the mortified matron demonstrates an *affective discontinuity* that points to a subtle loss of identity. Where the previous matrons gave weight to the former party girl's aims in current decision-making, the mortified matron, by contrast, gives these "objectives no weight at all."[331]

For the most part, Schechtman's insights in regards to these matrons seem right. The problem is that if the sort of disconnect as the result of alienation causes the formation of a new self, it seems that the agent might still be answerable despite repudiation. As we saw with Paul's personal transformation, I will argue that

[326] Ibid.
[327] Ibid.,250.
[328] Ibid.
[329] Ibid.
[330] Ibid.
[331] Ibid.,252.

Schectman's lack of empathic access is not going to be particularly useful in determining the limits of responsibility-aptness as agents seem to retain answerability despite the loss of this access. Instead, cases of alienation are not necessarily constitutive of personal transformation, but may be epistemic as a kind of additive change to the agent's evaluative profile that is consistent with answerability.

4.The Shameful Past: A Repudiated Expansion

Schechtman too sees the distinction between additive change and replacement as important, but does not extend the significance of this distinction to the mortified matron. In regard to the not-so-serious matron that is the same self on her account, she proposes that maturation and loss could be understood in terms of expansion and replacement as she states:

> The alterations in lifestyle and outlooks may be just as pronounced as those in the case of the serious matron, but these alterations are the result of an *expansion* of beliefs, values, desires and goals rather than a *replacement.* [332, 333]

Even if Schechtman believes these influences to be rejected in the case of the mortified matron, I would argue that such an explicit disavowal does not necessarily mean that the former perspective is entirely diminished or absent in current experience. Here, I would like to shift what is meant by alteration that preserves the same self to include the mortified matron and the like. Shameful episodes may still constitute and colour current experience; perhaps not to the same extent as aspects that are endorsed and sympathized with, but affect nonetheless.

---

[332] Schechtman. "Empathic Access", 246.

[333] In fact, it was her brief remark here that inspired this chapter. Although I used Paul to illustrate the concept, the seed of my proposed alteration/replacement distinction derives from Schechtman's remarks here.

It is clear that when agents lack sympathy for past states, they approach the world in a very different manner than they did before. The "less-serious" matron took her former motives, desires and thoughts into consideration and did not "reject them outright." [334] Her former self is still "alive" in influencing how she approaches the world.[335] The mortified matron by contrast does not feel the "need to give weight to former impulses" and is estranged from the past affect of those experiences.[336] Former motives and desires are simply overridden as a result. Through affective dissociation we are dealing with a very different subject of experience whose behavioural and emotional repertoire is changed from before. However, I would argue that to posit a new distinct self or even a new evaluative profile from cases of regret, shame and even fierce repudiation assumes that what the matron explicitly rejects is not influential and lacks any consideration in the present. Just as a friend's past unreliableness colours my current frustration at her late arrival, or embarrassment at a particular location infuses it with a consistent cringe, it is not clear why shameful episodes would not influence my experience in the same way. Moreover, the sense in which the mortified matron is a new self makes it difficult to explain why she feels such extremes of emotion at all. Can we really be embarrassed to the point of mortification for actions only tacitly acknowledged?

If we allow that alienated aspects of the self can still affect one's phenomenological experience, the differences between mortification and maturation seem to fade. The mortified matron may not reject the influences that cause her shame

[334] Schechtman, "Empathic Access", 255.
[335] Ibid.
[336] Ibid.

"outright", in much the same way as the not-so-serious matron, but she is still arguably able to access those not-so-former attitudes and desires.[337] In both cases considerations are weighed in what leads to ultimate rejection, so the difference between the matrons seems to primarily be of degree. Yet, being the same self should not depend on how quickly one rejects former attitudes and values, especially when those attitudes and values still influence who one is today. Each of the matrons is arguably phenomenologically connected to their past desires in a way that has relevance to their current perspective even if the mortified matron may act on those desire more infrequently than the others. To remain responsibility-apt, the matron need not be empathically connected, but just phenomenologically so.

Schechtman calls this sort of access "phenomenological access" as a kind of empathic access "minus the added sympathy." [338] However, I will be using a derivation of this term to refer to what happens when the evaluative profile in the past still inflects and comes to inform the agent's experience now. Thus, rather than phenomenological access, I will refer to this sort of connection as "evaluative access." The present self can 'access' the past on my terms only in a figurative sense in that they retain the much of the same evaluative aspects as before. In other words, the current agent is able to access and phenomenologically inhabit her former perspective because her evaluative profile is constituted in a similar manner. I would argue that this sort of evaluative access provides at least one necessary condition for being the same self and in turn forms a basis to

[337] Ibid., 246.
[338] Schechtman, "Mess Indeed", 26.

consider persons responsibility-apt over time.[339] Thus, while the mortified matron weighs former considerations in a different manner than she once did as a wild child, the fact that she feels such shame indicates that she is more connected to her repudiated past than Schechtman initially supposed. Given that she feels shame, her situation can be better understood as a kind of alteration that preserves evaluative access and, hence, answerability.

5. Answerability and Indifference

To highlight the importance and function of retaining evaluative access, let us return to the soup analogy introduced by Schechtman. I am arguing that in the cases of epistemic transformation and additive change, we are still working with the same recipe as before even if it is altered. Like making a soup from scratch, with additive change the proportions may vary, but the same soup is being made. We might even be adding a few new flavours and withholding some. If, however, we start replacing the ingredients we might begin to question whether the same recipe is being followed. It does not make sense to call a stew made of butternut squash, clam chowder.

Likewise, when evaluative aspects are lost, it is no longer fitting to require an answer. Brainwashed Alex is no longer responsibility-apt, not simply because he was brainwashed, but because of what that brainwashing says about his evaluative constitution. Just as Elie Wisel famously said, "The opposite of love is not hate, but indifference", sometimes we are unaffected by the past when certain formerly held

[339] I would note here that the way this sort of phenomenological access obtains will need to be further qualified and I will explore what this might entail in the following chapter.

evaluative aspects are seen as too insignificant to warrant much concern in the present.[340] This is why the brainwashed Alex, with no connection to the past, may no longer be responsibility-apt. His connections to his former evaluative profile are severed from who he is now. The past is neither "given any weight" in current decision-making nor does it produce affect in the present.[341] It is within this conception of indifference that Schechtman may be entitled to her loss of empathic access and subsequent loss of the self. Indifference signals that the former evaluative aspects no longer inflect or have space in one's evaluative profile, which in turn undermines answerability.

The idea of evaluative access as undermining answerability gives a new interpretation to Locke's notion of "distinct and incommunicable" consciousness we saw in chapter two.[342] At the time, the agent may have been answerable and the action itself was motivated by evaluatively sensitive influences and judgments, but once those former evaluative aspects are lost and no longer influence who we are, any demand for an answer is going to be met with an explanation without reference to evaluative aspects she currently holds. The agent's consciousness has not extended and may be as distinct as would hold for a different person altogether.

As long as many of the same ingredients are added (in a way that preserves answerability), the evaluative soup need not be constituted in the exact same way in order to be the same soup in ways that matter. Indeed, in a later paper, Schectman acknowledges these sorts of issues with empathic access as she states, "In my earlier

[340] Wiesel, Elie. "One Must Not Forget," interview by Alvin P. Sanoff, *US News & World Report* (27 Oct 1986).
[341] Schechtman, "Empathic Access", 246.
[342] Locke. *Essay*, II. xxvii.20.

paper I defined empathic access as involving phenomenological access *plus* endorsement; now I think it should be defined only in terms of the former."[343] Schechtman too is no longer committed to the cogency of empathic access for many of the reasons argued in this chapter. She states, "religious conversion, for instance, is often described not as a loss of phenomenological access to sinful impulses, but rather as a rejection of them. The convert does not claim to no longer *be* a sinner, but only to repudiate his sin." [344] Cases like the mortified matron do not show a loss of self because in order to feel these extremes of emotion requires "that the *person* who inhabits the point of view of the convert still has the point of view of the sinner in her experiential repertoire." [345] What matters is "phenomenological access" as the ability to "inhabit the first-person point of view", which I have renamed "evaluative access." [346] With this modification, incorporation of Schechtman's view into the framework I have thus far proposed is not difficult. Essentially, evaluative access is the opposite of indifference and in turn grounds the personal aspect of answerability.

Conclusion

Just as Locke argued what makes an experience one's own over time, will be "upon the same ground and for the same reason as it does the present", we see that with evaluative access we have the same structure.[347] Evaluative access importantly needs to hold both at a time and over time in order to say that the agent is or remains answerable. Whether this sort of access holds is best captured by the distinction between the loss of

[343] Schechtman, "Mess Indeed", 26.
[344] Ibid., 26.
[345] Ibid., 27.
[346] Ibid.
[347] Locke, *Essay*, II. xxvii. 26.

self and expansion by way of the alteration/replacement distinction. A personally transformative experience may change one's core values as we see with the first two versions of Alex, but there should be an extinction of those values in order to distinguish a new self as seen with Brainwashed Alex. The proposed distinction also sheds light on the situation of the mortified matron who represents a conflicted self that is affectively distinct from who she once was. She retains evaluative access despite approaching the world in a radically different manner. In some sense, she knows "what it was like" and psychologically inhabits her former motives even if she does not sympathize with them.[348] The retention of answerability through evaluative access begins to answer the dual problems of gappiness and patchiness by providing a fairly high threshold as to when moral selves are lost. What this distinction does not yet tell us is the specifics determining that threshold. In particular, I have only so far suggested that brainwashing and complete indifference may undermine answerability, but have not yet said anything about how much loss the evaluative profile can endure before it can no longer support answerability and responding to these questions will occupy the bulk of the next chapter. For now it is important to know how these evaluative aspects constitute the agent's experience in a way that renders a demand for a response intelligible. Any question posed is personal and pertaining to beliefs and attitudes she in fact still holds. There is no indifference.

[348] Schechtman, "Empathic Access", 247.

Chapter 6: Degrees of Transformation and Answerability

Introduction

In the previous chapter, I argued that if the evaluative profile undergoes alteration rather than replacement, persons might remain responsibility-apt because answerability persists. However, when introducing the alteration/replacement distinction, I treated responsibility-undermining change in an all or nothing manner. Either agents undergo complete replacement of their evaluative profile, such as in brainwashing, in which case they would cease to be personally answerable and responsibility-apt, or they do not undergo complete replacement, in which case evaluative access obtains and they still are answerable and responsibility-apt. However, there is a further question: could answerability fail to persist even in the absence of wholesale loss of evaluative aspects? This chapter will aim to resolve this problem and clarify the degree of change in the evaluative profile that corresponds to a loss of responsibility-aptness.

In §1, I reintroduce some of the problems raised in the last chapter and show how the alteration/replacement distinction does not address all the concerns one might have for the various versions of Alex. The notion of evaluative access is too general as it stands because it is susceptible to claims of indeterminacy. The rest of the chapter will then examine two related problems. In § 2, I identify the problems of *trivial aspects* and *degree,* illustrated by an example from Derek Parfit. In §3, I begin to respond to these potential objections by drawing from the work of Delia Graff. I argue that what matters for responsibility-aptness over time is the retention of certain relevant aspects of the

evaluative profile.[349] When this occurs the agent may be both radically altered and relevantly the same (for the purposes of responsibility attribution) all at once. In §4, I will apply the proposed framework for determining responsibility aptness to the proposed cases. Finally, in §5 I briefly review some epistemological worries raised by my approach.

Overall, the main purpose of this chapter is to provide a positive account that allows for the possibility of persisting answerability despite radical changes of the self. It is also possible that a person may not have undergone radical change, but nevertheless is no longer responsibility-apt. Although they appear somewhat paradoxical, these conclusions will have implications for the situation of rehabilitated criminals. It does not necessarily matter for responsibility-aptness how significantly the criminal has changed generally speaking, but whether certain evaluative aspects are similar enough to consider the agent the same for purposes of responsibility attribution.

1. A Bar Set Too High

As dramatically portrayed in *A Clockwork Orange*, a true change of self requires not just the modification of behaviour, but also a radically different set of evaluative beliefs and attitudes. As in cases of shame and repudiation, at a later stage in our lives we may value and weigh our options differently, and new desires may come to extinguish old ones, but this is usually better described as a change in behaviour, not a wholesale change in the moral self. In the last chapter, the only example discussed was that of Brainwashed Alex in which there is a wholesale replacement of the evaluative profile.

[349] Graff, Delia. "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 28, no. 1 (2000): 45-81.

Yet, at the end of the novel, Alex does not undergo any brainwashing, but seems altered in a way that questions his responsibility-aptness if these changes represent losses in his evaluative profile. I suspect that this will often be the case in cases of rehabilitation because the evaluative profile is seldom lost as extensively as in brainwashing. Instead, loss within the evaluative profile occurs gradually and in ways that might be indistinguishable from day-to-day fluctuations. Likewise, Alex's change lies somewhere between alteration and replacement and this indeterminate state is, more often than not, the norm. However, I have yet to discuss these indeterminate cases.

Previously I drew on Schechtman's analogy of narratives as *recipes* that would preserve the flavour of one's *phenomenological soup*.[350] I argued that persons could still be considered the same even if the ingredient list (so to speak) was altered, but not if they were thoroughly replaced in a manner analogous to brainwashing. Suppose however that the recipe is continually altered over time. For instance, if I gradually used less basil or swapped it out for another herb would we still have the same tomato-basil soup? Does it count if there is only a pinch of basil left? What about a smidgen? If not, when exactly did the soup change? This kind of gradual but steady change can be likened to a Sorites Paradox concerning predicates. Terms like 'bald' are thought to be vague predicates that carry no clear and universal conditions for satisfaction. When a single hair is lost everyday until a person is clearly bald, there is a question as to when the transition occurred.

It is not clear how much loss the evaluative profile can undergo before evaluative access is undermined. Like evaluating the number of single hairs that is required to apply

[350] Schechtman, *Constitution*, 143.

the predicate 'bald,' it is difficult to define a threshold point at which a change in evaluative profile constitutes a new self.[351] If Alex lost an evaluative aspect here and there, until his evaluative profile was entirely replaced, it is not clear at what point he became no longer responsibility-apt. Perhaps when he lost *most* of his former beliefs and attitudes, but what might we mean by 'most'? Like the head of a balding man, there is indeterminacy between two clear-cut cases. At the one end you have Alex as a violent teenager, while at the other a potentially reformed man with a whole set of new beliefs and attitudes. In between you have someone who is neither entirely the same nor entirely different. The answer to establishing a threshold, which I develop in this chapter, employs a solution analogous to that proposed by Graff to the Sorites paradox. First, I want to leave behind soups and baldness for a moment in order to illustrate the kinds of problems at issue by analysing an example of change given by Derek Parfit.

2. The Problem of Degree and Trivial Aspects

How might we understand subtle, responsibility undermining, change over time? Perhaps we could take a page from Derek Parfit's book and understand the persistence of the evaluative profile quantitatively. As Parfit might argue, responsibility would hold through time only by degree. He states:

> When some convict is now less closely connected to himself at the time of his crime, he deserves less punishment. If the connections are very weak, he may deserve none. This claim seems plausible. It may give one of the reasons why we have Statutes of Limitations, fixing periods of time after which we cannot be punished for our crimes. (Suppose that a man aged ninety, one of the few rightful holders of the Nobel Peace Prize, confesses that it was he who, at the age of twenty, injured a policeman in a

[351] If we prefer a different analogy, we could also imagine a changing Alex to be similar to the paradox of Theseus's ship whose planks are gradually replaced as they rot during multiple voyages. Eventually all the planks are replaced, calling into question the identity of the ship.

drunken brawl. Though this was a serious crime, this man may not now deserve to be punished.) [352]

Parfit's convict seems fairly close to the initial example that began my inquiry. Further, his solution seems right, at least to an extent. Responsibility seems to admit of degrees. Nevertheless, because he is primarily concerned with the different questions of re-identification, identity and survival, the quantitative solution he suggests may be insufficiently fine-grained to capture the persistence of answerability/responsibility-aptness taken in isolation.

To see this, consider the four primary concerns Schechtman identifies as motivating questions of the persistence of identity through change: survival, self-interested concern, compensation, and responsibility. My aim in this thesis has thus far been focused on explaining the persistence conditions required for moral responsibility. I have been silent on what it means to legitimately survive, be self-interested or compensated even if much of what is said about moral responsibility to some extent also speaks to these concerns. The notion of survival could very well be best served by something like Parfit's account of psychological continuity and connectedness (I will remain agnostic about this). But Parfit's concerns are much broader than mine: I am concerned with survival of the moral self only (for purposes of answerability and responsibility-aptness) and not a question of survival or ceasing to exist in a metaphysical sense. These responsibility determinations require a more narrowed focus than what is needed to determine survival more generally.

[352] Parfit, *Reasons and Persons*, 326.

To draw on some of Parfit's insights, and also shift the focus to the persistence of the moral self, the emphasis needs to be placed on whether this moral self can be retained by degree. The quantitative solution translated to the moral self would ask whether there is a quantitative threshold of components of the evaluative profile beneath which the convict would no longer be responsibility-apt. There are two related questions: the first, *the problem of degree*, asks about possible limitations of continued responsibility attributions over time due to the loss of a number of evaluative beliefs, while the second, *the problem of trivial aspects*, questions whether the retention of certain evaluative beliefs rather than others is more important in making this determination.

Consider Parfit's story of legitimate anticipation. He recalls a young socialist who will be inheriting a vast fortune. Fearing this wealth will cause him to lose his ideals, the young socialist signs a legal contract that requires his fortune to be handed over to local peasants if his fears turn out to be justified. He tells his wife never to revoke it, even if he should ask her to later in life. On Parfit's account, the young socialist believes that if he should lose his ideals, he will not have survived and hence will have ceased to exist. My question, however, is whether the potential future capitalist would be the same moral self that was once enraptured by socialist beliefs. On my view, he would have to have the same evaluative profile governing his experience of the world as the young socialist. The capitalist will retain many of his former beliefs and judgements, but before we can evaluate whether he remains responsibility-apt, we need to know how much of his former evaluative profile remains. Is there a precise amount of former attitudes that needs to be extinguished for the future capitalist to no longer be answerable for his former socialist self? Does it matter which evaluative beliefs are retained or

extinguished? I will argue that whether or not the young socialist (or the maturing Alex by extension) remains answerable can be understood quantitatively if sufficiently qualified.

I will argue that persons can remain responsible if a sufficient amount of relevant evaluative aspects persist. The focus on relevancy narrows the quantitative determination in a way that avoid the problem of trivial aspects and brings the focus on what matters for the moral self and determinations of responsibility, rather than survival of the self in general. Before arriving at this conclusion the following sections will introduce the concepts key to developing this solution, those of trivial aspects, significance, and finally salient similarity.

3. Potential Solutions

One way to deal with the problem of degrees of change is to reject the idea that responsibility-aptness depends on meeting a threshold. If this were right, we would talk only of degrees of being the same self and, by extension, degrees of continued responsibility. Holly Smith argues just that. She suggests that we should consider the extent of responsibility for an action in terms of the degree to which the evaluative profile (or "psychology" in her terms) is 'engaged.' [353] She states:

> Probably there is a continuum here: we should speak instead of degrees of blameworthiness, where an agent is more blameworthy if a substantial portion of his psychology was engaged in a decision, and less blameworthy if a less significant portion of his psychology was engaged.[354]

[353] Smith, Holly M. "Non-tracing Cases of Culpable Ignorance." *Criminal Law and Philosophy*. 5.2 (2011): 138.
[354] Ibid.

Like proponents of the 'whole self' views discussed in chapter four, Smith suggests that an action is more or less attributable to the person depending on the degree of involvement of one's psychology. In other words, we could posit the same moral self to a greater or lesser degree. Once there is a complete replacement, a different self emerges and the agent can no longer be morally responsible to any extent.

The solution suggested here may provide at least one answer to the problem of degree, but is open to the qualitative problem of trivial aspects. That is, the approach fails to acknowledge not only what matters to the agent, but also whether later persisting evaluative beliefs are sufficiently connected to what we would hold the agent responsible for. Consider the socialist again. We can ask: would the young socialist be relieved if he could peer into the future and assure his present self that, although his beliefs about socialism will have changed, many other psychological connections will hold to a sufficient degree and therefore he remains strongly connected to the moral self of the past? Understanding connectedness only in terms of degree without considering the qualitative aspects renders the concern of the young socialist unintelligible. It is not just a matter of preserving a certain number of psychological aspects but rather preserving the relevant psychological aspects. In particular, there may be a number of trivial aspects of the moral self that remain that do not speak to the socialist's concerns. I call these aspects trivial, not because they do not matter at all or could matter given varying contexts, but rather because they are peripheral to whether he will continue to act on his socialist ideals.

3(a). A Sense of Significance

To be a new moral self, the socialist needs to undergo not just any changes, but specific changes in his evaluative profile that relate to his political values. For an evaluative profile sufficient for answerability to persist, (what I will call) the *significant* aspects of the moral self have to be retained.

Significance, in the sense I am using here, does not correspond to aspects of the self with which the agent identifies. Identification may make it the case that the belief or attitude is featured more often in decision making, but does not make it significant in the sense I am looking for. It is possible that what we are most concerned with are aspects that the agent does not identify with. As we saw in chapter four, agents can in principle be responsibility-apt for implicit biases and other aspects of the moral self despite a lack of awareness or identification.

I would also like to distinguish significant psychological aspects from those that are integrated. Significance may be confused with foundational beliefs and attitudes that usually attract our attention, but the latter are not necessarily significant. Foundational psychological aspects may be well integrated in the sense defended by Arpaly and Schroder in chapter four. That is, they are the aspects on which many other psychological aspects depend. For instance, a belief in animal rights may permeate and support many other psychological aspects such as choice of profession, the feelings felt when animals are present or sensations of disgust when consuming animal products. It does not follow however that foundational or integrated aspects of the self are significant for purposes of being responsibility-apt.

Of course, when foundational aspects are replaced or otherwise lost, there is likely to be a very different moral self as other evaluative aspects may fall like a house of

cards. A loss of a foundational aspect could make it very likely that the individual would no longer be responsibility-apt, but such loss of integration does not assure it. The loss of a foundational belief in animal rights could affect the constitution of the moral self in many ways, but could also potentially be disconnected from what matters when asking if the individual is responsibility-apt for a particular action or crime. For instance, if a crime involved a release of animals for medical testing, then the belief in animal rights would be significant*;* however*,* it would not necessarily be significant if the crime were different.

Significance in my sense does not concern foundational aspects or identification with them, but relevance. If someone wanted to predict whether I would save an animal, it would not be relevant to question me on my colour preferences or whether or not I remain quite fond of poutine (unless the reward for saving the animal was given in this delicious Québécois dish). Perhaps it would be better to question my commitment to the moral stance of Peter Singer. In the next section, I propose that the question should be whether the person is similar enough to the person they were in the past and likely to commit relevantly similar actions.

3(b). Interest Relativity

Graff proposes that vague predicates appear to have no clear boundaries because their truth conditions are sensitive to our ever-shifting interests. The satisfaction conditions of which may shift depending on the contexts of their use. In other words, "… vagueness in language has a traceable source in the vagueness of our interests."[355] This connection to our interests will determine what is most salient in defining the

[355] Ibid., 5.

parameters of our inquiry. For instance, consider what happens when we compare movies. There are many dimensions along which to make a comparison. But if our purposes included contrasting the realism between two films, some aspects would be more relevant, and hence salient. We might want to focus on the details of the costumes, sets and plot instead of others such as lighting, pacing or visual effects. Equally, a friend might groan when someone points out the historical inaccuracies between films if her interest was simply in the entertainment value. Her concern might be with the visual effects and cinematography which is more salient and, hence, significant for her interests. So, although these other filmmaking aspects may be relevant in other comparisons, they are not necessarily most salient for the agent's specific purposes. Our interests thus define the parameters of our inquiry.

Of course, when we compare movies this is not a quantitative comparison as we might make for the application of vague predicates such as a 'heap,' 'bald' or even 'same moral self'. A movie is not necessarily made better by the quantity of visual effects whereas for Graff whether or not one is bald is determined by satisfying a certain quantitative threshold. Instead, our interests determine both the relevant comparison class and the threshold for satisfaction. Generally speaking we might not be able to locate the vague point at which a person may be considered bald, but this is not necessarily the case if we were to define baldness for a particular purpose. For instance, she asks us to consider baldness in the context of casting. She states:

> Suppose I am a casting agent auditioning actors for parts in a play. On one day I'm casting for someone to play the role of Yul Brynner, who had absolutely no hair. On a different day I'm casting for someone to play the role of Mikhail Gorbachev, who has some hair, but very little on top. When I turn away auditioners citing as a reason that they are not bald, I may be using different standards for 'bald' on the different days. I may say "Sorry, you're not bald" to an actor when he auditions to play Yul Brynner, and

may then say, to that very same man when he auditions on the following day to play Gorbachev, "Yes you look the part; at least you're bald." [356]

In casting, the director is concerned with the comparison class of bald men. Yet, what counts as being bald in this instance depends on what the director is looking for in casting the part. The threshold for the predicate 'bald' is thus a moving target that shifts according to the purposes of using the term.

I propose applying the notion of significance to narrow a comparison between the evaluative profile in the past and the evaluative profile in the present. I would also like to incorporate Graff's interest relative solution to Sorites paradoxes to determine the kind of threshold that is needed to determine responsibility-aptness.

3(c). Salient Similarity

Drawing on one last reference to Schechtman's soup analogy, I argue that persisting answerability allows for many of the ingredients in the phenomenological soup to be replaced as long as the most relevant ingredients are retained. What those ingredients are and how much is needed would in turn be determined by what we are trying to make. Asking of responsibility-aptness is not necessarily asking whether it is the same soup, but whether this soup has been made too garlicky or that it needs more pepper. In these cases it is not pertinent to cite whether the recipe as a whole was followed, but just a question concerning the amount of garlic or pepper added. Likewise, the persistence of the evaluative profile as a whole or the survival of a self in Parfit's sense is not sufficient to evaluate whether or not an agent is answerable or responsibility-

[356] Ibid., 12.

apt for particular past actions. We need a more fine-grained analysis that asks whether the relevant flavour of the evaluative profile has been retained.

My argument will be like Graff's in that responsibility-aptness "rests on the idea that two things that are qualitatively different in some respect, even when they are known to be different, can nonetheless be the *same for present purposes*."[357] So even though the individuals may change in many ways over time and experience the loss of certain evaluative aspects, what matters is evaluative access to certain significant aspects of their evaluative profile. Thus, if we asked whether agent A were responsibility-apt for action B, we would ask of the significant evaluative attitudes X, Y and Z and whether they remain to an agreed upon threshold (where it might be enough to retain X and Z if not Y, or X and Y and not Z, or etc.). If so, then A displays enough what I will call *salient similarity* for present purposes, and in being similar enough, A remains responsibility-apt.

The boundary as to whether or not the young socialist and Alex are saliently similar enough to their past selves to meet some threshold is determined by what is important for our purposes in defining that boundary. Graff uses the term "salient similarity" to mean that:

> …if two things are *saliently similar*, then it cannot be that one is in the extension of a vague predicate, or in its anti-extension, while the other is not. If two things are similar in the relevant respect, but *not* saliently so, then it may be that one is in the extension, or in the anti-extension, of the predicate while the other is not. [358]

I will use the term in the same way as Graff. When an agent is saliently similar, this marks the threshold to satisfy a claim to responsibility-aptness based on the

[357] Ibid., 24-25.
[358] Ibid.

constitution of his or her evaluative profile. This means that when either the Young Socialist or Alex are saliently similar to their past moral selves, their current outlook is similar enough for them to be treated as personally answerable and responsibility-apt.

Importantly, to continue to be responsibility-apt, there is no precise amount of evaluative aspects that need to be retained; it just needs to be reasonable that we might think the agent will approach the world in the same manner as before. Consider Graff's analogy to coffee making. She states:

> … [S]uppose a small child is watching me make a pot of coffee and, thinking she is being helpful, points out that a couple of grains have spilled from my coffee scoop. ... When she then wonders why I don't bother to replace the grounds I've spilled, I might explain to her that there is no need because the two amounts are the same for present purposes. … To say that the two amounts of coffee are the same for present purposes is to say that …my coffee-making purpose permits me to behave as if the two amounts were the same, since the purpose is in no way thwarted by my behaving as if they were the same.[359]

In Graff's sense, as far as coffee making is concerned, the scoop with the few grains spilled and the one without are both what she calls "live options" that will achieve my purposes.[360] Given that my purpose is to make coffee efficiently, stopping to discriminate between them will introduce costs to this purpose. There is no precise amount of grains that would be sufficient for coffee making and the act allows for a fair amount of variation without being "boundaryless."[361] The act of coffee making is tolerant of a few spilled grains here and there before it becomes too weak, strong, satisfying or unsatisfying. Graff argues that the amount necessary for the task shifts. Not only might I have:

---

[359] Ibid., 25.
[360] Ibid.
[361] Ibid.,4

> … inexact knowledge of the satisfaction conditions of my desire for coffee on a given occasion; it is also that the satisfaction conditions of my desire may subtly shift, so as to be satisfied by different amounts of coffee as different options become available to me and the costs of discriminating between different pairs of amounts change.[362]

I will suggest that, like coffee making, there may be some interest-relative threshold to be met when determining responsibility-aptness, but one that is generally tolerant of variation in the conditions for satisfaction.

Applied to the concerns of the young socialist, sameness for the purposes of whether or not the contract is satisfied need not be particularly precise. Indeed, there are many evaluative aspects that come to inform the young socialist's economic preferences and he does not need to retain them all in order for him to remain a socialist. What exact evaluative aspects these are does not matter as long as they generally support his socialist ends just as much as having a slightly bigger or smaller scoop of coffee is not going to deeply undermine one's particular interests in making coffee.

Overall, what matters for the socialist's concerns for the future, is not the exact recreation of his current beliefs, but a general cluster of similar enough evaluative aspects pertaining to his socialist agenda that would satisfy a claim to salient similarity. This sort of similarity in turn justifies a claim to answerability. Likewise, Alex remains responsibility-apt for past crimes will be determined by whether he is saliently similar enough to answer for them even if he has changed in many other respects. Graff's framework allows us to say that in determining similarity, the kinds of evaluative aspects under consideration and the threshold to be met are relative to our interests in determining responsibility. I will now apply these considerations to the examples of the

[362] Ibid., 28.

young socialist and Alex.

4. Survival and Responsibility

The notion of salient similarity provides some interesting results when applied to determining responsibility-aptness over time. In particular it is possible to no longer be responsibility-apt without radical change, just as much as it is possible to remain responsibility-apt despite radical change. For instance, it is possible that the young socialist will survive into the future in a Parfitian sense, yet he might find his worries confirmed by losing his socialist ideals. It may be the case that the impending young capitalist is, by all who knows him, very similar to who he once was. He may still be quite fond of reading, love his wife, and engage in many of the same activities he once enjoyed. He may approach the world in a generally similar way as before. However, he now fully and completely supports a capitalist agenda. He may have survived as the same moral self, but he may no longer be answerable for his socialist concerns. There is sure to be some degree of psychological continuity and connectedness to support a claim to survival. Yet, these overlapping chains of continuity and connectedness may be constituted by 'trivial' or non-significant psychological aspects, such as his love for his wife and books, that have little to no bearing on his socialist concerns. He may have survived, but not in the ways he may have hoped. He has lost salient similarity.

The loss of salient similarity might also characterize either Alex in Burgess' novel I called *Novel Alex* (who eventually matures) or the hypothetically *Cured Alex* (who represents a successful application of Ludovico's technique), but not Stanley Kubrick's version or *Movie Alex* (who under went a reversal of the technique). Movie Alex is clearly saliently similar to his past self insofar as he shows no indication that his violent

impulses have been quelled to any extent. Yet, it is possible that eventually Novel or Cured Alex would lose all connections to past violent preferences. If the hints at reform prove lasting, Novel Alex may continue to mature into a domestic life and away from the violent desires that once characterized his outlook. He may nevertheless retain many of his old preferences and survive in a more general sense. This is also true for Cured Alex as well. After forming adaptive preferences that push him away from his violent desires and guide a reluctant reformation, we could potentially call the procedure a success. In either case, we have situations in which persons do not undergo radical change, but nevertheless could potentially experience an eventual loss in responsibility-aptness. Maintaining our focus on years down the road, if the violent desires that once characterized Alex's youth are thoroughly replaced, salient similarity would not hold for a matured Novel or Cured Alex.

If, however, we look to when these changes began to occur as the first inklings for a domestic life or when the adaptive preferences were first formed, the answer here is more complicated. I would argue that responsibility-aptness holds during times of conflict and alienation. That is, salient similarity holds even if the agent otherwise repudiates the salient evaluative aspects. Before the young socialist or Alex experiences a loss of evaluative aspects, each many exist in a conflicted state in which least some salient evaluative aspects may still come to inflect their current perspective and constitute at least a portion of their evaluative profile. This current capitalist may experience a state of alienation not unlike Schechtman's mortified matron. His attitudes towards socialism are alive in some sense and represent an expansion, rather than a replacement. He could reason with his wife and argue that he is not a new self

altogether, but one who has matured from his old ideals. He could argue that the decisions of his capitalist self is in light of new and expanded information. Given his conflicted nature, she might even see a glimmer of hope for a return to his socialist ideals. Yet, she would still have cause for concern as the sort of repudiation seen here could signal an eventual loss of these relevant aspects in due time. Regardless, at this time he would still be answerable for his socialist concerns.

The impending capitalist's situation might also characterize Novel or Cured Alex as the changes begin to occur. At the end of the novel, Alex has not changed yet (in the relevant ways) and is generally surprised by what he is experiencing. At this time, the evaluative aspects that render him responsibility-apt still inflect and alter his experience of the world. This may be true of Cured Alex as well. Long before any adaptive preferences are formed, Alex might continue to feel anger at the kinds of behavioural changes that were imposed on him. If we think responsibility is tied to questions of whether or not the agent is liable to commit the same problematic actions, then our interests will be concerned with salient similarity of the relevant evaluative aspects that speak to whether he is liable to commit his violent acts again. The fact that Alex still contends with these evaluative aspects to some extent makes them deeply relevant to such an assessment even if it is unlikely (or impossible in Cured Alex's case) that he would indeed act in these ways. The fact that he holds these aspects nevertheless says something about who he is in the world.

Finally, using the notion of salient similarity, we might also say that it is possible to undergo radical change, but remain responsibility-apt. Consider the young socialist again. For instance, not only might it be possible that he no longer reads, loves his wife

and engage in the same activities as before, all of these changes could potentially occur without affecting his socialist ideals. As long as those evaluative aspects that pertain to socialism still inflect his experience he may be saliently similar. When the young socialist peers into the future, his worries concerning future affinities could be calmed if he retains a number of beliefs concerning equality and social justice even if he is generally unrecognizable in many other respects.

As for Alex, when it comes to being responsibility-apt, I doubt the prison administration would be too pleased if Alex was radically changed and lost the majority of his trivial beliefs but still maintained sadistic impulses. His love of violence is significant in a way other aspects are not and as long as these persist he would still be responsibility-apt. In either case salient similarity would hold despite what may otherwise appear to be radical change.

Overall, it is possible that the young socialist or Alex is no longer responsibility-apt despite a lack of radical change. It is also possible that they remain responsibility-apt despite experiencing deep changes in personality. To determine salient similarity, first we would need to locate which aspects are relevant to the determination. In the young socialist's and Alex's case, these would be those that pertain either to socialism or violence. Secondly, we would need to determine whether each retains enough evaluative aspects as to satisfy a threshold guided by one's interests. In the two cases, this threshold may be met when the young socialist's worries are allayed or the actions of Alex no longer suggest a threat of violence.

5. A Note about the Epistemological Problems

Whether we are discussing soup, coffee, a socialist agenda or responsibility-

aptness, we can derive an analogous response to each. What matters is whether each satisfies conditions of salient similarity. Thus, in all these cases, the problem of trivial beliefs and the problem of degree are answered with a resounding: "it depends!" The aspects under consideration as well as the threshold to be met are both determined in an interest relative manner. As I have argued, there is an answer to the problem of vagueness in this determination, but it is not exactly clear whether we can ever be assured of that answer. How it is even possible to know what evaluative aspects actually motivate the agent? We might not be sure as to whether A would be saliently similar if evaluative beliefs X and Y were replaced, but not Z. Or whether X and Y even influenced the agent in ways that should concern us. In this last section, I briefly consider some epistemological worries that are raised by my application of Graff's notion of salient similarity to the persistence of the moral self. I will suggest a couple of means to begin to allay these worries, even if a full treatment of this issue is beyond the scope of this thesis.

The possibility of knowing which evaluative aspects that bear on one's actions are relevant is not ever clear. We may only have an answer to these questions if we were able to fully understand which attitudes and motivations deeply influenced the agent's problematic actions, but practically speaking, it is not clear that we have or can ever have such knowledge. The problem here is similar to the one that may have led John Locke to introduce eschatological concerns into his account of the forensic unit. Locke noted the difficulty in truly knowing whether the now sober man could honestly not recall the actions committed while drunk. He states:

> But is not a man drunk and sober the same person? Why else is he punished for the fact he commits when drunk, though he be never afterwards conscious of it? … Human laws

punish both, with a justice suitable to their way of knowledge; because, in these cases, they cannot distinguish certainly what is real, what is counterfeit: and so the ignorance in drunkenness or sleep is not admitted as a plea.[363]

Locke's answer is to assure us that responsibility for one's actions will be fully known and assessed "on the great Day of Judgment" where all will "receive according to his doings, the secrets of all hearts shall be laid open." [364] Persons do not have the resources of a God to ensure that "no one shall be made to answer for what he knows nothing of", yet we can always strive to approximate this divine knowledge even if it is not complete.[365]

Similarly, even though we cannot have guaranteed knowledge of what aspects of the evaluative profile are actually retained by a person over time, this does not mean we cannot make an informed decision. For instance, it is clear that song lyrics, smells or thoughts that spontaneously occur to the person are transient enough not to be attributed to their ongoing evaluative profile. Yet it would be plausible to attribute these aspects to the evaluative profile if they were consistently repeated.[366] For instance, when a criminal

[363] Locke, *Essay,* II.xxvii.26.
[364] Ibid., II.xxvii.26
[365] Ibid., II.xxvii.22.
[366] Smell in particular can draw up vivid and affectively toned autobiographical memories in a phenomenon known as the "Proust Phenomena" (Toffolo et al. "Proust Revisited", 84.). Named after an often-quoted line in Marcel's Proust "*Swann's Way*", this phenomenon shows a deep connection between olfaction and memory. In particular remembrances due to smells have been shown to be more vivid than the other senses. Yet, on my account, flashbulb memories and other fleeting ways of recalling autobiographical memories do not forge a connection to a past self even if it is possible that the current agent is sufficiently similar in their evaluative aspects. After all, the smell "can only repeat indefinitely with a gradual loss of strength, the same testimony" which is not enough to say that it is the same self. The work in saying one is the same is being accomplished by similarity of evaluative profiles and not necessarily vivid recollection. See Toffolo, Marieke B. J, et al. "Proust Revisited: Odours As Triggers of Aversive Memories." *Cognition and Emotion*, vol. 26, no. 1, 2012: 83–92 and Proust,

begins to repeat poor behaviour he thought was lost, this may be taken to be significant or at least an indication of something deeper occurring in his usual motivations. It stands to question whether the evaluative beliefs connected to that behaviour have necessarily been replaced. This may be mere speculation, but the main point stands that when we are asking of whether the person before us is apt to display the same objectionable attitude. This here, is one way we might be assured in determining whether the same self persists – by determining his character over a period of time and judging whether the patterns of behaviour are consistently similar to the past self. We need a longer range to understand what is and what is not part of her character as the influences and particular evaluative beliefs are in continual flux from one specific moment to the next. The fact that it is in flux and is generally 'gappy' is not necessarily problematic in that we are not asking the question of strict identity, but character.

These epistemological concerns are representative of some larger methodological questions that I can acknowledge, but not fully address here. Perhaps one day we would have the technology and insight to determine the exact composition of an individual's attitudes and beliefs, but as Locke reminded us, we do not yet have the resources of a god to judge with such precision. We can at least say that being the same self for purposes of responsibility attribution, as I have argued, is a looser pursuit than traditionally and historically thought.

Conclusion

Marcel. *Swann's Way: The Moncrieff Translation*. Edited by Susanna Lee. Translated by C. K Scott-Moncrieff, W.W. Norton & Company, 2013.

I have argued in this chapter that the kind of indeterminacy that characterizes everyday change is not necessarily a problem for the account thus far. What is important to determine continued responsibility-aptness is not how much the agent's evaluative profile has changed, but what exactly has been changed and the purposes there are in making the assessment. The line of inquiry would then be sufficiently narrowed to only the evaluative aspects that relate to the act in question. Retaining the relevant evaluative aspects also retains evaluative access to a sufficient and interest relevant sense. In consequence, when salient similarity obtains, so does answerability (and responsibility-aptness in consequence), which means that in order to be answerable the agent need not hold all her former beliefs and attitudes, just the ones that are relevant to our line of inquiry. Agents can be answerable for certain particular past acts and this is consistent with much personality change even if more than a few epistemological worries are raised. Determining the extent of being responsibility-apt will never be exact and always be open to error. I have suggested a few means to make this determination more precise, but a full treatment of these epistemological worries is beyond the scope of the thesis. It is to this extent that we should heed the warnings from Locke and humble ourselves to always be open to changing our minds when determining continued responsibility-aptness.

What matters for rehabilitation?

Chapter 7: Extension of the Self through Narrative

Introduction

In the last couple of chapters, I have argued that replacement or indifference marks a loss in responsibility-aptness. This loss can in turn be quantified and determined by the degree of salient similarity the agent's current evaluative profile bears with her former one. In this chapter, I would like to return to Marya Schectman's narrative identity thesis, first seen in chapter five, in order to show how the articulation of a narrative thread extends answerability in a way that allows the agent to *take responsibility* for their pasts by rendering the agent saliently similar and, hence, answerable by her own accord. Of course, we might question why one would want to remain answerable in this manner and I will address these concerns in chapter eight. I will show that this process is advantageous to securing rehabilitative aims and warranting forgiveness, which is importantly distinct from simply no longer being responsibility-apt. For now, the focus of this chapter will not be on rehabilitation itself, but will be an effort to rehabilitate the concept of narrativity against some common criticisms that sees them as descriptively inaccurate and undesirable. I will then use this defence as an opportunity to advocate for a particular version of the narrative identity thesis that will highlight certain features that allow for salient similarity to obtain. The narrative identity thesis is thus deeply relevant for questions of answerability over time, not because it determines the extent of identity, but insofar as it can be used as a tool for appropriating lessons learned from one's past and extend a sense of salient similarity over time.

The chapter will proceed as follows: In §1 I will detail what I call the *epistemological problem* that sees narratives as inherently risky insofar as they provide

emotional or thematic meaning rather than an accurate description of events. The second problem or what I call the *conditional problem* explored in §2 sees narratives as always potentially being upturned in what Andrea Westlund terms, "trajectory frustration."[367] Section 3 will then look at Galen Strawson's objection I call the *necessity problem.*[368] Far from being detrimental to the account given thus far, I will argue that the features of the narrative identity thesis these critics see as problematic are instead what give narratives their value. Narrativity can extend evaluative access in a way that satisfies the salient similarity condition through its backward and forward looking functions: backward (as will be explored in §4), in the narrative's ability to highlight important episodes, and forward (§ 5) in the ability to bring those former values to conscious experience. Agents are thus able to take responsibility for the past through narrative because the articulation of such extends answerability by way of these two features.

The features of the narrative highlighted in this chapter will ultimately shed light on my initial questions concerning the responsibility-aptness of rehabilitated criminals. In the end I wish to show that, even though sufficient change in the relevant ways may amount to a loss of responsibility-aptness, rehabilitative aims require more than this sort of responsibility undermining dissolution. Asking if the criminal remains responsible and asking if he is rehabilitated are two separate questions with former involving the loss of salient similarity and the latter the retention of it through a narrative approach.

1. The Epistemological Problem

[367] Westlund, Andrea. "Autonomy and the Autobiographical Perspective". Oshana, Marina. In *Personal Autonomy and Social Oppression: Philosophical Perspectives*, (New York: Routledge 2015): 65.
[368] Strawson, Galen. "Against Narrativity." *Ratio* 17, no. 4 (2004): 428-52.

Self-narratives on Schectman's account, as we saw in chapter five, function as organizing principles that integrate experiences along emotional and thematic connections. David Velleman illustrates this connection by drawing on an example given by Aristotle, in which a statue of the murdered falls on the murderer. If we were seeking a causal explanation, then these series of events do not imply one another. Yet, a story of karmic revenge can be told in terms of a plot.[369] Velleman states, "Although these events follow no causal sequence, they provide an emotional resolution, and they so have meaning for the audience…" [370] Chronologically and causally speaking, the events are distinct and bear no particular relation to one another. It is in the ability to create coherence between two events that are not causally related and create an "appearance of a design" that allow the plot to carry a distinctive value beyond a simple chronological list of events.[371] Yet, this sort of reconstruction also provides the basis for criticism as well.

One persistent critic of narrative identity is Peter Lamarque, who argues that to see one's life as an unfolding story "is to aestheticize, if not fictionalize, real lives." [372] He maintains that narratives distort and detract from self-knowledge despite being positioned as a means to such an end. Real lives do not contain many of the elements that

---

[369] As usefully summarized by Paul Ricœur, the plot or "*muthos*" found in a narrative is " 'the combination [*sustasis*] of the incidents of the story'. The character is what confers coherence upon action, by a sort of unique 'purpose' underlying the action."(Ricœur quoting Aristotle:1450a 15, 1450b 7–9, *Rule of Metaphor*, 40). For a fuller treatment of Aristotle, see Ricœur, Paul. *The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language*. (University of Toronto Press, 1993).

[370] Velleman, David. "Well-being and Time" Metaphysics of Death. Ed. Martin Fischer. (Stanford: Stanford University Press, 1993): 6.

[371] Ibid., 8.

[372] Lamarque, Peter. "On the Distance between Literary Narratives and Real-Life Narratives." *Royal Institute of Philosophy Supplement* 60 (2007): 132.

make up a literary fiction and to apply these to our lives is to twist a non-fiction into a fiction. Galen Strawson also finds fault with narratives in their "form finding" or "story telling" tendencies.[373] These are problematic due to their close connection to a tendency towards revision insofar as memory already "deletes, abridges, edits, reorders, italicizes."[374] He continues, "The implication is plain: the more you recall, retell, narrate yourself, the further you risk moving away from accurate self-understanding, from the truth of your being.[375] If Lamarque and Strawson are right, then narratives risk a distorted sense of oneself through an incongruous analogy to literary works. The benefit of an emotional understanding of our lives advocated by Velleman is precisely what fictionalizes these narratives in a way that leaves the aim towards truth in question.[376]

Others, including Daniel Dennett, have embraced these story-telling aspects of narrative identity, but also do not deny the inherent epistemological risk associated with them. Following Hume, Dennett argues that the self is a construction of the imagination that is weaved into conscious experience when the latter is confronted with the flux and rapidity of perception. The self is a fiction imaginatively invented out of the necessity of conceiving ourselves as a unified whole. This sense of self is a fiction, but as Dennett notes, "it's a wonderful fictional object, and it has a perfectly legitimate place within

---

[373] Strawson, "Against Narrativity", 441,442.
[374] Ibid.,444.
[375] Ibid.,447.
[376] I understand "fictionality" here as defined by David Davies. To be fictional, a fictive utterance needs to be further guided by a general disregard of what Davies terms, the "fidelity constraint", which assumes that "the selection and temporal ordering of all the events was constrained by a desire, on the narrator's part, to be faithful to the manner in which actual events transpired" (Davies, *Aesthetics and Literature*, 46.) That is, a work would be fictional if were not held by this constraint. See Davies, David. *Aesthetics and Literature*. (London: Continuum, 2007).

serious, sober, *echt* physical science." [377] Like the notion of a center of gravity as it occurs in physics, these fictions set a frame that can be used to explain, predict and manipulate objects. As self-aware beings, we are not only aware of change over time, but also endeavour to make this change intelligible to ourselves. Given this need combined with an essential story-telling tendency, persons strive "to make all of our material cohere into a single good story. And that story is our autobiography." [378] The narrative is an organizing principle that integrates events and experiences into intelligible temporal sequences without a necessary eye to accuracy, but through a need to keep track of events and spin a good yarn.

There is also a protective function alongside the explanatory support the narrative provides. By distancing ourselves from behaviours that are destructive and upsetting though the way we edit our tales, we shield ourselves from unwanted and paralyzing truths. Dennett states, "Our fundamental tactic of self-protection, self-control, and self-definition is not spinning webs or building dams, but telling stories, and more particularly concocting and controlling the story we tell others—and ourselves—about who we are." [379] As a sort of defence mechanism, the emotional closure and connections between events delivered by narratives may shrink the boundaries of our self-concept and be a means of insulating ourselves from unwanted truths. The potential for

[377] Dennett, Daniel C. "The Self as a Center of Narrative Gravity." In *Self and Consciousness: Multiple Perspectives*, edited by F. Kessel, P. Cole and D. Johnson. (Hillsdale, NJ: Erlbaum, 1992).
[378] Ibid.
[379] Dennett, D. C. *Consciousness Explained*. (Boston: Little, Brown, 1991): 418.

falsification seems to be written into the very reason we articulate these narratives. It is a process of finding form where there may be none inherently.

Narratives place thematic constraints above a faithful representation of reality and this is the basis for the epistemological problem. The inclusion and exclusion of certain events serves to imbue our lives with unity and is not necessarily aimed at a faithful recreation of events. The narrative structure is thus susceptible to what Paul Ricœur calls "productive invention" as a means of ordering the world in a certain manner. In a process he calls "emplotment" actions undertaken in our lives can be transformed by plugging the events in one's life into something like a literary plot. He writes of emplotment, "It 'grasps together' and integrates into one whole and complete story multiple and scattered events, thereby schematizing the intelligible signification attached to the narrative taken as a whole." [380] Given that narratives "give[] form to what is unformed", Ricœur notes that they are "suspected of treachery" or seen with a "suspicion of interpretive violence."[381] He continues:

> … At best, [narratives] furnish[] the 'as if' proper to any fiction we know to be just fiction, a literary artifice. This is how it consoles us in the face of death. But as soon as we no longer fool ourselves by having recourse to the consolation offered by the paradigms, we become aware of the violence and the lie.[382]

Thus, if we are to think of our lives as genuinely like literature, this can mislead by creating a world filled with meaning and purpose and as a product of a grander design. Especially in the advent of a tragedy, persons may search for a deeper meaning in a series of events and prefer to tell themselves that the reason for this occurrence has yet

---

[380]Ricœur, Paul. *Time and Narrative: Volume One*. Translated by Kathleen McLaughlin, and David Pellauer (Chicago: University of Chicago Press, 1990): x.
[381] Ibid.,72-73
[382] Ibid.,72

to be discovered. However, as they excavate the events of the past, they often come up empty-handed and are left with a series of coincidences, random encounters and senseless heartaches. There is a large gulf between real lives and those dramatically depicted in a literary plot. Narratives, it would seem, are not in the business of providing a closer inspection of the world, but may be better seen as a turn away from it.

2. The Conditional Problem

The criticism so far is that narratives are inherently epistemologically risky. They tend towards fictionalizing knowledge of the self and the world one inhabits.[383] Yet, these criticisms say little of whether a narrative, when unhampered by self-regarding motives and ignorance, can be good for the agent. At most the argument could be that persons are generally bad storytellers, not that storytelling is necessarily bad. However, some may criticize narratives not just because they can be false, but also because they are essentially conditional. This is the second of the two issues I will explore in this chapter — *the conditional problem*.

Andrea Westlund notes that any meaning provided by the narrative is "at best provisional" due to the "trajectory dependent" properties inherent in the narrative.[384] Drawing on the work of Karen Jones, she argues that narrative reconstructions are

[383] The risk here may even be larger than first assumed as new studies are now coming to show that even those with what is described as possessing "highly superior autobiographical memory" (who have the ability to photographically and comprehensively remember past events) succumbed to the temptations of confabulation just as much as those without such advanced memory recall. See Patihis, Lawrence, Steven J Frenda, Aurora K. R LePort, Nicole Petersen, Rebecca M Nichols, Craig E. L Stark, James L McGaugh, and Elizabeth F Loftus. "False Memories in Highly Superior Autobiographical Memory Individuals." *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 52 (2013): 20947-0952.
[384] Westlund, "Autonomy", 86.

always unfinished in their telling in much the same way as certain emotions are. Part of what allows us to properly describe the meaning of events over time depends on their trajectory or how further events unfold. The assertion of whether or not I am in the proposed state is vulnerable to the "way things turn out." [385] The trajectory dependent events and episodes that concern both Jones and Westlund are those in which we have "(i)temporally extended truth makers such that (ii) whether it is correct to ascribe a trajectory dependent property to A at t depends on what happens elsewhen, whether at t+n or at t-n".[386] So even if the narrative provides a particular interpretation of events, that interpretation may always be vulnerable to frustration.

Whatever meaning we can gain from narratives seems to depend on how what we are narrating eventually unfolds. Before one's story can fully unfold, even if we could name a time when that could occur, there is no determinate fact about one's life story. What is important is that later episodes can change the meaning of earlier ones, not by changing the past but by changing how we come to see it. For instance, a romantic comedy might lose the 'romantic' adjective if the characters never felt later affection. We would just have a story of how two people once hated each other and continue to do so. If instead they do declare their love, this event forces us to reinterpret their intense feelings earlier. Maybe it wasn't hatred after all, but love in disguise. One interpretation is more readily supported than the other and past events get a new gloss given the audience's subsequent interpretations. This is how events that occur "elsewhen" can

[385] Jones, K. "How to Change the Past". In *Practical Identity and Narrative Agency,* edited by K. Atkins & C. Mackenzie (New York: Routledge. 2008): 271
[386] Ibid.,272

undermine what we once thought was true.[387] What counts as love depends, not only on certain recognized feelings and thoughts toward the person that are generally accepted as signs of love, but also how those feelings and actions are understood within the context of a larger temporal whole. The trajectory of how those feelings unfold constitutes a truth maker for being in love. Narratives are never decisively correct given the trajectory dependence of their interpretations. The identity of certain feelings and states are conditional on how events unfold.

3. The Necessity Problem

Narratives, as the first two criticisms show, are liable to be mistaken through their form-finding tendency. Even if that form is accurate, the interpretation of events is conditional. Strawson suggests that not only is the narrativity ethically dangerous, the inherent risk they pose is not necessary to live a life. He looks to his own experience to argue for the possibility of what he calls an "episodic" life as one that is not constrained by a form, narrative or otherwise.[388] Using himself as an example he states:

> I have a past, like any human being, and I know perfectly well that I have a past. I have a respectable amount of factual knowledge about it, and I also remember some of my past experiences 'from the inside', as philosophers say. And yet I have absolutely no sense of my life as a narrative with form, or indeed as a narrative without form. Absolutely none.[389]

Strawson suggests that the "fundamentals of temporal temperament" vary widely among people, so much so that these "episodics", as he calls them, do not consider themselves "as some- thing that was there in the (further) past and will be there in the

[387] Ibid., 272
[388] Strawson, Galen. "Episodic Ethics." *Royal Institute of Philosophy Supplement* 60 (2007): 106.
[389] Strawson, "Against Narrativity", 433

(further) future."[390] They can take stock and acknowledge the past, but simply do not see themselves as the experiencer in the moment, as something extended. Thus Strawson does not deny that people can maintain a narrative view of themselves, he simply claims that such a view is not necessary for self-constitution. I call this the *necessity problem*.

Indeed, even though Strawson provides us with little proof that such an episodic life is even possible (beside his own self-reporting), his criticism reveals that there is a deep need to explain why we might want to engage in such a project given the risks involved. Strawson states, "The aspiration to explicit Narrative self-articulation is natural for some – for some, perhaps, it may even be helpful – but in others it is highly unnatural and ruinous. My guess is that it almost always does more harm than good." [391] While I accept the first half of this claim and do not think that narrating one's life is necessary in order to live that life or provide a basis for identity, I also hold that there are some very good reasons to do so. If we at least grant the bare possibility of an episodic life (which I think can be supported by some real life cases that Strawson does not explore), living this sort of life without understanding it in narrative form is not desirable even if descriptively possible.

In what follows, I will argue that the apparent problems with narratives as form finding, conditional and potentially false are based on an interpretation of narratives that see them primarily as a means to recount the events of the past. If, however, we analyze the backward and forward-looking functions of the narrative, we may reframe the narrative thesis as a process to ground certain agential capacities. The three problems

[390] Ibid., 430, 431.
[391] Ibid.,447.

with narratives as a result become not only features of the account, but help explain why narratives support agential capacities. There is also an important upshot to this agential function of the narrative that is significant for our purposes. In particular, the backward and forward-looking functions of narratives help retain evaluative access to the past and, as a result, maintain salient similarity. Narratives thus not only support agency, but may also extend answerability via salient similarity as a result. In the following section, I will start by looking at the backward-looking function of the narrative to draw out the benefits of emotional and thematic meaning the narrative provides.

4. The Backward-Looking Function to Narratives

It would seem that any benefit derived from narrative explanation is purely aesthetic and will not lead to the truth of who we are. If we are seeking to know ourselves with any sincerity, understanding our lives as a kind of fiction seems to speak directly against this interest. Nevertheless, I would argue that the backward-looking function of narrative understanding does not necessarily aim at seeking truth through narrative, but rather at saliency through highlighting a particular narrative frame.

Consider an analogy given by Gordon Graham.[392] A chronicle of events of my past might function like a map. Just as we check a map against the actual geographical layout of the land, our general history can be confirmed or disconfirmed through third-person accounts and empirical evidence of the events in our pasts. This construction of a map of who I am and what I have done differs from a narrative, which is more like an artistic photograph. A map faithfully corresponds to the world showing the actual relation of

[392] Graham, Gordon. *Philosophy of the Arts: An Introduction to Aesthetics*. (London: Routledge, 1997).

each feature to one another, whereas a photograph might present a singular and contrived perspective. Both provide knowledge of a landscape, yet the photograph filters reality from a particular point of view and has special interest in what is shown. As a form of art, a photograph employs focus, composition and colour to achieve a distinct conception of reality. The narrative and photograph are alike in the sense that each offers a singular point of view in excluding the reality that lies beyond it, while using artistic means to relay meaning. We may be looking at a photograph of a singular flower surrounded by a wasteland hidden outside its frame. We are not privy to seeing what eventually happens to that flower and, depending on how these later events unfold, this can change how we view the initial picture: is it burgeoning life, or a story of extinction? As the epistemological objection goes, the form-finding tendency of both the photograph and narrative blind us to what is outside the specified frame.

If we are seeking knowledge or understanding of ourselves, it is not clear why we would favour a limited and possibly defective photograph over a map. I suggest, however, that it is not the purpose of an artistic photograph to accurately represent the entirety of a situation, nor is its whole value and merit given by function of imitating the world to us.

4(a). Divergence from Reality

When we appreciate artworks *qua* art or literature *qua* literature, it is their "artfulness" or "literariness" that "commands our attention." [393,394] If we wanted knowledge of our past, then we might think it best to turn to something like a map.

[393] Carroll, Noël. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *Journal of Aesthetics and Art Criticism* 60.1 (2002): 28.
[394] Ricœur, *Time and Narrative: Volume 1*, 52.

However, consider Jorge Luis Borges' cartographer who creates a map that "attained such Perfection that the map of a single Province occupied the entirety of a City."[395] Despite its perfection, the one to one correspondence made the map "Useless" as it entails a replication of reality that undermines the purpose of having a map in the first place.[396] As Paul Ricœur argues, to tell a history is to employ the form finding tendency for a different purpose.[397] Histories "abridge", "delete", and are always from a perspective in the same manner as a fictional narrative in order to render history comprehensible and useful.[398,399] Likewise, if we are looking to narrative to provide an exact recreation of the past, we are not going to find it. History uses elements of a fiction in service of rendering a series of events comprehensible, while fiction uses a series of events to give content to a different kind of truth found within the form finding tendency. In fact, as Ricœur argues, to look for that recreation in a fictional narrative at all is to miss the point. Indeed, the value of the narrative lies not in factual one to one correspondence as with Borges' ideal map. That is simply not the function of a narrative.

---

[395] Borges, Jorge Luis. *Collected Fictions*. Translated by Andrew Hurley. (New York, N.Y., U.S.A.: Viking, 1998): 325.
[396] Ibid.
[397] Indeed as Paul Ricœur notes, "A history book can be read as a novel" because there is a "gap" between what occurred and how those events are recounted. History is told from a perspective and given the fact that the perspective can shift gives reason to think the work of the historian is not entirely objective. Ricœur states, "We have not forgotten the gap between time of the world and lived time is bridged only by constructing some specific connectors that serve to make historical time conceivable and manipulable"(Ricœur, *Time and Narrative: Vol. 3*, 181). See Ricœur, Paul. *Time and Narrative: Volume 3*. Translated by Kathleeen Blamey and David Pellauer. (Chicago: University of Chicago Press, 1984): 181.
[398] Strawson, "Against Narrativity", 444.
[399] I would like to thank Frédérick Armstrong for drawing my attention to the similarities between fiction and history I had previously not explored and inspiring me to include references to Borges' cartographer.

Instead, its value can be found is in the artfulness of the form finding tendency in which the form itself is revealing. [400]

What I am suggesting here, alongside Ricœur, is that the narrative theorist may reply to criticisms like those given by Lamarque and Strawson (the epistemological problem) by claiming that their they simply miss why we engage in constructing narratives. The manner in which lives are organized into digestible wholes is not unlike literature because doing so allows us to gain a more general truth not graspable by viewing each incident in isolation or in chronological order. [401] The purpose of telling a narrative is not the same as telling a history. Historians aim to record the specific events of the past. Telling a narrative aims at truth as well, but without necessarily a faithful recounting of events. The connections made by the "emplotment" of one's actions into a narrative frame provides the "literariness" of literature, but also represents how time is experienced in the ebb and flow of conscious experience that does not necessarily follow

[400] Indeed, as noted by Ricœur, "History recounts what has happened, poetry what could have happened. History is based on the particular, poetry rises towards the universal: 'By a universal statement I mean one as to what such or such a kind of man will probably or necessarily say or do'" (Ricœur quoting Aristotle:1451b 9, *Rule of Metaphor*, 44).

[401] The idea here is similar to Ricœur's interpretation of Aristotle's notion of Mimêsis . As the "concept of *mimêsis* is narrowed down remarkably in passing from Plato to Aristotle", according to Ricœur, it is better understood as "the imitation of human action is an imitation that magnifies [and] ennobles." (Ibid., 42) It reconstructs what it recounts as it involves plot of the kind within narrative see earlier by connecting events that may not be causally related. Thus, mimesis involved in *muthos* or plot does not simply seek to recount because "the imitation is at once a portrayal of human reality *and* an original creation; on the other, it is faithful to things as they are *and* it depicts them as higher and greater than they are." (Ibid., 45) It imitates in order to better represent human action because the plot of a narrative does not just recount how things are in nature, but "serve[s] as an *index* for that dimension of reality that does not receive due account in the simple description of that-thing-over-there' (Ibid., 43) See Ricœur, Paul. *The Rule of Metaphor: The Creation of Meaning in Language*. (London: Routledge, 2003).

a specific succession of events.[402] Time is framed in a way that imitates how the subject experiences that time. In other words, a narrative best captures our phenomenological experience of the world, not only because of it is ability to unify disparate events into a coherent whole, but it also organizes along what is significant to the 'gappy' span of the self by tracing thematic rather than purely temporal connections (as a sequence of past present and future). Agents do not experience the world as only what has been experienced immediately prior, but is framed by what is relevant to one's current context. As Ricœur would argue, time, as faithful to how the present is generally inflected by past experiences and future expectations, thus becomes "human time to the extent that it is organised after the manner of a narrative."[403]

As I have argued before in chapter one, the moral self need not require a transitive relation because we are concerned with the constitution of an agent's evaluative profile as they stand, regardless of how that constitution came about. Responsibility-aptness concerns evaluative sensitivity alone. So a narrative need not trace the complete history of the person in order to capture what is important to the current moral self. Narratives are able to capture an agent-centric thematic truth that represents events how the agent experiences them. Narratives essentially provide an easily digestible snap shot of 'who the person is' that aims to be faithful to general thematic truths about the person without necessarily having fidelity to the specific temporal succession of events. It is this sort of self-knowledge that carries moral import because what the agent needs to know is not

[402] Ricœur, *Time and Narrative: Vol. 1*, 52
[403] Ibid., 52

specific events in one's life, but how these events related to one's evaluative complexity and current perspective.

If the concern is on thematic truth, which ties together the gappiness of the self and the complexity of the evaluative profile, certain factual inaccuracies are not necessarily going to undermine the integrity of a narrative. If the narrative can capture what it is like to be the person they are as a general thematic truth, then it is an accurate narrative even if the truth is stretched or omitted in some non-egregious instances. As Schechtman argues, what matters is whether the narrative expresses an accurate "general view of the person", such as the belief that one is a witty person even though the "narrative contains a specific recollection of making someone else's witty remark." [404] The memory of the witty remark is not identity constituting, even if the self-ascription is.

Yet, Schechtman is also careful to set a limit on how much the narrative can diverge from reality and I would be remiss not to borrow from these insights. Accuracy may be determined by both faithfulness to the facts and the cultural understanding of the trajectory of one's life. Thus, the structure of the narrative is limited by social norms and expectations. As Schechtman states, "To enter into the world of persons an individual needs, roughly speaking, to grasp her culture's conception of a person and apply it to herself." [405] Individuals constitute themselves into people in order to engage in person-related activities in a particular social setting. When culturally situated in this manner there is a wide range of typical narrative styles and standards that allows the general organization of a narrative to remain quite flexible. There are times, however, when

[404] Ibid., 128.
[405] Schechtman, *Constitution,* 95

divergence from this spectrum of generally acceptable person-constituting self-conceptions becomes too great and stretches the bounds of credulity to the point that the agent can no longer interact in the world of persons. Hence, inaccuracy in the narrative is generally only problematic when it no longer allows the agent to function within the world.

Minor inaccuracies are not necessarily a problem if the general, thematic interpretation is correct. Yet, even if we stand in the correct knowledge relation to our past, a narrative will always be contingent on how the events of our lives are framed. So, even when elements of the narrative are fully faithful to the facts, it is also possible that the interpretation is inaccurate enough that it is not identity constituting. For instance, no matter how much I try to frame certain actions as love, the description of abuse may be more readily applicable. Given what is already there, "the description that one accepts may not be able to get enough purchase within one's motivational and cognitive economy." [406] Likewise, an interpretation would be in question if one was self-deceived or ignorant of their past. Take Schechtman's example of a resentful sibling whose actions towards his brother betray seething jealousy despite his professed love.[407] It would not be okay if the jealous brother continuously misunderstood the significance of his actions towards his brother as something like sibling playfulness. He would then be blind to how his jealousy informs how he approaches the world and miss an important thematic truth about who he is.

[406] Jones, "How to Change the Past", 283
[407] Ibid., 95

Overall, the form finding tendency of the narrative is not necessarily detrimental. In fact, following Ricœur we might even say that the narrative better captures how events are felt and lived than would be given by a chronicle of these events. Narratives "abridge" and "delete", but in a way that highlights more general thematic truths. Yet, this does not mean that absolutely anything goes if we also follow some insights provided by Schechtman. [408] In particular, the narrative needs to have some eye toward accuracy of the events of one's life if the agent is going to live and interact within the world of persons. Narratives thus can accommodate some inaccuracy without stretching credibility and undermining thematic truths.

4(b). Saliency and the Narrative

Whether we are concerned with the facts or the interpretation, accuracy is key, but not so much as to force us to necessarily disavow some of the more aesthetic tendencies inherent in narrative self-construction. This is important because these tendencies towards thematic meaning may be key to fully understanding the backwards function to narratives. Following Gordon Graham once again, we see that art in general contains a different relation to the world that does "not need to be bound by the idea of correspondence." [409] That is, although we may not find a direct representation of the world, this does not mean no relation can be said to exist. To say that there is no connection is to rest on the assumption that the only viable relation is one that asks us to "look independently at reality and then at art in order to see how well the latter has

[408] Strawson, "Against Narrativity", 447.
[409] Graham, *Philosophy of Art*, 58

represented the former."[410] To see the value that the personal narrative, like art, has for agents, we need to reverse this order. We should look to art independently and subsequently see "reality afresh" and use art to properly become aware of our reality.[411] Once we reverse this relation between art and reality, we are no longer seeing particular aspects of ourselves within the narrative as aspects of our life history, but are seeing our lives as aspects of the narrative.[412] It is a way of bringing art to the world. The personal narrative is not a summary of who we are, but a way of "awakening our experience of the world." [413] Life consists of many encounters with different people, situations and circumstances. We can interpret these encounters with more or less imagination and find levels of meaning in everyday interactions.[414]

For Gordon, art plays a role in the "imaginative apprehension of experience" in ways where we might find ourselves relatively deficient.[415] Art, in this sense, suggests a way of coming to our experience with a ready schema rather than seeing it as a reflection

[410] Graham, Gordon. "Learning from Art." *British Journal of Aesthetics* 35.1 (1995): 34.
[411] Ibid., 34.
[412] This point is similar to Ricœur's thesis in that narratives are understood because we understand life, but by that same token, our understanding of narrative increases our understanding of life.
[413] Ibid.
[414] Indeed, breaking from a faithful recounting of events to tell a story might better serve to generate and highlight certain thematic truths. As Ricœur writes, "It is precisely when a work of art breaks with this part of verisimilitude [(as having the appearance of being real)] that it displays its true mimetic[(as the imitation of human action)] function" (Ricœur, *Time and Narrative: Vol. 3,* 191). The constraints of telling a history such as "documentary proof" are not in place for a fiction and it is this "freedom from" that gives rise to the fiction's "freedom for" artistic creation that depicts general thematic truths that cannot be captured by listing a sequence of events.(Ibid., 92) For instance, fiction captures the more emotional connections that would, as Ricœur would argue, "give eyes to the horrified narrator" when recounting events such as the holocaust. (Ibid.188). See Ricoeur, Paul, Kathleen McLaughlin, and David Pellauer. *Time and Narrative, Volume 1*. (Chicago: University of Chicago Press, 2012).
[415] Ibid., 35.

of it. It affords the 'careful reader' an opportunity to understand different ways of coming to the world. To use art or narrative to inform our experiences is not to impose it onto the world, but rather it allows us to be alive to aspects of the world that may not be readily seen. One benefit of the backward-looking aspects of the narrative is not based on accurately representing the past as it was, but in highlighting the values that are important to the agent as she moves forward in life.[416] This is why the conditional

[416] Overall, the process can be framed as a continual cycle of interpretation and reinterpretation. This may also be understood through Ricœur's three-step process of mimesis. Although based on an Aristotelian notion of art imitating nature, in Ricœur's sense, *mimêsis* refers to how narrative is imitative of action and this process can be separated into three stages he refers to as Mimêsis $_{1,}$ Mimêsis $_2$ and Mimêsis $_{3.}$ Mimêsis $_1$ is a kind of prefiguration as the general structure of elements the narrative will be organized around whereas mimêsis $_2$ concerns 'emplotment' as *configuration* in the organizing the various elements into an intelligible whole. This configuration is what draws out general and thematic truths as a structure is imposed to achieve some purposeful organization. Mimêsis $_3$ is *refiguration*, the act of reading whereby our understanding of the world is increased by the new interpretation provided by the narrative. Taken together, mimiesis$_1$ provides the basis for what is expected. Mimêsis $_2$ sees if these expectations configured into the general plot the story overall – the grasping together of these elements within this general prefiguration and mediates between Mimêsis $_1$ and Mimêsis $_3$. Then Mimêsis $_3$ as providing the new understanding will become the new prefiguration as Mimêsis $_1$. In narrative the preconfiguration is configured into the plot in order to achieve a new understanding, while the present experience is mediated by the past. Like the narrative process as well, this cycle of prefiguration and refiguration is never fully complete and meaning is never fully settled. As Ricœur argues, "Emplotment [as a thematic structuring of one's narrative] is never the simple triumph of order"(Ricœur, *Time and Narrative: Volume 1*, 73). To expect otherwise of a narrative is to miss "the "dialectical character of their relationship" because there is "never a time in which we can speak of a human life as a story in its nascent state, since we do not have access to the temporal dramas of existence outside of stories told about them by others or by ourselves"(Ibid.,73). As Ricœur argues, although this process is cyclical in nature, it is not vicious. He considers it a "healthy" circle because it increases our understanding with every pass (Ibid.,72). It is more like a spiral given that with each pass, new knowledge is gained. He states, "The manifest circularity of every analysis of narrative, an analysis that does not stop interpreting in terms of each other the temporal form inherent in experience and the narrative structure, is not a lifeless tautology. We should see in it instead a 'healthy circle' in which the arguments advanced about each side of the problem aid one another" (Ibid,.76). The importance of

problem is not necessarily a problem. Persons can benefit from the ability to frame and reframe their experiences.

4(c). Interpretation Sensitivity and Salience

The key benefit of seeing one's past through a narrative frame is to see its value outside and apart from a comprehensively accurate representation. It frames our experiences both in the past and into the future. The ready analogy to the backward-looking aspects of narratives have to art tells us that engaging in this process is cognitively valuable not because it tells us about the world, but is a tool for seeing the world within a particular frame.[417] Graham argues:

> The value of a picture lies not in supplying an accurate record of an event but in the way it enables us to look at the people, circumstances, and relationships in our own experience. The question to be asked of such a work is not, 'Is this how it really was?' but rather, 'Does this make us alive to new aspects of this sort of occasion?'[418]

Like Graham, Westlund too sees the value of the narrative, not as only that of answering questions concerning 'what happened?' but in its inherent aspirational nature.[419] She argues that narrative arcs are not just trajectory-dependent, but are

---

the healthy circle shows how the conditional nature of the narrative does not necessarily lead away from truth. Narrativity might involve continual interpretation and reinterpretation, but it is nevertheless a process that could potentially lead to a better understanding of one's current experiences. We can understand narrative because we understand life, and our understanding of narrative increases our understanding of life.

[417] Paul Ricœur argues that history borrows from fiction and fiction also borrows from history. But in recounting the past in order to draw out more general truths of the world, this "[v]ersimilitude" has been "confused with a mode of resemblance to the real that places fiction on the same plane of history" and thus leading to the error in thinking the purpose of fiction is the same as recounting a history (Ricœur, *Time and Narrative: Vol. 3*, 191.)

[418] Graham, *Philosophy of the Arts*, 60.

[419] Westlund, "Autonomy", 96.

"interpretation sensitive" as well.[420] If we consider love once again, we can see the importance of this sensitivity. What will count as love will be dependent on the availability of cultural scripts to name and identify my actions. Not just any actions, thoughts and feeling count as love. I cannot assert without question that it is love I feel if my actions do not cohere with generally accepted scripts that define love. Feelings that were once perceived as hatred might well fall under the umbrella of love-like feelings if those intense feelings of hatred evolved into passionate feelings of affection. So, there may be only certain feelings that correspond to culturally defined instances of love, which not only tells us what a series of events and actions amount to, but also that interpretation will help shape one's actions moving forward.

To be interpretation-sensitive in the manner suggested here does not just mean one's actions are open to interpretation, but that the fact that they are open may then help determine the events that unfold thereafter. Karen Jones states:

> An interpretation-sensitive trajectory has relatively structured rules governing the required kind of unfolding such that agents who conceptualize and endorse their activity under that description are more likely to bring it about that the resulting trajectory meets the relevant conditions than those who do not.[421]

Once the protagonist accepts that his actions were more readily indicative of an underlying love, it is more likely that his actions thereafter will conform to the script. Conceptualizing the activity under the relevant description brings it about that the activity can be properly described under that description in the future. We can only determine whether someone is in love by looking at "the unfolding sequence of states and events of which A's state is part", not just at whether certain states obtain at a

[420] Ibid.,90
[421] Jones, "How to Change the Past", 274.

time.[422, 423] If the agent does not endorse the interpretation that seems to fit, there is room to take corrective measures and "take steps to disrupt the patterns that are beginning to emerge in her thoughts, feelings, and actions." [424] We can correct or at least affect the further production of these mental states once they have been named and identified. As Jones notes, "Not having available a name around which to organize one's as yet inchoate feelings can stop them from assembling in the way that they would were that name available. In this way, our emotional vocabulary shapes what emotions we come to experience." [425] When we identify a feeling, we can work with it, give it a name and shape how we will proceed from this knowledge.[426]

---

[422] Ibid, 275.

[423] What structure this unfolding should take may be up to debate. Social context and further larger narratives circulating within the culture may provide a blueprint to what is an acceptable trajectory. And, given the social influence, there is much room to disagree on what constitutes an acceptable or unacceptable trajectory. As Jones clarifies: "Whether someone is correct in claiming that they are in love depends on whether the trajectory they are embarked on matches, or coherently extends, the socially available templates for love-trajectories. There can be dispute about this, and such dispute is often normative. For example, when homosexuality was classified as a mental illness, same-sex love was often denied and relabeled narcissism or obsession" (Jones, "How to Change the Past", 280). Although clearly false (as the this narrative frame did not fit reality), social norms nevertheless determined what was an acceptable love trajectory.

[424] Ibid., 281.

[425] Ibid.

[426] This process may also be understood through the concept of practical identities. Consider for the moment the work of Christine M. Korsgaard. Although her account of identity relies on a conception of endorsement I would deny, we might use her account to highlight the connection narrativity has to agency. Korsgaard argues that much of what we do is only made intelligible by the background of our larger projects, which require a unity of motives and coordination. We construct a unity within our various mental states in order to carry out any semblance of a rational plan of life along the guidelines of the various practical identities we inhabit. These practical identities are "description[s] under which you value yourself and find your life worth living" and become principles of choice in deliberation (Korsgaard, *Self-Constitution*, 20.). Depending on the practical identity one occupies, different sorts of reason-giving import are established for particular actions by limiting or favouring certain judgements that are constitutive of

How further events unfold is sensitive to what sort of interpretation we attribute to past events, feelings, desires and episodes and, by applying the narrative structure, we can affect a future outlook in much the same way. Depending on how the narrative is told, what is salient can shift. For instance, if Novel Alex, reflecting on his former criminal life, chose a career as a youth worker, the negatively viewed past may be held as central to this decision. Alex is in a position to reconceive his past and fit a different interpretation to the trajectory his life took. The narrative of redemption in turn provides "salience" that helps us organize, synthesize and properly orient ourselves towards facts of our lives moving forward.[427]

4(d). Responding to Mismatches

There is empirical evidence that agents do indeed use narratives to frame events in their past in these ways. As noted by criminologist, Shadd Maruna, when desisting offenders speak of events of their past, they often use certain literary techniques to foreshadow the coming reformation. They are "good guys" that were mistaken as "bad guys."[428] He states of the way these ex-offenders frame their past:

---

one's self conception. These practical identities may in fact be essentially contingent. Some practical identities can be given by our native country or religion. They may also be a product of pure chance. They nevertheless provide a basis in which to act and depending on the identity adopted, different routes and ways of seeing the world become available. Practical identities relate to narratives because to have a narrative does not mean having a single plot that traces each event of our lives, but multiple narratives that, when articulated highlight different aspects of our experience. See Korsgaard, Christine M. *Self-constitution: Agency, Identity, and Integrity*. (Oxford: Oxford Univ. Press, 2009): 20.

[427] Elgin, Catherine Z. "The Laboratory of the Mind". In *Sense of the World: Essays On Fiction, Narrative, And Knowledge.* Edited by John Gibson, Wolfgang Huemer and Luca Pocci. (Routledge-Taylor Francis, 2007): 44.

[428] Maruna, Shadd. *Making Good: How Ex-Convicts Reform and Rebuild Their Lives*. (Washington, D.C: American Psychological Association, 2001): 88-89.

…often, there will be one bad guy who will show the occasional glimpse of redeeming personal integrity. This may he conveyed in a moment of hesitation or a lingering look back at a victim, but it will be enough to foreshadow an ending whereby this particular bad guy aids our heroes in some way, ensuring victory for the good side. Such an ending is only believable because of the use of foreshadowing scenes.[429]

By reconceiving the past, the ex-offenders are able to modify the past in a way that supports a self-conception that is consistent with their new rehabilitated identity. Maruna continues:

After all, not all of the roles played by participants in this sample have been deviant ones. All of the narrators have played the role of the thief or the junkie, but they have also occasionally played the loving parent, working-class hero, loyal friend, and so forth. By falling back on these other identities, they are able to deemphasize the centrality of crime in the life history and suggest that they were just normal people 'all along.' [430]

Depending on how the narrative is told, different aspects of the past are made salient and emphasize the agent's current identity. More will be said on how offenders view their past in this way in chapter eight. For now, notice that there is a very deliberate sense in which the ex-offenders are "mining" deviant episodes in their past in order to find a trace of positive qualities that are now made salient due to the new frame in which they understand themselves.[431] This shows that self-narratives are in fact deeply contingent and may not be wholly accurate. This fact does not, however, degrade the value of the narrative. Each narrative interpretation gives us a new way to see the world and different ways to be 'alive' to change. In Catherine Z. Elgin's terms, it gives us a means to "exemplify" our past to create a "readily available, easily interpretable sample" of who we are and who we could be.[432] Each narrative interpretation provides a distinct

[429] Ibid.
[430] Ibid,. 89.
[431] Ibid.
[432] Elgin, "Laboratory", 47.

means of approaching and organizing our experience depending on shifting circumstances and the evolving projects we find ourselves in.

So narrative understanding, rather than a means of reflecting the world to us, is a frame in which we can understand our lives from different perspectives.[433] Under these terms, the contingency of a narrative is crucial to its value. The narrative is a mechanism that allows for change in how we see the world as opposed to the "tried and true." [434] As Elgin explains, "The familiar ways of conceptualizing and manipulating things have served us fairly well. But sticking to the tried and true has its costs. We overlook a lot, do not know what we are overlooking and often we are not aware *that* we are overlooking."[435] As new features of our world are brought to light, we change how we act and react in light of these new conceptions of who we are. It allows the agent to approach the world differently when consistent revision is required. This then calls for an embrace of the fact that personal narratives are not singularly correct. Personal narratives do not equip us with the ability to recognize a specific truth, but ways of being. We can

[433] Schechtman briefly argues in "A Mess Indeed" that the rejection of empathic access and criticisms from Peter Goldie has led her to conceive of the narrative self as something like a "perceiver self" (Schechtman, "A Mess Indeed", 31). She argues, "My claim is that this appreciation [of the ability to take on multiple perspectives] generates a metaperspective, a point of view of the person as a whole which is present throughout these vicissitudes Raymond Martin calls this perspective that of the 'perceiver-self'. We experience the world, he tells us, as if one part of the self was split off from the flux of events as an observer, watching and recording the stream of our experience (Ibid., 31). I see this chapter as in line with these suggestions even though Schechtman does not fully elaborate on what is meant by this perceiver self. I see it as a means to inhabit multiple perspectives in order to understand the past while conditioning our future and doing so facilitates continued answerability. See Schechtman, Marya."A Mess Indeed: Empathic Access, Narrative, and Identity". In *Art, Mind, and Narrative: Themes from the Work of Peter Goldie,* edited by Julian Dodd. (Oxford University Press, 2016).
[434] Elgin, "Laboratory", 46.
[435] Ibid.

agree with Lamarque that such narratives offer interpretations of a life and, as interpretations, do not guarantee the full truth of who we are. But it is also true that narratives bring new ways of organizing experience and this has a value beyond a chronicle of events.

Saliency fluctuates and in the way the narrative is framed, we see a reflection of the agent that framed it. There is also a kind of freedom that one achieves through the acceptance of what one cannot control. Freedom in the face of fate, according to Westlund, is freedom in the "joyful acceptance of necessity;" narrative recall can transform an experience.[436] We cannot change the events of the past, but we can work with them, interpret them and redeem them in light of an ongoing narrative. Westlund states "…adopting a narrative of self-realization or redemption over one of dashed hopes and lost opportunities might actually *allow* one to realize oneself instead of living a life of fragmented or diminished meaning." [437] Self-government, she argues, lies precisely in the acceptance of this fact and necessitates acknowledgement of the provisional nature of the narrative. Westlund continues:

> We need to know that the narratives, despite being necessary to understand our pasts, are only contingent and we must also be able to achieve some distance from the narratives we construct, open ourselves to alternative interpretations, and take responsibility for working and reworking our stories as our lives continue to unfold.[438]

There is an added responsibility to revise narratives when necessary, perhaps when the frame itself strains acceptability either because it flies in the face of the facts of reality or social acceptability, as Schechtman would argue. We are required as

[436] Westlund, "Autonomy", 92.
[437] Ibid..
[438] Ibid., 93.

responsible agents to pay attention to and be responsive to any lack of fit between one's actions and the evaluative profile one claims to be narratively true. We need to be able to respond to "mismatches between the narratives we project for ourselves and the ones that actually seem to be unfolding around us." [439] When we are responsible in such a manner, we are able to continue to guide and influence the sorts of evaluative perspective we will take thereafter.

Overall, as I have been arguing, narratives may be essentially conditional and there may be better or worse ways of organizing the self under a narrative frame. Narrative provides a means of organizing our experiences and evaluative aspects into a coherent whole and, when we organize in this way, there are benefits that carry into the future. Who one is and their evaluative profile may be given a particular interpretation that can shape and structure the constitution of that profile in what is retained and made salient to the agent. The backward-looking aspect highlights salient aspects that uphold a preferred trajectory, while the forward-looking aspect puts this interpretation into action.

5. The Forward Looking Function to Narratives

The narrative, rather than merely describing events, serves to make manifest to us patterns underlying our experiences and enable us to use such patterns as a basis to act within the world.[440] When we organize in this way, the shape and structure of our

[439] Ibid., 98.

[440] Likewise, and as suggested by Schechtman, the backward looking aspect represents just one of the roles we take towards our personal narratives. At once, we are authors insofar as we have a degree of agency in approaching the world, characters within the story enacting these choices and critics attempting to understand our actions and the direction our lives should take. Thus, life is different from literature because "we write it as we live it and engage in criticism as we go along rather than after the fact, and because

interpretation then influences how we approach the world thereafter. The value of the narrative is given in the way it not only colours and frames how we see our past, but can be also seen in the forward-looking function by influencing how we see the world. Through narrative understanding, we can work with our feelings, give them a name, and shape how we will proceed from this knowledge. It is more likely that the interpretation will be brought about in action given that we use this self-conception to structure our experiences and provide weight in decision-making. The articulation of the narrative then provides a means to assist agency in a way that answers Strawson's necessity problem. It is not that narrating one's life is necessary to live that life, but it is perhaps beneficial for agency to do so. In the following chapter I will show how this function of the narrative is also beneficial in mitigating blame and aiding rehabilitation. For now, I want to focus on the benefits narrative has for agency because I also see narrative appropriation as especially important in order to take responsibility for one's actions.

Appropriation of the past through narrative provides a basis for greater agential awareness of one's motivations, weakness of will, or overall psychological habits that otherwise may be missed as the agent moves forward in life. Actions on the basis of the narrative are, so to speak, *epistemically skillful* as knowingly guided. For instance, Catriona Mackenzie and Jacqui Poltera argue that devising a narrative can benefit the agent by making sense of jumbled and disparate experiences. In particular, they focus on the fragmented nature of lived experience of those with psychological disorders such as schizophrenia. If regularly experiencing delusional episodes, sufferers may be said to be

---

this forces us to take on different roles and perspectives" (Schechtman, "The Narrative Self", 414).

"genuinely episodic" and lose the capacity to order their experiences into a coherent temporal structure.[441] With little sense of the past or future they are "literally trapped in a ''stagnant present'." [442] However, a narrative can be told to integrate these episodes and provide a timeline between them in a way that makes the life of the sufferer as a whole more readily intelligible. Focusing on the experiences of Elyn Saks's memoir of her schizophrenia, Mackenzie and Poltera note an important distinction to be made between the content of one's delusional experiences that cannot be narratively integrated and the fact that the sufferer has such delusions. They state:

> As far as its content is concerned, [Saks's] mass-murderer delusion should not count as a self-constituting narrative. However, given that this and other delusions are central to Saks's subjective experience when she is unwell, the fact that she suffers from such delusions must be incorporated into her narrative self-conception, if it is to be accurate and genuinely self-constituting. It was only when Saks came to terms with the fact that she suffers from schizophrenia and that, when unwell, she experiences delusions—in other words, when she accepts that her illness is part of who she is—that she was able to form an accurate narrative self-conception.[443]

Not only should these experiences form part of Saks's self-conception despite estrangement content-wise, but they may even be vital to constituting her agency thereafter insofar as "incorporating the fact that she experiences delusions into her narrative self-conception is a condition for her being able to exercise autonomy to the extent that she does." [444] The narrative will, on Mackenzie and Poltera's reading:

> …include elements with respect to which she is passive—certain features of her embodiment and her genetic inheritance; the social, cultural, and linguistic practices through which her identity has been constituted; her historical and geographical circumstances; non-chosen relationships; contingent events in her life; and so on. It will

[441] Mackenzie, Catriona, and Jacqui Poltera. "Narrative Integration, Fragmented Selves, and Autonomy." *Hypatia* 25, no. 1 (2010). 39.
[442] Mackenzie and Poltera, "Narrative Integration", 39-40.
[443] Ibid., 45.
[444] Ibid.

also include elements that have arisen through the exercise of her agency.[445]

Saks's case shows us two things. First, it shows that an episodic life is at least descriptively possible. Secondly, it also shows that while the narrative may not be necessary, there are nevertheless some clear benefits to one's agency from engaging in this process. Reframing one's experiences and psychological activity draws out different salient aspects of the agent's self-conception that are best suited to her current projects.

We do not need to look only to those suffering from psychological disorders in order to see this benefit the narrative provides. Schizophrenia may create a debilitating sense of one's past and future, yet, as I argued in chapter four, the self in normal functioning may not be unified in any robust sense either. I say this not to diminish the experience of sufferers, but to highlight the fact that narratives may have this agent-constituting, backward-looking benefit in less extreme cases as well. Essentially, the point is this: because I narrate my life, circumstances and self in this particular way, I will experience the world in a way that favours doing X over Y. Or this can occur more explicitly (though not necessarily) in considerations that entail that because I am who I am I ought to do X rather than Y. Conversely, because I dislike and reject who I am, I should do X rather than Y.[446] In this way, narrative appropriation of the past is supportive of agency by the way it can guide the agent into the future with particular emphasis on what was salient in the past.

---

[445] Ibod.,49.

[446] However, if narratives change persons in this manner, how is it that my account differs from Schechtman? I argued that the narrative does not make the self, but merely tracks it and this can be done in a better or worse manner. The difference between my account and hers has more to do with the direction of fit. I suggest that the narrative tracks the self, rather than the self as tracking the narrative.

5(a). Evaluative Extension and Salient Similarity

Interestingly something else happens when agents narratively appropriate the past. In particular, salient aspects can be carried forward and brought into current experience (in narrative form) in a manner that, I would argue, is functionally equivalent to Lockean consciousness. The narrative is a lens through which our conscious experience is filtered and shaped in a way that provides a distinct subjectivity. As noted by Peter Goldie of narratives in general, we engage in narrative thinking both passively and actively. Narratives function passively by colouring our experiences. He states:

> When you meet your good friend for lunch, your perception of her is soaked with your knowledge of her past: with memories of all the times you have spent together, of her life when you were apart, and with thoughts of the myriad ways in which things might have been different. And your perception of her is equally soaked with the future, and with the branching possible ways in which things might turn out.[447]

Narratives colour one's experience of the world by conditioning and providing a backdrop to how the subject sees and interacts within that world. The narrative can also function actively when we choose to reflect and apply this reflection to current experience. Recalling in a narrative sense can create present feelings of nervousness and anger when we actively narrate an event to ourselves. Goldie states, "my immediate feelings of frustration and boredom are animated by my own experiences of her past and future latenesses, which I might express by asking myself why on earth I bother to turn up on time when I always just end up waiting." [448] Waiting for a friend can be felt in a whole host of ways depending on the sort of past one has with that person. Each moment can be felt with a sense of growing frustration due to the narrative I tell myself

[447] Goldie, Peter. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: (Oxford University Press, 2012): 119-120.
[448] Ibid.

concerning our friendship. Thus, narratives are able to both shape and transform our current subjectivity. They do not only describe a life, but also help produce a distinct phenomenological experience of living that life. As I will argue here, most importantly, narratives provide a kind of evaluative access and hence, a kind of answerability.

Attitudes, feelings and beliefs about and from the past are all brought under the current narrative frame with the result of inflecting and influencing the current subjectivity of the agent. More importantly, a continuous narrative is able to retain certain evaluative aspects of the self by keeping them alive within the narrative conception. This interpretation provides new meaning to the Lockean notion of "joining" with the consciousness.[449] Past episodes may no longer inflect one's experience; they may be like a severed limb "of whose heat, or cold, or other affections, having no longer any consciousness, it is no more of a man's self than any other matter of the universe".[450] What I am suggesting here however is that narrative articulation may allow persons to continue to experience the "heat, cold and other affections" of past values and experiences like a restored limb. This sort of narrative extension may be initially artificial, as the way in which persons are influenced does not occur unselfconsciously or without reflection. Yet a consistently articulated narrative puts these experiences and formerly held values on life support, so to speak, so they may still be 'alive' in some sense in the present. The narrated aspects may eventually be more fully united with consciousness and inflect experience without conscious articulation.[451] We have then a

[449] Locke, Understanding, II.xxvii.26.
[450] Ibid.
[451] Ibid.,27.

"vital union with that wherein this consciousness then resided, made a part of that same self".[452]

Narrative appropriation thus can provide a bridge to carry one's values into the present in a way that maintains evaluative access. In virtue of its connection to evaluative access, I will close this chapter by suggesting that narrative appropriation can be understood as a means to actively take responsibility for the past given though the way it extends personal answerability.

5(b). Agency and Taking Responsibility

Narrative appropriation may assist agency by offering a means to guide change with reference to one's self-conception, yet the forward-looking aspect of narratives is also important for responsibility-aptness and potential reform. According to Bernard Williams, there is an "aspect of responsibility, which comes out if we start on the question not from the response that the public or the state or the neighbours or the damaged parties demand of the agent, but from what the agent demands of himself."[453] What Williams seems to be describing here is more than simply being responsible. When one is responsible, the questionable action can be attributed to him and he can thus be open to demands for answers for why he did what he did. But responsibility in Williams' sense is different. He states, "apart from your effects on other people… and your attitude to their lives, there is a question of your attitude to your own [life]." [454] Williams suggests a notion of responsibility attribution that is active and more self-generative than

[452] Ibid.
[453] Williams, Bernard. *Shame and Necessity*. Sather Classical Lectures, V. 57. (Berkeley: University of California Press, 1993): 68.
[454] Ibid., 70.

that of responsibility-aptness, which is attributed to a person by others. I will suggest that the sense of responsibility emphasized by Williams can be captured by a notion of narrative appropriation. I call this *taking responsibility* as opposed to merely being responsibility-apt.

Narratives aid the agent in taking responsibility by mimicking evaluative access and maintaining salient similarity. A narrative of one's life highlights certain salient aspects of one's past and carries them forward and to current experience under a particular interpretation. In this way, as long as the same story is being told, an articulated narrative maintains consistency in one's values (salient similarity) by ensuring that certain evaluative aspects are carried forward. I call this taking responsibility because this process extends personal answerability. If we remember, I framed personal answerability as the necessary condition of the agent holding the evaluative aspects in which a response is requested. It is possible that when a narrative extends and maintains salient similarity that this necessary condition is thus satisfied. Past actions and episodes (including one's wrongdoing) and the values connected to them are not forgotten, but understood with the moral lessons learned and made to inflect on who the person is now. Accordingly, in an important sense agents *make* themselves responsible by forging a connection to past selves that could satisfy the answerability test. This captures the self-generative aspect highlighted by Williams because persons are answerable for a past in a way they would not be if the narrative connection had not been made. They take responsibility through narrative appropriation.

Given that narrative appropriation maintains evaluative access, this might result in the offender remaining responsibility-apt longer than if she simply changed without

engaging in this process.[455] Past values, even problematic ones, are brought to current experience in narrative form. Why then would the offender take responsibility in this manner? This sort of appropriation might seem especially puzzling in the context of a criminal past if the forged connection is to past offences. In the next chapter I will explore how such appropriation may be especially beneficial for criminal offenders. For now it is important to see that engaging in narrative appropriation can be understood as a means to retain a connection to the past that maintains deep attributability in some measure. Evaluative access is extended by narrative through maintaining one's values and past experiences in conscious thought. As a result problematic episodes and past wrongdoing do not fade into indifference. They are made central in how the world is framed (or reframed) for the agent. Thus, to be responsibility-apt might involve simply *being* saliently similar, while taking responsibility might mean making oneself saliently similar through the articulation of a narrative.

Conclusion

The problem with the initial criticism introduced by Lamarque is that it assumes value to be only within a backward-looking aspect, or what Elgin calls an "information-transfer." [456] It presumes that the narrative aims at producing a distinct piece of knowledge with propositional content. The narrative, in this sense, is no substitute for a correctly identified Strawsonian map. Assessing the narrative tendency on the backward-looking aspect alone however masks the reason why we engage in narrative construction.

[455] It should be noted that, simply due to this access, it does not mean the agent will indeed act in the same way as before. It is possible to retain evaluative access to past values without maintaining one's former behavioral repertoire. This was true of both the mortified matron and Novel Alex in chapter five.
[456] Elgin, "Laboratory", 44.

The map merely provides an assortment of our various past experiences, but it does not show what to make of these experiences when bringing them into present experience. In much the same way, understanding this reversal from art to the world allows us to conceive of the value of the personal narrative in a manner that does not reduce it to an ability to accurately represent the past. We can agree that when narrating the events of our lives, an accurate portrayal may not be guaranteed, yet this is not the primary means of assessing its value. If we are seeking factual knowledge of who we are, then a narrative retelling could potentially fall short. The narrative does not offer a mechanism for the proper identification of a person with their past as the importance of forming a narrative is more practical than metaphysical. Rather than offering knowledge of the world, it transforms the knowledge I already have.

Indeed, its trajectory dependence is what allows the narrative to be framed and reframed depending on what is currently important to the agent. As well, even if this process is not strictly necessary, as per Strawson's criticism, these narratives provide agents with a means of conceiving of themselves. Even if the narrative interpretation is not wholly correct, it is nevertheless agentially useful. The narrative does not faithfully and comprehensively recreate the events of our lives. Instead, it illuminates ways of conceiving of our past experiences that renders various aspects of our lives more salient than others. Our interpretations guide our future constitution. When we put our experiences into words, this story determines "what we focus on, what we tend to notice and how we are disposed to respond." [457] The narrative provides a readily accessible interpretation of experience that assists in understanding future events and experiences. It

[457] Velleman, "Well-being", 19.

allows us to make sense of the past, project ourselves into the future, and understand our own intentions, actions and beliefs. It may even allow one to create connections of salient similarity and therefore to *take* responsibility for the past.

Chapter 8: Making Sense of Blame, Forgiveness and Rehabilitation

Introduction

The last chapter saw how narratives function as organizing principles. They frame one's experiences and tie together events via emotional and thematic connections. I concluded by suggesting that narratives might also be a means to take responsibility for the past and, in this chapter, I will show why doing so would be advantageous. To warrant forgiveness and be considered rehabilitated each requires more than bare change within one's evaluative profile. The interpersonal conditions of these two concepts are only satisfied when there is a guided change that is aimed at righting the wrongs of the past and this involves taking responsibility in the manner I suggested. This also means that responsibility-mitigating change and rehabilitative change may be importantly distinct and require different conditions of satisfaction.

I start in §1 by describing two additional ways Alex may change to highlight some differing intuitions concerning rehabilitation. These versions reveal a central weakness in the account so far. In particular, it cannot intuitively distinguish between the *ways* a person can change. In §2 I focus on the concept of blame to suggest that this issue can be resolved once we distinguish between being responsibility-apt and blameworthy. These notions have distinct conditions of satisfaction and need not always coincide. So even though the manner in which persons change is treated fairly evenly, this does not mean that each case is equal in terms of blame. In §3 I show how articulation of a particular narrative satisfies the conditions required for the victim to forswear blame and forgive the offender in ways simple change cannot. Finally, in §4 I will show how taking

responsibility in this manner is also conducive to rehabilitative ends by drawing on some empirical research to support such claims.

Overall, this chapter provides a negative account to show where determinations of responsibility-aptness may be lacking. Yet, it also provides a positive account to establish why we might favour narrativity and the sense of taking responsibility that I am advocating. By considering the connections between blame, forgiveness and responsibility, we will see that although change may undermine the grounds of being responsibility-apt, it is not always sufficient to warrant forgiveness and to consider the criminal rehabilitated. To ask the question of whether the rehabilitated offender is responsible for past crimes is ambiguous as it masks two lines of inquiry implicit in the question. We may ask if the offender is actually rehabilitated or we may ask if he is still responsibility-apt and the answers to each question do not always converge.

1. The Ways Alex Changes

In Burgess's novel, some officials discuss what to do with the violent Alex following his involvement in another inmate's murder. One official proposes the introduction of Ludovico's technique. He states, "Common criminals like this unsavoury crowd … can best be dealt with on a purely curative basis. Kill the criminal reflex, that's all."[458] Of course Ludovico's technique 'cured' Alex in the sense that he could no longer act on his violent impulses. But as Burgess intended, we are compelled to question whether eliminating this 'reflex' is all that is required for rehabilitation. In chapter four I gave the example of *Cured Alex* (as an Alex who did not have Ludovico's technique reversed) whose apparent rehabilitation may be similarly questioned. The nausea Cured

[458] Burgess, *Clockwork,* 187.

Alex experiences might at first move him in ways that are alien to his evaluative stance, yet over time it has the potential to move him to form new preferences and desires. These certainly are adaptive preferences, but ones that still inflect and structure how he comes to view the world and act within it. When the technique is given enough time to overturn those violent preferences, it may be the case that Cured Alex no longer has a sufficiently similar evaluative profile to his former self to consider him responsibility-apt. The effects of Ludovico's technique result in an Alex who is no longer answerable. But is this sufficient for rehabilitation? That is, does the way the criminal changes matter for rehabilitation or should we be concerned just with "killing the criminal reflex" as was done with Cured Alex? [459] If we are left unsatisfied with calling the simple elimination of criminal desires rehabilitation, we need to ask ourselves why this is so. Here, I would like to consider two further ways Alex may have changed to highlight why we may be unconvinced in calling Alex rehabilitated even if the application of the technique could be eventually successful.

In chapter five, I argued that Alex would no longer be responsibility-apt if he had been the subject of a complete brainwashing that left him with an entirely new evaluative profile. However as brainwashing and other science fiction scenarios rarely have ready application in real life, let us focus on a more plausible way to conceive of the change in Alex. Consider Alex ten years from the end of the novel. The passage of time allows for the possibility that responsibility-aptness no longer holds. Imagine that this Alex is now happy within a domestic life and much of his earlier evaluative profile has been replaced. He can remember his previous exploits, but his former penchant for violence no longer

[459] Ibid.

finds a place in his evaluative profile. This change was gradual and simply the result of living a life. There was no effort to change 'who he was' and time did more work than active attempts to change. His former evaluative profile atrophied, as he became who he is today. Alex is now dissimilar enough (saliently so) that attributions of responsibility no longer hold. He is a changed man, but a man changed by *mere happenstance*. There was no intervention that forced this change, but there was no effort to change either. Let us call this *Happenstance Alex* in virtue of the lack of intention to change. Intuitively, there is something missing that might make us hesitant to describe Happenstance Alex as rehabilitated. This is because, as I will argue, the way Alex changes is relevant to our intuitions concerning blame and whether we treat him as rehabilitated.

There is also a second possible description of Alex's change. Despite sharing some common characteristics, I see this version as satisfying our intuitions about rehabilitation in a way Happenstance Alex does not. Again, let us focus on an Alex ten years down the road from the end of the novel who has changed to the same extent as Happenstance Alex. He is no longer sufficiently similar as to warrant responsibility attributions. Yet, the means by which he accomplished this change is different. He sought to understand his criminal wrongdoing and, in the process, altered his ongoing narrative in a way that not only took into account his negative past in its backward looking aspect, but projected an arc of change and redemption in its forward looking perspective as well. As I argued in the last chapter, through the narrative, this Alex remained saliently similar to his former self and, hence answerable for a longer duration than Happenstance Alex. But, it is not obvious that being answerable in this sense is sufficient to make him blameworthy. He may even be better considered rehabilitated due to the way his actions are guided by a

new conception of himself. Let's call this second Alex *Narrative Alex* to account for his process of change.

I suspect only Narrative Alex would receive the distinction of being called 'rehabilitated', but I have not yet outlined why this might be the case. After all, the evaluative profiles of each version of Alex have changed enough so that neither has a persisting evaluative profile sufficient to call the later self responsibility-apt. How should we distinguish between these two cases? Further, considering *Brainwashed Alex,* I assume most would not call him rehabilitated just on the basis of the break in his evaluative profile. I would argue that *how* the change occurs matters for blameworthiness and rehabilitation.

I will suggest that being responsibility-apt and being to blame (and rehabilitated) are importantly distinct and have different conditions of satisfaction that are not always satisfied at the same time. Once we appreciate this distinction, we can account for the intuitive difference between the two versions of Alex. In what follows, I will explain how it is the case that, after the period of change, Happenstance Alex is still blameworthy despite no longer being answerable, while Narrative Alex is answerable despite no longer being blameworthy.

2. Blame

Following T.M Scanlon and other contemporary theorists of blame, I will argue in this section that there is a sense in which blame is primarily a reaction to impairments to certain relationships. I will explore what it means to blame and why we blame, in order to understand how agents can be *blameworthy*. This interpretation of blame will show

how we may continue to blame despite a loss of salient similarity (and responsibility-aptness) or withdraw blame even though one may still be responsibility-apt.

2(a). Accounts of Blame

It is clear that even when one is responsibility-apt, the expression of the reactive attitudes (including blame) might not always be appropriate. Depending on who I am, my relationship to the transgressor and the significance of the fault may all moderate how we should respond to the agent. In addition, the appropriate situation in which to *express* blame is different from determining when someone is blameworthy, which consists in identifying the conditions needed in order to render one "eligible for reproach for her moral transgressions."[460] Here I would like to consider what it means to be eligible for this sort of reproach before I move to the kinds of reactive attitudes that follow.

Blame is usually construed as a response to damaging behaviour that typically involves the expression or activation of negative reactive attitudes. Yet, what makes blame distinct from other negative attitudes has been up for debate. It is not just anger or resentment toward another, but something that seems to carry some characteristic reprimanding force. According to Scanlon, there are many conflicting intuitions concerning what blame is, its force and when it is justified. What we believe about blame in these respects seems to point to an "inconsistent set" of beliefs that pulls in opposite directions.[461] Michael McKenna notes, "Despite the pervasiveness of the phenomenon in ordinary life, blame is an elusive notion. It is maddeningly hard to nail

[460] Smith, "Answerability", 108
[461] Scanlon, T. M. "Interpreting Blame". In *Blame: Its Nature and Norms*, edited by Justin D. Coates and Neal A. Tognazzini. (Oxford University Press, 2012): 84.

down a theory that gets the extension even close to right."[462] Blame has become a seemingly indeterminate term, a catchall for a number of attitudes.

Sometimes blame seems like just a negative evaluation of another. But if this is true, as Scanlon notes, then we are unable to explain why an agent's lack of control over the relevant action tends to mitigate blame. The idea is that "If blame is just a form of evaluation, there is no reason why causal explanations of our character and actions should undermine blame, any more than such explanations undermine appraisals of our intelligence or our athletic or aesthetic skills." [463] Blame therefore differs from basic evaluation.

Others have argued that the extra force of blame could be located in its punitive function. However, equating blame with a disposition to punish neglects a number of everyday cases in which blame is attributed without any accompanying overt acts that would be punitive for the target of blame. It seems reasonable that I could blame you, but due to my perhaps timid nature never express that blame even implicitly in behaviour. I still seem to be blaming, but this is private blame without the punitive effect.

Perhaps then blame is simply the judgement that one is worthy of punishment or a negative reactive attitude. Scanlon advocates that we should be "looking for an interpretation that lies between these two extremes" of private attitudes and judgments with overt punitive effects. [464] He argues that blame is deeply interpersonal in nature and responsive to the kinds of relationships persons find themselves in. Scanlon understands

[462] Mckenna, Michael "Directed Blame and Conversation" In *Blame: Its Nature and Norms*. Edited by D. Justin Coates and Neal A Tognazzini. (Oxford University Press, 2012): 120.
[463] Scanlon, "Interpreting Blame", 86.
[464] Ibid.,86.

these relationships as "a set of intentions and expectations about our actions and attitudes toward one another that are justified by certain facts about us." [465] It is an abstract concept that encompasses the kind of ideal attitudes one should have within a relationship with another. As Margret Urban Walker notes, our relationships are governed "by a particular scale of values, set of imperatives, or system of role-bound obligations" that create a general cluster of expectations and intentions that track each of these roles in which I could be judged when I fail to conform.[466, 467]

Different relationships involve many different expectations and, often, when these play out in real life, many fall short of them. Scanlon uses friendship as an example. When I am considered to be a friend of another, this relationship involves expectations of good will, time investment, mutual-aid, confidences and many more. These intentions and expectations constitute an abstract normative ideal of friendship that if one's actions "conform closely enough to the normative ideal of friendship, then they count as friends even though their relationship may be flawed as measured by this ideal." [468] A judgment of blameworthiness, for Scanlon, is a judgment that a relationship has been impaired in

[465] Ibid., 86.
[466] Walker, Margaret. *Moral Repair: Reconstructing Moral Relations After Wrongdoing*. (Cambridge, UK: Cambridge University Press, 2006): 23.
[467] Walker's characterization of the moral relationship is right and quite similar to Scanlon's. I would resist, however, characterizing these relationships as particularly moral. Depending on the kind of violation it involves, blame does not always and necessarily seek moral violations, but simply violations in the expectations one has in a particular relationship. I can blame you for actions that do not necessarily carry moral significance as long as there is some normative component. It is normative insofar as there is an expectation, but this need not specifically refer to a moral relationship or actions that have moral significance.
[468] Scanlon, "Interpreting Blame", 87.

some manner, usually by the violation of a general expectation or the presence of intentions contrary to the normative ideals of the relationship.[469]

Once a violation has been perceived, the expectations and intentions pertaining to the relationship are withdrawn. Someone who was once considered a friend may not show the kind of compassion that is expected or could more egregiously undermine the relationship by betraying the confidences of the other. According to social norms governing the relationship, a friend is supposed to care about one's troubles and not harm

[469] Like interpretation sensitive emotions and narratives, I would argue that the parameters of when a relationship counts as friendship, for instance, might be defined by and measured against social and cultural norms on what constitutes friendship. Depending on the particular social ideal governing the relationship, different expectations and intentions will be prominent. The relationship of friendship given here is voluntary and often involves a one to one interaction, but that isn't the case with all kinds of relationships. Some hold in virtue of our projects and indeed simply in virtue of being moral creatures. Scanlon, in particular, notes that there may be more wide-ranging relationships based on shared commitment to a group such as relationships between citizens, corporations, assemblies and the like. Yet, he argues that an even more general relationship obtains between all persons holding certain capacities of rationality. In these further cases, Scanlon describes persons as standing in interconnected moral relationships in an almost Kantian sense. He argues that, "…the normative concept specifies that we should have certain general intentions about how we will behave toward other rational creatures, namely, in my view, that we will treat them only in ways that would be allowed by principles that they could not reasonably reject. (Scanlon, "Interpreting Blame", 87.) He continues, as rational creatures "… we have not only intentions and expectations regarding our interactions with friends and associates but also intentions and expectations that define a relationship with other people in general. We have, for example, views and intentions about the care one should take not to injure strangers and the duties one has to aid them should we be in a position to do so." (Ibid.,88.)The kind of relationship specified here is odd as it may even hold between strangers, but it nevertheless fulfills Scanlon's general definition of what relationships are. Even if we might disagree of whether another can hold us to these expectations and disagree on what capacities ground them, there is at least the descriptive fact that such expectations exist and this might be all that is needed to allow for blame at such a general level.

the other in their actions. When the relationship is questioned or undermined in this manner, the offended party forms a judgment or judgments that she has reasons:

> …to modify one's understanding of one's relationship with that person (that is, to alter or withhold intentions and expectations that that relationship would normally involve) in the particular ways that that judgment of blameworthiness makes appropriate, given one's relation with the person and the significance for one of what that person has done.[470]

In other words, the transgressor would no longer be a friend to the offended party and, as a result, would be treated in a different manner than before.[471] Walker elaborates, "Among our normative expectations are expectations that others, with whom we think we are playing by rules, not only play by them, but also rise to the reiteration and enforcement of those rules when someone goes out of bounds."[472] When these expectations are violated, such norms also govern our reactions to the violations. Walker argues, "The responses can be immediate and expressive (an angry scowl), articulated ('How dare you!'), or elaborately institutionalized in custom or law. In these responses and by them we participate in the reiteration and enforcement of shared norms and the normative expectations they entail."[473] Likewise, when reactive attitudes such as resentment are felt, this acts as an invitation to others to see the violation as we do and

---

[470] Ibid., 89.
[471] Interestingly, while it may be impossible to actually withhold intentions and expectations for strangers and persons we have never met in any substantial way, we can still blame on Scanlon's account even if it does not involve us personally. Third party judgments are common and may only differ from first-person relationships due to an increased and added feeling of resentment and emotion for this failure in expectation. This should not be surprising as the relationship itself is much more substantial involving a wider array of expectations and intentions that can be altered. Thus there is a space for the reactive attitudes and the emotional force of blame, but the content of blame does not lie solely in this.
[472] Walker, "Moral Repair", 27.
[473] Ibid., 25.

share our interest in reaffirming those normative relationships. The reactive attitudes, rather than being the sole basis of blame, are a testament to "the normative expectations that define the scope and nature of our senses of responsibility." [474]

On the above Scanlonian analysis, blame is a move to withdraw certain expectations in a relationship. Blame may be the alteration of expectations, while blaming activity involves the communicative attempt that precedes this sort of blame. When a friend has betrayed us, we do not automatically alter the relationship, but may try to repair it. We might first want to communicate that a transgression has occurred and that a modification in the relationship is pending if the communicative attempt is disregarded or ignored. Thus, I would argue that if we follow Scanlon, then the reactive attitudes and other blaming activity could be seen as attempts to halt the modification of the relationship.

Indeed, a number of theorists have recently begun to expand on Scanlon's account to argue that blame is a means of communicating these normative expectations to others. For instance, Angela Smith argues that blame is a means of "protesting rather than merely adjusting to what I regard as relationship-impairing attitudes on your part." [475] Others have contended that it is a particular type of protest with a particular aim. Michael McKenna states, "the relation between a morally responsible agent and those who hold her to account for her blameworthy conduct … can be usefully illuminated on an analogy with a conversation."[476] Blaming activity thus can be analyzed as a means of engaging

[474] Ibid.
[475] Smith, Angela. "Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by D. Justin Coates; Neal A Tognazzini. (Oxford: Oxford University Press, 2012): 39.
[476] McKenna, "Directed Blame", 120.

and attempting to convince the wrongdoer prior to a modification in the relationship. Understanding blame in this sense could also help to explain the multifarious instances of blame. McKenna continues "An altered pattern of behavior by one person as a means of manifesting her indignation could very well be taken to have a salience by a blamed party that it would not and could not have for another person… two individuals might very well blame another in ways that are equally warranted and fitting, but do so in wildly divergent ways." [477]

Miranda Fricker also argues that this communicative interpretation of blame represents a paradigm case of our blaming practices that is present in all other subsequent iterations (including private blame, third-party blame, self-censure, etc.).[478] She argues that blame is best understood as "finding fault with the other party, communicating this judgment of fault to them with the added force of some negative emotional charge."[479] As Fricker states, "I wrong you, and in response you let me know with feeling that I am at fault for it."[480] Blame, she argues, is a type of illocutionary act aimed at achieving an effect or state of affairs in the world.[481] Illocutionary acts such as

[477] Ibid., 130.
[478] Because the concept of blame is significantly dis-unified and diverse in its manifestations, Fricker opts for a paradigm approach in which the analysis seeks "a paradigm of the phenomenon we want to understand, not only in the sense that it constitutes a clear and central exemplar but also in the sense of being a candidate for an explanatorily basic form." (Fricker, "What's the Point of Blame?", 165) This way of highlighting the paradigm case could also be achieved through a genealogical (state of nature) explanation and argue for the most explanatorily basic form of blame that is required for human relationships to be what they are. See Fricker, Miranda. "What's the Point of Blame? A Paradigm Based Explanation." *Nous*. 50.1 (2016): 165
[479] Fricker, Miranda. "Point of Blame", 172.
[480] Ibid., 171.
[481] Private blame, as Fricker argues, is not an ideal form blame, but an iteration of blame nonetheless as there may be further external factors that halt the communication of

saying, "I do" in a marriage ceremony, constitute the relevant act (marrying) when spoken in the right contexts by the right person. The illocutionary force (or what the speech does) is the act the speaker intends to perform by the speech act or illocution. For blame, it might be the act of pressing for moral realignment. The illocutionary point is to make the target "feel sorry for what they have done", inspire remorse and achieve a sort of "moral psychological calibration" and repair the relationship.[482] Like on Scanlon's account, it includes a judgment that there is an impaired relationship, but here blame is framed more like an expression with added purpose of moving the one blamed to amend the transgression. The emotional force is the "wronged party's attempt to jolt the wrongdoer into seeing things more from their perspective."[483]

I would like to put the elements of blame proposed by each of these theorists together to form a coherent picture of the practice of blaming. First, let us call expressions of blame, *blaming activity*. These expressions are analogous to attempts to initiate conversations (as suggested by McKenna) that convey one's protest (Smith) to another's actions in order to address the impairment in the relationship (Scanlon). Thus, blaming activity is the communicative attempt to move the offender to moral realignment (Fricker) and this is fitting when the agent is appropriately blameworthy. I mean the term *blameworthy* to be taken in the Scanlonian sense as apt for a modification of the target's expectations and attitudes about the relationship. One is blameworthy

---

blame. She states, "Non-communicated blame is therefore readily understood as derivative of Communicative Blame in just this simple way: sometimes it is better all things considered *not* to communicate a judgment even while it is of a type that is best understood as essentially apt for communication."(Fricker, "Point of Blame", 179)

[482] Ibid.,172-173.

[483] Ibid.,173.

when their actions are shown to be "faulty by the standards of the moral relationship." [484]

In what follows we will see that when the agent commits an act that undermines the social norms of a particular relationship (making her blameworthy), the blamer wants this blamee to see her actions as wrong by way of protest that threatens modification of the relationship. The blamer then might engage in blaming activity as an attempt to bring the impaired relationship back into realignment. Blameworthiness in this sense does not necessarily require blaming activity because is possible to blame (alter the relationship) without seeking the communicative ends of moral realignment. Nevertheless, when this blaming activity is successful, it is not because it inspires feelings of remorse or contrition alone. When those moral emotions move the blamee to take action toward reparation, the relationship can return to its initial footing. Blame as the alteration of the relationship can then be withdrawn. When the blaming activity fails, the relationship is thus modified, as the blamer is given no reason to resume the relationship as it once was.

2(b). Grounds for Blame and Responsibility

There are sure to be criticisms to framing blaming activity as communication, but I will not rehearse them here. I take the general idea to be plausible enough for our immediate purposes. What is important is the connection communicative blame has between being responsibility-apt and blameworthy that I will now explore.

I have argued that one is responsibility-apt for aspects of the self that it is intelligible for a person to answer for. Persons may be responsibility-apt for a great many things, not all of which would intuitively deserve blame as a specific illocutionary act. Holding answerable is, as Andrea Westlund would describe it a "summons to moral

[484] Scanlon, "Interpreting Blame", 86.

dialogue" or and "opening gambit in a conversation—a gambit to which others respond with further moves, which themselves invite yet further responses." [485] Blame on the other hand, when understood in the communicative sense, seems to have extended conditions of satisfaction, including the act as a transgression of a particular relationship. As Scanlon argues, "[a] person is blameworthy, in my view, if he does something that indicates intentions or attitudes that are faulty by the standards of a relationship." [486] The subject of proper blame should be the proper target of a modification of intentions and expectations of the relations between the blamer and blamee.

This suggests that one could be answerable without being blameworthy. Consider morally benign acts. I could answer for why I chose to take one route to a friend's home rather than another. My friend may ask why I didn't take a quicker route and I could justify my actions for any number of reasons. It seems intelligible to request an answer, as the act was evaluatively sensitive. But surely I would not be *blamed* unless I was late due to taking my preferred route after promising I would be on time. If I was late for these reasons, then it would be fitting to blame because I transgressed an expectation of friendship: that of promise keeping. My friend might question me for being late, but she may only blame me when I promised not to be.

I would suggest that the opposite might be true as well, namely one could be blameworthy without it being appropriate to hold her answerable. The expectations and intentions inherent in a relationship might be worthy of modification even if the person being blamed (i.e. the subject of the request for moral repair of the relationship) is not

[485] Westlund, Andrea. "Answerability Without Blame" *Social Dimensions of Moral Responsibility*. (Oxford: Oxford University Press, 2018): 262- 263.
[486] Scanlon, "Interpreting Blame", 89.

strictly answerable. For instance, Happenstance Alex may no longer be responsibility-apt as there is little reason to *hold him answerable* for his past crimes. Nevertheless, the illocutionary point of blaming activity has not been satisfied. The blaming activity is unrequited. Change occurred without communication and resolution of the transgression and as a result, there are no practical reasons to think that this change is lasting. Hence, even if there is little reason to continue blaming activity, as he is no longer personally answerable, there is still reason to consider him blameworthy in a Scanlonian, relational sense (apt for an alteration of a relationship).

It may be objected that blaming activity could be justified only when the agent remains answerable. In other words, a persisting evaluative profile (and therefore responsibility-aptness) is a necessary condition of engaging in acts that are "faulty by the standards of a relationship." [487] However, it is not clear that the agent needs to remain answerable in order to be blameworthy over time. When radical change has occurred, it is possible that the blamer might continue the communicative attempt of blame in hopes of a return to the relationship. That hope of the return to the relationship, however, would diminish as salient similarity diminishes. This is because when an agent is sufficiently dissimilar to her former self, there is little reason to think the relationship can return. Instead, modification of intentions and expectations are warranted. When answerability-undermining change occurs, such change gives reasons to cease the blaming activity and alter the relationship in ways advocated by Scanlon. The change would signal the start of a new relationship, not a return to a former one.

Given that the conditions of blameworthiness "are always relative to some

[487] Scanlon, "Interpreting Blame", 89.

relationship or relationships", it seems possible that a changed Alex could still be blameworthy even without being appropriately answerable if we take blameworthiness as equivalent to a situation in which modification of a relationship is warranted.[488] In what follows I will suggest that the difference in our intuitions between the various versions of Alex is explained by considerations about the illocutionary point of blaming activity. As I have shown here, Happenstance Alex warrants modification in one's intentions and expectations despite no longer being personally answerable. Yet, because Narrative Alex engages in narrative appropriation, it would no longer be appropriate to blame him. He was able to address the wrongs committed in the past that Happenstance Alex simply ignored. Despite being personally answerable, he may not be blameworthy because he has not satisfied those interpersonal conditions that will also make attributions of rehabilitation and forgiveness appropriate. Each of these conditions is responsive to the *ways* change occurs and not *how much* change has occurred.

3. Forgiveness

In the last section I proposed that persons might be responsibility-apt without blame if we think of blameworthiness as the failed communicative attempt of blaming activity. That is, when blaming activity fails at its intended communicative purpose, one is blameworthy by being appropriately subject to a modification of intentions and expectations. Blame as a communicative attempt has different conditions of satisfaction from answerability. In this section, I explore what those conditions might be. Here we will see the strength of these communicative accounts of blame: they can explain the relations between blame, forgiveness and narrative in a direct way.

[488] Ibid.

3(a). To Forgive and to Forget

It could be argued that a return to one's former relationship would be best served by forgetting. Forgetting would be consistent with Strawson's episodics (discussed in chapter seven) who may easily forgive because they do not see the past as involving them. Despite being able to cognitively acknowledge what they did in the past or what they plan to do in the future, those who take on an episodic perspective do not consider themselves "as something that was there in the (further) past and will be there in the (further) future."[489] That means that there may be a loss in "opportunities to forgive" because the episodic "had no memory for insults and vile actions done to him and was unable to forgive simply because he—forgot ... Such a man shakes off with a single shrug many vermin that eat deep into others."[490] Episodics feel no need for blaming activity to jolt another into reparation (either for themselves or another) because they have already moved on. They might not have literally forgotten, but they are indifferent to that episode in the past. The same is true of Happenstance Alex, who has sufficiently changed to undermine answerability even if he did not seek out that change. Yet, does this sort of change alone warrant forgiveness – that is, the forswearing of blame – of either the episodic or Happenstance Alex? I argue in this section that forgiveness is not warranted in these cases because the changes 'just happen' and are not aimed at reparation of the relationship. Blaming is appropriate despite the fact that they are no longer answerable because they have not taken responsibility for the past in any

[489] Strawson, "Against Narrativity", 430.
[490] Strawson, "Episodic Ethics", 111.

substantial sense. Change seen in these instances is accidental and not aimed at reparation of the relationship.

Taking responsibility and being responsible generally overlap. Yet, persons can take responsibility for more than is appropriately attributed to them. Jeffrey Blustein argues that "taking responsibility for *x* does not presuppose being responsible for *x*, in the sense of being "open to creditworthiness or blameworthiness for it," so there is no conceptual bar to taking responsibility for something that one is not responsible for.[491] In this sense it is possible for Happenstance Alex or the episodic to recognize their responsibilities even if they do not generally think the past involves them now. Happenstance Alex in particular, may not be saliently similar enough to recognize his former actions as his own, but could still be moved by the communicative impetus inherent in blaming activity. One can accept responsibility and see the reasons for another's blaming activity without thinking the past personally involves them. There is however, another sense of taking responsibility that I introduced in the last chapter that the episodic and Happenstance Alex might not be able to accomplish.

I argued that taking responsibility amounted to developing a narrative that included one's past wrongdoing. The process highlights the moral lessons learned in the past and brings them into one's current perspective. So while Happenstance Alex and the episodic are able to *accept* responsibility, they do not *take* responsibility in a more personal sense. As Blustein elaborates, "*Taking* responsibility, unlike *accepting* responsibility, connotes an action that is not undertaken grudgingly or merely in response to pressure or threats

[491] Blustein, Jeffrey. "On Taking Responsibility for One's Past." *Journal of Applied Philosophy*. 17.1 (2000): 63.

from others or strictly according to some script that specifies what one is to do in situations like this." [492] There is a willingness on the part of the agent to take some initiative in meeting one's apparent responsibility as not what others demand of an agent, but, in Williams' terms, what the agent "demands of himself." [493]

In what follows I will argue that Narrative Alex satisfies Williams' demand placed on oneself in a way Happenstance Alex cannot. Arguably narratives function to facilitate this process of taking responsibility and thereby warranting forgiveness from others. The episodic and Happenstance Alex may be able to accept responsibility for the past, but they would not be able to take responsibility for it.

3(b). Narrative and Forgiveness

Happenstance Alex (like the episodic) may have satisfied conditions to no longer be responsibility-apt, but as I will argue here, he has not satisfied conditions for forgiveness because he has not taken responsibility for past wrongdoing. If we think blame is a judgement of warranted modification of the expectations within a relationship, then the victim needs to see a sincere commitment from the offender that the relationship will not be disrupted in this manner again. Blame, as a modification of attitudes, would only be unwarranted if the offender's "repudiation of her 'past self' would become credible." [494] Indeed, there is an onus on the offender to show good reasons that favour reconciliation. I would argue that even if they were able to offer compensation, and apology or some measure of remediation, these efforts would be hollow without some assurance that restoration of the former relationship is possible.

[492] Ibid., 64.
[493] Williams, *Shame and Necessity*, 68.
[494] Ibid., 50.

Certain aspects of the restorative justice processes in victim/offender mediation may be illustrative of the kind of assurance necessary to warrant forgiveness. Consider first this notion of restoration as it is used in arguments for restorative justice. For instance, in the 'Truth and Reconciliation Commissions' of South Africa and also within local criminal mediation programs, the process of restorative justice mediation after an offence consists in the bringing together of individuals who have been affected by an offense or crime (victim(s), offenders(s) and other interested parties) or whole populations of people who were affected by injustices within the society. In any of these cases, the aim is to help the victims and offenders to agree on how to repair the ruptures between persons or within the whole culture. Together a decision is made on how best to restore the relationships due to the wrongs committed.

Restorative justice models differ from punitive models that emphasize punishment and desert and focus on social restoration. Yet this restoration of the relationship requires much more than simple remediation or monetary compensation, as neither of these addresses the victim's perspective and how they may have been specifically wronged. If blaming activity is the communicative attempt to restore the relationship, then the offender's response needs to specifically address the reasons the victim has to modify the relationship.[495] Through the focus on direct conversational mediation, restorative justice

[495] The aims of restorative justice are not always best served with these punitive means. These methods may be pursued for reasons of desert, social stability through deterrent, contractual obligation, moral education or any of the other reasons that are usually cited to justify punishment and the like. All these reasons for punishment do not necessarily address the reasons for the ruptured relationship.

models offer the offender a means to respond to legitimate blame and answer for wrongdoing.[496]

By confronting the past through mediation, conferences and meetings with the victim, the offender may be moved to a broader sense of responsibility. They are said to gain a "deepened sense of the reality, extent, and consequences of what they have done to another human being." [497] As Walker describes, "restorative justice practice may be the way to discover, induce, deepen, extend and clarify responsibilities that are unnoticed, resisted, or denied at the outset of a process, or have been reassuringly assigned to some small number of target individuals." [498] Offenders require not only full knowledge of the wrongs caused in order to move toward restoration, but also need to address all those who were affected by the transgression for a proper and full realignment between the affected parties.[499] Walker continues:

---

[496] During the South African Truth and Reconciliation Commissions, many victims could not even begin to reconcile with the past given that they did not know what even happened to their loved ones or where their bodies could be found. The giving of truth and statements of the offenders were in this sense central to addressing the specific harms done to the people.

[497] Walker, *Moral Repair*, 384.

[498] Ibid., 386.

[499] The restorative justice approach nevertheless remains a highly contested means of offender/victim mediation. Some worry that reviewing and rehashing traumatic events risks re-traumatization where the victims are used as props in order to secure offender rehabilitation. The victims may be pushed to speak before they are ready. The worry extends for the offenders as well as the while process could be seen as a 'shaming machine' that breaks down criminals as the procedures are subject to much manipulation given the lack of standardization. Further, when such techniques have been utilized in a wide scale, as in South Africa, many saw the processes of narrative retelling in exchange for amnesty as a means of trading 'justice for truth'. There remains the larger question of whether or not reconciliation was ever achieved or whether it was just an expedient version of justice? For our purposes, however, the aim of these sorts of projects is not only consistent with, but also indicative of the process of taking responsibility and not

Without that acknowledgment, reparative actions are charitable, compassionate, or generous, even dutifully so, but they do not "make amends." Making amends involves taking reparative action, but only action that issues from an acceptance of responsibility for *wrong*, and that embodies the will to set right something for which amends are *owed*, counts as making amends.[500]

The emphasis here is on the necessary condition of knowing the past wrongs and working to specifically address them is positioned as required to warrant reconciliation.

It may be true that acts of forgiveness are unlikely to recapture the pre-transgression state of affairs between a wrongdoer and victim in a way forgetting might. As Walker notes, "Repair cannot mean return to the status quo, but must aim at bringing morally diminished or shattered relations closer to a morally adequate form." [501] Forgiveness as restoration in Walker's sense does not necessarily signal an actual return to the former relations as such a relationship may have already been inadequate or non-existent. The offender must work to "restor[e] confidence in shared moral standards"[502] At the very least, the kind of reconciliation involved in forgiving may comprise nothing more than the cessation of resentment alongside the restoration of civility and basic respect for one another. It is a move toward "moral adequacy."[503]

This move toward adequacy, however, is only warranted when there are reasons to forgive. Walker argues that to justify forgiveness, the actions of the offender should inspire some measure of confidence, trust, and hope for the future of the relationship to the extent that "unacceptable treatment will not prevail, that unacceptable behavior will

---

necessarily whether such practices should be widely adopted within the criminal justice system.

[500] Walker, *Moral Repair*, 191.
[501] Ibid., 27.
[502] Ibid.,191
[503] Ibid.,27.

not be defended or ignored where it occurs, and that victims will not be abandoned in their reliance on our shared commitment to our standards and to each other." [504]

Likewise, Robert Roberts argues that the "teleology of forgiveness is reconciliation" as the "restoration or maintenance of a relationship of acceptance, benevolent attitude, and harmonious interaction." [505] But in order to forgive, there should be a "readiness to forgive the offender" that remains as only a possibility until "repentance becomes evident." [506]

Charles Griswold offers six necessary criteria for an offender to warrant forgiveness that seem to represent the same sort of necessary steps for realignment. His criteria involve the offender both owning and repudiating her actions, feeling contrition and sympathy, and most importantly offering an account of the wrongdoing that would ensure that the victim is "right to forgive the offender for these deeds." [507]

According to each of these theorists, forgiveness is only satisfied when there are good reasons to return the relationship to moral adequacy. This "requires that the offender not only take responsibility for her past wrong-doing but for emendation."[508] Likewise, blame, as I have argued, can be understood as the communication of social transgression that will result in an alteration of intentions and expectations when left unresolved. If blaming activity is essentially communicative in this manner then, it seems that some instances of forgiveness may be justified when the communicative

---

[504] Walker, "Moral Repair" 384.
[505] Roberts, Robert C. "Forgivingness." *American Philosophical Quarterly*, vol. 32, no. 4, 1995, 229.
[506] Ibid., 229.
[507] Griswold, Charles L. *Forgiveness: A Philosophical Exploration*. (New York: Cambridge University Press, 2007): 51.
[508] Ibid.

purpose of the blaming activity is successful and the damaged relationship can return to where it was before. Forgiveness, as Walker argues, is aimed at reconciliation and repair as the task of restoring or stabilizing the "basic elements that sustain human beings in a recognizably moral relationship." [509] Thus, to offset the force of blame and warrant forgiveness is not just to quell a reactive attitude, but also to seek repair in one's relationship in order to signal a return to something approximating its original status. This requires some assurance from the offender that will motivate the offended party to initiate this return. One way to achieve this end, I will suggest in the next section, may be through narrative appropriation.

3(c). Taking Responsibility through Narrative

As we saw in the last chapter, narratives are able to assist agency even if they are not strictly necessary in order to live a life. They provide a means by which the person can achieve some measure of evaluative access to past wrongdoing. This is why I associated this process with *taking* responsibility. Narrative not only accounts for the past, but can also aid in shaping one's evaluative profile thereafter. This narrative thread may mean that persons remain responsibility-apt longer than if they simply changed without engaging in this process. The thread keeps values alive and frames one's experiences. Narratives provide a measure of salient similarity by unifying and organizing one's experiences and evaluative profile in an enduring way. Yet, even if this similarity would technically mean the agent is responsibility-apt, it does not mean he or she is still to blame.

[509] Walker, *Moral Repair*, 23.

The narrative process offers a means for *guided change*. This process then informs any change that occurs afterwards without necessarily incurring further blame because conditions for responsibility-aptness have been maintained without having maintained the conditions for blame.[510] Narrative Alex, during the process of change, may still be answerable and sufficiently similar but not an appropriate target of blame. As I mentioned in the last chapter, ten years down the road, it is possible that salient similarity has faded to the point that he is no longer personally answerable. But how he got there matters. In those ten years, Narrative Alex kept past wrongdoing in a narrative in a way that retained evaluative access as to maintain a continuous evaluative profile. So for an extended period, Narrative Alex would be personally answerable. Perhaps he would even be personally answerable for longer period of time than Happenstance Alex. Yet, the narrative process, in this instance, at least signals a potential return to the relationship, not the start of a new one with different expectations and intentions that might otherwise be the case when changed. He would no longer be blameworthy.

---

[510] There are cases in which the victim forgives without change in the offender such as "Gifted Forgiveness". Yet, as Fricker argues, this is an offshoot of the paradigmatic case of blame as communicative. The forgiveness is not a gift without strings, but one that is still aimed at jolting the offender into remorse or change. It may also be for the sake of the one forgiving. Nevertheless the point is that this sort of case is exceptional because the offender has done nothing to deserve blame. Otherwise it would not be such a gracious gift. She states, "The variety of Gifted Forgiveness I wish to focus on here is exemplified in the much cited literary example from Victor Hugo's *Les Misérables*. The Bishop forgives Jean Valjean for betraying his trust and stealing the rectory silver, despite the fact that Valjean expresses no remorse. This is an archetypal case of Gifted Forgiveness, but (here's the point) we can only make sense of it as forgiveness by thinking of it as the Bishop giving Valjean something that would *normally* need to be earned through remorse but on this occasion *isn't*." See Fricker, Miranda. " Explaining Blame and Forgiveness" Peasoup.com. May 26, 2015. https://peasoup.typepad.com/peasoup/2015/05/explaining-blame-and-forgiveness-by-featured-philosopher-miranda-fricker.html

Overall, as an answer to Burgess' question as to whether "just killing the criminal reflex" is enough, we see that it is not in these cases. It would not be enough to simply make the offender indifferent to his past because the ruptures in the social relationships would not have been addressed. The seemingly cured criminal would no longer be personally answerable and as a result, there is a necessary modification of expectations and intentions because they are different enough as to think of the change as a start of a new relationship. Yet if we would like to return the relationship to some degree of moral adequacy, the blaming activity need to hit its mark, be taken into consideration and change be understood as mending to ruptured relationship. The offender, like Narrative Alex, would then no longer be blameworthy because the relationship could return to where it once was or at least to some degree of moral adequacy. One means of accomplishing this I suggested was through devising a narrative that tracks one's wrongdoing. I will leave it open as to whether there are other ways to provide this kind of assurance. What I wish to highlight nevertheless is how the narrative is beneficial insofar as one may be more easily forgiven and no longer be considered blameworthy (as being apt for a modification of the intentions and expectations of the relationship), as there is at least the potential for a return to the former relationship. In the following section, I will expand on the benefits of the narrative to argue that providing a kind of evaluative bridge to the past, not only helps alleviate blame, but, as we will see, allows us to consider offenders who engage in this process as rehabilitated.

4. Rehabilitation

Narrative articulation provides a means to guide change. As a result, some initial benefits to rehabilitative processes should be clear. An offender that aims to right the

wrongs of the past and appropriate his wrongdoing in an ongoing narrative is likely to betray a picture of lasting change over one's whose change is coincidental. In this section, however, I want to focus on another benefit that might not be as readily apparent. In particular, I argue that the revisionary and interpretative aspects may be beneficial by encouraging positive change. We know that narratives assist in excavating memories of one's former actions and applying new narrative frames to see if they reveal meaning that the agent may have never appreciated before. They are essentially 'interpretive' as multiple frames can be applied to understanding a single episode. This fact of potential interpretation and reinterpretation can be seen as supportive of rehabilitative aims by helping offenders to reconceive their pasts in a more positive (redemptive) light. To support this possibility, I would like to consider some studies on rehabilitation and recidivism from criminology to argue that narratives of reform help to make sense of one's experiences in a way that aids taking responsibility for the past and ultimately discourages recidivism.

4(a). Criminology and Recidivism

There has been some dispute in the psychology literature concerning whether changes in behaviour are "structurally induced" or "agentic." [511] Structurally induced identity change explains 'desistance' from criminal behaviour through contexts and social institutions (such as marriage or vocational circumstances) that increase social bonds previously unavailable. In their quantitative study on the effects of pro-social

[511] See Rocque, Michael, Chad Posick, and Ray Paternoster. "Identities through Time: An Exploration of Identity Change As a Cause of Desistance." *Justice Quarterly* 33, no. 1 (2016): 45-72.

identity formation for desistance, Michael Rocque, Chad Posick and Ray Paternoster argue that:

…[F]ormer offenders find themselves in conventional social roles, most often without their intention, and the social role changes them for the better, usually by restricting their opportunity to commit crime…[512]

Recidivism is reduced because strengthened social bonds "binds" the individual to the community and encourages conformity.[513] This is thought to happen "without any intention or agency on the former offender's part." [514]

On the other hand, agentic identity change, like that I am proposing with narrative articulation, focuses more on notions of the self and intentional self-change to explain desistance. In long term studies on the causes of change, Rocque et al. also found that there is interdependence between agentic identity change and structural aspects; neither is sufficient on its own to reduce recidivism. Changes in identity conception are strongly related to decreases in crime over time and help to support change when shifting into new contexts because in order for "social control processes to have an impact on behavior, the individual's identity must first have become sufficiently pro-social (and believe themselves so). In other words, changes in social control without changes in identity are unlikely to be enough to effect behavioral reform."[515] Rocque et al. concede that while their findings do not directly support this conclusion, they nevertheless point to conceptions of one's identity as "a strong and robust predictor of desistance from crime", which in turn should helpfully guide the kind of correctional programing that is

[512] Ibid., 50.
[513] Ibid.
[514] Ibid.
[515] Ibid., 65.

needed in order to rehabilitate offenders.[516] They explain, "In addition, our findings are supportive of the 'redemption' policies that have been recently advocated by criminologists. According to this line of thinking, rites of passage ceremonies, indicating to the individual and to the community that the offender status has been shed, are integral in reintegration."[517] The point is that successful rehabilitation seems to require at least some change on the offender's self-conception in order to reintegrate them and find a place within society that coheres with their new self-conception. These studies also show that changes of self-conceptions may be insufficient on their own, unless there are structural and circumstantial means (such as strengthened social and communal bonds) to support the new identity conception.

Unfortunately, according to criminologist Shadd Maurna, if shifts in self-conception require strengthening of social bonds, such social integration is not always available for offenders due to how they are generally perceived by the larger community. Offenders tended to truncate their narratives though a denial of the past in order to fit their story into what is more readily available. In the "Liverpool Desistance Study", researchers analyzed life stories of desisting offenders.[518] One finding was that ex-offenders would frequently construct their past in a manner that rationalizes their deviation from social norms. When telling their stories, offenders would adopt "neutralizing techniques" that denied their full involvement in past criminal action.[519] Past episodes were described theoretically as 'it was happening' or, more subtly, using

[516] Ibid.
[517] Ibid., 65-66.
[518] Maruna, Shadd. *Making Good: How Ex-Convicts Reform and Rebuild Their Lives*. (Washington, D.C: American Psychological Association, 2001): 38.
[519] Ibid., 94.

the pronoun 'we' rather than 'I' to disperse responsibility and neutralize the reports.[520] Without a sense of ownership, the action is made to be a purely causal product of external factors. They would make excuses or argue that their past criminal action did not involve the "true me" in manner not unlike an episodic.

According to Maruna, ex-offenders relied on these techniques, not because they disregarded social norms, but because they were heavily invested in them. Many offenders in the Liverpool study found the process of fully accepting their pasts challenging (often leading to shame and depression) once that acceptance meant owning a social identity that labeled them as 'irredeemable' or 'evil.' This could partially be due to the way potentially desisting offenders may find their stories of rehabilitation and change unarticulated within the larger social narrative.[521] Faced with a negative public

[520] Ibid..

[521] The prison system may compound the problem of incorporating one's criminal past. As Craig Haney notes, "…many people come to prison already having begun to think of themselves as marginal, as outlaw, or 'other'" (Haney, *Reforming Punishment,*178.) Incarceration, he argues, "tends to foist such an identity on new arrivals and then 'fix' or harden it by virtue of the way prisoners [are] subsequently treated, referred to, and looked upon by many staff members" (Ibid.,178). Part of the way these labels are fixed may indeed be due to the singular and institutional interpretation of their lives. As John McKendy recounts after as series of interviews with convicted offenders: "Imprisonment involves not just physical confinement, but also discursive or ideological confinement. What men in prison are prompted to say, the sorts of discursive opportunities they are afforded, the kinds of stories that are officially ratified – all of these are severely restricted." (Mckendy, "'I'm Very Careful About That', 496) The prison system, it is argued, silences the self told stories provided by the inmates in a number of ways. After nineteen years of working in maximum and medium security prisons, Sociolinguist Patricia E. O'Connor, notes that there is not only a lack of public interest for the stories of these offenders, but the isolation and self-imposed silencing through codes of prison culture serve to restrict narrative innovation and exploration. She states, "the intentional isolation along with the self-censuring that arises through prison violence combine to form a large and dangerous silencing of voices and issues" (O'Connor, *Speaking of Crime*, 141.) This sense of otherness experienced by the prisoners, do not leave them once stepping out of the prison yard either. Each will likely encounter such social

interpretation of their identity, it is not surprising that many criminals seeking reform find themselves in a state of denial as a result. The criminal who sincerely wants to re-join the moral norms of society is left to "knife off" his past and deny that he ever was that criminal. [522]

Indeed, the rehabilitated offender's circumstances seem to mirror the stateside release of Burgess's novel. Alex's prospective redemption as seen in the last chapter of the novel was not even accepted when it was released in America - known for its emphasis on retribution - much to Burgess's distress. When *A Clockwork Orange* was imported for American audiences, this final chapter was excluded and consequently the redemptive lesson of the novel was removed. According to Burgess, no longer was the novel "art founded on the principle that human beings change." [523] Rather it became a sensationalist picture of incurable evil that is not a "fair picture of human life." [524] Alex is made to be inherently immoral, as the perpetual and irredeemable sinner. Burgess's narrative of redemption stood in contrast to the more prevalent narrative in America that contained a more pessimistic view of criminality.

Of course, it is not unusual for an artistic vision to go unappreciated when adapting it for a larger audience. General expectations of what a narrative should hold often alter

---

stigmas in the larger society as a result of their ex-convict status. See Haney, Craig. *Reforming Punishment: Psychological Limits to the Pains of Imprisonment*. (Washington, DC: American Psychological Association, 2006), Mckendy, John P. "'I'm Very Careful About That': Narrative and Agency of Men in Prison." *Discourse & Society* 17, no. 4 (2006) and O'Connor, Patricia E. *Speaking of Crime: Narratives of Prisoners*. Lincoln Neb: University of Nebraska Press, 2000).

[522] Ibid., 4.

[523] Burgess, Anthony. "A Clockwork Orange Resucked" *The Floating Library*. 20 April 2009.

[524] Ibid.

the kinds of stories that can be told. However, I propose that something similar happens to personal narratives of desisting offenders as well. Sometimes, stories that would be truly apt to capture these individuals' unique experiences, motivations and indeed *who they are*, are not stories that are generally accepted by a wider audience. They are instead edited to fit within given expectations in a manner similar to the editing of Burgess's work.

Personal narratives are edited to fit larger social narratives in a way that undermines what Westlund calls "fluency" in regards to answerability demands. [525] Westlund argues that if holding answerable is a "gambit to which others respond with further moves", not all persons may be fluent in offering a response.[526] In fact she suggests that certain forms of blame may even undermine therapeutic aims for those with certain cognitive and emotional deficiencies. She argues that there are those at the margins "who cannot be drawn into our responsibility practices, just as there are those who cannot be drawn into linguistic practices, because they lack underlying cognitive and (perhaps) emotional capacities."[527] Just as with speech, there are those who may be more or less socially fluent in knowing "their way around complex practices of praising, blaming, excusing, repenting, apologizing, forgiving, and so on and so forth" when faced with an answerability demand.[528] I would suggest here that although rehabilitative offenders do not necessarily have cognitive or emotional deficiencies that would

[525] Westlund, Andrea. "Answerability Without Blame" In *Social Dimensions of Moral Responsibility*, edited by Katrina Hutchison, Catriona Mackenzie, and Marina Oshana. (Oxford: Oxford University Press, 2018): 262.
[526] Ibid.
[527] Ibid
[528] Ibid

undermine their capacities, their fluency in responding to answerability demands and facilitating agentic change is undermined because their responses are limited due to social stigma.

As noted by Maruna, "Criminals and delinquents become dishonest because of the words available to them."[529] Including the negative past and promoting agentic change may be beneficial in order to reintegrate the offender into society, yet doing so can be more than simply distressing, but leave one in a profound identity crisis due to the way that criminality is perceived. For Maruna, forcing an offender to own and take responsibility for a past is good for first-time offenders, but not necessarily for those whose criminal behaviour became a lifestyle. As she states, "being ashamed of an isolated act or two is one thing, but it is quite a different thing to be ashamed of one's entire past identity, of *who* one used to be." [530] Sometimes excuses, Maruna argues, are needed to properly align oneself with the social norms. Neutralizations may be adaptive in dealing with an overpowering amount of shame for one's life that in turn mitigates the ability to take responsibility and lead one's life with the full acknowledgement of past wrong doing. These criminals were not unlike Happenstance Alex as they try to forget who they once were. This denial has consequences for agency and rehabilitation.

Social stigma can undermine the criminal's ability to "find[ ] their way around these practices with the right sort of support, guidance, or prompting from others." [531] They are as a result unable to fully answer for and take responsibility for their

[529] Maura, Shadd. "To Tell the Tale". In *Handbook of Restorative Justice : A Global Perspective*, edited by Sullivan, Dennis, and Larry Tifft. (London: Routledge, 2006): 132.
[530] Ibid., 143.
[531] Westlund, "Answerability Without Blame", 262.

troublesome pasts. If I am correct that owning and undergoing a guided change through narrative appropriation is required to warrant forgiveness, these criminals may be hampered in engaging in this process. Worse yet, they may also suffer some agential harm due to such denials and undermine their rehabilitative aims. As Jeffrey Blustein notes, "The person who fails to take responsibility for an important aspect of his past, even when wrong-doing is not at issue, might very well have only a superficial understanding of himself, his abilities, interests, and concerns, and a life led under these conditions is not a life well led." [532]

Indeed a study of seventy-three offenders living in and around Dublin, conducted by Deirdre Healy, identified three kinds of narratives ex-offenders used to speak of their former lives. She labeled these "rejection", "integration" and "stability" narratives. [533] The "rejection" narrative dismisses the past as having no relevance to the current self, while "integration" narratives, by contrast, look at past episodes as foundational as it reframes negative episodes into something positive moving forward.[534] Those with this narrative saw the past as a resource for education as lessons learned in the past are carried into the future. The smallest group was described as having a "stability" narrative that acknowledged their criminal past as an ordinary part of life, not to be lamented, but an everyday aspect of adolescence.[535] Healy found that:

Ex-offenders were more likely to adopt an agentic self-narrative. They were more likely to seek meaning in their criminal pasts and try to derive wisdom from their negative experiences. They were forward-looking and their aspirational identities centered

[532] Blustein, "On Taking Responsibility for One's Past", 87.
[533] Healy, Deirdre. *The Dynamics of Desistance: Charting Pathways through Change*. International Series on Desistance and Rehabilitation. (London: Routledge, 2012): 120.
[534] Ibid.
[535] Ibid.

primarily on conventional adult pursuits, which they felt confident they could achieve. They did not endorse generative concerns or demonstrate evidence of agency as measured by self-mastery, victory, achievement or empowerment but were committed to desistance and developed clear strategies to address any barriers they expected to encounter.[536]

However, Healy also argues that while subjective factors, like narrative reframing, are "likely to be found at the forefront of change but, given their unstable nature and liability to change, they are unlikely to be strongly associated with long-term desistance." [537] So it may be the case that a shift in narrative is only beneficial once it can be socially sustained. Indeed, Marieke Liem and Nicholas J. Richardson argue, "the distinguishing factor between desisters and non-desisters is agency, or a lack thereof, rather than other parts of the transformation narrative, such as a good core self or generative motivations."[538] Passivity, much like the perspective of Happenstance Alex, was characterized as seeing change simply as the result of time and external factors acting on them. Passive change did not lead to long lasting desistence.

Overall, what mattered for many was not the formation of the narrative alone, but the way the narrative was framed and the sense of agency such reframing provided. Liem and Richardson, who interviewed re-incarcerated and paroled offenders within Boston and Philadelphia metropolitan areas, found that what was "strikingly different between the desisting and non-desisting groups was their sense of agency." [539] Recidivism was generally linked to passivity in the way they understood their actions and motivations.

---

[536] Ibid.,124.
[537] Ibid.,125.
[538] Liem, Marieke, and Nicholas J. Richardson. "The Role of Transformation Narratives in Desistance Among Released Lifers." *Criminal Justice and Behavior*. 41.6 (2014): 709.
[539] Ibid., 705.

Many would cite failure due to "external forces such as God or their parole officer" instead of themselves. Liem and Richardson continue:

> Those from the desisting group, however, expressed a strong sense of agency and control over their own lives. While recognizing that a variety of social and environmental factors (e.g., substance abuse, lack of financial resources, and low educational achievement) influenced their behavior and prospects upon release, the desisters still displayed high levels of agency as evidenced by their belief that they are able to act independently and make their own choices.[540]

What we see when we look at the literature is not desistance due to simply articulating a narrative of transformation if that narrative did not emphasize the agency of the subject. The narrative needs to inspire confidence in the offender and frame their experiences in a way that makes change achievable. Contrary to pessimist narratives that frame them as irredeemable, the offender, in order to find lasting desistence needs to see themselves as someone who can change. This may require reframing their past in a more positive light that emphasizes their agency.

4(b). Narratives of Reform

I would argue that the way that narrative modulates saliency and bridges connections to a troubled past is an essential component of these stories of change due to the sense of agency narrative reframing can provide. That is, narratives can reframe past events and recast the offender's self-understanding in a more positive light. They are able to incorporate troublesome episodes in a way that might help mitigate the profound feelings of shame by shifting the saliency and meaning of former episodes.

Recall the feature of interpretation sensitivity explored in the previous chapter. This term referred to how the emotional vocabulary we assign to a set of experiences can

[540] Ibid,. 705.

come to shape what emotions we come to experience later. In other words, our interpretation of events can make it more likely that persons will act according to that interpretation thereafter. With a more redemptive narrative, it may be possible for offenders to better understand their past while equally acknowledging themselves as persons who can adhere to the moral norms in the future. It is not narrativity alone doing the work, but the sense of agency the articulation of the narrative provides. Essentially, it matters *how* the agent changes and whether that change is supportive of agency.

These reflections from criminology provide further grounds to think Narrative Alex may be better considered rehabilitated than his Happenstance counterpart. In cases of criminal rehabilitation, offenders are put in a better position to deal with aspects of who they are when engaging in this sort of narrative reframing. By reconceiving the past, the ex-offenders are able to modify how they understand it in a way that supports working toward a rehabilitated identity and long-term stability in a new identity. So, we can grant that sometimes it may be a "good idea to put the past behind us or the future out of play." [541] The offender may say "that is not me anymore", but it is a further question of whether that is the correct attitude to take towards one's past. Owning the past has been seen to be useful for one's agency and even morally obligatory in some cases (in order to right a wrong in a deeper sense than by providing compensation). Extending the self through narrative can be restorative of the offender's relationships with others and themselves.

Conclusion

[541] Schechtman, Marya. "Stories, Lives, and Basic Survival: a Refinement and Defense of the Narrative View." *Royal Institute of Philosophy Supplement*. 60 (2007): 176.

As I suggested at the end of the last chapter, I hope it is now clear why it is the case that to be responsibility-apt involves being saliently similar, while taking responsibility might mean *making oneself* saliently similar. When a house is returned to its former shine, we rehabilitate it. We might care for an injured animal and rehabilitate it so they may walk again. An addict may return to her prior health having succeeded in rehabilitative aims. Likewise, generally restorative change in criminal rehabilitation should aim to return someone to a position in society that they formerly held, which also means being the kind of person who warrants forgiveness. I have argued that this aim may be best accomplished through a narrative approach. Narratives provide a stable and non-contingent basis for change that also helps to distinguish between the changes the different Alexes undergo. Responsibility concerns answerability whereas blame is sensitive to damages in interpersonal relationships.

For Happenstance Alex, his change did not necessarily address the reasons for which he might be blamed even if the change sufficed to make him no longer responsibility-apt. He may no longer be the same, but why he was blamed in the first place plays no role in how and why he changed. Not just any change will do. There is nothing in the way he changes that would provide a reason to return the relationship to where it was before the offence nor would we have any confidence in calling him rehabilitated. Part of the reason why we might intuitively want to continue to blame him is because the victim still has more than enough reason to continue his or her blaming activity. He may also drift into a person who brings the relationship into moral alignment and hence someone the victim could forgive. Both cases are possible, but any change for

Happenstance Alex is contingent and arbitrary. So neither change would offer good reasons for the victim to foreswear blame and engage in a process of forgiveness.

Narrative reframing by contrast provides a kind of responsible agency by structuring a self that is sensitive to one's past wrongdoings in a manner conducive to rehabilitation as a kind of restoration. We may want to call Narrative Alex rehabilitated, not because the conditions in being responsibility-apt have run their course and no longer hold, but because the change he underwent was a guided process in which he took responsibility for the past and incorporated past wrongdoings into his narrative self-conception. Narrative articulation provides assurance that the broken relationship can be mended, while equally offering a means to reframe problematic episodes of one's past that may mitigate profound shame and emphasize one's agency. It is a means of taking responsibility by accounting for the past as well as projecting that interpretation of the past onto the future.

Thesis Conclusion

> There was nothing outside of self-obstruction that could stop my progression. …Likewise, for a lowly person such as myself, there was a harmonized order, a dimly glimpsed path I could take to alter my negative existence. The isolation designed to emasculate me and cripple my spirit had failed. I was not the same man they had marched through the entrance to the Hole years earlier.[542]

Starting from Locke and ending with blame, forgiveness and rehabilitation, I hope to have identified not only what we might call the moral self, but also what it means for that self to persist over time. I started this thesis by asking the question of whether offenders, such as Stan 'Tookie' Williams, remained responsibility-apt for their earlier crimes after what appears to be a change of self. By pulling from numerous sources in hopes of uniting the many disparate discussions on the topic of the self and responsibility, I aimed to shed some light on what it would take to fill out these potential stories of redemption. In the introduction I suggested that addressing the questions of responsibility and rehabilitation requires developing answers to four related sub-questions. These were: "What are the conditions of a self at a time?", "What makes one responsibility-apt?", "What are the conditions of responsibility-aptness over time?" and "What matters for rehabilitation?" I hope that the thesis has provided an answer to each of these and has generated an overall answer to the question of Williams' responsibility-aptness in the face of radical personality change. In what follows I will review my answers to these questions in the context of Williams' reported redemption.

[542] Williams, Stanley. *Blue Rage, Black Redemption: A Memoir*. (Touchstone, 2004): 280.

Due to the finality of his sentence, like the ending of "A Clockwork Orange", we will not see what could have eventually happened to Williams. All we have is Williams' professed innocence and the state's clear denial of it. What we do know is that Williams maintained his innocence until the end. He argued that he was set up by "career criminals" who had no "compunction about ruining [his] life to save their own scrawny necks," as he put it.[543] But the question of guilt or innocence went deeper than whether he had been the man to commit the murders. His work with similarly situated youth and the ways in which he considered himself to be a new self raised the *characterization* question. He may have been properly *re-identified* as the same man, but the question is whether he was the same moral self.

Did Williams acquire a new moral self on my account? To this I would answer, quite unsatisfyingly, *it depends*. Responsibility-aptness and continued responsibility-aptness are complicated and depend on a number of facts concerning the kind of change that Williams underwent. Adding to this empirical complication is the fact that much of the controversy surrounding Williams' case involves what I see to be an entanglement of two separate questions: a question of responsibility-aptness and one of rehabilitation. I will close by showing that on my account Williams remained responsibility-apt for his crimes. Yet, my account also shows that given the way in which he took responsibility through narrative appropriation for his crimes, there is also a good argument that he was rehabilitated and therefore not blameworthy at the time of his execution.

In response to the first question ("What are the conditions of a self at a time?"), I outlined the shape of the self as a forensic unit by defining it as one's ongoing and

[543] Williams, *Redemption*, 233.

unique phenomenological perspective or motivational profile. Then, to answer the second question ("What makes one responsibility-apt?"), I narrowed this profile to what I called the evaluative profile that constitutes the moral self. The latter includes only those aspects of the self for which persons are answerable. The question is whether in identifying the later Williams we identified the proper forensic unit to initiate an inquiry, and also whether the later Williams was answerable for his past crimes and therefore was the same moral self.

There is little reason to think that basic attributability was violated in Williams' case. By most accounts, he seemed to be the same forensic unit or persisting consciousness; he retained the same first-personal perspective or motivational profile constituted by a collection of psychological features. I argued that for a past crime to be deeply attributable to the current offender, the question is whether the current offender retains the relevant evaluative aspects so as to render him answerable. When evaluative access and answerability over time obtains, the person may be said to be the same moral self.

There is an argument that Williams was the same criminal that first entered the prison, that is, that he retained the same evaluative profile. Law enforcement officials and victims' rights activists argued that the changes in Williams were overstated. After all, he had numerous violations in San Quentin and this behaviour seemingly implicated him as the same violent criminal he once was. One persistent reason given to deny him clemency was not only that his behaviour in prison was riddled with citations and aggression, but also that he repeatedly refused to inform on his former gang associates. Gov. Schwarzenegger questioned his change as "hollow" because Williams had

remained "loyal to the gang member street code of ethics" and refused to be debriefed by the prison authorities and provide information on the way the gang operated.[544, 545] In a "60 Minutes" interview, Williams stated, "I have to say that the word 'debriefing' is a euphemistic term for snitching. And my--my convictions won't allow that." [546] If the self over time is best understood as one's continued evaluative perspective, as I have argued, then Williams actions and responses to investigators seemingly show him to be the same self constituted by the same sort of motivations and values.

However, it is not fully clear that Williams maintained evaluative access to his past to a sufficient enough extent that would render him answerable. Some may want to argue that his poor formative circumstances on the streets of Los Angeles could mitigate his current responsibility-aptness to some extent. However, as I argued, although considerations such as these might help excuse his acts and render them reasonable given his situation, determinations of responsibility-aptness are different. What matters for responsibility-aptness is answerability, not the potential excuses one might give. Yet, Williams purported to be distanced from his past in another manner. Despite remaining loyal to some of the values from the life prior to his life in prison, Williams also underwent much personal change. He seemed conflicted about the values of his past and exhibited optimistic narratives to reframe his future. Could such changes potentially undermine the evaluative access required to consider him responsibility-apt in the relevant sense?

---

[544] Schwarzenegger, *Statement of Decision*, 5.
[545] Williams, *Redemption*, 274.
[546] Leung, Rebecca. "Rewriting the Past: Former Crip Leader teaches Children how to Avoid Gangs". *60 Minutes*. May 21, 2004.

As an answer to the third question concerning responsibility-aptness over time, I argued that to determine whether a person remains answerable we need the test of *salient similarity*, which involves a comparison of the evaluative profile in the present with significant evaluative aspects of the past. So the question is not simply whether Williams is similar generally speaking, but whether he is similar in the relevant ways.

Was Williams saliently similar? Much of the values and the "street code of ethics" of the former self seemed to remain in Williams.[547] But despite retaining some loyalty to his former criminal life, Williams also considered himself to be "a student of sociology and psychology" and memorized words in the dictionary to improve his vocabulary.[548] He would draw portraits and sketch family and famous figures, in a process that he claimed to have a "halcyon affect" as a means of "calming the beast within." [549] This process and a newfound inner calmness moved him to write books that targeted at-risk youth. He stated:

> I discovered that writing the book had a sublime effect on me. It seemed to melt away the years of being desensitized and callous. I felt a sense of genuine purpose: to create a book that might tap into the social pathology affecting black children. Though I held no academic degree, I had created my own college curriculum through years of study, extrospection, and hard-knock experiences both on the streets of hell and in San Quentin. Though a role model I could never be, I could act as an African griot or Paul Revere, warning youths about what is coming down the crooked path.[550]

Those arguing on Williams' behalf might also point to his repudiation of his former gang associations as a change in the relevant evaluative aspects. Williams writes "In a cold sweat I shook myself out of this awful reverie, consumed by sadness—not for

[547] Ibid.,274.
[548] Williams, *Redemption*, 224
[549] Ibid.
[550] Ibid.,207.

Crippen, but for the lives of all the Crips who had died, for the innocent black lives hurt in the crossfire, for the decades of young lives ruined for a causeless cause."[551]

The change seen here may represent an alteration that occurred with "day to day improvement." [552] Many of his evaluative beliefs shifted priority or started to wither away. I suspect that his case is not unlike the other cases of alienation and conflict we saw previously in this thesis. Williams felt shame about his criminal life and experienced inner conflict; he was loyal to his former gang associates, but repudiated this loyalty at the same time. Consequently, the changes Williams underwent should be described as alteration rather than replacement. It may be the case that his newfound passion for writing, art and the spoken word is telling of an eventual loss of his former self, but until that happens it seems arguable that there is sufficient salient similarity to his former self. Since evaluative access is retained in the relevant ways, his former life is deeply attributable to him. Thus, Williams may be genuinely conflicted and experience a deep affective break with his past, but nevertheless remain responsibility-apt.

Finally, I turn to the question as to whether Williams should be considered rehabilitated. On my account, repudiation of one's former values does not guarantee absolution, especially if what is repudiated still inflects and forms one's experience. Moral selves do not rupture very easily. Neither change of perspective, internal conflict, nor even fierce repudiation guarantees a loss of responsibility-aptness. Yet, even if Williams was most likely the same moral self as before and answerable for his former crimes, this does not undermine the possibility that he was indeed rehabilitated. This

[551] Ibid.,274.
[552] Ibid.,295.

question concerning his rehabilitation, as I have argued, is a different and trickier question, which perhaps generates a different response.

Gov. Schwarzenegger said that Williams' regret was communicated "only through innuendo and inference."[553] He implied that Williams was unrepentant due to a lack of an explicit apology. As Williams maintained his innocence until the end, he swore he would "never apologize for capital crimes that [he] did not commit—not even to save [his] life."[554] What is important in an apology is not the act of apologizing, but how an apology affects interpersonal relations. An apology is forward looking and involves a repudiation of the offending activity. This acts as a kind of promise that a modification in the relationship that might be justified due to the offending activity will be unwarranted. A sincere apology gains traction, importantly, because it gives insight into what we can expect in future of the offender's character.

Williams may not have apologized for his past crimes, but he could have achieved the same effect by a different means (assuming the changes he underwent were sincere). To satisfy Schwarzenegger's concerns, he would have needed to recognize his responsibility for these crimes through narrative appropriation and give reason that certain relationships could be restored in a manner like an apology. I argued that narrative appropriation allows one to take responsibility for the past by providing a means for the moral lessons learned from past wrongdoing to come to inflect that agent as they currently stand. In this way, I argued that the narrative provides a sense of guided change. Like an apology, narrative appropriation therefore provides a reason to withdraw

[553] Schwarzenegger, *Statement of Decision*, 5.
[554] Williams, Redemption, 25.

blame: there is a reason to think that a relationship can be restored, not necessarily to where it once was, but to something morally adequate for moving forward in the relationship.

Williams could be said to have taken responsibility through narrative appropriation in a way that mitigates blame and encourages rehabilitation. We would need to know if he simply denied his past or integrated it into his future narrative conception and whether these changes are not just those of happenstance, but aimed at righting the wrongs of the past. For instance, does his work in guiding youths away from the life of crime he regretted provide a reason to modify expectations of his future character?

Williams's memoirs provide a narrative of this sort. In Williams' own words:

> There was no defining moment that marked my redemption, no voice of reason from the sky, no jolt of energy. The path of education and introspection enabled me to reason and to develop a conscience that rejects criminality, drugs, and senseless violence. Redemption allowed me to acknowledge and atone for my past indiscretions, vow never to repeat or create new ones—and extend an olive branch to youths and adults who desire peace.[555]

The narrative Williams professed to maintain was one of redemption that might have continued if his appeal for clemency had been granted. I maintained in the thesis that rehabilitation involves more than a question of strict responsibility-aptness, but also requires that we consider the conditions of blame and forgiveness. Blame and forgiveness, I argued, should be best addressed by means of narrative appropriation.

However, it is possible that Williams was not rehabilitated even in light of his redemptive narrative self-conception; at least, not yet. Like Alex at the end of Burgess' novel, there is hope of lasting change. But also like the novel, that story of potential

[555] Ibid.,295.

redemption is left untold. Perhaps then it is on this point that we may now begin to raise some larger questions about the justice of his treatment. The argument in the thesis also gives us reason to say that the choice to execute Williams was unjust because it pre-empted the possibility of the rehabilitation that likely would have occurred. If he was a changing self, guided by a different narrative self-conception, should he have been subject to this treatment? The account I have laid out in this thesis allows us to answer in the negative. Even if a person remains responsibility-apt and the same moral self, due to their rehabilitative efforts, they may no longer be blameworthy. We should be wary about interrupting these rehabilitative efforts by applying a punishment that is so final. Yet, questions concerning the justification of punishment are beyond the scope of the thesis. Nevertheless my argument does show that although Williams is the same self and remains responsibility-apt during the personality change, this does not render the change morally inert. Rather, it is indicative of his attempt to take responsibility for the past and become a new moral self. The state of California may not have executed an innocent man, but they likely executed a rehabilitated one.

Bibilography

Arpaly, Nomy. "Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph K. Clarke and Angela Smith, 141- 156. Oxford: Oxford University Press, 2015.

---. "Praise, Blame and the Whole self." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*93, no. 2 (1999).

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

Ashby Plant, E., and B. Michelle Peruche. "The Consequences of Race for Police Officers' Responses to Criminal Suspects." *Psychological Science* 16, no. 3 (2005): 180-83. doi:10.1111/j.0956-7976.2005.00800.x.

Blum, Lawrence A. *Moral Perception and Particularity*. Cambridge: Cambridge University Press, 1994.

Blustein, Jeffrey. "On Taking Responsibility for One's Past." *Journal of Applied Philosophy* 17, no. 1 (2000).

Brink, David O. "Situationism, Responsibility, and Fair Opportunity." *Social Philosophy and Policy* 30, no. 1-2 (2013): 121-49. doi:10.1017/S026505251300006X.

Brownstein, Michael, and Jennifer Mather Saul, eds. *Implicit Bias and Philosophy*. Volume 2, Moral Responsibility, Structural Injustice, and Ethics . Oxford: Oxford University Press, 2016. 2016. Accessed November 1, 2018.

Brison, Susan, "Trauma, Memory and Personal Identity." In *Feminists Rethink the Self*, edited by Diana Tietjens Meyers. (Boulder: Westview Press, 1997).

---. *Aftermath : Violence and the Remaking of a Self*. Princeton: Princeton University Press, 2002.

Burgess, Anthony. *A Clockwork Orange*. New York: W.W. Norton, 1986.

Carroll, Noël. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *The Journal of Aesthetics and Art Criticism* 60, no. 1 (2002): 3-26. doi:10.1111/1540-6245.00048.

---, “Learning through Fictional Narratives in Art and Science”. In *Beyond Mimesis and Convention:P Representation in Art and Science*, edited by Roman Frigg and Matthew Hunter, 51-70. Dordrecht: Springer, 2010.

Craig, Jared N. "Incarceration, Direct Brain Intervention, and the Right to Mental Integrity - a Reply to Thomas Douglas." *Neuroethics* 9, no. 2 (2016): 107-18. doi:10.1007/s12152-016-9255-x.

Correll, Joshua, Sean M Hudson, Steffanie Guillermo, and Debbie S Ma. "The Police Officer's Dilemma: A Decade of Research on Racial Bias in the Decision to Shoot." *Social and Personality Psychology Compass* 8, no. 5 (2014): 201-13. doi:10.1111/spc3.12099.

Coventry, Angela, and Uriah Kriegel. "Locke on Consciousness." *History of Philosophy Quarterly* 25, no. 3 (2008): 221-42.

Davies, David. *Aesthetics and Literature*. Continuum Aesthetics. London: Continuum, 2007

Dennett, Daniel. *Consciousness Explained*. Boston: Little, Brown and Co, 1991.

---. “The Self as a Center of Narrative Gravity”. In: F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness: Multiple Perspectives.* Hillsdale, NJ: Erlbaum, 1992.

Devine, Patricia G, E. Ashby Plant, David M Amodio, Eddie Harmon-Jones and Stephanie L Vance. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice." *Journal of Personality and Social Psychology* 82, no. 5 (2002): 835-48.

Doris, John M. *Lack of Character: Personality and Moral Behavior*. Cambridge, U.K: Cambridge University Press, 2002.

Duff, Anthony. “Choice Character and Action” In *Readings in the Philosophy of Law*, edited by Culver, Keith C, and Michael Giudice. *, 392-409.* Peterborough, Ontario: Broadview Press, 2017.

Egan, Andy. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*140, no. 1 (2008): 47-63.

Elgin, Catherine, Z. “The laboratory of the Mind”. In *A Sense of the World: Essays on Fiction, Narrative, and Knowledge*, edited by John Gibson, Wolfgang Huemer, and Luca Pocci,43-54. New York: Routledge, 2007.

Fisher, Martin and Neal A. Tognazzini. “The Triumph of Tracing”. In *Deep Control: Essays on Free Will and Value*, edited by Fischer, John M, 206-234. Oxford: Oxford University Press, 2012.

Frankfurt, Harry G. *The Importance of What We Care About: Philosophical Essays*. England: Cambridge University Press, 1988.

---. *Necessity, Volition, and Love*. Cambridge, U.K: Cambridge University Press, 1999.

---, “Reply to Gary Watson.” *Contours of Agency: Essays on Themes from Harry Frankfurt*. Edited by Buss, Sarah, and Lee Overton, 160-165.Cambridge, Mass: MIT Press, 2002.

Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007.

---. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (2016): 165-83. doi:10.1111/nous.12067.

---. “ Explaining Blame and Forgiveness” Peasoup.com. May 26, 2015. Accessed January 27, 2019. https://peasoup.typepad.com/peasoup/2015/05/explaining-blame-and-forgiveness-by-featured-philosopher-miranda-fricker.html

Gallagher, Shaun. *How the Body Shapes the Mind*. Oxford: Oxford University Press, 2005.

Goldie, Peter. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press, 2012.

Graham, Gordon. "Learning from Art." *The British Journal of Aesthetics* 35, no. 1 (1995): 26-37. doi:10.1093/bjaesthetics/35.1.26.

---. *Philosophy of the Arts: An Introduction to Aesthetics*. London: Routledge, 1997

Graff, Delia. "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 28, no. 1 (2000): 45-81.

Griswold, Charles L. *Forgiveness: A Philosophical Exploration*. New York: Cambridge University Press, 2007.

Haney, Craig. *Reforming Punishment : Psychological Limits to the Pains of Imprisonment*. First Edtion ed. The Law and Public Policy. Washington, DC: American Psychological Association, 2006.

Healy, Deirdre. *The Dynamics of Desistance : Charting Pathways through Change*. International Series on Desistance and Rehabilitation. London: Routledge, 2012.

Holton, Richard. "How is Strength of Will Possible?" In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, 39-67.Oxford University Press, 2003

Holroyd, Jules. "Implicit Bias, Awareness and Imperfect Cognitions." *Consciousness and Cognition* 33 (2015): 511-23. doi:10.1016/j.concog.2014.08.024.

Holroyd, Jules, Robin Scaife, and Tom Stafford. "Responsibility for Implicit Bias." *Philosophy Compass* 12, no. 3 (2017). doi:10.1111/phc3.12410.

Husserl, Edmund, and John B Brough. *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*. Edmund Husserl Collected Works, V. 4. Dordrecht: Kluwer Academic, 1991.

Jones, K. “How to Change the Past”. In K. Atkins & C. Mackenzie (Eds.), *Practical identity and narrative agency* , 269–288. New York: Routledge. 2008

Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford Univ. Press, 2009.

Kubrick, Stanley, director. *A Clockwork Orange*. Warner Bros, 1971.

Lamarque, P. "On the Distance between Literary Narratives and Real-Life Narratives." *Royal Institute of Philosophy Supplement* 60, no. 60 (2007): 117-32.

Levy, Neil. "Culpable Ignorance and Moral Responsibility: A Reply to Fitzpatrick." *Ethics* 119, no. 4 (2009): 729-41. doi:10.1086/605018.

---. "Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action." *Analytic Philosophy* 52, no. 4 (2011): 243-61. doi:10.1111/j.2153-960X.2011.00543.x.

---. *Consciousness and Moral Responsibility*. New York: Oxford University Press, 2014.

---. "Consciousness, Implicit Attitudes and Moral Responsibility." *Noûs* 48, no. 1 (2014): 21-40. doi:10.1111/j.1468-0068.2011.00853.x.

---. "Neither Fish nor Fowl: Implicit Attitudes As Patchy Endorsements." *Noûs* 49, no. 4 (2015): 800-23. doi:10.1111/nous.12074.

Leung, Rebecca. “Rewriting the Past: Former Crip Leader teaches Children how to Avoid Gangs”. *60 Minutes*. May 21, 2004.

Liem M, and Richardson N.J. "The Role of Transformation Narratives in Desistance among Released Lifers." *Criminal Justice and Behavior* 41, no. 6 (2014): 692-712. doi:10.1177/0093854813515445.

Locke, John, and Pauline Phemister. *An Essay Concerning Human Understanding*. Oxford World's Classics Paperback. Oxford, England: Oxford University Press, 2008.

Mackie, J. L. *Problems from Locke*. Oxford: Clarendon Press, 1976.

Mackenzie, Catriona, and Jacqui Poltera. "Narrative Integration, Fragmented Selves, and Autonomy." *Hypatia* 25, no. 1 (2010): 31-54. doi:10.1111/j.1527-2001.2009.01083.x.

Mandelbaum, Eric. "Against Alief." *Philosophical Studies : An International Journal for Philosophy in the Analytic Tradition*165, no. 1 (2013): 197-211. doi:10.1007/s11098-012-9930-7.

---. "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias." *Noûs* 50, no. 3 (2016): 629-58. doi:10.1111/nous.12089.

Maruna, Shadd. *Making Good: How Ex-Convicts Reform and Rebuild Their Lives*. Washington, D.C: American Psychological Association, 2001

--- and Derek Ramsden, "Living to Tell the Tale: Redemption Narratives, Shame Management, and Offender Rehabilitation." In Healing Plots: The Narrative Basis of Psychotherapy, edited by Amia Lieblich, Dan P. McAdams, and Ruthellen Josselson, 129-150. Washington, DC: American Psychological Association, 2004.

Mckendy, John P. "'I'm Very Careful About That': Narrative and Agency of Men in Prison." *Discourse & Society* 17, no. 4 (2006): 473-502. doi:10.1177/0957926506063128.

McKenna, M. "Directed Blame and Conversation." In *Blame : Its Nature and Norms*, edited by Justin. D Coates and Neal A Tognazzini, 119-140. New York: Oxford University Press, 2013.

Mill, John Stuart. *Utilitarianism*. Edited by Andrew Bailey. Peterborough, Ontario, Canada: Broadview Press, 2016.

Miller, Christian B. *Moral Character: An Empirical Theory*. Oxford: Oxford University Press, 2013.

Morse, Stephen J. "Hooked on Hype: Addiction and Responsibility." *Law and Philosophy* 19, no. 1 (2000): 3. doi:10.2307/3505173.

Moskowitz, Gordon B, and Peizhong Li. "Egalitarian Goals Trigger Stereotype Inhibition: A Proactive Form of Stereotype Control." *Journal of Experimental Social Psychology* 47, no. 1 (2011): 103-16. doi:10.1016/j.jesp.2010.08.014.

Nagel, Thomas. *Mortal Questions*. London: Canto, 1991.

Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1987.

Parnas, Josef, and Louis A. Sass. "The Structure of Self Consciousness in Schizophrenia." In *The Oxford Handbook of the Self,* edited by Shaun Gallagher, 521-546. Oxford: Oxford University Press, 2011.

Patihis, Lawrence, Steven J Frenda, Aurora K. R LePort, Nicole Petersen, Rebecca M Nichols, Craig E. L Stark, James L McGaugh, and Elizabeth F Loftus. "False Memories in Highly Superior Autobiographical Memory Individuals." *Proceedings of the National Academy of Sciences of the United States of America*110, no. 52 (2013): 20947-0952.

Paul, L. A. *Transformative Experience*. Oxford: Oxford University Press, 2014.

Payne, B.Keith, Alan J Lambert, and Larry L Jacoby. "Best Laid Plans: Effects of Goals on Accessibility Bias and Cognitive Control in Race-Based Misperceptions of Weapons." *Journal of Experimental Social Psychology* 38, no. 4 (2002): 384-96. doi:10.1016/S0022-1031(02)00006-9.

Perry, John. *A Dialogue on Personal Identity and Immortality*. Indianapolis: Hackett Pub, 1978.

Perry, John. “Selves and Self-Concepts.” In *Time and Identity*, edited by Campbell, Joseph K, Michael O'Rourke, and Harry Silverstein, 229-248. Cambridge: MIT Press, 2010.

Proust, Marcel. *Swann's Way: The Moncrieff Translation*, edited by Susanna Lee. Translated by C. K Scott-Moncrieff, First edition. New York: W.W. Norton & Company, 2014.

O'Connor, Patricia E. *Speaking of Crime: Narratives of Prisoners*. Stages, V. 17. Lincoln: University of Nebraska Press, 2000.

Oslen, Eric. T. “Personal Identity.” In *The Blackwell Guide to Philosophy of Mind*, edited by Stich, Stephen P, and Ted A Warfield, 352- 368. Maldon: Blackwell Pub, 2003.

Quine, W.V. and Ullian, J.S. *The Web of Belief.* New York: Random House, 1979

Radden, Jennifer. *Divided Minds and Successive Selves : Ethical Issues in Disorders of Identity and Personality*. Philosophical Psychopathology. Disorders in Mind. Cambridge, Mass.: MIT Press, 1996.

Railton, Peter, "Practical competence and fluent agency" *Reasons for Action*. Edited by Sobel, David, and Steven Wall. Cambridge, UK: Cambridge University Press, 2009.

Ricœur, Paul. *The Rule of Metaphor: The Creation of Meaning in Language*. (Routledge Classics. London: Routledge, 2003

Ricoeur, Paul, Kathleen McLaughlin, and David Pellauer. *Time and Narrative, Volume 1*. Chicago, IL: University of Chicago Press, 2012.

Roberts, Robert C. "Forgivingness." *American Philosophical Quarterly* 32, no. 4 (1995): 289-306.

Rosen, Gideon. "Skepticism About Moral Responsibility." *Philosophical Perspectives*, vol. 18, no. 1, 2004, pp. 295–313. doi:10.1111/j.1520-8583.2004.00030.x.

Rosenblatt, Kalhan. *Hot Car Deaths: Scientists Detail Why Parents Forget Their Children*. NBC news. Jun.27.2017 / 11:18

Sartre, Jean-Paul. *Being and Nothingness: An Essay in Phenomenological Ontology*. Translated by Sarah Richmond. Abingdon, Oxon, UK: Routledge, 2018.

Saul, J. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy What Needs to Change?,* edited by Katrina Hutchinson and Fiona Jenkins, 39-58. New York: Oxford University Press, 2013.

Sher, George. *Who Knew?: Responsibility Without Awareness*. Oxford: Oxford University Press, 2009.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, Mass: Belknap Press of Harvard University Press, 1998.

---. "Interpreting Blame." In *Blame : Its Nature and Norms*, edited by Justin. D. Coates and Neal A Tognazzini, 84-99. New York: Oxford University Press, 2013.

Schechtman, Marya. *The Constitution of Selves*. Ithaca, NY: Cornell University Press, 1996.

---. "Empathic Access: The Missing Ingredient of Personal Identity." In *Personal Identity,* edited by Raymond Martin and John Barresi, 238-259. Oxford: Blackwell, 2003. 238-259

---. "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View." *Royal Institute of Philosophy Supplement* 60 (2007): 155-78. doi:10.1017/S1358246107000082.

---. "The Narrative Self". In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 394-418. Oxford: Oxford University Press, 2011.

---. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. New York: Oxford University Press, 2014.

--- "'The Size of the Self": Minimalist Selves and Narrative Self-Constitution." In *Narrative, Philosophy and Life*, edited by Allen Speight. 33-47. Dordrecht: Springer, 2015.

--- "A Mess Indeed: Empathic Access, Narrative, and Identity". In *Art, Mind and Narrative: Themes from the Work of Peter Goldie*, edited by Julian Dodd, 17-34. Oxford University Press, 2016.

Schwarzenegger, Arnold. Statement of Decision: Request for Clemency by Stanley Williams, (Dec 12, 2005): 3. https://graphics8.nytimes.com/packages/pdf/national/Williams_Clemency_Decision.pdf (accessed July 25th 2018)

Shoemaker, D. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121, no. 3 (2011): 602-32. doi:10.1086/659003.

----. *Responsibility from the Margins*. Oxford: Oxford University Press, 2015. 2015. Accessed November 1, 2018.

---. "Ecumenical Attributability." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph K. Clarke and Angela Smith, 115- 140. Oxford: Oxford University Press, 2015.

---, "Personal Identity and Ethics", *The Stanford Encyclopedia of Philosophy* last modified Winter 2016, Edited by Edward N. Zalta, https://plato.stanford.edu/archives/win2016/entries/identity-ethics/.

Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236-71. doi:10.1086/426957.

---. "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11, no. 4 (2007): 465-84.

---. "Control, Responsibility, and Moral Assessment." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*138, no. 3 (2008): 367-92.

---. “Book Review: Who Knew? Responsibility Without Awareness.” Social Theory and Practice 36 no. 3 (2010): 515-524.

---. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122, no. 3 (2012): 575-89. doi:10.1086/664752.

---, “Blame and Moral Protest” In *Blame: Its Nature and Norms*, edited by Justin D. Coates and Neal A Tognazzini, 27-48. Oxford University Press, 2012.

---. "Attitudes, Tracing, and Control." *Journal of Applied Philosophy* 32, no. 2 (2015): 115-32. doi:10.1111/japp.12107.

Smith, Holly M. "Non-Tracing Cases of Culpable Ignorance." *Criminal Law and Philosophy* 5, no. 2 (2011): 115-46.

Strawson, Galen. *Locke on Personal Identity: Consciousness and Concernment*. Princeton: Princeton University Press, 2014.

Strawson, Galen. "Against Narrativity." *Ratio* 17, no. 4 (2004): 428-52.

---. "Episodic Ethics." *Royal Institute of Philosophy Supplement* 60 (2007): 85-116. doi:10.1017/S1358246107000057.

Stuart, Matthew. *Locke's Metaphysics*. Oxford: Clarendon Press, 2013.

Talbert, Matthew. "Moral Competence, Moral Blame, and Protest." *The Journal of Ethics*. 16.1 (2012): 89-109.

---. “Akrasia, Awareness and Blameworthiness.” In *Responsibility: The Epistemic Condition, e*dited by Philip Robichaud, Jan Willem Wieland, 47-63. Oxford: Oxford University Press, 2017.

Thiel, Udo. *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford: Oxford University Press, 2011.

Toffolo, Marieke B. J, et al. “Proust Revisited: Odours As Triggers of Aversive Memories.” *Cognition and Emotion*, vol. 26, no. 1, 2012, pp. 83–92. doi:10.1080/02699931.2011.555475.

Vargas, Manuel. "The Trouble with Tracing." *Midwest Studies in Philosophy* 29, no. 1 (2005): 269-91. doi:10.1111/j.1475-4975.2005.00117.x.

Velleman, David. “Well-being and Time” In *Metaphysics of Death*, edited by Martin Fischer, 327-362. Stanford: Stanford University Press, 1993.

---. “Identification and Identity”. In *Contours of Agency : Essays on Themes from Harry Frankfurt,* edited by Sarah Buss and Lee Overton, 129-159. Cambridge: MIT Press, 2002.

---. *Self to Self: Selected Essays*. Cambridge: Cambridge University Press, 2006.

Vincent, Nicole A. "Restoring Responsibility: Promoting Justice, Therapy and Reform through Direct Brain Interventions." *Criminal Law and Philosophy: An International Journal for Philosophy of Crime, Criminal Law and Punishment*, no. 1 (2014): 21-42. doi:10.1007/s11572-012-9156-y.

Warren, Jennifer and Mura Dolan. “Tookie Williams Is Executed”. LAtimes.com (Dec 13, 2005). http://www.latimes.com/local/la-me-execution13dec13-story.html. (accessed July 25th 2018)

Watson, Gary. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227-48.

---. “Volitional Necessities” In *Contours of Agency : Essays on Themes from Harry Frankfurt,* edited by Sarah Buss and Lee Overton, 129-159. Cambridge: MIT Press, 2002.

---. *Free Will*. Oxford: Oxford University Press, 2003.

Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations After Wrongdoing*. Cambridge: Cambridge University Press, 2006.

Williams, Bernard. *Shame and Necessity*. Berkeley: University of California Press, 1993.

Williams, Castle A, and Andrew J Grundstein. "Children Forgotten in Hot Cars: A Mental Models Approach for Improving Public Health Messaging." *Injury Prevention* 24, no. 4 (2018): 279-87. doi:10.1136/injuryprev-2016-042261.

Williams, Stanley. *Blue Rage, Black Redemption: A Memoir*. New York: Touchstone, 2004.

Weinberg, Shelley. "Locke on Personal Identity." *Philosophy Compass* 6, no. 6 (2011): 398-407. doi:10.1111/j.1747-9991.2011.00402.x.

---. "The Coherence of Consciousness in Locke's ‘Essay’." *History of Philosophy Quarterly* 25, no. 1 (2008): 21-39.

Westlund, Andrea C. "Rethinking Relational Autonomy." *Hypatia* 24, no. 4 (2009): 26-49.

---. "Autonomy and the Autobiographical Perspective." In *Personal Autonomy and Social Oppression: Philosophical Perspectives*, edited by Marina Oshana, 85-102. New York: Routledge, 2015.

Wiesel, Elie. "One Must Not Forget," interview by Alvin P. Sanoff, *US News & World Report* (27 Oct 1986)

Wolf, Susan. *Freedom Within Reason*. Oxford University Press, 1994.

Zahavi, Dan. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass: MIT Press, 2005.

---. *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press, 2014.

---. "The Unity of Consciousness" In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 316-338. Oxford: Oxford University Press, 2011.

Zimmerman, Michael J. "Varieties of Moral Responsibility." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph K. Clarke and Angela Smith, 45-64. NY: Oxford University Press, 2015.