MONASH University

Desire, Subjectivity, and the Self: A Relational Account of Lived Experience

James Alessandro Wells
Bachelor of Arts with Honours., Diploma of Languages in German

A thesis submitted for the degree of Master of Arts at

Monash University in 2021

School of Philosophical, Historical, and International Studies

Copyright notice

© The author 2021.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

This thesis is an investigation of contemporary philosophical accounts of the self. Beginning from the groundwork laid by the canonical debate regarding self-formation between Locke and Hume, I identify that what is central to a notion of 'self' in the Western tradition is a robust, explanatory relationship between a subject of experience and their beliefs, desires, and intentions. This background informs the dominant account of the self currently endorsed by philosophers, narrative theory, which claims that subjects are connected to the contents of their lived experience, their motivations, and their actions by enacting a unifying, teleological narrative.

In binding these features of the self together, narrative theory provides a framework for describing self-formation, however, I find that it also suffers from serious limitations. As it relies on a troubling form of holism, which undermines the epistemological basis of narrative accounts of the self, narrative theory cannot satisfactorily account for the presence of self-constituting beliefs, desires, and intentions in every subject's life.

In light of these problems, I propose an alternative account of self-formation, grounded in a capacity to take on and respond to relations of care. Not only does this 'relational' position fit better with the wide variety of forms that human lived experience can take but it also allows for a clear limit to be established on how the notion of the 'self' can intelligibly be used.

As self-formation is a relationship between a subject of experience and their beliefs, desires, intentions, and values, I define its limit in terms of how a person's capacity to relate the world can be destroyed. An example of this is provided by clinical analysis of the effects of torture, which strips away a subject's ability to coherently desire and thereby connect with sources of meaning and value in the world that make their lived experience intelligible.

From this basis, I am able to conclude that narrative theory provides an unsatisfactory account of self-formation as it draws an arbitrarily narrow boundary on which kinds of subjects

and lives legitimately count as displaying and constituting a self. Additionally, in describing the destruction of the self in terms of the destruction of a coherent capacity to desire, I raise a number of open questions regarding how persons who enjoy atypical forms of lived experience ought to be treated ethically and counted as selves.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Print Name: James Alessandro Wells

Date: 25/03/2021

Acknowledgements

I wish to thank my supervisors, Graham Oppy and Monima Chadha, for the support and guidance they provided me with throughout the course of composing this thesis. The clarity of both my thinking and writing has been changed immeasurably for the better by their input.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Table of Contents

Introduction – Understanding Lived Experience: The Question of Self-Formation

Chapter One – The Narrative Account of Self-formation
- Introduction: The commitments of narrative theory
- 1.1 Narration as the ground of intelligible action, narrative as the ground of an intelligible life
- 1.2 The narrative present is explained by reference to a further past and future
- 1.3 A single, unified self is grounded in the wholeness of a life narrative
- 1.4 The limits of narrative and narration make lives intelligible only in terms of an end
- Conclusion: The narrative point of view

Chapter Two – Problems with the Narrative Account
- Introduction: The issue of indispensability
- 2.1 The limits of narrative intelligibility
- 2.2 The limits of narrative temporality
- 2.3 The limits of narrative teleology and wholeness
- Conclusion: Everything right apart from the narrative?

Chapter Three – The Character Account and its Sceptics
- Introduction: The self as character and style
- 3.1 Camp's character account
- 3.2 The situationist challenge
- 3.3 Extending self-reference: Howell's extension of virtues
- Conclusion: The capacity to care and respond as relatedness

Chapter Four – A Relational Account of Self-formation
- Introduction: From characterisations to relations
- 4.1 Susan Wolf's account of meaningful lives and worth
- 4.2 Second-personal care: a relational account of need
- 4.3 Classical Confucian accounts of relational self-formation
- Conclusion: A sketch of the limits of the self

Chapter Five – The Boundaries of the Self
- Introduction: Desire as the boundary of the self
- 5.1 The thread view of meaning
- 5.2 Torture and the loss of desire
- 5.3 Value fanaticism
- 5.4 Hypocrisy and rehabilitation
- Conclusion: Desire and value

Conclusion – Idealised Lives and Ideal Ways of Living

Introduction – Understanding Lived Experience: The Question of Self-Formation

My thesis is an attempt to answer the following question: if there is a self which explains human experience and action, what is the best account that can be given of it? Specifically, I will deal with the relationship between the notion of 'the self' and the beliefs, desires, and intentions that constitute agency. The best account of *self-formation*, that is, of how persons define themselves and are defined by others as agents in terms of a thick notion of who they are, will unify the concept of the self and beliefs, desires, and intentions in the most coherent and explanatorily rich way.

The foundation of this question can be traced back to Locke's (1690/1997, p. 312) description of the "forensic" self in his *Essay*, that is, the notion of a "personality" which is "concerned and accountable," and "owns and imputes to it*self* past actions". This self plays a central explanatory role in describing the particular human being who it identifies as both an *agent* and a *subject of experience* by describing what they do in terms of a particular *way of being*. As such a self, a person is *constituted* by what they care about and value, defined by their beliefs, desires, and intentions.

Of course, this foundation has been challenged and the modern notions of self-formation I will examine owe just as much to Hume's (1739-1740/1878, p. 543) famous reduction of the self to a "bundle of perceptions" as they do to Locke's thick conception of it. This disagreement establishes the stakes of the debate around self-formation, as it were, by confronting philosophers with the problem of what accounts of agency and lived experience stand to lose if the self is reduced to a flux of simple impressions or isolated moments.

Roderick Chisholm (2013, p. 39) illuminates this problem clearly with his rejection of a purely impression-based account of "mind" or "self". "It is not the idea of the perception of

love or hate" or any other experience or attribute but "an idea of *that which* loves or hates" that defines what the self is (Chisholm 2013, p. 39, emphasis added). On this view, agents' acquaintance with their own experience and grounds for action does not take the form of disparate conceptions of sense-data but that of some kind of coherent *experiencing*.

If it is implausible for Hume to claim that "he doesn't find himself" in his experiences and yet to also "find himself stumbling" across various perceptions, then philosophers are left needing to define what form this 'finding' or self-formation takes (Chisholm 2013, pp. 39-40). How ought philosophers to account for the connection between the 'idea of that which experiences,' of who lives a person's life, and the actual experiences that constitute that life and person?

A similar concern regarding the "liquidation of the self" into "a series of unconnected episodes" within a life has prompted Alasdair MacIntyre (2007, p. 174) to connect the traits that constitute it to the structure of an extended, enacted, historical narrative, a theory which has grown to dominate the literature on self-formation.[1] For philosophers like MacIntyre (2007, p. 175) and those who follow him, a strong notion of "unitary life" which forms a conceptual "whole" is necessary to explain human lived experience and bridge the gap left between lived experience and the self-that-lives left by Humean scepticism.

Paul Ricœur (1992) has also offered a narrative account of self-formation, in which the beliefs, desires, and intentions of an agent are made sense of by being fitted into the plot of a life story, however, with a greater emphasis on the role played by others in narration. Specifically, Ricœur (1992, p. 336) attempts to establish what constitutes the self's "capability of responding" to the world and others, which grounds the ability to give a narrative account of agency and lived experience.

[1] Philosophers now offer narrative accounts of everything from personal identity to free will to psychotherapy. See: Schechtman, (2015), Fiorello (2019), and Lumsden (2013a).

This difference between a focus on unity and capability in the two major narrative approaches to self-formation suggests that questions regarding the constitution of the self and the task of providing a convincing answer to Hume are far from settled.[2] What narrative theories "purport to capture [is] the right connection between a subject and an experience or action" (Lumsden 2013a, p. 6). However, there seems to be a significant disconnection, and very weak entailment, between the ability to *unify* all of a person's experiences into a *whole*, self-constituting life narrative and the capacity to respond to the world, let alone narrate an account about it.

It is a recognition of the need to bridge this divide that has prompted Elisabeth Camp (2011, para. 1, https://nonsite.org/wordsworths-prelude-poetic-autobiography-and-narrative-constructions-of-the-self/) to hold that while "we do need rich, substantive selves of the sort delivered by narratives…both in order to evaluate our past actions and guide future ones," life stories by themselves are not what is important to self-formation. In a similar vein, Galen Strawson (2004; 2016) denies that both the long-term temporal connectedness and meaning-making that are characteristic of narrative theory actually capture every human subject's ability to care about and respond to the contents of their lived experience.

Even more pressure has been put on thick notions of self-formation by the situationist scepticism about character traits advanced by Gilbert Harman (2009) and Mark Alfano (2013). In a move reminiscent of Hume's, these philosophers maintain that positing a self as a locus of agency and factor that makes sense of experience over time flies in the face of the facts. Rather than seeking a connection between a life and the self living it, situationists instead argue that the relationships between a life and the milieus in which it is lived generate the traits which produce and explain behaviour.

[2] Not only are MacIntyre and Ricœur the historical forefathers of modern narrative theory but contemporary theorists typically model their accounts either one or the other of these thinkers' positions. Compare: Rudd (2009, p. 61) and Lumsden (2013a, pp. 6-7)

It is my hypothesis that this kind of scepticism and many of the core features of self-formation that narrative theorists posit are wrong: a thick concept of the self still plays an important role in theories which seek to explain human agency and lived experience. The mistakes implicit in both of these positions can be remedied, however, by a retreat from a narrative to a capability or care-based configuration of life, in which a person's intersubjective relations ground beliefs, desires, and intentions.

To demonstrate this, in chapter one I set out the general commitments of narrative theory and the rich notion of self-formation they are taken to generate. This response to a Humean liquidation of the self will serve as the model for explaining agency and experience, to which I shall develop an alternative over the course of my thesis.

Chapter two begins this process by interrogating which elements of life stories provide a plausible basis for self-formation. The key criterion critics have established is that for an account of the self to be plausible, it must guarantee a capacity for self-formation at any point in a subject's lived experience during which they display even weak beliefs, desires, and intentions. This I define in terms of the capability to respond and care about the contents of one's life as described by Ricœur.

Having separated the capacity to respond from narrative form, I consider how an alternative configuration of agential traits might bridge the gap between lived experience and the living self. Considering Camp's suggestion that self-formation retreat from the narrative plane to focus instead upon the 'characterisation' of a person in terms of their traits, I offer a defence of this position from situationism by demonstrating that even Harman's strict scepticism about character traits itself relies on agents having a relational capacity to care.

This work concludes the defence of the first half of my hypothesis: that a non-narrative, thick account of self-formation is both viable and necessary for theories which seek to explain human agency and lived experience. My final two chapters focus on its second claim: that a

relational capacity to care does in fact provide an adequate answer to Humean scepticism regarding the self by providing the basis for self-formation.

In chapter four, I advance a theory of relationality derived from Susan Wolf's (2014) account of meaningful lives. By establishing an intersubjective basis for meaning, in which both objective and subjective attitudes are made sense of in a realm of interpersonal values, this theory allows Ricœur's notion of responsiveness to others and the world to be fully decoupled from the vocabulary of narrative theory. I demonstrate this by aligning the ontological conclusions that Ricœur draws regarding the self with two different relational theories which support Susan Wolf's model; Katharine Wolfe's (2016; 2017) relational theory of the self, grounded in feminist care ethics; and Classical Confucian accounts of relational self-formation (Ivanhoe and Van Norden eds. 2005).

With a relational basis for self-formation established, I turn to focus on the self's relationship with agency and experience that lies at the heart of the disagreement between Locke and Hume. My final chapter considers how self-formation can reasonably fail in terms of how agency can be lost and how the subject of experience who lives a life be reduced to less than a self by extreme trauma. Drawing on Leanh Nguyen's (2007) work on the effects of torture, I make explicit an underlying theme that permeates the discussion of self-formation all throughout my thesis: the connection between desire and agency. Losing their capacity to desire renders a subject of experience incapable of intelligibly formulating a response to what they live in relational, agential terms. Actions and responses to the world that require a rich concept of self to be made sense of also require this minimal capacity to desire – as well as to believe and intend – which the trauma inflicted by extreme torture robs persons of.

On this basis, I conclude my discussion by noting the questions that this relational account of self-formation leaves open, specifically those regarding the morality of rehabilitation. If a *particular* nexus of beliefs, desires, and intentions, as well as other character

traits, is constitutive of a particular self, then engaging in a practice which neuters or replaces some of these may lead to the destruction of a self and the formation of a new one, rather than smooth personal development. In light of this, I suggest that philosophers ought to prefer a more eclectic picture of lived experience and human subjectivity rather than accounts of the self which place a strong emphasis on ideal, moralistic forms of life.

Chapter One – The Narrative Account of Self-formation

Introduction: The commitments of narrative theory

What is a narrative and why turn to this concept to describe human life?[3] To begin with, it is important to distinguish two key terms that underscore narrative accounts of self-formation: narration and narrative. Narration is a particular attempt at communication and description; at *giving an account* of something. Narratives are tales that can be told about something in general, which, although they themselves must be described through communication, are not necessarily either a description or communication of anything themselves.

The form of narratives is useful, even when they do not have a real subject. Even Daniel Dennett (1992, p. 103), who claims narratives about the self are ontologically equivalent to "a theorist's fiction," acknowledges that it is a useful schema for *unifying* the experiences of a person's life. Stories bind together important elements in terms of a central plot that plays out over time and, vitally, exclude or downplay other facts that might take place alongside the main action of their account.

This kind of a way of structuring lived experience is indispensable for providing reasons for why agents act and for explaining those reasons to themselves and others. As "we…find ourselves engaged in all sorts of behaviour, more or less unified, but sometimes disunified," people need some way of making sense of what agents actually experience (Dennett 1992, p. 114).[4] As they organise actions and explain their significance over the whole of their subjects'

[3] MacIntyre, Ricœur, and other narrative theorists typically limit plots to describing *human* lives, given the centrality of the ability to self-narrate to their accounts. For further discussion of what makes human lives and *self-formation* distinctive beyond species membership, see Singer (2012, p. 73).
[4] For another less-ambitious narrative account of the self, see: Velleman (1996).

lived experience, life stories provide a clear schema of intelligibility and standard of unity for human life.

However, many philosophers go far further than Dennett's commitments regarding the explanatory role narratives play and the kinds of selves who live it. Particularly, Alasdair MacIntyre and Paul Ricœur have claimed that human life *must* take a narrative form and that it is really lived by a subject of experience whose behaviour requires explanation in terms of their personal history, context, and motivations to be made intelligible: a self.

The reasons they provide for this connect with the other core aspect of narrative theory: the act of narration. Human lived experience, both philosophers claim, is legitimately open to being narrated at any time. That is, at any moment an agent could ask of themselves or be subjected to a demand from another to give an account of what they are doing. Narrative theorists claim that in order to respond to this, agents must be able to describe their reasons for acting. This is taken to involve not only an appeal to the present contextual factors which influence them but also to the end they intend to bring about by acting and the background information from their past that informs their decisions.

Both the further past and further future are also narratively-mediated, according to MacIntyre and Ricœur, as they too can only be made sense of in terms of a descriptive account that takes the form of narration. Moreover, as the schema that fixes the intelligibility of each of an agent's actions and descriptions is derived from the narrative accounts that they or another agent are able to give of what they are doing, self-formation must be explained in terms of a story.

Understood in these terms, an agent's self is both constituted by the plot of a whole life narrative, which unifies their lived experience, and also present in every particular episode and action that comprises the totality of their life. They are governed by ideals but faced with the

challenge of living up to them at each moment as they proceed towards these future goals from a past, whose traditions they must choose to sustain or break with.

This also makes who an agent is a matter of constant interpretation. As self-formation is constituted by the stories that can be intelligibly told about oneself, then, according to this position, one must always be ready to engage with new interpretations of who one is either by accepting or rejecting the kind of person they describe one as being and the kind of principles they describe one as prescribing to.

Narrative theory also offers a remarkable notion of stability in its account of self-formation. For as it maintains that self-understanding its limited to the narrative plane, life is made intelligible and coherent by the intelligibility and coherence of the overarching plot which governs it. This renders the whole life that an agent lives, if not every single moment of it or event within it, *purposeful*. It also explains self-formation in terms of a teleological, historically-extended concept of the 'good life'.

As agents act for reasons that aim their actions at projected outcomes and, as these outcomes themselves can only be made sense of in terms of a whole life story, agents' conduct is directed towards living a good life, defined in terms of these reasons. That is, by preferring certain outcomes and conceptions of themselves over others, agents have reasons to pursue courses of action which they deem likely to lead to the realisation of those outcomes and the development of those ways of being.

For narrative theorists, a whole life story not only opens self-formation to a world of possible conclusions and alternative paths for an agent to take but also prescribes one or more ideal ways of being for them. This, however, is always tempered by what can actually be narrated about a given person and what others can be convinced to believe about them. This imposes an array of strict demands on self-formation, whose explanations must mirror the form

of narration and whose overall description of the self must mirror the form of a unifying narrative.

My aim in this chapter is to present these core claims that constitute MacIntyre's and Ricœur's narrative accounts of the self. Namely, that a) intelligible human action is constrained by the form of narration; b) human life is temporally extended so that the present cannot be explained without reference to a further future and past; c) to describe a single, unified self, human life must take the form of a narrative whole; and d) the limits of intelligibility that narration and narratives describe prescribe a purpose for each person's life.

The reason that I constrain my focus to exegesis here is to identify what narrative theorists take to be the fundamental features of self-formation. As will become clear in the following chapter, I do not endorse the narrative model, however any better alternative to it will need to respond to the commitments that MacIntyre's and Ricœur's theories set out. Thus, I confine myself to noting the more problematic aspects of these positions here in anticipation of the more detailed analysis of them that is presented in the next chapter.

1.1 Narration as the limit of intelligible action, narrative as the limit of an intelligible life

According to Alasdair MacIntyre (2007, p. 180), "successfully understanding and identifying what someone is doing" cannot be separated from "placing a particular episode in the context of a set of narrative histories". The reason that narrative theorists give for this is that since human beings not only live in social contexts but develop and inherit social identities within those milieus, the understanding of human selves cannot be separated from the ongoing history of their environments.

On this model, actions themselves, as well as agents, can only be made intelligible in terms of their connectedness to a present context and past background. When related to actions and events in this way, human lived experience is made intelligible as a “hierarchy of units of praxis…each unit on its own level containing a specific principle of organization” (Ricœur 1992, p. 153). Agents act for reasons that are made sense of in terms of the end they to which contribute, as defined by the temporally extended socio-historical context in which they arise.

How does this connect lived experience and agency to the structure of narratives? According to these philosophers, the inescapability of some historical background, which *must* be understood in practical, narrative terms, makes the process of narration necessary for making sense of both an agent’s motivations and the life into which they fit. Agents not only derive meaning from this background but, it is claimed, also contribute to it and develop alongside it in such a way that while their way of being might deviate from its norms, the life they live cannot escape its form.

“To a large extent,” Ricœur (1992, p. 121) writes, “the identity of a person or a community is made up of these identifications with values, norms, ideals and models”. And MacIntyre (2007, p. 183) claims that “[w]e enter human society…with one or more imputed characters,” whose position in their own imputed narratives and the relatedness of this position to others make their experiences and interactions intelligible.

An agent’s understanding of who they are, it follows, is at least partly derived from a pre-existing structure of meaning. Yet what makes this narrative? Both MacIntyre’s and Ricœur’s positions acknowledge from the outset that there are limits on what human beings can coherently and genuinely communicate to one another and that these boundaries are contextually grounded.

For instance, it makes sense for someone to ask a fellow commuter at a bus stop for the time. There is a connection to the timetable and arrival of the bus in so far as the commuter has

an intention and desire to catch it and the relevant beliefs regarding whether it will arrive, which explains why they are waiting for it and why they have a reason for knowing the time. An account such as this can easily be attached to their utterance to make it intelligible in light of their motivations and the its context.

By contrast, it appears unintelligible for one commuter to instead say to another: "The name of the common wild duck is *Histrionicus histrionicus histrionicus*" (MacIntyre 2007, p. 178). What makes this utterance an unintelligible action is the lack of an account which connects the context and the commuter's motivations in a coherent way. Given the description of the situation, they do not appear to have any reasons for offering this fact to anyone.

What would make this action intelligible, according to narrative theorists, is exactly what distinguishes it from the first example I offered: an appropriate account. Moreover, both MacIntyre and Ricœur maintain that this is also what makes an agent's actions intelligible to themselves as their agent. Consider, for example, what would make sense of this action, according to MacIntyre (2007, p. 179); namely, the fact that the speaker is "a Soviet spy waiting at a prearranged rendez-vous and uttering the ill-chosen code sentence which will identify him to his contact".

As a spy, the commuter might have a wide range of reasons for saying the Latin name of the common wild duck. What makes it make sense for them to utter those words will depend on the history of the conflict between the Soviet Union and whichever country the spy finds themselves in, which in turn extends more broadly to the history of Russia, of Marxism, and the twentieth century, to name but a few further related explanatory accounts.[5]

However, on the model this example supports, these accounts will also interact with ones which explain the agent's own life and trace their development up to the moment in which

[5] This explanation assumes a perspective from which an interpreter understands both the *function* of the code, *that* it was chosen by the spy or their superiors to identify them, and the *reasons* why it was chosen. The epistemological problems that attend narrative explanations such as this will be dealt with further in the next chapter.

they speak the code sentence. The reason they have for being at the bus stop will be explained by their relationship with the USSR, with the values with which they grew up, and perhaps a passing interest in ornithology that inspired the code, to name but a few further accounts.

What narrative theorists claim an agent cannot do is make sense of an action without accessing some further personal and social histories. Indeed, in addition to providing an intelligible account of themselves, Ricœur also stipulates that an agent be able to offer a similar account of the actions they intend to perform. Just as with agents, "the unit of configuration constitutive to a practice is based upon a particular relation of meaning, that expressed by the notion of a constitutive rule," by which something "counts as" or "constitutes" an action (Ricœur 1992, p. 154).

For instance, if someone heard the commuter say the name of the common wild duck to them, they might well think that person suffers from a "possible form of madness" and, hence, that they are not really acting at all (MacIntyre 2007, p. 178). Otherwise, they might understand that the commuter is doing something, but be incapable of determining exactly what action they are performing and intend to perform.

This account of action and agency places a narrative limit on intelligibility as it requires both what is done and the person who does it to be open to description in terms of a historical account structured in terms of a story.[6] An action and the life of the agent who performs it are both made sense of in terms of a plot that extends over time and unifies its various contents into a coherent whole, while also being constituted by those particular actions and events.

This view states that it is only by recounting the histories of a person and practice that the values and other motivations which might inform them can be made sense of. Ricœur (1992, p. 158) holds that to ascribe something like "an ethical character to…life taken as a whole" it must be "gathered together…in the form of a narrative". That is, for a person's lived experience

[6] Even though, as my last footnote suggests, this account may not actually be grasped by the agent.

to be explicable in terms of overarching principles and traits, these features of self-formation must be made to fit into the structure of a plot.

It is from this basis that MacIntyre (2007, p. 182) claims that for human agents “a history is an enacted dramatic narrative in which the characters are also the authors”. Just as actions themselves are defined by constitutive rules, by which they *count as* something, agents *count among* others as a specific kind of self by dint of the role which they take up in their life. For Ricœur (1992, p. 114), “narrative theory finds one of its major justifications in the role it plays as a middle ground between the descriptive viewpoint on an action and the prescriptive view”.

Simply by virtue of defining actions and their agents in terms of broader practices and histories, aspects of which they must *enact* in order to intelligibly connect their behaviour to a source of meaning, narratives both define and stipulate limits to the self and what it can author. Acting intelligibly not only requires an agent to *take up* motivations from traditions and practices as part of a narrative description of who they are but also requires that agent be able to support that overarching account of their self-formation with specific acts of narration at any given point in time that it covers.

This description requires more to be said regarding the relationship between the narrative account of self-formation and the self’s connection to a broader temporal background. In particular, the manner in which traditions and practices are temporally extended and apprehended requires more explanation in order for it to yield a description of how narratives and specifically narrative form extend across a whole life.

1.2 The narrative present is explained by reference to a further past and future

By relying on traditions and practices to form a past which makes the present intelligible, narrative theorists seek to explain how present experience of the self both develops in reference to but is also capable of breaking from its historical background. This further past is itself full of changes, as, indeed, the present is from context to context. Traditions change over time, altering what they prescribe and describe and leaving those who rely upon them to make their lived experience intelligible through the choice of embracing or rejecting these differences.

For MacIntyre (2007, p. 188) "a tradition in good order…is always partially constituted by an argument [that is, a disagreement] about the goods the pursuit of which gives to that tradition its particular point and purpose". Partly, this statement is meant to hit back against conceptions of traditions and practices which are largely static and dogmatic. The very usefulness of this concept for narrative theory is derived from the fact that traditions present and defend claims which those faced with them must respond to, either by taking them up or rejecting them.

One who adheres to a tradition and its description of the world, or at least its presentation of some contexts and information within it, *uses* that tradition as an interpretative tool. In so far as human life is full of the unexpected, this will also involve carrying the elements of the traditional point of view one bears into the unknown. Radical shifts, such as the sudden emergence of COVID-19, force agents to readapt to new circumstances and present them with the choice of attempting to preserve or break with the traditions that guided them.

"A living tradition then is an historically extended, *socially embodied* argument …about the goods which constitute that tradition" (MacIntyre 2007, p. 188, emphasis added). 'Goods' here takes on two important but distinct meanings. First, and integral to the description of narrative intelligibility above, this term describes the constitutive standards by which something is identified as *fitting under* a description. That is, a tradition is identified by what, why, and how it distinguishes things from one-another.

Secondly, it is prescriptive. As a tradition describes things as being a certain way, it also stipulates how to be such a thing. Here, the notion of a 'constitutive rule' put forward by Ricœur is important to understanding this distinction. Traditions are constituted by particular descriptions, however these descriptions themselves describe a certain way of being that is independent from traditions which carve up the world in a different way.

This is what helps MacIntyre (2007, p. 186) distinguish that "[w]hat the good life is for a fifth-century Athenian general will not be the same as what it was for a medieval nun". Agents in both situations would be separated by different descriptions of what roles they should fill *as well as* different descriptions of how best to go about doing so. Both, for instance, might involve prayer but describing piety as an integral trait or virtue for someone in either role will nonetheless yield two vastly different images of how to *be pious*.[7]

The notion of an historically extended series of disagreements within MacIntyre's conception of traditions links with a fundamental aspect of Ricœur's account of self-formation. Namely, that self-formation can be expressed in two different forms, each describing a different relatedness across time. According to Ricœur (1992, p. 121) there is a significant distinction between physical measures of character (idem), such as "a *trait*…by which a person is recognized, reidentified as the same," and the "fidelity" a subject feels towards those traits (ipse), which motivates the "maintaining of the self" or self-formation in terms of them.

Although one of these aspects may dominate an agent's way of being, the role that they play in making the self intelligible makes it impossible to do away entirely with the other. Likewise, the two 'goods' of tradition must coexist despite the different roles they play in defining the whole they constitute. The similarities do not end there, though: what Ricœur describes is essentially two different ways of measuring and describing the good *for someone*.

---

[7] Note that it would not necessarily be impossible for either person to *try* to be pious after the fashion of the other but simply be very difficult practically and likely very difficult for them to make sense of doing.

That is, what is ‘good for one’ in terms of what traits are seen as beneficial, and in terms of what the person in question has a positive attitude towards.

Just as with the general and the nun, *idem* and *ipse* might both highlight the same feature of an agent, however the manner in which they do so will differ as it does in MacIntyre’s description. For a *characterisation* of both persons as pious to be true, it would require that they *display* piety, however, for them to genuinely appear as such a person *to themselves*, they must not only personally acknowledge but be guided by a principle of piety.

Narrative accounts of self-formation make selves similar to traditions in another way by recognising that both are subject to change and interact with the world across time. Part of what Ricœur’s distinction between *idem* and *ipse* captures is that just as traditions involve an ongoing ‘argument,’ so too does self-formation; new information might force or prompt an agent to reconsider whether they still relate to their motivations in a particular way, or whether they still possess the same traits that they did before.

However, this change is not only past- but also future-oriented, even when it comes to the concept of a tradition. Understanding self-formation in narrative terms not only involves unifying a past description with present contexts, but also determining what kind of future an agent’s present actions describe and the trajectory of their past leads towards. The reasons that narration makes intelligible become goals when understood in terms of a narrative.

Similarly, a tradition or practice intuitively aims at an outcome or endpoint. Especially in the case of the latter, what constitutes something as a kind of action is its fittingness for *causing* something. In a practical context, this might be something as simple as chopping onion in order to cook a meal; in the context of a game, it might be moving a pawn forward one square. Committing to the tradition of Marxism, by contrast, involves fitting one’s behaviour towards establishing local worker communes; an overarching endpoint in terms of which actions are made sense of.

Responding to an opponent's *en passant* by thumping half an onion onto the board and proceeding to dice it, like simply uttering the name of the common wild duck at a bus stop, does not qualify as an intelligible action according to this model. However, if practices and traditions are constituted in a similar way to the selves that they describe, then the actions of a person must also, narrative theorists argue, be directed towards some end.

The "attempt to envisage each human life as a whole" involves, for MacInytre (2007, p. 174), conceiving it as "a unity, whose character provides the virtues with an adequate *telos*". An aim echoed by Ricœur's comments concerning the role narrative plays as guaranteeing the ethical character of a whole human life.[8]

Both of these statements, while illuminating the next important claim at the heart of narrative accounts of self-formation, can only be made sense of in terms of the temporal extendedness selves display in taking on and reacting to practices and traditions. For in so doing, narrative theorists claim that agents not only require an openness to the further past but also to one or more further futures.

Additionally, the intelligible description of conduct I presented in this chapter's first section, which grounds narrative theory's explanation of the subject of experience, essentially plays out between these two poles. Selves are, then, on the narrative model, temporally extended in such a way that their whole life is made sense of in terms of the contributions of past, present, and future.

In presenting the temporal schema which undergirds the narrative account of intelligibility, I suggested that this theory of self-formation mirrors personal development with the development of the 'argument' within a tradition. However, for this parallel to itself make sense, the subject of experience that narrative theory describes must meet the same standards

[8] As I will discuss when I critique narrative theory, commitments like this risk being far too strong. The life of a person as monstrous as Hitler might well possess a 'unity' but certainly fails to present an 'adequate telos' from which virtues might be derived.

of coherence and wholeness as a tradition or practice does. Recognising this, the following section sets out accounts of what a *whole* life narrative derived from these claims must look like and what consequences this has for the *whole self* who lives it.

1.3 A single, unified self is grounded in the wholeness of a life narrative

What makes the way in which an agent acts not only in individual episodes but across their whole life coherent and intelligible in such a way that their experiences are unified? That is, what describes present character traits and relationships with oneself as following from a past? If traditions play such a fundamental role in narrative intelligibility, this kind of stable development over time seems necessary to any account of self-formation which they sustain.

For MacIntyre, the demands that the historicity and sociality of human lived experience place upon interpretations of it require a distinct kind of narration about the self in order to be adequately responded to. “The unity of a human life is the unity of a narrative quest,” that is, a life spent in pursuit of an overarching end which informs and binds together each discrete episode within it (MacIntyre 2007, p. 185).

This makes the intelligibility of a human life impossible to grasp “without at least some partially determinate conception of the final *telos*,” which guides its quest and determines its progression (MacIntyre 2007, p. 185). Likewise, this account of self-formation also accepts that human life is not only subjected to but must be measured by standards of success and failure. Just like a narrative itself, MacIntyre’s position disqualifies lives that cannot appeal to a strong enough unifying account to fit all of what they claim is important into a single coherent plot.

For MacIntyre, though, the importance of wholeness and unity not only covers the descriptive plane of lived experience but also extends to the prescriptive judgements that narration about the self generates. One element that is fundamental to account of self-formation is the notion of "accountability of the self," without which "those trains of events that constitute all but the simplest and barest of narratives could not occur" (MacIntyre 2007, p. 185).

Once again, the connection to a background constituted out of traditions and practices plays a key role for narrative theorists here. This claim, that what a self does must be measurable by some kind of account given of it, demands that there be some common, intelligible basis from which agents can interrogate themselves, each other, and their behaviour. When connected to the model of the 'narrative quest,' which describes the wholeness of a life, this sets out strict success and failure conditions which descriptions and prescriptions of a self must abide by.

MacIntyre's model requires agents to actually attempt to respond to the principles that are communicated through the common life defined by traditions, practices, and quest in order for these background features to provide a genuine description of them. It is not enough to simply *inherit* an identity as a person from one's socio-historical position. While an agent may be born into a certain society and role, they must nevertheless satisfy the demands placed on them as a member of that community and an individual agent for this description of their self to develop beyond an imputation and become *actual* (MacIntyre 2007, p. 183).

The alternative to following the demands that an agent is born with nonetheless involves developing as a different self; that is, into a different kind of person, who must be understood in terms of different kinds of traditions and practices. Simply rejecting one's starting point here does not lead to rejecting the notion of a unifying quest or the schema of intelligibility provided by a whole life narrative. Rather, renunciation merely leads one's life down a different path, making sense of it in light of a different plot, in MacIntyre's terms.

The notion of a ‘quest’ or search after an overarching goal takes a slightly different form in Ricœur’s setting out of narrative theory. Chiefly, Ricœur (1992, p. 170) develops an understanding of the mores and principles by which human life is understood in terms of a distinction between ethics, “that which is *considered to be good*,” and morality, “that which *imposes itself as obligatory*”.

According to this view, “the aim” of a life is primary over and against “the norms” which govern it (Ricœur 1992, p. 171). Given the practices and traditions that an agent inherits, they are presented first and foremost with competing ideas and models of how to *live well*, which, in turn, generate precepts for the agent to follow in pursing the particular form of life those principles describe. As with MacIntyre, an account of the ‘good life,’ understood as a ‘telos’ is central to this position.

“Whatever the image that each of us has of a full life,” Ricœur (1992, p. 172) explains, “is the ultimate end of our action”. On this narrative account as well, life requires an *overarching goal* in order for its constituent contents to be made sense of. As the relation between traditions, practices, and narratives is supposed to demonstrate, some kind of socio-historical account, which takes the form of narration, is required to render lived experience intelligible on both of these models. Hence, it is claimed, any description or prescription that is generated from the contents of a human life must take a narrative form.

The relationship that Ricœur emphasises between plot, teleology, and ethics arises most strongly here. Human life is always sensitive and always opens the subject living it to “the sphere of meaning in which appraisals of an evaluative (and subsequently normative) nature are attached to precepts of doing (something) well” (Ricœur 1992, p. 176). This turns once again on the conception of life derived from a background of practices and traditions and further developed and refined through interactions with others’ interpretations of it.

Both formulations of a narrative account of self-formation commit themselves to a description of life in which the agent who lives it is a single, unified, whole self, following the temporal trajectory described by the plot of their life story towards an overarching end. Life itself, then, as well as the self, are only intelligible in so far as they themselves are *teleologically mediated* by the influences that the life story exerts on them. This in turn derives its own teleological intelligibility from the normative background of contemporary interpretations of life and the subject's inherited background of tradition and practices.

Accompanying this setting out of a teleological conception of life is a more subtle point, however. As both MacIntyre and Ricœur maintain that in addition to requiring some conception of an overarching end in order to be made sense of, the selves which life stories describe, as well as those narratives themselves *must be whole*. That is, there is no room for fragmentation or discontinuity of the meaning-giving intentions, practices, and traditions which inform the goals which make its plot intelligible.

Neither narrative theorist claims that these components of a life cannot change; recall that both allow for an agent to break with the precepts of their normative background. What is central to their positions is the claim that there must be a *coherent trajectory* that these 'breaks' with a principle or ideal take. It is intelligible to seek to change one's projects upon being confronted with new information or a new set of circumstances. For instance, when one realises that one was mistaken about an important self-constituting interpretation. It is unintelligible, perhaps impossible on this view, to simply break with a background on a whim.

The overall guarantor of intelligibility and meaning for narrative theory is this normative background, constituted from practices, traditions, and projected conceptions of the form of life. For both MacIntyre and Ricœur (1992, p. 176) these "cooperative activities whose constitutive rules are established socially…originate much further back than the solitary

practitioner” and generate “the standards of excellence” by which an agent can be described as living well or poorly.

More importantly, this appeal to a constitutive normative background sets strict standards of *success* and *failure*, by which accounts of self-formation must abide. It is this point that the final section of my chapter focuses on, both as a means of bringing the other claims that underpin narrative theory together into an overall picture and also to introduce the basis from which I shall construct my critical response to it. What I present as most central to a narrative account of self-formation is the claim that due to the limits of intelligibility that agents’ socio-historical situatedness generates, human life can only be understood in terms of a purpose that it seeks to fulfill.

1.4 The limits of narrative and narration make lives intelligible only in terms of an end

The connection between the narrative account of intelligibility, its account of temporal extendedness, and its account of wholeness, when taken together, suggest that human agents are born with a purpose, which they spend their lives discovering and pursuing. The notion of a ‘good life’ plays a vital role in both MacIntyre’s and Ricœur’s theories, providing an ideal form for a person’s life to take.

Surprisingly, the definition of this term appears circular: “the good life for man is the life spent in seeking for the good life for man, and the virtues necessary for the seeking are those which enable us to understand more and what else the good life for man is” (MacIntyre 2007, p. 186). However, when understood in terms of the structural, limiting role that traditions and practices play in narrative theory, explanations like this can potentially be understood in a non-vicious way.

Considering that narrative theorists maintain that persons *inherit* a normative background which forms the basis and boundaries of what they can make intelligible, the notion of a 'search' or 'quest' for a good life can be interpreted in a developmental fashion. That is, the model of whole life that is put forward here is one in which an agent, starting from a particular socio-historical position, must square their understanding with a much broader, dynamic world than the one presented to them by their own traditions and practices.

Nevertheless, the notion of a teleological life quest also emphasises the importance this background has in setting out the 'constitutive rules' by which evaluations are made possible. Especially on Ricœur's model, the subject of experience requires some grasp of a practice in order to understand their actions as better or worse than one-another and some grasp of a tradition in order to understand the trajectory their life, as a whole, could take as better or worse.

Narrative theorists insist that it follows from this that an account of self-formation must describe of whole lives, whose unity is understood in terms of an overarching goal. In addition to the reasons I have presented for this so far there is also a stronger ontological claim entwined with this teleological account of lived experience: that only a whole life can give a coherent account of an actual self.

Consider what narrative theorists claim is required for understanding selves. The notion of accountability plays a fundamental role here for, according to MacIntyre and Ricœur, the reality of the self is guaranteed by the ability to give an account of it and its features. If it were impossible to narrate about something, then it would lack the features which permit it to be fitted into an intelligible, temporally-extended structure like a life and, hence, a narrative.

Ontologically, though, what form does this reporting take? *Is* the narrative self the accounts that are given of it, or is it necessarily something which can be reported about? Ricœur (1992, p. 22) offers a particular model of accountability, which he calls "attestation," meant to settle these questions, which describes the self in terms of "trust". According to this, a self must

be something in which someone can *believe* and be presented as an appropriate object of belief by a sufficiently coherent and intelligible testimony.

The fundamental role that narratives and narration play in making sense of human lived experience is by providing a model in which this "assurance – the credence and trust – of *existing* in the mode of selfhood" can be provided (Ricœur 1992, p. 302). In attesting to being a certain way, an agent must, for narrative theorists, be able to describe the background against which they act, which guarantees the intelligibility of their conduct.

However, as the interrogation of the concept of a tradition in terms of Ricœur's *idem* and *ipse* identity demonstrates, this relationship between the self and intelligible world is itself dialogical and dynamic. According to this model, understanding who someone is, is never as simple as identifying a particular trait or a particular way of being; it is the interactions between these two poles which generate the content about which an intelligible account of self-formation can be given.

Understanding who someone is thus becomes the *apprehending something as being a certain way*, that is as fitting into an ontological category. One cannot have attestation and its object, the self, without their *being* some such self to attest to. This self, though, does not exist in a detached, abstract sense. Rather, it dwells in the traditions, practices, actions, and histories of agents and their environments, which are both open to observation and open those who come into contact with them to an intelligible world.

What is attested to, trusted, and believed in by narrative accounts of self-formation is the actual embodiment of practices, histories, and traditions, and, thereby, the *enacting* of the appropriate narratives by the agent in question. For narrative theorists, "[t]he being of the self presupposes the totality of a world that is the horizon of its thinking, acting, feeling – in short, of its *care*" which is communicated through and contributes to narrative histories (Ricœur 1992, p. 310).

Conclusion: The narrative point of view

Narrative theories of self-formation bridge the gap between lived experience of a life and the self which lives that life by asserting that selves are not only made intelligible but constituted by the narratives of which they form a part. The emphasis narrative theorists place on action makes this evident. The reason that stories are claimed to be indispensable for making sense of lived experience over time is that in order for an agent to be able to grasp their experience of their actions and differentiate them from one another, they must be able to determine which beliefs, desires, and intentions gave rise to their behaviour. The plot of a life story explains this by describing not only the past background from which the grounds for action arise but also projecting an ideal future goal at which those actions are aimed, which, when taken together, provide adequate context for present experiences to be understood.

It is for this reason that the 'enactment' of a teleological 'quest' or 'search' for the good life is so important to both MacIntyre's and Ricœur's narrative theories. An agent must actually take up and exhibit the appropriate character traits or virtues that are fitting for the protagonist of their life story in order for its plot to both make intelligible and constitute their self. Without this connectedness to a temporally-extended narrative context, subjects actually experience a fiction; they present themselves as being such a self, when, in fact, the story which they act out describes them as being someone else.

The dangers of fictionalism are somewhat muted by the important clause, written into both narrative accounts I have described in this chapter, that all life stories are essentially co-authored with the other characters that an agent comes into contact with and influenced by the contexts an agent inhabits. For MacIntyre in particular, the historical background that traditions

project into the present places a significant limit on the intelligibility of life stories, making agents accountable to a set of principles and values which define the understanding of the world they inherit.

This places a distinct limit on how subjects can legitimately account for their own behaviour and experiences. Whatever narratives, traditions, and histories they appeal to must be accessible to others and plausibly represent themselves and the world. Because even the background against which life stories play out is co-authored, each narrator must recognise this shared system of measuring values and establishing normative evaluations for the self.

From this point of view, narratives establish the limit of intelligibility and, hence, a limit on self-formation. This makes narration and self-formation inextricable; an agent's beliefs, desires, and intentions cannot be grasped or even *exist* independently of a story which unifies and explains them in terms of its subject's history, their present context, and the future goals at which their behaviour aims.

While providing a clear account of how the self relates to and explains the contents of a subject's lived experience, this conclusion also generates a problematic tension between the former and latter two claims that support narrative theory. For it seems possible to claim that some limit on intelligibility and temporal connectedness is required by theories of self-formation, while also rejecting that these must describe a teleologically unified whole life.

Notions such as Ricœur's horizon of care support drawing this distinction. For an agent to give content to their capacity to be open to and *care* about the world, is an overarching narrative unity really necessary and will such a monolithic evaluative framework genuinely capture and express *every* care an agent feels and everything that they are open to?

Much of what narrative theorists say regarding the self and its connection with agency risks being made incoherent if they cannot provide a defence against problems like this one. The main reason philosophers have come to endorse accounts of self-formation and narrative

theory in particular is that they provide both a coherent link between agency and experience and a coherent account of how agents *actually do* care as well as how they respond to the world.

What strikes me as unsatisfactory in the narrative point of view is the way that it confuses the force of the *demand* that subjects respond to the world by giving an account of their beliefs, desires, and intentions and the limits of most persons' actual capacity to do so. If a strict, narrative standard of accountability, rather than a simple measure of openness to the world is made central to accounts of self-formation, then many subjects will risk losing the ability to account for their relationship with their own lived experience. The reason for this being that in order to narrate about one's life from the point of view of a protagonist enacting a narrative quest, one must already be sensitive and open to sources of meaning and value in the world.

Given this, I propose focusing on the specific elements of lived experience about which narrative theorists aim to give accounts in order to better fix on what the subject of experience is and how this ought to influence descriptions of agency. My aim is to see whether narrative theory actually offers a way of connecting self, life, and action by MacIntyre's and Ricœur's own lights.

Doing this requires separating the assertions that narrative theorists make along the lines I previously suggested; dividing this account of self-formation into two distinct domains, represented by the former and latter two claims described in this chapter. The question this raises is whether the explanatory basis of self-formation that narrative theory endorses actually entails an account of agency grounded in overarching unity. More specifically, it questions whether a strong notion of unity is *indispensable* for self-formation.

In spite of these problems, the kinds of questions that narrative theory poses and the concepts it makes use of have proved useful for developing thick accounts of the self. Even in resolving the tensions between openness and unity, it is difficult, if not impossible to escape

the framework that Ricœur has established regarding the constitution of selves. Specifically, it will still be essential to determine what gives content to and constitutes the 'horizon' against and from which selves are formed and whether care alone is central to self-formation.[9]

Consequentially, while I ultimately disagree with all of narrative theory's stronger claims regarding the indispensability of life stories, I also believe that any alternative to it must rehabilitate the conceptual basis it establishes. As my following discussion of the problems with narrative theory and of their solutions will show, the notions of a horizon of and capacity to care will still play a central role in this thesis.

[9] Fundamental to the kind of caring that both Ricœur and I focus on is its inextricability from *desire* and value. To care for something, whether that be a self, a principle, another being, or an object, is to *desire the good* for it and to value it positively. I separate desire and care throughout this work to indicate not their separation but to draw attention to the fact that care is a specific and important manifestation of a subject's desires.

Chapter Two – Problems with the Narrative Account

Introduction: The issue of indispensability

In my first chapter, I outlined the four central claims of narrative theory. These are that: a) the intelligibility of action is constrained by what can be narrated in terms of a life story; b) human life is temporally extended in a way that means it can only be explained by reference to a further past and future; c) each single life is constituted in terms of a narrative whole; and d) the intelligibility must take the form of a unifying telos or purpose. Having surveyed the reasons narrative theorists provide for accepting these claims, my aim here is to critically analyse whether they actually provide a workable account of self-formation.

The tension between the former and latter two of these claims raises a number of doubts regarding whether narratives adequately explain the subjective experience that every subject has of their life. To demonstrate this, I will consider each of narrative theory's four claims in turn in order to determine which elements of them are open to dispute, which require reformulation, and which should be rejected. My concern is that too strong a focus on meaning-making and significance *within* the plot of a life story risks arbitrarily excluding important but difficult-to-narrate information that can be involved in the constitution of selves.

I consider that, suitably revised, the first two claims of narrative theory are acceptable, but reject the latter ones. That is, none of the aspects of agency that these earlier claims themselves identify necessarily demand that agents *must* unify all of their experiences in terms

of the character of the protagonist of a narrative quest. The most significant issue I find with narrative theory is that it requires each story to be explicable in terms of other narratives, leaving no room for there to be facts about subjects of experience which do not take a narrative form. I believe that without some way of appealing to extra-narrative further facts, it is impossible to provide both a sound epistemic basis for self-formation and to actually fix a single coherent narrative account of the self.

Another worry I shall raise regards how some aspects of narrative theory not only risk arbitrarily ignoring important information about agents but risk leaving some without a self. I maintain that the differences between these individuals and those whose lives easily fit into a narrative structure are trivial, further demonstrating that narration requires the support of further facts and information to yield a consistent account of self-formation.

The final point I make at the end of this chapter is that although narrative theorists claim that life stories are indispensable for making sense of lived experience, they only do so when supported by extra-narrative further facts. Without these the bridge between lived experience and the subject which lives it cannot be bridged and no life story can demonstrate that it is not trivial, not indeterminate, or that it is concluded in a way that actually describes a self.

2.1 The limits of narrative intelligibility

Narrative theorists assert that the ability to give an intelligible account of an action is required in order to make sense of what an agent is doing. In addition to this, they make the further assertion that what defines an intelligible account is the ability to locate an action and its agent in a narrative context that stipulates the success and failure conditions of their behaviour. I believe that the former element of this position is essential to accounts of self-formation. By

contrast, the strict limitations that the second assertion places on the first makes the narrative formulation of it untenable.

There are two major reasons for this. One is that if virtually anything can be *narrated* or fitted into a narrative whose plot simply defines intelligibility, then the demand for narrative indispensability becomes so weak as to be trivial. The other reason is that the intelligibility guaranteed by a whole life narrative may be too much to ask of agents, even of those who do see their lives in somewhat narrative terms. Overall, both of these reasons put pressure on narrative theorists to explain how things that make sense 'locally' or 'discretely' for persons at a time time contribute to the meaningfulness of a whole life over time.

In many cases, it seems that the attitudes and relations that are constitutive of a lived episode or moment do not endure across the whole of an agent's life or even contribute significantly to the development of a whole life. Consider a child's desire for a new toy car. While that desire lasts, it seems strongly constitutive of who they are. However, looking back on their whole life after a few months, then years, then decades this desire would likely not be worth mentioning.

Galen Strawson criticises this aspect of narrative theory, believing not only that narrative intelligibility is not indispensable but that it might be harmful to impose on many subjects' lived experience. He seeks to demonstrate this by distinguishing between the long-term "Diachronic" experience that narrative theorists describe with a much shorter-term "Episodic" attitude, according to which "one does not consider oneself…as something that was there in the (further) past and will be there in the (further) future" (G. Strawson 2004, p. 430).

An Episodic account of self-formation will only extend across a local timeframe and neither rely on nor connect to an overarching life narrative in making sense of behaviour. Nevertheless, Episodic agents can still render what they do intelligible. They are still called upon by facts and features presented to them by others and their environment, however they

only recognise those which arise “in the present” (G. Strawson 2004, p. 432). Just as an adult agent will fail to recognise their eight-year-old self’s desire for a toy as intelligible to them in the present, so too will they fail to recognise more recent episodes of their life as relevantly connected to who they are unless some of those episodes’ influences still pervade their present experience.

While these episodes might themselves be understood as small narratives, they do not set a limit on intelligibility at the level of narration. The whole life of an Episodic cannot be made sense of by fitting all of its smaller episodes into an overarching story and likewise purely present, self-constituting experience does not require any long-term grasp of narrative histories. Even though each episode might form part of such a history, the Episodic life, characterised by a *different form* of experience of the self throughout each one, does not.

For instance, while the eight-year-old agent and present-day agent might both experience their lives in terms of the general historical context of contemporary Australia, this alone does not make both episodes part of the same whole life narrative. This case can be reversed, however, to put serious pressure on the narrative ground of intelligibility. Things like desires for – as well as beliefs and intentions that relate to – toys or “making coffee” might occupy certain episodes of my life, however when viewed through the lens of a whole life narrative, they cannot help but seem “trivial” (G. Strawson 2004, p. 439).

Moreover, much of what constitutes who an agent is at a given point in time will be involved in pursuing these activities. They may have longer-term goals but these can easily fail to figure in their immediate lived experience when they are occupied with the day-to-day minutiae of living. The problem here is that these activities themselves can be described in narrative terms: “you have to think ahead, do things in the right order, and so on” (G. Strawson 2004, p. 439).

If this is correct, then one can narrate *parts* of one's life but not form a significant narrative about *all* of it from those parts. Since many 'trivial' and disconnected episodes within a life can be narrated, it seems likely that "any human life, even a highly disordered one, can be the subject of an outstanding biography that possesses all the narrative-unity-related-virtues" (G. Strawson 2004, p. 440, italics removed). *Moments* in a life can still be highly significant; however, it takes more than this to make any whole which they might constitute meaningful.

This, however, only goes part of the way to addressing the stipulative element of narrative intelligibility which I find problematic. Narrative theorists claim that the broader historical-social context that narration taps into is what sets success and failure conditions for behaviour at any given point in time. Hence, according to this model even determining basic intelligibility at a point in time requires the narrative configuration of a life and its episodes in relation to a further set of narrative contexts.

Anthony Rudd (2009, p. 61) phrases this claim in the following way: "what is crucial about narrative is that it links episodes over time in such a way as to *make sense* of them". That is, the form of a plot allows interpreters to understand agents and their attributes as developing and arising from a past and moving towards a future. Typical human experience is, after all, marked by both moments and periods of formation, such as schooling, which leave marks on who subjects are.

The question is whether these 'marks' themselves constitute the self. Specifically, it is necessary to settle whether moments of development set a plausible, stable limit on intelligibility and hence success and failure conditions for self-formation. One immediate issue with the developmental model has already been raised by Galen Strawson: what about those 'trivial' moments that comprise a large portion of day-to-day existence?

Surely the quality of someone's school life and the effects it and other significant developmental episodes have had on them do not genuinely bear on activities such as coffee-

making. Specifically, they do not appear to contribute to making these ‘trivial’ episodes *intelligible*. However, narrative theorists believe that drawing such a sharp distinction between profound developmental themes and mundane, trivial ones ignores the rich texture of narrative connections between *all* of the elements of a single life.

Rudd (2009, p. 64) claims that any strict separation of these elements of lived experience “is tendentious insofar as it suggests that these are discrete narratives that can be understood separately” since “any particular narrative is always embedded in a wider narrative”. This embeddedness means that agents can only explain their behaviour at a moment in an episode by narrating further back and projecting its future consequences.

The intelligibility of this narration relies upon a background of historical narratives. Any narrative account essentially extends into the past and future according to this model, whether or not it is interpreted episodically. Moreover, when it comes to following the thread of intelligibility of a plot, “[w]here one stops doing this is a pragmatic, contextual matter…the wider the narrative, the more context that is given, the more intelligible the action becomes” (Rudd 2009, p. 64).

This might suffice as a condition for writing narrative histories, however its representation of human lived experience leaves much to be desired. Does the childhood desire for a toy car really *make sense* in terms of how it extends to the past and constructs a future image of a self? In a very limited sense, this desire projects a future, but not one which relates to the *self* in the significant sense required by a thick notion self-formation. That is, a child might imagine that they will enjoy playing with the car but this activity alone is surely not significant enough to contribute to their self-formation in terms of a whole life narrative.

Moreover, especially when it comes to childhood memories and desires, the retention of the relevant attitudes and relations which support them as marks of self-formation seems particularly limited. Although they might remember their childhood fondly, the adult agent

simply does not have an actual connection with what informed them when they were young and cannot genuinely take up their old psychological attitudes. This raises a tricky question for narrative theorists: at what stage in their life does the agent lose these attitudes and why does this loss count as coherent *development*?

Galen Strawson also draws attention to lives which lack even this ‘normal’ connectedness over time. “What if ‘what one lived’ was dominated by intense pain, misery, terror, trauma, humiliation” (G. Strawson 2016, p. 126)? At any moment a person living that life could describe *that* they are suffering and *why* they are suffering, making causal sense of those experiences. However, beyond this what prospect do they have for narrating an intelligible whole life narrative explaining their behaviour in terms of overarching developmental unity?

I believe that the chances of this occurring are poor in most such cases and that appealing to a background set of narrative histories is unlikely to help matters. The issue for these individuals is that even in cases where their present is intelligible to them as a subject of experience, their potential to act and develop as an agent is severely hampered. Say, though, that they were somehow released from this unhappy existence, why would this count as narrative development?

One answer might be that they have transitioned from not being able to determine their trajectory through life to enjoying more control over their fate. However, this only describes that their lived experience is of a better quality than it was beforehand. The development of a narrative might, after all, describe a worsening of the quality of a life: say the commuter at the bus stop heroically leaps in front of a bus and gets badly injured while trying to save another keen ornithologist who has been drawn onto the road by the sudden appearance of a common wild duck.

A stronger response might be that the liberated person can now figure their life more easily in terms of extended narrative histories. However, this move suggests that at least some subjects' lives are either only minimally narratively connected or not narratively connected at all if put under too much pressure. As Elisabeth Camp (2011, §5, para. 5) urges "we want to allow that someone could not only once have been but continue to be an amazing, robust, interesting person" if their life is suddenly subjected to these difficulties.

These problems make the stipulative element of narrative intelligibility untenable. Even if it is always possible to tell a story about what one experiences and to connect that story with a further past and use it to describe a further future, this does not mean that one *should* do so. Nonetheless, some account which makes sense of subjective lived experience is necessary to make sense of the self; for cases of extreme misery and perfectly normal stages of life such as childhood to make sense, something apart from the rigid, overarching structure of a narrative is needed.

This intersects with the second commitment of narrative theory that I believe requires modification: its view of temporality. I have attempted to show that narrative intelligibility's commitment to development over time is problematic, an issue which I will now consider specifically in terms of investigating how self-formation is interpreted across different contexts within a life. In doing so, I consider how modifying the temporal commitments of an account of self-formation allows for the retention of the useful aspects of the narrative account of intelligibility.

2.2 The limits of narrative temporality

The first two commitments of narrative theory I have outlined are intertwined in an important way as both essentially deal with meaning-making over time. Narrative intelligibility is

temporally extended just as the lives the subjects of narration are. According to narrative theorists, these lives also play out across a *historical* context; connected to a further past and projecting a future in terms of the backgrounds that persons inherit.

I believe that some version of this claim is required in order to make sense of self-formation. However, figuring it in solely narrative terms seems to detract from rather than emphasise the centrality of how temporal contexts mediate lived experience. Specifically, I hold that the narratable past need not be present or central in lived experience, even if it might be central to a narrative told about its subject.

An example that highlights the concerns I have with the claim that an account of a further past is *required* to make sense of lived experience is provided by amnesia. Consider waking up one morning without coherent memories of one's past but the same capacity to think and feel as any other person. This, I take it, is not an uncontroversial account of persons faced with this condition or similar ones. As Camp (2011, §5, para. 11) has noted "children and adults with Asperger's and high-functioning autism" also "display strong, specific personalities, interests, and commitments" yet lack a *narrative* relatedness to a further past.

Fundamentally, those who are "significantly impaired when it comes to narrative generation and comprehension" can still respond to the world in a coherent, meaningful way (Camp 2011, §5, para. 11). This suggests that an *ability to respond* rather than an *ability to narrate* is central to making lived experience intelligible both in the present and as connected to the past and future. Similar to this is the example of a life dominated by trauma, in which an agent's ability to act is significantly curtailed even though they might be able to narrate what is happening to them.

It is useful to consider how this example differs from amnesia. The amnesiac is unable to remember but still able to care. Traumatised, one is able to remember but severely constrained in what one can intelligibly attach oneself to in one's life. The problem for narrative

theory is that both cases might be fitted into a coherent, overarching narrative whole by an observer even though neither seems fit to constitute a meaning-making story and hence a self at all from the point of view of the subject of experience or, indeed narrative theorists themselves.

This potential for excluding the mentally ill and traumatised from qualifying as selves is worrying. However, narrative theorists might respond to this concern by claiming that these persons still have the *potential* to develop narrative unity. "It is through hearing stories," MacIntyre (2007, p. 183) holds, "that children learn or mislearn both what a child and what a parent is, what the cast of characters may be in the drama into which they are born and what the ways of the world are".

This claim suggests that what gives both content and form to temporally extended understanding of the self is the very structure of narratives. Experiences themselves are differentiated according to how they are narrated and what stories they fit into. Rudd (2009, p. 64) emphasises that "a person is not understandable at a single point in time; one needs to understand where that person has come from and where s/he is trying to get to". Hence, one requires a narrative to position a self relative to not only their past historical background and traditions, but also relative to their remembered past.

According to this view the amnesiac, the high-functioning autistic, or the traumatised person still *require* a narrative in some form to make sense of where they currently are. They may not have one initially available to them, like the children MacIntyre describes, however, in order to navigate the world, they will rely on some narrative schematisation of the facts about who, what, and where they are.

The claim narrative theorists defend here is the same with which I took issue above: that all intelligible accounts take the form of narration. This may well be true, but it is not clear that these reports being narrative gives them any special claim to intelligibility, especially if

every report is an act of narration in some form of another. If narratives *constitute* intelligibility, and everything for which an account can be given of can fit into a narrative form, then by transitivity everything about which an account is capable of being given is intelligible.

The issue, though, is that not every narrative is intelligible to the person who is its subject. More specifically, I hold that intelligibility is conditional on the factors which I identified as common among amnesiacs and persons suffering from mental illnesses and pathologies, as well as those who have easy access to life stories: the ability to respond to the contents of their lived experience. It is this which separates an intelligible narrative from an unintelligible one; if an agent cannot figure themselves as responding to anything or be motivated to respond as a particular narrative dictates they should, then that narrative cannot intelligibly be one about who they are.

A related problem stems from how humans appear to remember the past and represent it to themselves. Camp (2011, §5, para. 15) observes that even for those who do fit into the 'normal' parameters of narrative theory, "subjects frequently mould their memories to fit a narrative form". True details of their lives that do not appealingly conform to the ideal pattern they project upon them actually risk being discounted.

Furthermore, should it be the case that one "'recodes' one's memories in the light of the context of recall," narratives may well distort interpretations of the self rather than constitute actual selves and describe actual lives (Camp 2011, §5, para. 15). A number of psychologists have proposed models of memory that support this criticism. Michael Ross (1989, p. 341) argues that "personal recall involves an active constructive process that is guided by people's knowledge at the time of retrieval". Subjects are strongly inclined to remember *from* the present, reconstructing the past to *fit* with their current expectations.

More recent work by Stan Klein (2018, p. 120) defines memory as a "special mode of presentation" that must be both "*about*" and "*from* the past" if it is to adequately distinguish

between memories and other past-related mental events. As long as the "content presented to consciousness" has "a feeling of pastness" and 'is directly felt as "mine,"' then a phenomenon could masquerade as a genuine act of remembering, even if these features that qualify it as a memory are purely constructed (Klein 2018, p. 121).

What this shows is that even if a representation of the past is necessary for making agency and lived experience intelligible, it need not be temporally extended like or constituted by a genuine narrative. More worrying still is the role that MacIntyre and others appear to give to a strong notion of development in narrative temporality. The claim that stories are necessary for 'learning one's place in the world' entails that self-formation is constituted as much by the ability to interpret the past as the ability to respond to the present.

What are the consequences of 'mislearning' who one is? Questions such as these will occupy the following section, however the distinction they raise between responsiveness and narratability highlights the issue being dealt with here. I think that the chief problem for narrative temporality is still that it cannot account for genuinely normal moments in human life such as childhood. Is a child not a 'self' until they are adequately saturated in narrative accounts and explanations and if so to what degree do they require this initiation before they become one?

I submit that there are plenty of persons like this; those who feel either disconnected from or simply lack the awareness of further narrative pasts and futures, yet who still possess a robust capacity to respond to the present. Furthermore, narrative theory seems incapable of accounting for persons who suddenly gain or lose this sense of further temporal connectedness within their lives.

What happens to the self in these cases? Is it created or destroyed at the moment of transition? Likewise, what happens if someone's story changes radically? Does a *different* self arise to take the place of the one which no longer neatly fits their present narrative? I believe

that the best way to confront these issues is to clearly distinguish intelligibility and responsiveness from narratability. The contents of a life about which a narrative can be told strike me as necessarily prior to any act of narration which encompasses them. This is due to the simple reason that one can have an awareness of these features without having an awareness of any narrative configuration of them.

I believe that narrative theory's worrying description of childhood also highlights the elements of its first two claims which should be retained by accounts of self-formation. A child is defined here as *lacking* a fully-developed relationship with a temporal background which informs their present in a coherent way. However, what they do possess independently of this is the ability to act and react to what they experience; a capacity which narratives are supposed to mediate and make conform with established conventions.

However, this demonstrates that what motivates the telling and imposition of narratives on lived experience in the first place and that what makes it possible for someone to take up the principles that stories describe is this capacity to react and care, not the capacity to narrate. Much of what agents react to and care about occurs beyond their immediate spatiotemporal location, often arising from the past and ingrained present contextual factors such as social institutions and mores. Nevertheless, reacting to or caring about these in an intelligible way does not require a person to be able to trace the development of these factors in narrative terms.

Indeed, given the multiplicity of such factors within a single life or context and the multiple interests which colour how their development and influence should be interpreted, it seems incredibly difficult to fix on one *true* narrative account of them. These problems only become worse when applied to the phenomenon of a whole life with all of its complexities and connections. Given this, I suggest retreating from fixing intelligibility and temporality in narrative terms and instead interpreting them in regards to responsiveness and care.

Not only does this capacity appear prior to narrative formulations of events and available to all subjects but also survives the most serious problems narrative theory faces. Although it is always possible to narrate about lived experience, it is not always possible to determine which narrative interpretation of it is correct or whether there is a single true interpretation of it at any time. Establishing these concerns and rejecting the final two commitments of narrative theory firmly cements the necessity of world-relatedness for accounts of self-formation and paves the way to formulating one grounded in this capacity.

2.3 The limits of narrative teleology and wholeness

I explore these two commitments in one section due to their interwovenness. As has been noted both by narrative theorists and their critics, many actions that an agent will perform in their life may be teleological. However, this alone is not enough to guarantee that life itself is defined by a determinate telos. What is needed to support that claim is the notion of a whole life narrative, which both describes and prescribes the ends at which an agent's actions should aim and unifies them coherently.

I do not believe that either of these claims can support a tenable theory of self-formation. The reason for this is that the developmental unity of a narrative whole that binds teleological elements together cannot be determinately applied to a human life in a way that makes sense of it. There are simply too many possible narrative interpretations of a subject's life for a single one to constitute the sole undisputable self that explains its behaviour coherently.

Additionally, as narrative theorists limit self-interpretation to a purely narrative plane, in which only further stories can be accessed to corroborate facts, their position faces a

troubling form of regress. Because narratives can appeal to no extra-narrative, further facts to explain their organisation of events, so too does their organisation of events fail to capture any further facts about their subjects. This is deeply at odds with the way in which humans typically experience their lives and hostile to the notions of intelligibility and temporal extendedness at the heart of the notion of self-formation.

These problems can be demonstrated by two examples drawn from two widely different applications of narrative theory. Huiyuhl Yi's (2020) recent account of narrative identity and value captures the essence of the indeterminacy problem. According to Yi (2020, p. 285) "the characteristics that constitute a person's true self are the ones that are incorporated into her self-told life story as represented in a narrative structure". This explanation appears dangerously circular: what constitutes the 'true self' is a story told by the 'self'.

If there must already be a self telling the story of its life in order for that story to constitute who they are as a self, then narrative identity appears to fall foul of a similar kind of objection that Bulter raised against Locke. The existence and prior constitution of a self is *assumed* in Yi's account as the source of self-constitution. As Camp (2011, §5, para. 18) urges, though, when it comes to grounding self-formation it cannot simply be that "thinking makes it so". In presenting this kind of narrative chain as the basis for self-formation, though, Yi and other theorists seem to beg the question rather than answer it.

I have already referred to the second example I will use to demonstrate problems facing narrative theory: the notion of a *centre of gravity* appealed to by Dennett's account of narrative self-formation. In order for a theory to employ a centre of gravity to explain something, it must already be able to define its other component elements: "an atom or subatomic particle" for physics and the experience of beliefs, desires, and values for lives (Dennett 1992, 103).

It would make no sense to talk of centres of gravity without a sufficiently developed model of physics, just as it would make no sense to talk of narrative self-formation without

already having a sufficiently developed basis for making ascriptions and descriptions about the self. Narrative theory appears to ignore this fact, however, arguing that the stories that can be told about agents' lives are the basic building blocks from which selves are constituted.

This not only excludes the information which narratives report as unimportant or insignificant but thereby disqualifies agents who fail to live out whole life from counting as selves. On this view, the actual events of a life and actual experiences themselves are subsumed by an overarching plot and telos. Yet, like a centre of gravity, this explanatory abstracta requires the elements it unifies and describes to be able to be grasped independently of the whole they describe in order for it to make sense itself.

This is the most troubling problem that the final two commitments of narrative theory generate. Above, I cited Rudd's (2009, p. 64) claim that narrative accounts can always be extended further back and forwards in time to build in more detail due to the embeddedness of each narrative account within another. This, however, stands in stark contrast with the demands he later places on accounts of self-formation to provide genuine ethical and moral knowledge for agents.

Arguing against Episodicity, Rudd's (2009, p. 69) aim is to demonstrate that self-formation as a being capable of moral and ethical evaluation of the contents of their life "will prevent me from thinking of them as entirely separable and unrelated". That is, as Ricœur and MacIntyre also hold, narrative form is essential for ranking and ordering beliefs, desires, and intentions in such a way that an agent can intelligibly be motivated by them.

The outcome that narrative theorists seek to establish here is fundamental to accounts of self-formation, however their method entails a number of problematic conclusions. Chief among them is the fact that narratives by themselves have no way of guaranteeing a genuine organisation of lived experience. Consider Rudd's (2009, p. 69) setting out of morally coherent behaviour: "I cannot think of myself as a serious ethical agent if I am concerned to live

according to principles of justice in one aspect of my life, but content to act with cruelty and treachery in another".

For a narrative to sustain this claim it must demonstrate that it coherently distinguishes between justice and cruelty and describes its subject in such a way that clearly demonstrates that their conduct accords with the one rather than the other. What, though, are they to do when confronted by two or more competing narratives, each of which appears to provide good reasons for interpreting their behaviour in different ways?

The narrative method tells the subject to look back and project forward to the further narratives in which the competing accounts are embedded in order to fix the relative authenticity of each. An immediate worry this raises is that rather than accessing any actual facts, the interpretation of the subject which accesses a more *convincing* background will be selected, a point which represents human interactions well but strikes me as a worrying basis for a theory self-formation.

The fundamental issue that instances of competing narration points to is that narrative theory actually rests on "a kind of holism; events do not exist in their own right, prior to and independently of their narration" (Rudd 2009, p. 62). Although this provides an incredibly strong, albeit flawed basis for intelligibility, it leaves narrative theorists with almost no room to provide an epistemic basis for self-formation. For if selves are only who they are once a narrative about their conduct has been provided and if more than one competing and equally intelligible and authentic narrative exists, then who the self is, is *indeterminate*.

According to this model, narrative theory cannot even settle the matter of the constitution of a subject's self by appealing to a historical background or projected future because it is apparently possible to continue to extend narrative interpretations indefinitely in this way. Although it is likely, as Rudd himself appears to acknowledge, that once the most

immediate contextual factors have been fixed, further narration is likely to add less and less useful information to an account.

Despite this, some form of infinite regress through a presumably infinite set of narratives always seems possible. This leaves the general coherence or some other such epistemic measure of the most important narratives in a chain as the measure by which accounts are compared. However, this might be even more problematic than seeking a foundation in an original narrative as on this account the quality of interpretations of events cannot be evaluated separately from the narratives in which they figure.

As Camp (2011, §5, para. 16) points out, "[i]f narratives are to constitute epistemically respectable selves, it must be possible for them to be checked against some external standard," a standard which it seems impossible for narrative holism to ever meet. Even though narrative theorists maintain that one's narrative accounts of oneself are always open to dispute from others, the indeterminacy of any purely narrative account means that even these cannot provide an appropriate epistemic standard for judging accounts of self-formation.

This problem cleaves not only against the wholeness and coherence of life stories but also against narrative teleology. For the holistic model is meant to provide a stable way of measuring development over time in such a way that fixes each of an agent's actions in a hierarchy aimed at an end.

I contend that given the problems that arise when competing self-interpretations face an agent, narrative theory also cannot provide a coherent measure of development necessary for grounding its strong teleological commitments. Each narrative interpretation of a life may provide *a* telos for it and the trajectory of a quest or plot for its protagonist to follow, however the indeterminacy that faces agents confronted by competing accounts makes it impossible to genuinely determine which purpose actually exerts a unifying normative force over their life.

Even when dealing with only a single narrative interpretation of a life this problem still lingers. The constraints that narrative holism places on self-formation demand that only narrative elements be considered in evaluating and constructing a teleological hierarchy. Yet in doing so how is one supposed to determine every salient aspect of one's experience has been captured? If an interpretation can only appeal to narrative figurations of events, then anything that the narratives have missed must be discounted. More worryingly, an interpreter appears to have no way of telling whether they have missed something important; whether the quest upon which they see the subject of the narrative embarking on is actually a fiction.

Rejecting holism while maintaining narrative form only provides a partial solution to these problems. The core issue is that no matter what stories end up being told they must be evaluated in terms of extra-narrative, further facts. If this is the case, though, then surely it is these facts, rather than whatever stories that are told about them, which may conflict irreconcilably with one another, that are important to self-formation.

Another upshot of rejecting holism and attempting to fix on what these pre-narrative facts are is that very few lives will likely end up with a teleological endpoint informed by a hierarchy of beliefs, intentions, desires, and motivations. This is because the facts and information that would inform this structure do not themselves constitute a narrative; rather, the story with its plot and goal are imposed after the fact, constructing a teleology.

Finally, it is not plausible that every person's lived experience is teleologically-mediated at *every point* during their life. Once again, childhood offers a useful counterexample against narrative form. What long-term, overarching, unifying goals do these persons who, MacIntyre himself accepts have not yet 'learned their place' within the narratives that introduce them to the world, have? Similarly, how should philosophers account for amnesiacs and certain pathologies or life events which radically disrupt the plot of a life?

I think that these demands demonstrate the fundamental problem with narrative teleology: it's insistence on *development* throughout life. It is possible to quantify and qualify change in the nature of my experiences but enjoying 'better' experiences relative to earlier or later ones does not necessarily constitute a development or devolution of the self. Indeed, such an evaluation of experiences over time may well be impossible. For instance, the comparison of the quality and quantity of the experiences an eight-year-old and the adult they become would only make sense if there were enough common conceptualisations and structures of meaning that they shared.

I do not deny that both are capable of feeling pleasure and pain; they both share that crucial capacity to care and respond I have identified as central to self-formation. What I do deny is that they are both capable of intelligibly representing *how* we feel called upon by this capacity and respond to it in the same first-personal way, which would make a significant comparison between their experiences intelligible.

By contrast, consider the same process occurring between that agent now and themselves a year ago. In that case, the representation and sharing of meaning and experiences would be much more feasible. However, this case would become more complicated were they to have fallen in love, or suffered a horrible accident during that timeframe, which drastically shifted their outlook.

Coupled with the problem of indeterminacy, these problems that confront the construction of developmental unity within and across a life make narrative teleology unattainable for most if not all people. This is an issue which a final criticism of Camp's highlights neatly: from which point in a life are 'development' or 'unity' to be measured? The worry here is that "because that arc [of our overall life narrative] is itself a trajectory towards the story's end, no event can be assigned a determinate significance until that life is over" (Camp 2011, §5, para. 6).

Despite all of these problems, though, there is still clearly something occurring which moves agents to act and helps them make sense of why they are acting. Whether or not there is a story about their lives, agents actually *do* things; they care about others, and respond to the world. Given the serious issues regarding the coherence of the form of whole life narratives, I propose acknowledging the sensitivity to these two aspects of lived experience that MacIntyre, Ricœur, and Rudd have shown by attempting to construct and alternative interpretation of them.

Ultimately, what such an account of self-formation will need to do is avoid imposing an overly strong demand for unity and development on subjects of experience. The crucial error that narrative theorists have committed is sacrificing a simple notion of *being a person* in favour of the stronger notion of *being Someone*; that is, taking a particular image of self-formation as prior to the activities which would confirm or deny one's existence as such a kind of self. I do not hold that some agents' lives may not take such a form but, rather, that it is an implausible criterion against which to measure the coherence of all persons' lives.

Conclusion: Everything right apart from the narrative?

The main problem I have identified here is a tension between the capacity to narrate and the capacity to respond to and care about the contents of lived experience. Narration is a finer-grained and, hence, more exclusive concept, imposing strict measures on intelligibility, agency, and self-formation. As a consequence, theories which rely upon this rather than more basic capacities risk admitting only a narrow portion of persons into the ranks of selves.

This is at odds with much of what narrative theorists themselves have attempted to demonstrate. As I have identified, Ricœur (1992, p. 310) in particular is deeply concerned with the role played by "care, in its ontological dimension," considering this to set a crucial limit on

the being of the self. Without care, the convictions, ideals, and the relations a self holds arise as mere coincidences or basic facts for the persons involved in them, rather than as intelligible sources of motivation.

I believe that focusing on this aspect of narrative theory provides the best basis from which to both develop a more accommodating account of self-formation and defend a thick notion of the self against sceptical challenges. Considering both of these issues, my next chapter follows the work Elisabeth Camp has done in attempting to derive a stable basis for self-formation from the criticisms levelled against the claims of narrative theory.

In addition, a number of philosophers, particularly situationist sceptics about character, are likely to view the case I made against narrative self-formation in the third section of this chapter as a basis for rejecting all notions of the self as a ground for agency. The aim of my project as it moves beyond the narrative paradigm will be to demonstrate that this scepticism goes too far. However, as I have argued here, accounts of self-formation must also incorporate more environmental factors if they are to present plausible explanations of lived experience.

Chapter Three – The Character Account and its Sceptics

Introduction: The self as character and style

I concluded my critical analysis of narrative theory with the claim that a capacity to care and respond to the world is what is essential to accounts of the self. In this chapter, I attempt to better define this capacity and the effects it has on persons' self-formation by considering the alternative to narrative theory proposed by Elisabeth Camp. This account identifies characters rather than stories as central to describing the self and holds that the unifying process of characterisation provides a more epistemically respectable and less cognitively burdensome means of representing lived experience than narration.[10]

While Camp provides a useful schema for understanding how non-narrative self-formation might be understood, her strong appeal to character invites a number of sceptical challenges. Specifically, situationist sceptics such as Gilbert Harman and Mark Alfano have raised worries regarding the role that character traits play as a source of motivation. These philosophers consider character to be too weak and fraught a concept to determine behaviour, instead claiming that it is the environmental factors of an agent's situation which overwhelmingly determine what they do and how they live.

Developing a response to the situationist challenge is vital not only for Camp's view but for thick accounts of the self-formation in general. Even on the less-severe theory that Alfano offers, there is little room for a coherent, unifying self at the heart of human action and

[10] For more on Camp's broader approach to characterisation, fiction, and perspective, see: Camp (2009; 2017) for her work on perspectives and socially-constructed identity, slurs, and character tropes see: Camp (2013).

life. Resisting situationist conclusions, however, requires the character account to be modified so that it encompasses more environmental factors than Camp's initial sketch of it suggests.

Expanding accounts of self-formation so that they incorporate much of what situationists understand as impersonal, situational factors is possible thanks to the work done by Robert Howell.[11] Maintaining that character traits can arise and be sustained externally from the physical boundaries of the person, Howell's criticisms of situationism suggest that the distinctions these sceptics draw are often arbitrary and risk leading to a category mistake.

Confronting situationist scepticism in this way suggests an interesting conclusion regarding how non-narrative accounts of self-formation should be understood. When treated as partially external to an agent, the capacity to care and respond, which endures from Ricœur's theory and informs Camp's view, opens subjects of experience to a relational horizon of meaning. It is the capacity to interact with others, institutions, and even traditions as they live that provide the necessary texture and resources to both constitute and sustain the self.

3.1 Camp's Character Account

Elisabeth Camp (2011, §2, para. 4) frames accounts of self-formation as answers to the question: "What is my true self, such that I should pursue and cultivate it". Demands such as this constitute and constrain the choices that an agent makes throughout their life, paralleling the two questions central to MacIntyre's (2007, p. 175; 183) narrative theory: "What is he doing" and "What am I to do".

[11] What these factors are actually called will vary depending on the empirical research that describes them. What I mean here by terms such as 'environmental factors,' 'impersonal situational factors,' 'contextual influences,' and influences which act 'externally from the physical boundaries of the person' is the same: they are forces which pressure an agent to act and pursue a certain form of self-formation, which they do not wholly generate within themselves.

In order for these queries to make sense for an agent to respond to, both narrative theorists and Camp agree that they must have a self to which they can pose them and as which an agent can reflect upon them. However, these positions disagree when it comes to determining the constitution of lives and the subjects who live them. Unlike narrative theorists, Camp (2011, §6, para. 4) rejects the need for any overarching teleological unity or for a single life story for self-formation; rather, she suggests a less-ambitious measure of "the stylistic coherence of character".

Genuine coherence of this sort is guaranteed by a number of standards. Understood in terms of style, "a self is a distinctive way in which a particular 'I' inhabits, interprets, and engages with the world" (Camp 2011, §6, para. 4). The way one is and wants to or feels one ought to be plays a strong role in constituting who one is. It is by striving to be a certain way and make the contents of life cohere in accordance with that aim or passion that an agent attempts to make a particular self-conception true of them.

Note that this still leaves plenty of room for mistakes and failure, especially given that Camp is cautious not to endorse purely personal characterisations by themselves. Instead, she introduces two important criteria into her account. Primarily, "that self-interpretations are subject to the usual standards of judging theories: they should explain as wide a range of features as possible as simply as possible" (Camp 2011, §6, para. 6). That is, accounts of self-formation should be evaluated based on their *explanatory breadth* and *depth*. Additionally, in order to accomplish this, "we are free to appeal to other people's reports, along with any other epistemically respectable source of evidence" (Camp 2011, §6, para. 11). Ideally, a broad and profound interpretation of the self will be supported by a number of different sources.

The approach that is suggested by these standards appears a less problematic and more easily satisfiable measure for self-formation than that provided by narrative theory. Most importantly, it leaves room for error: an agent *should* subject their self-interpretations to a

degree of scrutiny that reflects how important they are to them; however, they may not do so. It is possible to be deluded on this account; however, this self-delusion is not a result of the epistemic constraints that Camp has set but is rather due to agents' own decisions and situation.

This still leaves open the question of exactly how the information about every person is made available for interpretation in an intelligible way. Without narration and narratives to set a limit of coherence on accounts of self-formation, what makes certain features *justifiably* arise as important and others as trivial? Camp's answer is to figure human lived experience in terms of characterisation and characters rather than plots and stories.

In order to determine the importance of a feature or trait, one must determine how it fits into the "structures of relative prominence and centrality" which arise from the information about me and "interact to produce a coherent organisational gestalt" (Camp 2011, §6, para. 5). Notably, this still leaves room for competing, mutually exclusive interpretations, however these are generated from the agent's interactions with others in such a way that they can attempt to judge the evidence that supports them.

In addition to this, one is always free to 'move on' from a self-interpretation by identifying new characteristics to make prominent and central to one's style of being. The vital distinction between this position and narrative theory here is that on Camp's account, changing one's self does not amount to changing one's life as the two are held to be separate. As "selves exist at each moment" the decision to stay the same kind of self or try to change is always available to an agent (Camp 2011, §6, para. 2).

This makes it possible for someone to fail to identify with who they were as a child if their experiences and motivations were radically different. Likewise, it makes it possible for them to fail to identify with who they were last month if some event that significantly altered the structures of prominence and centrality that they understand themselves by occurred between then and now. Vitally, agents do not need to have some overarching quest or telos to

make sense of this change precisely because what is appealed to in order to engage in intelligible self-formation is the agent's present situation and the limits which they perceive it placing on them.

According to Camp (2011, §6, para. 7) "a self…can know who she is at each moment by reflecting on the commitments and memories that matter most to her, and by investigating how well her actual habits cohere with that self-conception".[12] One question this raises is how demanding an epistemic standard the character account sets on subjects. This is especially important considering Camp's own concerns regarding the status of the mentally ill as selves.

Knowledge by reflection typically suggests that the interpreter actively takes a step back and analyses a set of information; in this case memories. However, I think that Camp's model only requires that some basic sense of a set of 'actual habits' and 'desires,' as well as beliefs and intentions, be available to an agent in order for them to self-interpret. These are the necessary ingredients for the pursuit and cultivation of anything; for moving from how one is to what one wants to be.[13]

An interpretation need not be able to provide a particularly complex articulation of why a subject desires something or how they have formed their actual habits. Consider again the contrast between an eight-year-old who wants a toy car and the adult agent they become; I doubt that the child would either be interested in or able to provide a detailed account of why they desired that plaything. Nonetheless, this largely unreflective desire provides an intelligible source of motivation for their behaviour both to the child and, assumedly, their parents.

What is interesting here is that from the point of view of its parents, a far more complex account of the child's desires might be available. They might have worried that their child had developed an alarming and violent fixation on toy cars due to an underlying need to compete

[12] This also leaves more room for self-formation than Episodicity does; there might be enduring memory chains and perspectives within a subject's life which make multiple episodes cohere in terms of a robust, long-term conception of the self.
[13] I discuss the consequences of losing this basis further in chapter five.

with their friends at school. Trying to explain this to them and work through interpretations of why their child might and whether they should want the toy car would help them develop as persons and expand the child's perspective regarding who they are.

Camp's account also holds that self-formation need only be temporally discrete but can be extended over long periods of time, even across a whole life. It determines the coherence of interpretations by considering how well they represent the actual behaviour and traits a subject has displayed and intend to display given their current self-presentation. What is essential to self-formation in the character account is the pursuit of *coherence at a given time* rather than coherence over time; the choice regarding which important commitments to endorse is one available to agents at any given time.

As Camp (2011, §6, para. 10) describes it "for most selves, multiple overlapping but distinct strands of centrality will collect subsets of habits, commitments, and memories into kernels of personality". Unlike a narratively-mediated self, which requires an overarching plot to make these strands cohere, the character account leaves the choice and responsibility of self-formation squarely in the individual's hands. Likewise, while many of the sources of information that inform reflections about the self and constitute personality might stem from the historical background an agent inherits and the peers with whom they interact, the character account incorporates this information at the level of their personality itself rather than at the level of a historical narrative.

Ultimately, for "the character conception, having a self merely requires having a stable, coherent 'self-presentation': a way of presenting ourselves to, and more generally engaging with and responding to the world" (Camp 2011, §6, para. 10). It is precisely because this coherence need not be long-term but can nonetheless still be informed by further past and future images and facts that it provides a more plausible, stable basis for self-formation than narrative theory.

On Camp's view, one does not need to build up and inherit momentum to get one's motivations and personality rolling, so to speak. Rather, an agent simply needs dispositions to act and respond in a certain way that make it intelligible for them to act. Essentially, all that is required by the character account is the capacity to care, which informed Ricœur's conclusions regarding the ontology of the self.

This outcome and the more general view regarding self-formation are challenged, however, by scepticism about the coherence of cares, commitments, and capacities guaranteed by a personal 'character' or 'characterisations' of the self. The lynchpin of the character account is the ability to choose and endorse a particular way of being; of pursing and cultivating a self that one presents to the world. This explanation of self-formation grounded in the capacity to care and respond fails, though, when at least a number, if not all, of the choices and commitments an agent appears to endorse are actually the products of their environment rather than a self that connects their lived experience and beliefs, desires, and intentions.

Situationism provides the clearest example of such scepticism. My next section presents its central claim that most agents' individual choices are in fact the product of external, environmental forces rather than personal motivations. The point of this is to demonstrate that even the most open forms of situationist scepticism and similar positions do not leave adequate room for self-formation as Camp, or indeed narrative theorists, have conceived of it. It will be the work of my third section to show that most of the claims I have endorsed here can be defended against this sceptical challenge.

3.2 The Situationist Challenge

Situationism shares many features with the character account. It agrees with Camp's conclusions insofar as it holds that there are structures of prominence and centrality which determine the fittingness and meaningfulness of traits. Its view differs, though, concerning the access subjects have to this information. According to situationists, social psychology shows that what appears to be important, and indeed what a person would typically attest to as being important to them, is often mere fanciful self-delusion.

Experiments, such as the infamous Milgram and Stanford prison examples and the less-controversial Good Samaritan Study, show that "situational influences swamp dispositional ones, rendering them predictively and explanatorily impotent" (Alfano 2013, p. 3). On the basis of this evidence, situationism denies the coherence and effectiveness of 'character,' and claims that self-constituting personality traits are produced by environmental factors and therefore are not a "metaphysically robust property" (Alfano 2013, p. 3).

Although these central commitments guide situationist theories in general, both stronger and weaker accounts of it have been formulated. In this section, I examine one of each kind of situationist challenge to self-formation. First, I consider what I will call the 'hardline' position taken by Gilbert Harman.[14] This holds that "character" is overwhelmingly "a fleeting feature of an act," rather than "an enduring characteristic of a person" (Harman 2009, p. 241). As what an agent does is strongly shaped by situational influences outside of their control, hardline situationism concludes that ascribing genuine character traits to persons is almost always mistaken.

What I will call the 'conciliatory' position is held by Mark Alfano (2013, p. 83; 82), who offers an account of "*factitious virtue*" generated by "plausible public attributions of virtue traits". According to this view, recognising the influence situations have over agents,

[14] For more extreme statements of this hardline position, from which I take him to have somewhat retreated, see Harman (1999; 2000).

philosophers should turn their attention to how best to wrangle these factors into producing benign, productive conduct. Essentially, Alfano's situationism seeks to rehabilitate theories which rely upon thick concepts of character by engineering situations to produce the beneficial results they claim self-formation should achieve.

What differentiates these two forms of situationism is what they take the findings of social psychology to mandate. What, both ask, is philosophy to do given that a plethora of experiments seem to show that 'character' is a concept either founded in pretence or so highly malleable as to lack genuine integrity? How, Alfano (2013, p. 36) asks, can one attest to having a stable character and style of being when, as the Good Samaritan study shows, a "degree of hurry" rather than professed "religiosity" or the priming material of a "speech subject" affected seminarians' behaviour towards a "distressed confederate slumped on the ground"?[15]

What is of particular interest in the Darley and Batson study is the fact that its participants appear to have been primed by two separate sets of stimuli. Firstly, Alfano (2013, p. 35) describes how the seminarians who participated in the study "were…asked to prepare a talk either on job prospects for seminarians or on the New Testament parable of the Good Samaritan". This stimulus generated the reasonable assumption that if one were on one's way to give a seminar on that parable, one would be more highly motivated to act in accordance with its message.

What the study actually recorded, though, was that the second set of stimuli, provided by the researchers, was a more effective motivator. Of both groups of seminarians "[s]ome were told that they had time to spare, others that they were just on time, and still others that they were running late" (Alfano 2013, p. 36). The fact that "a huge majority (90 percent) of the rushed participants failed to show compassion" speaks heavily in favour of a situationist interpretation of events (Alfano 2013, p. 36).

[15] See: Darley and Batson (1973).

When faced with the competing calls upon oneself to attend a seminar on time, or to aid a distressed stranger, the motivations presented by the first won out. A large number of subjects quite literally failed to practice what they preached, vindicating Harman's (2009, p. 236) concern that "[p]eople merely pretend, *sometimes even to themselves*, to be one sort of person rather than another" (emphasis added).

Inspired by Sartre, Harman (2009, p. 236) denies "that people have fixed characters in the sense that they actually are in themselves any of the sorts of people they present themselves as being". Rather, an agent is always 'pretending to be' who they are, since they are capable of becoming someone else at any given moment and, hence, lack the genuine integrity and stability that would *make* them that character as opposed to another kind of person.

Typical theories of virtue ethics and accounts of self-formation stringently deny that it is possible to simply 'choose to be someone else'; what about one's standing desires, commitments, and values? If the social psychology that provides the empirical basis for situationism is correct, however, the real mistake is thinking that an agent has any great degree of control over how they act and what they convince themselves of in the first place.

According to situationism, one cannot even be sure of one's own self-presentation, let alone others' testimonies about one's self. All such reports are liable to be motivated by factors external to the agents who provide them. If seemingly trivial factors, such as hurriedness, and not high-minded or noble intentions are really what cause behaviour, then unless they were superhumanly astute an agent's self-reporting – or indeed their reports regarding others – should be accepted only with further empirical corroboration.

The conclusions of the Good Samaritan study suggest that the seminarians *lacked* agential control over their behaviour. Despite their professed religiosity and their study of scripture other features of their situation ultimately dictated how they would behave. Given

this, a thick notion of the self appears to fail to explain lived experience adequately and to actually obscure the real reasons why persons behave the way they do.

Two elements of the hardline position suggest that this problem with self-formation obtains only at the level of appearance. The first is that while Harman offers a robust criticism of character within virtue *ethics*, he does not necessarily offer such a criticism of character traits themselves. Secondly, and connectedly, his position presents agents' environments as opaque in a way that might be problematic for *moral* action, but does not provide such a barrier to *rational* action, which is what Camp and others attempt to describe.

I believe that accounts of self-formation can largely avoid the problems that Harman raises because what they seek to establish is the basis for intelligible action, not a theory of ethics. As Aristotle (2000, 2.6) himself warns, "[w]e must…not only describe virtue as a state of character, but also say what sort of state it is"; namely "a state of character concerned with choice…determined by a rational principle, and by that principle by which the man of practical wisdom would determine it".

At its very core, Camp's position focuses on this very kind of choice, considering how the various decisions and actions that an agent authors shape who they are and give them a means of making sense of the world and who they are. What Harman appears to deny here is that committing an action or decision entails a special, first-personal *commitment to* the outcome or motivation which informs it.

Only happy circumstance, brought about when the environmental features of one's milieu happen to align with one's projected character traits, actually produces this kind of a relationship between agent and action. Even then, this relationship would not explain 'who one is' in any broad or counterfactually stable sense; the confluence of personal attitudes and environmental factors could be so brief that the support it appears to give to a self-presentation would vanish the moment something else captures one's attention. Nor is it clear, in Harman's

account, whether such an agreement between self and environment need occur at all during a life.

An issue that arises here, which I shall expand upon in this chapter's third section, is exactly how being a person at all is intelligible. The conclusion I understand Harman (2009, p. 235; 236) as defending is that "attributions of character traits tend to be inaccurate," a much weaker claim than what he derives from Sartre: "nothing that is now settled can make it the case that one *is* such a person". The former is a concern regarding the stability of virtue (or vice) ascriptions, the latter amounts to full-blown scepticism about the existence of any 'personal' character, personality, or thick notion of the self.

Distinguishing between these two positions is important; I doubt that Camp or Aristotle would deny that there is either a tendency or possibility for trait-attribution to be mistaken. Both of these philosophers' positions aim to offer a means of distinguishing mistaken attributions from well-supported ones and, most importantly, seek to establish a basis from which both kinds of attributions can be made intelligible. On the other hand, scepticism about the ontology of selves, about what it takes to *be* someone, amounts to a denial of any ground from which a personal commitment or relationship to the contents of one's life could be established.

Carrying situationism to this extreme conclusion does not appear to sit well with elements of the social psychology that supports its position. Given that environmental factors can exert an influence on persons' behaviour and potentially even 'train' them to behave in certain ways, it seems that it is entirely possible for someone to take these factors up in the first-personal way which can clearly be settled in empirically observable facts. Indeed, Harman's own appeal to the work of R. Hardin (1995, cited in Harman 2009, p. 237) further suggests that he endorses the weaker 'mistakenness' view rather than the stronger, globally sceptical one.

Put to work in this way, situationism describes human interaction in terms of concrete, environmental factors. For example, rather than being motivated to join conflicts purely by personal animosity towards others, persons are naturally pushed to form "coalitions" to more effectively compete with others for "limited resources," in turn posing a "threat" to other groups and resulting in a "struggle" (Harman 2009, p. 237). Animosity, xenophobia, and other traits result from these struggles and agents are still defined in terms of them, rather than existing in a free-for-all provoked by bad faith.

Given this, it seems best to adopt a more modest reading of Harman's position, especially when it comes to regarding the 'fleetingness' of virtues and character traits. Rather than holding that all such self-determining information is *necessarily* momentary, this interpretation of hardline situationism suggests that it simply could be so as the environmental factors which ground personality can change from one moment to the next. After all, for the interpretation of Hardin to work, there must be some pretty stable, long-term dispositions entrenched in and between coalitions.

Retreating from a radically sceptical position regarding character traits also leaves more room for elements of situationism to coexist not only with virtue ethics but more specifically with the concept of character and traits that is central to Camp's account of self-formation. In so far as situations *impose* dispositions or attributes on agents, then an explanation of what these features of lived experience are and how they are intelligible is also still required. It seems, then, that the hardline position must be somewhat qualified.

Alfano attempts to do just this by developing a position that allows for agents to at least partially regain ownership of their decision making from the contextual influences acting upon them. Given that subjects are influenced by situational reasons, the aim of conciliatory situationism is to create environments that promote the right kind of reasons and thereby

inculcate the right kinds of patterns of response in agents, training them to eventually adopt moral behaviour independently.

This solution to situationist scepticism concerning character suggests that it is potentially compatible with the development of “full virtue,” albeit under the direction of external stimuli (Alfano 2013, p. 105). How, though, does an agent achieve ‘factitious’, let alone full virtue, and what kind of a character would they end up with on this account? Noting the power of both placebo effects and self-fulfilling prophecies, Alfano (2013, p. 85; 86) argues that virtue-ascriptions can be constructed after these models to become more effective. Both of these phenomena provide *reasons* for a subject to expect certain outcomes, which, while paradoxical, often lead to the effects they suggest will come about.

For instance, by calling someone virtuous, then even if they am not *necessarily* so they may perceive a reason to behave as if that were the case; just as giving someone sugar tablets and telling them that they are a strong course of medicine might cure chronic pain. By generating enough pressure on an individual from within themselves and by making them perceive that equally strong reasons to accept and act upon the virtue-ascription exist in their environment and are held by *others*, this process controls situational factors to produce stable dispositions.

This raises two significant questions: why and how might factitious virtue-ascriptions fail, and how and why might they succeed? If the conditions that govern their effectiveness turn out to not be the result of situations, then the challenge that intrusive environmental factors pose to self-formation would be severely hampered, if not dispelled.

Alfano (2013, p. 91) argues that virtue-ascriptions must be “*plausible*” and “*public*”. They must generate adequate agreement between their target and the predicate being ascribed to them, provide adequate evidence for this predication, and predicate with sufficient authority, moreover, the target must know that they are the object of the predication (Alfano 2013, p. 92).

This is very similar to elements of Ricœur's thought; subjects are presented as bearers of traits, along with an injunction, tacit or explicit, to *attest* to their possession of them. Lacking a trait, they might change to justify this ascription. Perceiving an ascription as bad for them or wrong, they might reject it.

Furthermore, in arguing that "[p]eople must know the reference of labels and predicates to be motivated by them," Alfano (2013, p. 93) seems committed to claiming that in order to successfully induce factitious virtue, one must be careful to prime persons in such a way that they are appropriately referenced by it. However, notions of reference and responsiveness of this sort do play a pivotal role in both Camp's character account and also Ricœur's description of the self.

The interplay between *idem* and *ipse* suggests that a subject of experience has such a refernce-based *relationship with* the traits and predicates which constitute the self. More crucially, the central role that care plays in constructing a Ricœurian horizon of experience also suggests that a notion of *conscience* plays the same role in thick accounts of self-formation that virtue ascription does for Alfano. This notion of conscience, "[t]o find oneself called upon in the second person at the very core of the optative of living well…[and] then of the search for the choice appropriate to the situation," is what makes virtue ascriptions intelligible for an agent in the first place (Ricœur 1992, p. 352).

The suggestion that all such appears to an agent bypass or fail to constitute a self and, hence, fail to guarantee agents a personal, action-guiding conscience by which they make sense of their lived experience, is further called into question by Camp. The likelihood that "other people represent at least some aspects of our character more accurately than we do" makes self-formation on the character account *public* and thereby *plausible* in the same way that Alfano mandates effective ascriptions must be (Camp 2011, §6, para. 11).

Developing a fully-fledged cognitive awareness of a virtue is supposed to bypass the worst influences of situationism by giving agents a robust enough internal and external self-conception to resist situational factors' ability to make "people…fail to attend to morally relevant cues" (Alfano 2013, p. 94). As factitious virtue-labelling works by generating plausible concepts for a person to take up, and makes these effective by attributing them publicly, it influences subjects' development by exposing them to an environment in which the demands of the labelled virtue and the demands upon themselves are presented as congruent.

Saying: 'You are honest!' comes, however tacitly, with the corollary demand: 'So *be* honest or *change*!' if it is genuinely taken up by the addressee. Factitious virtue labelling, then, is essentially an attempt to construct an identity, a character that is robust enough to withstand the influences that typically undermine agents' integrity. For example, as a degree of hurry was identified as a strongly intrusive factor in the seminarians' decision making, cultivating environments in which audiences are more forgiving of late speakers would help support the expected moral standards for which the Good Samaritan Study tested.

Making this move comes with its own difficulties, however, which demonstrate that the kind of subjectivity assumed by self-formation, questioned by Harman and which Alfano attempts to rehabilitate, is not so easily done away with. What factitious virtue-formation actually draws attention to is the fact that in order for self-formation to be effective it cannot purely concern a single agent by themselves. Rather, it must also extend to and from their milieu, encompassing their peers and the institutions they interact with.

Should this be true, then extending the terms of reference used to describe selves from the mere subject themselves to a *relational account of subjectivity* would dispel the worst of the problems situationism presents for self-formation. If the agent who is called upon by Alfano's factitious virtue labelling is understood as part of an environment, and that

environment is a source of reasons for them to act, then that agent themselves appears to play an effective role in cultivating a robust identity through their choices after all.

3.3 Extending self-reference: Howell's extension of virtues

Responding chiefly to Harman, Robert Howell (2016, p. 147) has argued that situationists and a number of other philosophers assume a problematic form of internalism: that "the person does not extend beyond his or her skin".[16] The distinction Harman draws between character as the property of actions and as a property of persons is a clear example of an attempt to separate the properties of character or motivation from the self by drawing this distinction between an agent and their environment.

Not only does Alfano's conciliatory situationism, if true, show that this distinction does not obtain but, according to Howell, the very psychology that situationists rely on also demonstrates that virtues and character traits in general are extended. Should "some of the prominent results of situationist psychology support a picture in which our virtues are undergirded by systems beyond our skins," then it seems that denying the effectiveness of purely individual control over these systems involves some form of category mistake (Howell 2016, p. 148).

This response to situationist scepticism about virtue ethics and self-formation also raises another problem, which, while tangential to the issue at hand, helps frame the debate around the constitution of the self nicely. Namely, both Alfano and Howell appear to rely upon some – quite likely pretheoretical – sense of 'normal situations' in order to discount or account

[16] Howell calls this 'skindividualism', however, as it is effectively a form of internalism and because Howell endorses an *extended* virtue thesis instead of it, I have chosen the latter term.

for variations in agents' behaviour. For example, the Milgram Experiments do not appear to count as a typical situation: it is not normal (I hope!) for one to go about testing others' knowledge with hundreds of volts of electricity. The encounter in the Good Samaritan Study, on the other hand, seems one much more likely to befall a typical person.

Perhaps this suggests, as I imagine Ricœur would maintain, that what is required is a strong contextual basis for judgements concerning character, which can factor in a variety of different situations and permit a subject to project themselves beyond their current position. Situations, then, like narratives – which some might suggest are in fact particularly potent situational factors – seem to be powerful sources of justification and checks against self-formation but which should not, without considerably more evidence, provide the final word about which self-presentations are justified and which are not.

Noting how the possession of a phone or other external data-storing device is taken to extend memory, Howell (2016, p. 149) considers that "it would seem that in principle states like virtues can be extended as well". Instead of a phone, the possession of a handbook on etiquette would also modify behaviour in the appropriate way. One could consult, memorise it, or even be prompted to behave courteously by its weight in one's blazer pocket.

A propensity to behave well would not only be fostered and made more available to an agent by this prop's presence but as courteousness would arise for them as a live option, it would also be a characteristic which they could pursue and cultivate. According to Howell (2016, p. 154), the experiments that situationists rely on do little more than remove certain *constitutive* elements of a person. This method is similar to taking someone off antidepressants then concluding that the drugs were not important determinants of the patient's character in the first place.

A tweaked version of the medication example further vindicates this view. Imagine that instead of relying on pills a person turned to therapy. In this case they do not rely solely on

internal resources to overcome their issues as they appeal to something, in this case someone else, 'outside of' themselves in order to regulate their behaviour. If this is correct, then influences such as those imposed on an agent in the Milgram experiments *could* count as a plausible source of self-formation. In fact, in Howell's account, these influences likely would do so if they were taken up by a subject and incorporated into their life. However, by and large, this is not the case.

Situationists problematically "assume that whatever dispositions one had outside the experiment one also has within the experiment," "and only on this assumption can behaviour within the experiment indicate that the disposition was not robust" (Howell 2016, p. 154). On the extended virtue thesis, however, the dispositions displayed in these experiments became part of the participants contingently, but *could not be sustained outside of that milieu*.

Cases like the Milgram experiments, in which "people are moved into relative isolation where the rules of the game are unfamiliar" in fact target weaknesses that are fostered by typical process of socialisation, rather than demonstrating a lack of integrity (Howell 2016, p. 155). Indeed, if the traits which contribute to self-formation are extended, then social psychology generally "does not test whether a single enduring disposition is elicited in the experimental conditions; rather, it smothers the disposition by replacing it with a new one that is corrupted by experimenter suggestions and expectations" (Howell 2016, p. 156).

This fact alone provides self-formation with a robust defence against situationism. A position that relies on internalism assumes that some form of self-formation is possible in order to deny it. This challenge identifies that subjects appear to conceive of themselves as related to the world in significant ways but argues that because elements of these relations are either driven by or located within the environment, these situational factors are what determines their bearer's character.

This is true to an extent. However, the claim that the environments one encounters contribute to one's self-formation is quite distinct from the more radical conclusions situationists draw. Indeed, the fact that the base of evidence situationists typically draw upon describes scenarios that are expressly set up to dampen agential control leaves many of their arguments against self-formation question-begging.

Ultimately, as Howell's critique of this position has shown, most persons' lives are defined by less radical cases, in which their environment supports rather than dominates their self-formation. Nevertheless, in invoking the problem of internalism to defend self-formation it is also important to note that these positions do not need to fully endorse the extended virtue thesis Howell advances.

All that a relational account of the self needs for it to escape the worst of situationism is to claim that the subject of experience is not wholly contained in or adequately described by what goes on within their skin. A description of self-formation in terms of Camp's character account, as well as Ricœur's horizon of care, vindicates this point. What stands out as prominent and central for an agent, what demands their attention by arising as something worthy of care can be wholly external to one's body and, hence, demand incorporation into one's life in terms of the relations of care one builds between oneself and this other thing.

Conclusion: the capacity to care and respond as relatedness

What both Camp's account of the self as character and Howell's defence of character traits have shown is that self-formation necessarily extends beyond the ambit of the agent's own body and first-personal point of view. In recognising this, both of these positions preserve the

horizontal aspect of Ricœur's ontology of persons, according to which care about and the cares of another constitute the self.

I believe that expanding upon this concept and the relationships it establishes between oneself, others, and the world is essential to providing a clearer account of the notion of 'pursuit and cultivation' which Camp identifies as central to self-formation. As the confrontation between the character account and situationism has shown, any characterisation of a self must extend beyond the physical individual who identifies themselves as the person it describes. Given this, it is vital to not just acknowledge the possibility and intelligibility of extended virtues, but provide a clear and precise definition of what form these traits actually take and how they affect their bearers.

What is vital to this claim and to understanding how accounts of self-formation can incorporate both pursuit and cultivation and extended traits is the claim that "otherness is not added to selfhood from outside" but "belongs instead to the same tenor of meaning and to the ontological constitution of selfhood" (Ricœur 1992, p. 317). That is, accounts of the self must allow for the agent to be able to account for themselves appearing to and addressing themselves as another person. They must explain both the relations which obtain between the self and the world, as constituted by others and milieus that arise against a horizon of care, and the self and itself, constituted by an individual's awareness of their responses to the world.

Camp's work on character goes part of the way to providing a basis for making sense of these aspects of human lived experience; however, this foundation can be made stronger by moving from a focus on style to a focus on the relations which constitute it. One of the most important aspects of the character account, which highlights this, is its application of a scientific standard of theory judging to self-formation.

Central to Camp's positive position and her critique of narrative theory is the demand that accounts of self-formation be epistemically respectable. As Howell's work has shown, part

of what any self-presentation as character will derive from exists in an interpersonal space which it is impossible for one to access as an isolated ego. Acknowledging this, my fourth chapter focuses on what gives content to these intersubjective relations by demonstrating that these are particularly important external features of subjective lived experience, which play a unique role in the process of self-formation. The sift in focus, then, will be from the 'domain' of the external, discussed here, to the particular kinds of interactions and encounters which require agents to extend themselves beyond their 'skins' as social, relational beings.

Chapter Four – A Relational Account of Self-formation

Introduction: From characterisations to relations

According to Camp (2011, §6, para. 4) "a self is a distinctive way in which a particular 'I' inhabits, interprets, and engages with the world – a particular nexus of dispositions, memories, interests, and commitments that flesh out the bare skeletal ego". As my last chapter has attempted to show, this nexus and the self at the heart of it cannot exist in isolation from the broader social milieus in and with which persons interact. That is, self-formation can tolerate neither purely impersonal nor purely egoistical detachment from the world and others.

Part of what motivates situationism is that both purely personal and impersonal points of view *appear* to provide the same opportunities for constructing a nexus of traits as an intersubjective perspective. After all, a purely detached self-presentation, "sanctioned by metaphysical investigation," that "would employ only general rational and normative principles to determine how to act at any moment" still constructs a nexus out of a set of dispositions, commitments, and interests (Camp 2011, §3, para. 2).

Likewise, as Camp's critique of narrative epistemology has shown, it is possible for agents to simply make up or construct an account of themselves from a purely personal point of view. However, both of these extreme poles merely describe the utter *limits* on how a self can be constructed, rather than providing a basis for accounts of self-formation in general. While it is true that a nexus of traits can be generated from purely personal or impersonal perspectives, it would not be one which provides the coherent, stable, basis for a character.

My focus in this chapter is to demonstrate first that the self *must* be constructed in the middle ground between these two poles by showing that meaningful experiences and values

can only arise intersubjectively. Here I draw heavily on Susan Wolf's work on how meaning and values arise in life to supplement Camp's (2011, §3, para. 4) claim that while lives lived to a purely subjective or objective extreme "might be morally exemplary" or ideal in some other way they "would not produce a recognizably human self".

My reason for focusing on Susan Wolf here is not to draw a conclusion regarding the scope of the moral standing of selfhood or agency. Rather, I find that the distinctions that she draws between meaningfulness and meaninglessness provide the basis for the kind of relational account that Camp suggests ought to replace the narrative approach. Specifically, my approach concurs with hers in holding that what makes human life genuinely valuable and open to normative evaluation is the fact that meaning is generated intersubjectively at the point of confluence between subjective attraction and objective attractiveness.

Camp (2011, §6, para. 1) adds that accounts of self-formation must provide a means "[t]o evaluate our past actions and guide future ones in a meaningful way". While Susan Wolf's theory gives content to the notion of 'meaning' in this process, the relational accounts this chapter explores describe different aspects of the 'way' in which this evaluation occurs. Camp's (2011, §6, para. 9) account defines how "to make oneself into an ever more coherent, rich, and aesthetically pleasing whole" by considering what array of features matters in a prominent and centrally stable way and forming a self precisely by acting upon them and taking them up.

According to this model, self-formation need be no more complicated than trying to make sense of "the contradictory impulses we encounter competing within ourselves" in an active way (Camp 2011, §7, para. 2). The important caveat here is that features such as richness and complexity, while generally available to agents, need not be actively pursued or, indeed, possible for everyone to pursue. Plenty of persons might simply not be in a position to identify such a fixed measure of significance for themselves. Indeed, as the relational theories I discuss

below will show, such measures are also, crucially, dependent on the conduct of other agents and the milieu in which one lives.

I discuss two different and equally important accounts of the intersubjective relations which support self-formation. Firstly, I will focus on relational theories developed from care ethics, which concentrate on the shared nature of needs and values. This approach proposes a model on which the elements of a life to which the subject of experience is emotionally open, and indeed emotional openness itself, are essentially intersubjective. Secondly, I consider the focus of Confucian relational theories, which describe the need for agents to regulate their conduct and account for others and the shared social reality and climate in which persons exist.

Both of these relational approaches describe important limits of self-formation. For an agent to be motivated to regulate their conduct, they must care about the contents of their life and the relations they enjoy with others. Equally, in order to coherently and properly display and respond to their cares, an agent must align them with broader intersubjective principles and the milieus they govern.

By considering how these limitations interact, I aim to describe what counts as an actual characterisation of a self. To do so, I distinguish this from characterisations which share many of the features of Camp's model but fail to provide an adequate account of how the self is constituted out of its relations. Ultimately, what is required by an account of self-formation is not just the ability to construct a stable and coherent nexus which presents a certain set of traits. Characterisations must also provide a stable and coherent ground for self-presentation (Camp 2011, §6, para. 10).

The importance of this limit will not only become apparent here, where I confront issues of meaninglessness and value in life, but is also vital for the practical applications of my account that I explore in my final chapter. Understanding how self-formation is limited also provides a model for how individuals can be robbed of agency and even have their selves

destroyed. Most importantly, as the development of my position across these last two chapters will seek to show, describing the intersubjective basis for and the boundaries on self-formation defines what exactly agents are owed and responsible for *qua* selves.

4.1 Susan Wolf's account of meaningful lives and worth

At the core of Susan Wolf's (2014a, p. 12) position lies the claim that "there are important values other than those of morality and self-interest". That is, for a person to strive to be either only a "morally perfect person – a moral saint" or to seek to fulfil what is subjectively "meaningful to *her*" both fail to appreciate and, worse, risk excluding aspects of human life which make living worthwhile (Susan Wolf 2014a, p. 12; 2014c, p. 118, italics in original).

The first of the two extremes that Susan Wolf (2014a, p. 14) describes a life in which an agent "must have and cultivate those qualities that are apt to allow him to treat others as justly and kindly as possible". That is, a life which is governed by the ideal of being as good as possible, *regardless of one's own dispositions, interests, and commitments*. What one finds personally valuable, or interesting, or worthy of devoting oneself to must be put aside when one is faced with the option of doing something either kinder or more just.

The problem with moral sainthood and other forms of fanaticism about values is that having "morality" as one's sole personal, self-defining goal, "is particularly disturbing, for it seems to require either the lack or the denial of the existence of an identifiable, personal self" (Susan Wolf 2014a, p. 17). Someone of this persuasion must rescind any commitments beyond those to a single ideal, which they make the object of their existence.

The loss of a personal self that Susan Wolf describes here is fundamentally a loss of the elements which constitute personal self-presentation for Camp. As long as one can pursue and

cultivate 'saintly' traits, then one must accede to the demands of whatever ideal they contribute to and accept them over other personal ones. The issue with this, though, is that while it is true that taking these traits on is essential to being as moral as possible, "some of the qualities the moral saint necessarily lacks are virtues, albeit nonmoral virtues"; they are qualities that persons have reasons to endorse (Susan Wolf 2014a, p. 18).

The point here is that "striving toward achieving any of a great variety of forms of personal excellence are character traits it is desirable and valuable for people to have" (Susan Wolf 2014a, p. 19). While these nonmoral traits or virtues might sometimes serve a moral cause or end, they need not always do so, just as the selves of those agents who pursue and cultivate them are intelligible independently of a conception of moral sainthood. For instance, it seems desirable to have members of society who are concerned to become excellent cooks for nonmoral reasons, even though good cooking can contribute to moral projects. It is perfectly *humanly* coherent to strive to be a good chef for nonmoral reasons and it seems quite worrying that space should not be left to persons to behave in this way.

Once these reasons and values enter play, though, the development and conduct of human agents are subjected to more than one kind of scrutiny; it is possible to ask both whether what they are doing is morally valuable and also whether it is non-morally valuable. Furthermore, a theory which does not account for this latter kind of value will not give a particularly good account of how agents genuinely make decisions. For Susan Wolf (2014a, p. 29) "our values cannot be fully comprehended on the model of a hierarchical system with morality at the top" for precisely this reason.

A subject may be called on by an aesthetic or other non-moral value in just the same way as or more strongly than a moral value or principle. A commitment to achieving mastery of hobby, an interest in something as banal as "Nicole Brown Simpson's sex life," or a disposition to react with awe to and hence value a sculpture might constitute who a subject of

experience is just as strongly as a commitment to supporting a charity (Susan Wolf 2014c, p. 114).

This point, that there is something valuable about the nonmoral reasons subjects are confronted by and interact with, which makes endorsing a morally unsaintly existence a worthwhile option, leads to Susan Wolf's account of what a middle, intersubjective ground for self-formation is. Specifically, this aspect of her thought raises significant worries regarding how meaningful lives lived from an absolute, non-intersubjective extreme are *to* the persons living them.

Worthwhile, meaningful lives require, for Susan Wolf (2014c, p. 108), "a substantial conception of self-interest". This is something that is lacked by persons who act and conceive of themselves from either a purely subjective or purely objective point of view due to the fact that, like the moral saint, they abrogate any personal, self-related relationships and values from their perspective. At the core of this position is the claim that "the very idea that activities can make a life meaningful without the subject's endorsement is a dubious one," clearly ruling out moral sainthood as a path to a personally meaningful life for most agents (Susan Wolf, 2014c, p. 109).

The point here is more nuanced, though, than a simple rejection of ascriptions of meaning from an impersonal, objective point of view; it also questions the role that subjective endorsement plays in generating meaning. A meaningful life requires more than a purely personal desire for or belief in an activity or ideal, rather it requires, where possible, the recognition of a broader notion of significance which accounts for the possibility of one's subjective meaning-ascriptions being false.

Susan Wolf demonstrates this by contrasting the intersubjective position on meaning that she endorses against radically subjective or objective alternatives. The first: "someone whose sole passion in life is collecting rubber bands" may have good personal reasons for doing

so but appears unable to communicate or represent the meaningfulness of their conduct to anyone else who does not share their perspective (Susan Wolf 2014c, p. 113). The second: a "corporate lawyer who sacrifices her private life and health for success along the professional ladder" gives up on personal attachments and commitments in favour of an ideal, like the moral saint (Susan Wolf 2014c, p. 113).[17]

Admittedly, both of these cases represent extremes on the spectrum of attitudes someone might adopt regarding meaning in their life. However, they highlight the degree to which mere subjective or objective perspectives do not satisfactorily ground significance by themselves. What is at issue here is the fact that if one genuinely recognises there are real "distinctions in nonsubjective value," then only endorsing a set of subjective values seems not only to make one miss out on some things that are worthwhile but flies dangerously in the face of how the world is and what the things that one values in it actually are (Susan Wolf 2014b, p. 101).

If by endorsing purely subjective preferences over all others – to such an extent that an agent ignores the meaningfulness of anything else – an agent aims at living a life that is meaningful, then they appear to be committing an error. On the other hand, failing to appreciate that some of what they experience has value to it, regardless of whether or not that agent might subjectively benefit from or enjoy it, seems to force them into adopting an overly narrow perspective on their life, making their experience more or less equivalent to the rubber band collector's.

The collector has purely personal reasons for behaving as they do, which are so strong that they constitute their 'sole passion'. When faced with such conduct, it seems obvious to

[17] It is important to note that Susan Wolf understands this kind of commitment to impersonal values in a very strong sense. The lawyer's devotion to her career is taken to be equivalent to David Wiggins' account of "the pig farmer who buys more land to grow more corn to feed more pigs to buy more land to grow more corn to feed more pigs" (Susan Wolf, 2014c, p. 113). Without a degree of subjective desire for this outcome independent of the objective value it produces, it becomes a viciously circular account of meaning.

conclude that the person in question "is wasting or misdirecting her life" precisely because they fail to grasp the other extra-subjective sources of value in the world (Susan Wolf 2014b, p. 99). They may be happy, satisfied, or somehow fulfilled by an activity like collecting but their life would be *better* if they were to incorporate non-subjective values into their self-presentation.

Indeed, in the face of the broader perspectives on value available to persons, adopting a purely subjective attitude seems not only mistaken but *incoherent*.[18] So too, though, does a purely objective point of view on meaningfulness as Susan Wolf's worries regarding moral sainthood demonstrate. What the purely subjective position lacks is the recognition of a broader set of values within the world than those which the agent subjectively endorses. However, included in this set is a vast and important array of *intersubjective* values, which are also unavailable from a purely objective point of view.

According to the example of the corporate lawyer, it seems that she must necessarily sacrifice most if not all of these shared and all of her personal values in favour of an impersonal ideal. Moreover, even if the lawyer does find any of the success-related values of that ideal personally or intersubjectively praiseworthy, she nonetheless cannot endorse them as such. All that matters, that is, the only good reason for accepting something as meaningful from this point of view, is generated by the purely impersonal, objective features of those values.

The fundamental point behind the critical evaluation of these extremes is that in order for most of what makes meaning and is recognised as valuable in human life to be made sense of, values must be interpreted and adopted from a middle ground. As Susan Wolf (2014c, p. 113) puts it "[m]eaning arises when subjective attraction meets objective attractiveness". This

[18] Unless, of course, the person in question suffers from a pathology or other such limitation on their ability to engage with the world and values, in which case devoting themselves solely to an activity such as rubber band collecting might well be normal and genuinely beneficial and meaningful behaviour for them. Susan Wolf's account of meaninglessness requires there to be a genuine option for a person to appreciate and engage with intersubjective sources of meaning, which they fail to take up.

is what I understand to be the essential ingredient to what makes a self recognisably human for Camp; the ability to balance the personal and impersonal sources of meaning within a life.

There are also more pragmatic reasons for endorsing the middle ground that Susan Wolf describes. As the relational theories that I present in the following two sections show, in order to actually pursue and realise meaning in my life, an agent needs opportunities to genuinely interact and connect with other persons as well as the various environments in which they live alongside them. Excluding values in the ways that the extremes I have considered here do makes this kind of relational self-formation impossible and, I believe, would make it impossible for an individual to genuinely develop a self.

It is from this basis that I suggest that theories of self-formation must adjust their focus from the more general domain of the *external* to the more specific arena of the intersubjective. It is this *relational* value, which develops from the perspective that Susan Wolf describes between purely personal and impersonal points of view, that makes lived experience significant and intelligible. While appreciating this requires an acknowledgement of the extendedness of self-formation, focusing solely on the general category of 'external features' will not provide an adequately fine-grained account of what legitimately constitutes the self. Rather, as Susan Wolf claims, theories which operate at the level of and attempt to describe and evaluate the actions, beliefs, desires, and intentions which constitute human lived experience must be able to describe the difference between the purely subjective and objective features of it and require a middle ground between them to function.

Given this, the question of how intersubjectivity affects the self now arises, as well as one regarding how agents should regulate their conduct in order to properly respond to the demands intersubjectivity places on them and situations in which it puts them. I shall consider two different responses to these questions in the next two sections; feminist and Confucian relational accounts of self-formation. Once coupled with these relational theories of self-

formation, the full significance of the relationship between Susan Wolf's account of meaning and the character account will become apparent, suggesting what limits need to be imposed on self-formation.

4.2 Second-personal care: a relational account of need

Susan Wolf's repudiation of subjectively-endorsed impersonal or universal perspectives on meaning leaves theories of the self needing to fill the gap that this point of view leaves behind. The pertinent questions asked by narrative theorists help clarify the contours of this space and what will neatly fit within it. The positions I described in the first two chapters essentially ask both *how* one understands who one is, and *where* that understanding comes from.

The ability to provide an account of who one is in the form of a story answers these demands by fixing one's self in relation to a background of traditions or *praxis*, which informs the limit of what one can do and fixes the limits of intelligibility for one's self-formation. For narrative theorists, a person is always related to one or more stories which track, contest, and inform their experience. However, as I have aimed to show, when a narrative is stripped away, a certain number of vital relations still persist and still connect the self to certain valuable objects and ideals.

Katharine Wolfe offers an intricate insight into the nature of this aspect of lived experience from the perspective of feminist ethics of care.[19] The core issue that she identifies with debates regarding self-formation is that this concept is too often construed in radically *first-*, or *third-personal* terms. Rather than being purely grounded in one's personal, voluntary

[19] For similar work regarding the relationship between needs, values, and care, as well as the finitude of the human perspective, see: Miller (2020) and Simola (2010). For a perspective on care ethics regarding the debate surrounding the 'Other' and Lévinas in Continental Philosophy, see: Xin (2020). For a comparative reading of feminist ethics of care and Confucian ethics, see: Li (2010).

choices, commitments, and desires, however, the self is far more strongly constituted out of one's relations with others.

A conclusion this critique of conventional approaches to self-formation yields is that the 'first-personal' basis upon which such accounts rest implicitly relies upon an extended, relational form of self-awareness. Katharine Wolfe (2017, p. 1) equates this with the logic of the grammatical second-person: "the mode of address through which we apprehend another not as a 'he,' 'she,' an object or as an 'it,' but as a 'you,' that is, as a subject in relation to ourselves".

In order not only to make sense of, but also to enjoy the full richness of subjective lived experience, this kind of recognition is essential to and, it appears, operative even in the lives of agents who ignore or reject it. This is due to the fact that the lived experience of persons is defined by "vulnerability to harm," which Katharine Wolfe (2016, p. 137) describes as an "objective component need that can be recognized and assessed by others".

As beings who are susceptible to events in the world, produced by environments and by others, persons are faced both with desires that they wish to see fulfilled, and needs that they *must* see fulfilled not only to survive physiologically but to survive as the same person. What persons live out are lives and episodes that take place among others and social institutions, practices, and traditions, each of which play a role in determining who they and what their psychological needs are.

Katharine Wolfe (2016, p. 144) concludes that these "investments in the world" not only contribute to one's "own sense of well-being," but that in order to aid in "maintaining one's very sense of self" they must acknowledge the role that others play in making these valuable investments meaningful. Just as an agent cannot choose whether to be dependent on food and water, they cannot voluntarily choose everyone and everything in their environment upon whom and which they depend.

More importantly, the way in which agents voluntarily cultivate specific relationships demonstrates that the cares and vulnerabilities these relations require them to take on as their own cannot be coherently confined to only the connections they choose to maintain. In accepting another as a friend one commits to supporting their wellbeing to a certain degree, which, in turn, places a limit upon one's own sense of wholeness. The stronger the relation, the more fundamental a role it plays in affectively anchoring one's identity as a self: "When I try to imagine my life uncoupled from my partner's, for instance, my life ceases to be recognizable to me as *my* life" (Katharine Wolfe 2016, p. 130).

Most of these strong relations of care that Katharine Wolfe describes are voluntary. However, consider others that appear equally fundamental to the lives of persons: those between parents and children. It seems genuinely impossible for a child to choose who their parents are going to be or for a parent to have complete voluntary control over who their child will be and how they will mature. Nonetheless, unless something goes horribly wrong, the role these persons will play in each-other's lives is constitutive of who they are to such an extent that losing them would radically change the character of their lived experience.

"The special relationships we have with others make strong moral claims upon us," Katharine Wolfe (2017, p. 4) observes, "not because we have elected to enter into them voluntarily, but because our special relationships with others make them uniquely vulnerable to harm as a result of our actions or neglect". Even putting the moral claims these relations support to one side, the vulnerability to harm that relatedness engenders and the manner in which it constitutes one's self fundamentally speaks to the fact that subjectivity cannot be reduced to either purely personal or impersonal lived experience.

A person must essentially care about others by bearing a second-personal relation to them in order to care for themselves, even if they actually ignore these important, self-constituting relations. This point, that vulnerability and care are not "endogenous," further

supports a relational reading of Camp's and Howell's views regarding trait externalism (Katharine Wolfe 2016, p. 141). It seems fair, then, to conclude both that "[s]econd person needs occur when one's experience of the needs of others gives rise to certain needs of one's own," and that person's own experience as a self is fundamentally constituted by the needs of others (Katharine Wolfe 2016, pp. 143-144).

While Katharine Wolfe's position draws awareness to the necessary and involuntary way in which subjects of experience incorporate the experiences of others into their lives, it does not have much to say about what appropriate self-regulation might look like and how it might be cultivated. This aspect of relational theory is addressed, however, by classical Confucian accounts of relational self-formation, which focus heavily on how this regulation should be developed in terms of one's confrontation with others and environmental influences.

4.3 Classical Confucian accounts of relational self-formation

Another thinker who has noted the tension between radically egoistical and impersonal accounts of the self is Kirill Thompson, who recognises that in order to deal with the inadequacies of both of these extremes a third path is required. However, rather than turning to the Western tradition, Thompson considers the issues that attend theories of self-formation from a Confucian perspective, aiming to balance the excesses and defects he identifies in other theories on a middle ground between them.

Thompson begins his discussion by distinguishing two Western notions of the self that parallel the subjective/objective and first-personal/third-personal distinctions that concern Susan Wolf and Katharine Wolfe. First is what Thompson (2017, p. 887) likens to a 'modernist self', which can be traced back to "the Pythagorean and Platonic soul" and be smoothly equated

with the kind of identity that Locke was interested in affirming and Hume in denying. Being "the centre of each person's consciousness and conscience…descriptively *given* and prescriptively *required*," the modernist self stands, then, at the heart of the moral life of subjects that MacIntyre's and Ricœur's theories aim at describing (Thompson 2017, p. 887).

The contrasting extreme to this, the 'postmodernist self' is grounded, like situationism, in observations regarding "the multiplicity, passivity, and constructed nature of personal selfhood" (Thompson 2017, p. 888). These positions hold that the phenomenon of "selfhood" is more a "reflection of one's sociocultural milieu," than one's personal traits and choices (Thompson 2017, p. 888).

Taken together, both explanations of subjectivity appear to fall on one side or another of the clash between theories of self-formation and the sceptical challenge presented in the last chapter. 'Postmodernists' argue that the unity that narrative theorists demand is a phantasm, and a pernicious one at that. Inversely, defenders of a thick notion of self question what assumptions about the fragmented subject their opponents smuggle in and whether they in fact rely on some equally reductive description of the self, such as the radical internalism with which Howell takes issue.

Thompson finds neither view satisfactory. "However effective they have been in breaking up the modernist conception of the self," he notes, "postmodern accounts have tended not to square adequately with the psychological facts or be sufficiently explanatory" (Thompson 2017, 888). As Ricœur (1992, p. 20; 22) concluded forcefully, the human mode of "being" is "as act and potentiality," supported by the "*assurance*" one derives from attestation "*of being oneself acting and suffering*".

There seems to be a truth to this that the postmodernist fragmentation of the self either disregards, or outright denies and in so doing renders any account of subjectivity which might proceed from it untenable. For "if there is a being of the self," which can in any sense be

described as an agent as well as a subject, then “this is in conjunction with a ground starting from which the self can be said to be *acting*” (Ricœur 1992, p. 308).

Indeed, if, as the previous chapter and this sketch of postmodernism have shown, subjects are utterly determined by and dependent on their environment then the active existence by which subjects are identified cannot obtain and all talk of self-formation in relation to agency collapses into a category mistake. Selves would, rather than constructing a horizon of care, simply be determined by the strongest affective demands on their consciousness; be *utterly constructed* by their cares themselves.

As Ricœur (1992, p. 315) claims: “it is on this very difference between *energeia* [acting] and *dunamis* [suffering, passivity], as much as on the primacy of the first over the second, that the possibility ensues of interpreting human action and being together as both act and potentiality”. However, despite the issues that ‘postmodernist’ criticisms of self-formation face, it seems that the essence of their concerns still has some justification. After all, as the critique of narrative theory has shown, a “self-contained, hermetically sealed” account of the self is equally unappealing (Thompson 2017, p. 887).

A robust, further connection between the subject and their milieus, which captures the lived experience of the full spectrum of selves is required. And, as Thompson’s juxtaposition of the two traditional Western positions on this issue shows, this ideal description requires philosophers to tread a careful line between radically personalised and de-personalised accounts of subjectivity. In short, it requires a third approach, typified by classical Confucianism: that of relational self-cultivation.

Thompson (2017, p. 889) describes this view as standing in contrast to traditional Western theories,[20] which describe “people influencing each other more or less on the model

[20] Given the tensions Thompson himself identifies within the Western tradition, calling this approach ‘traditional’ might be something of a stretch. These tensions do, however, seem implicit in both the egoistic and impersonal views that pose such a problem for the relational account of meaning.

of mechanical causation". Rather, in a Confucian model, the elements of subjective lived experience "constitute confluences in the flow of characteristically *human* life, in which the human relata are interactively co-formed and characterized" (Thompson 2017, p. 889).

Two vital themes that run through the *Analects*, and the works of Kongzi's two immediate successors, Mengzi and Xunzi, ground this model.[21] Namely, all three of these thinkers endorse the claim that in order to properly order oneself internally, one must pursue and cultivate the appropriate external relations. "Desiring to take his stand, one who is Good helps others to take their stand; wanting to realize himself, he helps others to realize themselves," Kongzi (6.30) writes (Ivanhoe and Van Norden eds. 2005, p. 20).

The second important theme, which in many senses is a caveat to the first, is that proper relational self-cultivation can only be pursued in *society*. "A person cannot flock together with the birds and the beasts," they require the strictures, support, and structure of social relations in order to truly cultivate themselves (Kongzi 18.6) (Ivanhoe and Van Norden eds. 2005, p. 52).

A significant element of this approach is its recognition of the variety of demands that relatedness places upon subjects. As a member of a community, one must ensure that one relates to others in way that ensures not only that one's own needs are met, but that the needs of the community are properly met and regulated. However, membership of this community also grounds the vital relations of care that constitute oneself, encompassing the family – the most immediate sphere of one's shared needs – friends, and colleagues.

This contrast between the radically individual – and hence radically unhuman – person, and the relationally-sensitive, cultivated one highlights another tension Thompson identifies between the Confucian and Western understanding of the self. Unlike in the analytic or

---

[21] I cite the translations of these works provided in Ivanhoe and Van Norden, 2005. For the *Analects* and *Mengzi*, I provide the book number followed by passage number. For the *Xunzi*, I provide page number references.

continental traditions, the concept of self described in the *Analects* is not and cannot intelligibly be "ontologically and ethically individual" (Thompson 2017, p. 887).

"The gentleman,"[22] Kongzi (12.24) notes, "acquires friends by means of cultural refinement, and then relies upon his friends for support in becoming Good". This is not meant to absolve subjects of individual responsibility for their conduct – "The key to achieving goodness lies within yourself" according to the *Analects* – but is rather a recognition of the co-operative nature of self-cultivation (12.1).

Finding the right means of acquiring friends – cultural refinement – involves cultivating oneself in a way which attracts the best kind of people to aid one in pursuing and achieving the Good of benevolence (*ren*). Such refinement, however, is only possible in communities, and the pursuit and cultivation of it within a social milieu will necessarily have an effect on that environment, its inhabitants and institutions, just as the abandonment of it would.

Mengzi states this tension between relational, and egotistic and impersonal approaches – Yangism and Mohism respectively – most clearly. "Yang is 'for ourselves'. This is to not have a ruler. Mo is 'impartial caring'. This is to not have a father. To not have a father and to not have a ruler is to be an animal" (Mengzi, 3B9). Refusing the strictures of society and consistently putting myself first jeopardises the basis I have for pursuing my own good, while refusing to attach particular significance to my relations of care requires me to ignore that good.

What the description of those who break the fundamental relations that constitute human society and human good as 'animals' is meant to show is that the mere presence of a body in a physical position or social situation is not enough to describe what the embodied character of the agent in that state is. What is vital to doing so is determining *how* the person

[22] It is important to note that by 'gentleman' Confucians mean someone who possesses the *character* of one of noble bearing, regardless of the station they acquire or are born with. After all, as much of the *Analects* notes, plenty of men of high station in ancient China did not measure up to their positions.

inhabits that position; how they conduct themselves, and what relations do and should mediate their conduct.[23]

Furthermore, in the *Analects* a clear distinction is consistently drawn between "utterly private inner possessions of self," and "the unique set of relations in which each person is born and constituted in his or her flow of life" (Thompson 2017, p. 893). Indeed, it is argued that the open or public possession of traits and the relations which constitute them renders these elements of subjectivity in fact more effective in describing and motivating behaviour than a private or inherent possession of them would.

By making subjects accountable to something beyond themselves in a relational way, not only does Confucian theory suggest that "perceptive others" can know one better than one does oneself but *demands* that one responds to these extra-subjective influences (Thompson 2017, p. 893). For such forces are fundamentally constitutive of the social, affective, and even epistemological milieu which one inhabits.

Xunzi was deeply concerned about how to ensure that these relations survived the Warring States period, formulating the forceful injunction: "To pursue [learning] is to be human, to give it up is to be a beast" (Ivanhoe and Van Norden eds. 2005, p, 258). Understood in this sense, learning is much more than abstract knowledge of facts, it is essentially a form of *phronesis*, specifically the knowledge of how to comport oneself in the most benevolent and effective way.

Mastery of this kind of expertise would, therefore, ensure that one's relations of care are appropriately met and appropriately regulated. However, the way to accomplish this itself requires one to forge the right kind of social relations; as Xunzi argues "[i]n learning, nothing is more expedient than to draw near to the right teacher" (Ivanhoe and Van Norden eds. 2005, p. 259). That is, one cannot rely on social mores and structures by themselves. "Rituals and

[23] For a Confucian critique of situationism along these lines, see: Mower, (2013).

music provide proper models but give no precepts" (Ivanhoe and Van Norden eds. 2005, p. 259).

It is on these grounds that classical Confucians draw a distinction between 'cultivated' and 'uncultivated' subjects. Indeed, in setting out this relational account Thompson (2017, p. 893) even claims that "mental states" themselves cannot be understood "in the sense of a private inner world of self". Partly, this is due to the fact that, according to classical Confucian thought, an individual by themselves, 'cut off' from relations, would essentially lack the resources to cultivate and construct this world.

More forcefully, however, Xunzi notes that it is fundamentally impossible to escape these relations. If "he who criticizes me is my teacher, and he who rightly supports me is my friend" then one always has teachers and friends, and is always faced with the choice and responsibility of how to respond to them (Ivanhoe and Van Norden eds. 2005, p. 261). Unless, of course, one is without relations and, therefore, less than human.[24]

This gives further content to the Confucian view that an uncultivated, and hence relationally unbalanced, subject cannot exert full control over the contents of their mental lived experience. Specifically, this "account of…states of mind is *descriptive* of the highly cultivated person, who normally keeps his mind tranquil, but is *prescriptive* for others, in the sense of a psycho-emotional goal to emulate and realize" (Thompson 2017, p. 893).

The aim of Confucian self-formation is, then, to achieve this degree of tranquillity, a robust and genuine recognition of the relations which are constitutive of oneself, others, and the common life one shares with them. It is also important to note that according to this account, self-formation is not regimented, but rather that one's uniqueness as a self ("*du*") "refers to the unique set of relations in which each person is born and constituted in his or her flow of life"

[24] This claim naturally carries a number of seriously problematic undertones. However, it also carries a great deal of positive normative force. The issues that attend the loss of relations and 'humanity' will be dealt with further in Chapter 5.

(Thompson 2017, p. 893). What makes a person different from others, that is, what makes them the particular self that they are, is what they do and experience; how they act and what they suffer.

What a focus on the power and, Confucians would argue, the *necessity* of co-operative self-formation shows is that "relational accounts" do not concern "just…the person, but all sorts of human practices, ethics, forms of life," and demand a recognition of the variety of interactions that these support (Thompson 2017, p. 898). Specifically, in order to effectively pursue one's own good or engage in self-formation, one must do so with an awareness of the role of others in my life, of extra-personal sources of meaning and value.

Moreover, leading a life in society, governed by but also engaged in a number of relations, requires an agent to develop a certain degree of self-control. As the parallel drawn above between Ricœur's and Thompson's positions demonstrates, acknowledgement of one's ability to *act* as well as *suffer* through one's relations is essential for making sense of lived experience in the ways in which both 'modernist' and 'postmodernist' theories have attempted to demonstrate.

By acknowledging that self-formation cannot be conducted in isolation, Confucian thought helps to lay an important element of groundwork for accounts of the self. Supplementing the concept of self-formation with a notion of "co-formation" mediated by "the nexus of all the confluences of relations," these relational accounts tap into a powerful source of intersubjective justification (Thompson 2017, p. 890).

This aligns strongly with Katharine Wolfe's account of second-person needs and shared relations of care, suggesting that there is a large segment of human life that is experienced intersubjectively and relationally. In order to provide an adequate account of subjective lived experience, *all* relevant experiences must be accounted for. Doing so requires an appreciation

of the strong emotional and psychological forces that bind often disparate experiences and interests together in a distinctively human, personal, and meaningful way.

Conclusion: A sketch of the limits of self-formation

What the supplementation of the character account with relational theories of self-formation and Susan Wolf's account of meaning has shown is that developing a nexus of dispositions, interests, and commitments which can be pursued and cultivated, all things being equal, requires a number of criteria to be met. First, I presented reasons why self-formation requires an intersubjective perspective, given the untenability of a purely subjective or objective point of view regarding the vast set of intersubjective reasons which texture human lives. Secondly, I presented reasons why this texture requires agents to adopt second-personal relations of care with not only others but also themselves. Thirdly, I presented reasons why the middle ground that theories of self-formation seek requires a degree of relatedness to one's own lived experience, understood in terms of a Confucian notion of refinement.

I do not believe that even when taken together these criteria place either a particularly burdensome intellectual or practical demand on agents. Indeed, I do not see any reason why children or those suffering from mental illness cannot engage in self-formation in these terms. To appreciate this point, it helps to consider these three elements of self-formation as an *injunction*; a demand to which one is susceptible insofar as one has the ability to respond to the needs and values that one is confronted with in a meaningful way.

This injunction will differ in strictness and force regarding the area of an agent's conduct that provokes it and the capacities of the individual at whom it is directed. An agent's childhood self faces a much less strict injunction than their present one; amusing oneself by

running around in the mud all day is a very meaningful thing for the younger self to do. However, were the agent to respond to those injunctions as their present self, they would likely not only fail to satisfy their actual personal beliefs, desires, commitments, and intentions, as well as to live up to the ethical standards they ought to, but also to meaningfully and intelligibly sustain and develop the self-presentation they endorse.

Similarly, from this point of view, spending one's time collecting rubber bands and devoting all of one's energy and passion to that activity would seem grossly mistaken. However, from the point of view of someone who has suffered a great deal of trauma, or who is affected by a pathology or other condition, this is far more likely to adequately satisfy what their way of being demands of them. Indeed, in such cases I follow Camp in believing that it is eminently possible for these persons to develop and cultivate the kind of nexus that underpins the character account.

With this said, I must also be careful not to extend the charity built into this concept of an injunction too far. Although it is important to allow for agents to sometimes, perhaps even often, fail to live up to their self-presentations – indeed, consistent failure might be what gives rise to a *new* style of being – the possibility of inconsistency and incoherence must nevertheless be preserved to guarantee some characterisations as actually coherent. Likewise, it is vital to understand exactly how much damage and disruption a person can endure before it becomes incoherent to level any self-related injunction, born out of their capacity to care, against their conduct at all.

Before I engage in a more thorough analysis of the limits of self-formation and the consequences of the damage to an individual's subjectivity, though, I must further clarify one element of the relational account I have proposed here. That is, does the criterion I establish of second-personal care for *oneself* not beg the question of both the existence of a self and, practically, exclude a large number of individuals from the community of subjects?

Part of the evil of the harms I will describe in the following chapter stems from the fact that extreme forms of trauma, isolation, and discrimination destroy a subject's capacity to relate to themselves in a first-, second-, or even third-personal way. It is important to understand, though, the way in which a relational theory of self-formation describes this issue; what is lost or suppressed in these circumstances is the *capacity* to meaningfully relate to oneself, as well as others.

The vital distinction I aim to draw is between loss and destruction The former recognises that a person still possesses the capacity, yet perhaps does not have access to the resources or the right kinds of relations or relational support, to easily engage in self-formation. The latter describes cases in which even this capacity has been overwhelmed and extirpated by extreme trauma, cases in which it seems no amount of aid will rehabilitate the self.

It is also important to note that I consider these three criteria for relational self-formation to be fairly easily accessible *all things being equal*. That is, I mean to point out that there is nothing mysterious, particularly cognitively burdensome, or practically challenging about self-formation through relations, given the account of human subjectivity I present. This is not to say that, in specific circumstances, for specific subjects, this account may not be more or less opaque, burdensome, or challenging. As the discussion that will follow will show, there are plenty of fairly commonplace cases in which persons face particular difficulties in engaging in self-formation. However, these must be distinguished from those extreme ones in which no self can be formed or salvaged at all.

With these concerns in mind, my final chapter considers exactly how far a thick notion of the self can extend before the conditions for its formation are lost. Once again, Susan Wolf's work regarding the fanaticism about values that is emblematic of moral sainthood shall be my concern. In addition to this, in examining what it takes to reconstruct or reformulate a way of being, I also present a different kind of limit on self-formation, exemplified the extreme,

character-destroying trauma that constitutes torture. With these limits established, I hope to demonstrate not only exactly what a self intelligibly is, but also what it *is not*.

Chapter Five – The Boundaries of the Self

Introduction: Desire as the boundary of the self

One of the most attractive features of narrative theory, which has also been the cause of much of the criticism directed against it, is that a life story explains and makes intelligible an agent's reasons for acting. The coherence of a plot in which an event or action takes place is directly related to the intelligibility of any explanation can be given of what occurs and the life and character of the person who experiences it. This has led to two broad problems, which I shall summarise again here in order to highlight the issues that attend any attempt to describe the boundaries of the self.

First, because the intelligibility of a life story determines the intelligibility of the self and life of its subject, persons *ought* to do their best to make it as coherent, rich, and well-rounded as possible in so far as these narrative features contribute positively to intelligibility. This issue lies at the heart of Camp's critique of the narrative treatment of the mentally ill, children, and other 'narratively inopportune' lives. No matter what they do, some persons simply cannot answer this normative injunction to desire *in a particular way*, at least in strictly narrative terms.

Second, a focus on narratives rather than relations leaves many interpretations of events and lives indeterminate. Given that memory is constructed and that narrative interpretations of the present often rely on ingrained prejudgements, an account of self-formation requires a measure by which these configurations of events can be explained independently of their own assertions. For a life story to genuinely constitute a self, it must be supported by something beyond its status as a narrative; its intelligibility – and thereby the intelligibility of whatever

life and self it describes – is guaranteed by further facts that would be true whether or not a story was told about them.

I believe that the relational account that I have described need not be concerned by these worries. Relations themselves *are* the further facts about selves which narrative theory lacks and which guarantee the intelligibility and coherence of self-formation. Furthermore, relations are highly context- and agent-sensitive, generating specific normative injunctions for specific persons regarding their actual circumstances, understood as a horizon of care, rather than a rigid, overarching demand for unity.

Nevertheless, my relational account must still be able to exclude some lives from being led by selves in order to avoid falling foul of the triviality complaint levelled against narrative theory. While any event in a life might be narratable, I contend that only some of what persons experience can genuinely be forged into the meaning-giving relations that undergird self-formation. In order to demonstrate this, this chapter focuses on how some events and experiences can not only fail to form robust relations but also on how some might instead damage and undermine a person's existing ones.

One of the strengths of the relational account of self-formation is that it allows for change, even fairly radical change, over the course of a life in a way that narrative theory does not. However, there is a significant difference between positive change, which preserves the ability to pursue and cultivate a self-presentation, and negative change, which undermines this capacity.

In order to explain this difference, I begin by constructing a measure for positive change within a life that establishes the minimal degree of connectedness between the subject of experience and their life. A basis for this has been provided by David Lumsden's (2013a) narrative account of the self in psychotherapy. Lumsden suggests a narrative approach which rejects overarching unity in favour of discrete meaningful connections, which track parts of a

person’s life and character. Although this offers a much better approach than those discussed in chapters one and two, I nonetheless propose modifying it so that rather than employing Lumsden’s concept of ‘narrative threads’, the intelligibility and meaning-making of self-formation is instead supported by ‘relational threads’. This retains the practical applicability of Lumsden’s position while dropping the more problematic elements that attend narrative theory.

This modification of the thread view further demonstrates that ‘overarching’ coherence and unity are not important for self-formation. Rather, many persons will not only survive but thrive as selves with relatively disparate and unconnected structures of meaning in their lives. The reason for this is that the threads of meaning that connect them to the world allow them to coherently express and respond to desires and define what a ‘thing *that* desires’ is.

Lumsden’s account of how either disorganised or disrupted lives can still be relationally sound and coherent establishes the limit of self-formation by measuring how basic the connection between a self, life, and its beliefs, desires, and intentions can feasibly be. When stretched too far, when the presence of motivators such as desire can no longer be explained in reference to the self, these threads ‘snap,’ causing a person’s capacity for self-formation to critically break down.

The second section of this chapter aims to describe the other side of this boundary: the point from which a self has been too badly damaged to survive. Following Leanh Nguyen’s (2007) account of treating cases of torture, I define the destruction or loss of the self as a loss of the capacity to coherently desire from either a first-, second-, or third-personal point of view.

This is consistent with the account of meaningless lives I appealed to above in constructing my relational position, offering a new perspective on Susan Wolf’s work. Rather than living ‘meaningless’ lives due to falsely endorsing an inadequate or hollow sense of satisfaction over happiness, those who have suffered extreme trauma such as torture are incapable of fully appreciating either passion.

Instead, I explain the meaninglessness of these individuals' condition in terms of functions. Since they cannot adopt either a purely personal, relational, or even impersonal perspective on their lived experience, both who they and others are is reduced to a series of functions that cannot be coherently connected to broader threads of meaning.

Once again, the connection between desire and agency plays a critical role in defining the boundaries of self-formation. Desires may be present in these individuals' lives, yet not as something that they themselves can either make sense of or structure their experience around. Here, the position Chisholm described is macabrely reversed: there is stumbling upon and after perceptions and memories but really no self to be found, no coherent nexus of beliefs, desires, and intentions which connects a subject to their experiences.

A return to Susan Wolf's account of moral sainthood and value fanaticism further vindicates this observation. Although the moral saint (or the pure egoist) still manages to adopt a purely impersonal (or personal) point of view, something which is in principle unavailable to someone who has been extremely traumatised by torture, they nonetheless appear to adopt a similar function-based view of life.

The key difference between these cases is that meaningful connections are still available to the value fanatic. What is problematic about moral sainthood or rubber band collecting as a personal ideal is that both forms of value fanaticism ignore sources of meaning and value which it would be good for an agent to endorse. Indeed, both do so to such an extent that their lives appear meaningless because they lack such goods within them.

On the thread view, there is a great risk that little or no coherent connectedness will obtain within these fanatics' lives for this reason. However, it still remains possible, in principle, for these fanatics to broaden their horizon and expand and forge new connections among the threads they care about. This is the key difference between them and patients

Nguyen describes: the latter have *lost* any ability to relate meaningfully whilst the former fail to exercise this ability in a coherent way.

These observations paint quite a clear picture of the limits of a relational concept of self-formation. In doing so, they vindicate it from suffering from the significant limitations I have identified in narrative accounts. Having established the boundary of self-formation as an ability to genuinely *desire* in a coherent way, I also further emphasise the importance of Camp's urgings to recognise persons from a broad and diverse array of lives and ages as selves.

Making desire a central element of self-formation also requires me to address a certain tension between elements of Susan Wolf's account of meaning and the thread-relational view I will develop in this chapter. Specifically, it appears that some persons can coherently adopt a radically objective or subjective attitude towards meaning, such as that of a moral saint, and also coherently desire. Following Lumsden, I believe that recognising these persons as selves is vital, even if they relate only minimally to intersubjective values, or do not do so at all.

As my conclusion will express in more detail, and as I noted at the end of the previous chapter, the reason for this is that these agents can still desire in a robust, meaningful way. Moreover, many of them, due to pathologies or other circumstances, cannot take up the kind of second-personal standpoint that relational theories describe as the 'norm' for self-formation'. For torture to describe the meaningful boundary I aim to establish, my account must recognise anyone who can desire, even in quite a minimal, disparate way compared to what is typically understood as the norm for living and experiencing as a self. In asserting this, I do not claim that either Susan Wolf or the relational views I described in the previous chapter are wrong but rather aim to extend their account of meaning and self-formation as charitably and coherently as possible.

5.1 The thread view of meaning

Recognising that providing "a narrative of an entire person's life" is "an implausible requirement," David Lumsden (2013a, p. 1) instead suggests a more limited application of narrative theory that focuses on discrete elements of lived experience rather than whole lives.[25] Meant to provide "a notion of a person that does not necessarily require a complete narrative unity, but is a bundle of narrative threads," this position recognises that the lives of many persons both cannot and should not be forced to fit an overarching narrative mould (Lumsden 2013a, p. 1).

Given the criticisms typically advanced against narrative theory, Lumsden's adaptation of it and his subsequent attempt to apply the narrative thread view to psychotherapeutic contexts is an interesting development. However, while I agree that endorsing something like a thread view is a necessary step for theories of the self to take, I still believe that Lumsden does not go far enough in disassociating his position from the more problematic claims advanced by MacIntyre and Ricœur.

By fitting Lumsden's account of threads of meaning into the relational account I have outlined, I aim to demonstrate that the narrative component of his view is not necessary for reaching the conclusions it sets forth. Specifically, I intend to show that a relational thread view should be preferred over any narratively-mediated alternative. This should also make clear the important role that *desire* and the ability to occupy multiple different perspectives regarding it play in undergirding self-formation.

Lumsden begins his account by limiting his use of the concept of 'narrative' to how threads of narrative meaning arise and run through a person's life. "Within a person's mental

[25] For further discussion of the thread view and narrative theory, as well as of narrative theory, Parfit and personal identity, see: Lumsden (2013b) and Lumsden and Ulatowski (2017). For an overview of the relationship between narrative theory and cognitive science, see: Gallagher (2000).

life are some narrative threads," he claims, "which can be large and small; threads can be intertwined to form larger strands" (Lumsden 2013a, p. 6). Considering Galen Strawson's coffee-making example, Lumsden accepts that this is a very small, or minor, thread but one which can nonetheless be interwoven with others to form a rich tapestry of meaning.

Likewise, it appears that having a significant goal, suitable of a narrative quest, available to oneself would count as a large thread; likely one large enough to support multiple others being woven about it in a particular pattern. In setting out narrative thread theory in this way, Lumsden appears to invite the same problems that I outlined at the beginning of this chapter, though. For, while it might be the case that "[w]e are involved in projects of different sizes, some of which could last virtually a lifetime," nothing about the way that the thread theory is set up suggests that 'narration' is indispensable for understanding these (Lumsden 2013a, p. 6).

However, to simply dismiss this position on these grounds would be to ignore a vital distinction for theories of the self. By disassociating his position from the teleological and holistic commitments of narrative theory, Lumsden presents an account of intelligibility in which the constitutive elements of a self might be scattered quite disparately both throughout the mental and physical milieus they occupy.

Continuing his analysis of the coffee-making example, Lumsden notes that brewing a drink does not seem enough by itself to make sense of me as an agent. Nevertheless, he also rejects "the other extreme in which the broader picture incorporates an account of where I went to primary school, when I got married, or the fate of my last article submission" (Lumsden 2013a, p. 7). In other words, the account of meaning appropriate to an account of self-formation should not depend on either extreme actually obtaining in a life in order to function.

As it is constituted out of the threads themselves, rather than what those threads describe, meaning must be connected to what "provides a sense of [there being] a subject of

experience and action" in this view (Lumsden 2013a, p. 7). It is at this stage where the thread view and narrative theory can most clearly be distinguished. As the relational account I have developed so far and the critiques of narrative theory I have advanced have established, it takes more than the ability to narrate an action or experience into a story to sustain self-formation. What actually plays this role is the way in which a subject navigates the strands of intersubjective meaning with which they are presented.

Nevertheless, the description of meaning that Lumsden offers makes clearer the way in which subjects of experience engage with this intersubjective, relational web of meaning. Specifically, the thread view explains how confluences of actions and experiences come together and relate to the same self in different ways. "For some people, the major…threads interweave tightly, so that…activities interact frequently and consistently" – a picture suggestive of, but not necessarily reliant on strong narrative unity (Lumsden 2013a, p. 8). In cases such as this, the various relations that constitute lived experience comingle, although not necessarily in a homogenous way.

On the other hand, for others, "activit[ies] occur in different locations with different groups of people and generally do not connect in a rich way" (Lumsden 2013a, p. 8). These persons might have a number of different and potentially conflicting dominant interests or commitments which require them to maintain a degree of 'distance' between the relations which support the different things they care about. What is particularly appealing about this element of the thread view is that it captures the kinds of lived experience characteristic of children, the mentally ill, and others typically excluded by narrative theory.

The reason why I believe that adopting something similar to the thread view is essential to helping a relational theory set limits on self-formation is that it incorporates the full range of types of experience as a self I have considered so far. Moreover, Lumsden (2013a, p. 9) notes that when it comes to treating disorders of the self, the "tool of narrative creation" that

the thread view describes "could be effective even when it does not duplicate the state of the healthy person". An ideal of coherence and meaning-making, not of *unity*, is identified as vital to the process of self-formation.

In providing such an account of meaning-making and the intelligibility of the self as the subject of experience, the thread view provides the final element of the account of what a self *is*. By locating meaning in threads, which may or may not be disparately spread throughout a person's life, Lumsden's position, once separated from any general commitments to narrative theory, further reinforces the claim that the ability to meaningfully relate to *something* is what is vital for self-formation.

This 'something' need not last throughout even a whole episode, let alone a whole life; as long as it is replaced by something else or as long as there are other relations to sustain meaning-making when it is gone, the subject who experiences them will have a self. This caveat describes precisely the boundary I have aimed to draw in this chapter: as long as there is some ability to relate to and pick up even disparate strands or individual threads of meaning, there is an intelligible account of self-formation to be had.

This might not be a particularly pleasant or 'rich' account; sustaining a variety of disparate relations might well involve a degree of flagrant hypocrisy or disregard for others, and these relations themselves might be quite simple. What it does reinforce, however, is that suggestions that there is a 'normal' way of being a self, which agents ought to exemplify, are likely deluded at best and extremely dangerous at worst.

What does the opposite of this position, the other side of the boundary, look like? In order to answer this question, an account of a life which lacks any meaningful relations and is devoid of the very *capacity* to meaningfully relate to anything is needed. The damage sustained by the self through suffering extreme trauma, such as torture, is a paradigm instance of this,

clearly identifying what particular aspects of lived experience need to be lost in order for the self to go with them and suggesting what existence might look like without it.

5.2 Torture and the loss of desire

Reflecting on her experience treating it, Leanh Nguyen (2007, p. 57) claims that "[t]he most terrible, and intractable, legacy of torture is the killing of desire". Essentially, what Nguyen describes is the loss of the ability to *meaningfully relate* to either oneself or another. A person's connection with themselves, let alone the broader living world they inhabit, is poisoned by the arbitrary, extreme violence that has disrupted their lived experience.

The account of meaningful lives presented across the last chapter and previous section strongly vindicates these findings. Like Susan Wolf's rubber band collector or Sisyphus, these patients find themselves in a position in which their life is occupied by meaningless tasks. However, what makes the Collector's project incoherent for someone who does not suffer from a mitigating pathology or other influences, is the fact that it makes sense to level some kind of a normative injunction against them to engage with the world and their peers, rather than just collect rubber bands. What makes the lived experience of those who have suffered extreme trauma incoherent is their inability to personally cherish a broadly meaningless activity like the Collector does.

The destruction of desire that results from torture makes it impossible for these individuals to do anything *for themselves*, even in the purely egoistical sense with which Susan Wolf takes issue. Moreover, it is equally impossible for them to occupy the other end of the spectrum this account of meaning sets out by giving themselves to a purely objective set of values or principles. Essentially, in this case the loss of the capacity to desire results in the

inability to pursue or cultivate self-constituting relations between oneself, the world, and others.

This aspect of the psychological damage characteristic of torture is further borne out by Nguyen's (2007, p. 63) observations regarding one of her patients: "he limits himself to interactions with healthcare providers, where the roles are prescribed, the interpersonal grounds are circumscribed, and desire, and ambiguity and meaning are minimized". This describes an inability to meaningfully relate to the world beyond the most minimal and highly-regulated ways.

The patient seems incapable of relating with his healthcare providers *as other persons*; what makes their company tolerable appears to be the role that they play, not the features which make them persons. Contrasted with "[m]ost of us, the fortunate ordinary ones," who "inhabit desire in spite of, because of, the margin between need and reality, between the original loss and putative future recovery," those who have suffered extreme trauma cannot tolerate this openness to new possibilities and relations (Nguyen 2007, p. 63).

Desiring as a brute fact, let alone the more refined forms of it, make no coherent sense for the patients Nguyen describes. They cannot make sense of the grossness of what they have suffered and, hence, can no longer make sense of what it means to resist pain and embrace positive enjoyment. "Our aliveness," Nguyen (2007, p. 63) claims, "is confirmed and honoured by stepping into desire," which is exactly what is attacked and broken beyond repair through torture.

In this way the loss of desire stands in direct contrast with the thread view. Although meaning and sources of desire can be very broadly dispersed in Lumsden's view, they nonetheless still offer the kind of capacity for self-repair and development that trauma blocks. Even if what an agent desires is taken away from them, the capacity they have to value themselves, the world, and others, does not necessarily go with it. In principle, the agent can

still forge new relations and connect themselves with new threads of meaning and, hence, have a chance to recover from the disappointment which initially disrupted their relations.

Contrasting this process with the condition of someone like Sisyphus demonstrates just how vital the capacity to desire is for self-formation. Bound as he is to his boulder, Sisyphus cannot reach beyond his predicament to become someone else in spite of what he suffers or seek out new sources of meaning to reinforce his personality. Not only who Sisyphus is, but all that he could be or want is constrained by his suffering to such an extent that he cannot exist beyond it or make meaningful sense of his predicament.

Recognising this brings the discussion that began with the Ricœurian subject of experience around full circle. The definition of 'that which' experiences its life as a self accesses as thick a notion of desires as of self-formation. This requires a distinction between the capacity to experience and the capacity to care and respond to the world to be drawn. In so far as a human being can access sense data, it will have experiences, yet only in so far as one can access and inhabit *desires* can one care about and relate to the contents of those experiences.

The limit of Nguyen's patients' responses to a functional level vindicates this point. Without thick desires, which they can *inhabit*, these unfortunate individuals cannot develop coherent relations or threads of meaning that sustain robust agency. They can make choices but not genuinely choose *for* any cause or sake, let alone for the sake of themselves.

To emphasise this, it is also important to consider how losing the corollary grounds of agency impact an individual's ability to engage in self-formation. For instance, someone who suffers from Alzheimer's might well be able to desire and intend in a robust way but have lost an important part of their capacity to believe. As with one subjected to extreme trauma, their ability to coherently link what they do and experience to who they are is disrupted, so that connecting the two becomes impossible.

Additionally, severe alienation from the contents of one's experience, such as is brought about by some forms of depression or the frustration of one's goals might constitute a critical loss of the ability to intend. Someone who has discovered their partner cheating on them and whose life was so centred around that relationship likely still has coherent desires and beliefs about the world but cannot bring themselves to genuinely intend anything. Whatever they think and feel is made meaningless and neutered by the frustration they have suffered.[26]

Nonetheless, understanding this also provides a way of making sense of recovery from other forms of trauma and of how not every episode of tremendous harm or cruel reversal that is visited on a person need destroy their subjectivity. Likening living to a Sisyphean struggle, Nguyen observes that the kind of isolation which comes to consume the lives of torture victims is present, to an extent, in everyone's lived experience as are the resources to confront it, albeit to varying degrees:

> What distinguishes the true survivor from the walking dead is the capacity to bear the anxiety and emptiness of that paradoxical space – that is, the ability to find creativity from within the knowledge of death, to hold onto desire in the face of loss, to stay in relatedness and aliveness in spite of annihilation, to keep rolling [the Sisyphean boulder] while in acceptance of uncontrollability, futility, and inevitability. (Nguyen 2007, p. 65)

The difference between those who have suffered the depredations of torture and 'survivors' is that only the former are constrained by their suffering. As the thread view and Susan Wolf's account of meaningful lives suggest, so long as a capacity to desire and engage in

[26] These are clearly *extreme* cases of pathology and suffering. I do not mean to suggest that persons with milder forms of these conditions or in less-severe situations of this sort will suffer a collapse of the self.

intersubjectively-generated meaning is preserved, a person can pursue and cultivate genuinely meaningful projects and live a flourishing life. For them it is still possible to live "in a way that is significantly focused on, engaged with, and concerned to promote or realize value whose source comes from outside oneself" (Susan Wolf 2014b, p. 105).

What Nguyen's work shows is that while everyone is potentially at risk of having their capacity for self-formation destroyed, this nonetheless takes suffering an extreme degree of trauma to come about. Vitally, this establishes a clear distinction between persons who have fewer threads of meaning, say, because of a mental illness or simple bad luck, and those who have no hope of recovering and generating any self-sustaining meaning in their lives at all.

Contrasting this account of torture with Susan Wolf's work raises another important point, however, which needs to be resolved in order to fully vindicate the boundary of self-formation that I propose. For it seems that those persons who live meaningless lives or fail to recognise the worthlessness of their projects are worryingly close to torture victims. Regarding some of Susan Wolf's remarks in *Moral Saints*, it seems worrying, if not downright wrong, to suggest that Mother Teresa was just like Nguyen's patients because of her selflessness.

While the value fanaticism emblematic of moral sainthood strikes many of the same chords as the loss of desire through torture, I believe that the two states are not equivalent. What fundamentally distinguishes them is the presence of a capacity to desire in the fanatic, which is visible in their appeal to extreme values, and which the experience of torture destroys in a person. Consequentially, moral saints can still construct threads of meaning to pursue and cultivate, albeit only in a very limited sense.

5.3 Value fanaticism

While torture and extreme trauma describe the critical collapse of the self, a similarly unpleasant fate that might befall a subject is presented by an *overabundance* of value – more specifically, an overzealous relation to values. In the case of the Collector, Susan Wolf describes someone whose singlemindedness regarding a particular activity or outcome consumes and renders meaningless their life and experiences. However, this critique, and the normative injunction that attends it, can also be extended to cases in which a *moral* fanaticism dominates a subject.

For these persons, doing the right thing morally *to the exclusion of all else* is the guiding principle of their lives, and, crucially, of their self-formation. Interests, commitments, and values that appear to have no moral worth are left uncultivated or only given attention when there is no 'worthy' alternative. In short, such subjects only foster wholehearted, genuine connections to moral projects and characteristics and, at best, enjoy a contingent connection with nonmoral goods and values.

This account still operates within the paradigm of self-formation and, vitally, still carries with it the same kind of normative injunction that the thread view preserves. Like persons who suffer from a pathology, whose lives have been disrupted, or children and the elderly, it is still intelligible to suggest to these persons that they should consider endorsing a different set of values from the ones that they actually do.

The key issue that Susan Wolf identifies, and which makes value fanaticism appear dangerously close to a form of self-inflicted torture, is that in pursing moral sainthood *'who* one is' risks being sacrificed in order to promote '*what* good one might do'. When adhered to stringently, whatever constitutes the self is prone to be given up to serve a purely moral end. What contributes to self-formation can, and in many cases must, be separated from what makes a moral saint. The strict principles and demands of morality generally do not care whether

someone has personally valuable friends, family relations, or hobbies if these things do not substantially contribute to achieving morally good consequences or duties.

I believe that this tension can be resolved by returning to the basis of the relational account of self-formation, at which the capacity to desire and *care* play a fundamental role. What I propose is that *sometimes* pursuing moral sainthood might be self-destructive but only when the overarching principles which it demands that an agent follow cannot genuinely align with or be developed from the threads of meaning that constitute who they are.

In some ways, this is quite similar to Alfano's account of factitious virtue. Some persons might be able to 'train themselves' and modify the threads of meaning from which they form themselves over time in order to adopt moral rather than more personal values. This process would be successful if they were able to genuinely take on a broader moral perspective on values as their own.

This same process would, however, fail and draw dangerously close to torture, if not collapse into a form of it, should a person take the leap from personal to moral values without the right background of relations and meaning-generating threads. In this case, adopting a purely moral standpoint would involve attacking their desires and the very basis of their self-formation and values.

As Susan Wolf has suggested, it is unclear whether the world, society, or individual lives would be better if everyone were a moral saint; much of what is extraneous to moral conduct is what makes life deeply valuable. What is clear, though, is that many of these non-moral goods are fundamental to certain persons' selves and capacity for meaning-making. Removing them in favour of a purely moral principle, which fails to sustain that capacity, opens the door to deeply problematic consequences.

The crux of the issue and, I believe, of the boundary on self-formation is whether or not the changes which affect a person still permit for some normative injunction, constituted in

terms of their beliefs, desires, and values, to be intelligibly levelled against them. As MacIntyre discerned in making the 'What is he doing?' question central to narrative self-formation, some kind of accountability to one's own values is essential for any intelligible account of them to be given.

The trauma inflicted by torture makes this impossible. However, moral sainthood only reaches this stage in some cases. Mother Teresa's conduct, for instance, has clearly struck many – Susan Wolf included – as exemplary of a certain, deeply moral kind of accountability. For better or worse, adopting such a strong relationship with values is neither possible nor healthy for everyone. It is important for theories of the self to recognise this and respect the possibility for successful self-formation as a moral saint as well as to be wary of the risks involved in pursing such an ideal.

This also holds true for the other kind of problematic example that Susan Wolf describes; the Collector. Here, the thread view also allows for persons who focus narrowly on subjective attractions to count as selves, even though they clearly ignore intersubjective relations in the same way that moral saints do. Consider, for example, what the life of the Collector would look like if they suffered from a pathology. In this case, any kind of injunction to participate further in intersubjective, strongly-relational self-formation would make little sense. However, this Collector can nonetheless pursue and cultivate purely subjective relationships with the contents of their lived experience, developing and refining certain ways of being.

In this case, the subject's capacity to care about and respond to the world is not destroyed as it is in the lives of Nguyen's patients. Nevertheless, there is no guarantee that the Collector's personal interpretations of them either must or ought to be publicly available. Meaning, on the thread-relational view, can be idiosyncratic, however, for many persons it cannot be purely personal or impersonal. For instance, most persons who do not suffer from a

pathology in the way that Susan Wolf suggests the Collector might would likely be incapable of coherently engaging in self-formation in the same way as they do. Due to their greater degree of awareness of and connectedness to interpersonal relations, rubber band-collecting would be very difficult for them to genuinely endorse as a meaningful personal ambition.

The contrast between extreme perspectives regarding value and the relational-thread view suggests that while second-personal self-formation presents something of an ideal, enough relations and enough threads can be maintained by first- and third-personal positions as well. What all of these different accounts of the self have in common is a capacity to care about and relate to certain things in the world. Hence, some kind of coherent injunction to behave in a certain way, to be a certain kind of person, can be levelled against the agents in all three examples I have covered.

Nonetheless, the discussion of moral sainthood in this section also suggests that extreme third- and likely first-personal standpoints can collapse much more easily than the second-personal one I described in chapter four. This is caused by the conscious decision on the agent's part to try to become a moral saint when they are not ready to give up their personal, intersubjective relations.

A parallel case at the other end of the spectrum would be someone trying to leave behind their second-personal cares and commit to behaving in a purely egoistical fashion when a strong connection with others is still central to their life. However, someone born without or who loses or simply has less of an ability to comprehend and engage in this intersubjective realm of meaning would inhabit these extremes far more easily.

By placing the capacity to make normative injunctions alongside the thread and relational views I have proposed, I believe that a clear boundary for self-formation can be drawn. Using examples of torture as a limit case, I have demonstrated not only how the loss of the capacity to care and respond, expressing one's beliefs, desires, and intentions, makes self-

formation impossible but also the important distinctions that must be made between the kinds of lived experience that characterise mental illness, different ages and stages of life, and moral sainthood.

With these categories in mind, I conclude this chapter by introducing a number of open questions that they pose. Chiefly, the above discussion of torture and pathologies gives a new importance to the notion of 'rehabilitation' in various different contexts. There seems to be a *prima facie* case for attempting to rehabilitate and restore the capacity for self-formation that torture inflicts, where possible. However, in cases where persons might have less of such a capacity than their peers, but still enough to genuinely desire and value projects, this question becomes far more vexed.

As this last section has highlighted, imposing ideal forms of life on persons can lead to significant harms, While the application of the thread view that Lumsden proposes notes that the 'healthy' person is not necessarily a useful or good paradigm for treatment to appeal to. Given this, I worry that in some cases rehabilitation that focuses too heavily on removing 'abnormal' behaviours or outlooks might also constitute a form of torture, in that it necessarily involves destroying the capacity for meaning-making emblematic of a person's self in order to turn them into someone else.

5.4 Hypocrisy and Rehabilitation

One final aspect of the thread view I will consider is the manner in which it supports account of lived experience in terms of "pictures of the ideal forms of life" given by Peter Strawson (2008, p. 29). The main concern this position captures is that "it is…easy to exaggerate the unity of the personalities of those we say we know when we really know them only in one or

two particular connections" (P. Strawson 2008, p. 30). Absolute stability is not a necessary or even natural component of self-formation. Rather, what furnishes persons with their own *ethics* can change perfectly naturally, and indeed quite rapidly.

Given this, it might be tempting to suggest that this view merely points to the suspicion that selves are, by and large, mercurial and that there is no sense to be made in trying to measure or observe what ideals they really hold to. Yet, the mere fact that persons can and do change their opinions on and develop new measures of how they should live does not by itself undermine the claim that they can live coherent lives as selves.

What is crucial to Peter Strawson's (2008, p. 33) account is the distinction he draws between a personal ethic – represented by one's individual ideals – and what he calls the "sphere of morality". In this view morality, what one might call the extra-subjective injunctions a person faces, is essentially social and relational (P. Strawson 2008, p. 34). What others can call upon an agent to do goes beyond, but might conflict with, what they are personally attached to but as long as they also have some minimal interest in these social institutions, they cannot remain insensible to their demands.

Shared interest, understood in this robust, relational sense, is crucial for understanding how persons develop the consistent goals and concerns beyond their own which characterise them. As Susan Wolf and the relational theorists, as well as Lumsden, have claimed it also seems impossible to adequately configure a robust individual ideal without the further support of a community. This is not to deny the claim that there could be a subject "who recognized his interest in the system of moral demands and resolved to merely profit by it as much as he could," rather, it acknowledges the "practice of hypocrisy which this policy would necessarily involve" (P. Strawson 2008, p. 40).

Hypocrites of this kind are not rare but their hypocrisy is only made possible by the relational background of interests which informs their resolutions. Indeed, it also appears that

this kind of problematic attitude can manifest on the ethical plane as well as on the moral. Wholly devoting oneself to an ideal only to betray it whenever it is convenient is just as striking a form of hypocrisy as devoting oneself to a moral law and breaking it whenever one feels inclined to.

It is also on this ethical plane, rather than the moral one, where the tensions which undergird the thread view are best expressed and met. The account of individual ideal acknowledges the changeability of persons' attitudes regarding how they should live their lives – and, insofar as they satisfy the criterion of minimal interest in an extra-subjective institution, how they should live their lives *morally*. Yet, leaving space for this multiplicity of ideals, Peter Strawson grants subjects the room to experiment with, encounter, and experience a number of different ways of being.

According to this model, it seems perfectly natural that ideals can be changed or replaced by new experiences or spontaneous reconfigurations of old ones. Indeed, it is likely that many subjects will develop a repertoire of sorts, collating specific arrays of personal *ethics*. It is also likely that many lives will be characterised by a number of tensions between these ideals, which strikes me as the major motivation for rejecting them and demanding a stronger account of personal unity.

Here, then, lies the crux of this issue: does entertaining a number of conflicting personal ideals, or finding oneself torn between them, constitute a kind of worrying hypocrisy or the kind of disunity that narrative theorists are concerned about? A core consideration, which relational theories embody, is that there is a genuine limit to what one can observe both about oneself and others when it comes to these personal ideals. Another, which the debate regarding narrative theories has raised again and again, is that this charge of hypocrisy becomes most troubling, and in a sense most incoherent, only when juxtaposed with an overarching demand for unity.

The first consideration demands that subjects be given room to explore their identities through pursuing and cultivating elements of their self. Self-formation is not and cannot be static. What Ricœur's test of suspicion aimed at addressing and what the care-derived relatedness argued for by Katharine Wolfe acknowledges is that even when one has a stable set of individual ideals, they are liable to be tested as one's circumstances change. This strikes me as a necessary condition for making any claim that a subject *should* change.

Narrative theory, however, risks taking this injunction to a different radical extreme. Requiring that one's individual ideals be unified in such a way that they might be explained in terms of a plot or quest gives this *ethical* element of self-formation a highly, and potentially dangerously, *moral* tone. The overarching teleological demands of the story do provide a strong measure for settling a test of suspicion. Yet this comes at the expense of providing room for subjects to explore 'weaker' configurations of ideals or entertaining a broad multiplicity of them which are unlikely to cohere in the strong way that a plot demands.

The possibility of displaying such moral hypocrisy on the ethical plane of individual ideals only appears to arise if one demands that lives must be 'whole' in order to be lived by coherent selves. Numerous issues with this have been raised already, as have the number of consistent alternatives to such a view. Perhaps Camp best states the crucial point that this disagreement raises in suggesting that subjects strive for a coherent style rather than the holistic coherence of a life story.

It is also important to note that tolerating this degree of *disassociation* between ideals and threads on the ethical plane is not consistent with tolerating a moral disunity of the self. As the interests which Peter Strawson, Susan Wolf, and relational theorists describe as underpinning moral injunctions are extra-subjective, they stand in a distinctive, independent relation to the self and their individual ideals. A useful way of describing this relationship is in

terms of the difference between one's *way* of proceeding, and the demand that one proceed *somewhere* in particular.

Nguyen's patients are incapable of personally cultivating or exhibiting any real control over either facet of this distinction. Naturally, they cannot answer the call of a whole life narrative to proceed towards a particular goal or end but most significantly, they cannot pursue or cultivate a personal, personally valuable, way of proceeding through life in general. A question this raises is whether the account of torture entails that persons always should strive to find a better way of proceeding if they can.

One way of understanding this is in terms of a spectrum on which pleasant stylistic coherence might form one pole and the destruction of desire another. On such a model, getting as far away as possible from the risk of extreme trauma is of fundamental importance to persons. This suggests that persons *ought* to pursue ever-stronger positive efforts at self-formation and that making efforts to help others in doing this is morally required.

Another interpretation of the facts suggests quite a different schema. Rather than viewing the potential ways that persons can live in terms of a positive-negative spectrum, along which one should progress, this second position suggests a binary account of the self. Either an agent qualifies as having something to commit to and make meaning for themselves with, or they do not.

I believe that this second schema best captures the insights of the relational position that I am proposing. In its earliest form, developed from Camp's character account, all that was important for discussing self-formation on this view was whether or not a person genuinely had access to a 'nexus' of character traits. According to Susan Wolf, what is minimally required is the ability to be aware of meaning and value in one's existence, even if one is somewhat mistaken about how it is intersubjectively generated. Likewise, for relational accounts of self-formation and the thread view itself, the ability to connect with other persons, ideals, events,

and things in the world is the starting point from which further demands might be imposed on the subject of experience.

The specific relational circumstances a subject finds themselves in might, for instance, demand a particular response according to Confucian ethics, but what makes this demand itself intelligible is the fact that they are an agent, constituted by relational connections. It is a mistake to conflate the specifics of the demands that they face as a self with the basic fact, which I have been trying to defend throughout my project, that as a self a subject of experience is open to demands and can make them intelligible to themselves.

It is for this reason that I believe Peter Strawson's distinction between morality and individual ideal is so important. The first of the two schemas that describe the boundaries of self-formation seems to not properly incorporate this notion. The 'ideal' in this view is to not lose one's self, however this is imbued with a significant degree of moral force, becoming: 'one ought not to have one's capacity to desire destroyed and help others avoid this fate'. This is achieved by enjoining persons to push themselves and others as far as possible towards the positive end of the spectrum.

A return to some of the psychological issues discussed earlier in this chapter makes the significance of the mistake here apparent. What vindicates Lumsden's conclusion that the ideal of narrative 'wholeness' should not be the goal of psychotherapeutic treatment is that value itself requires only the presence of some threads of meaning, not a whole tapestry. Worse, forcing this on some persons is liable to do more harm than good.

In analysing cases in which patients are presented with 'cures', Lumsden (2013b, p. 172) remarks that "consideration about identity at a time clearly has significance for identity through time". Citing an example from Oliver Sacks' psychological work, he notes the worries a sufferer of Tourette's Syndrome expressed concerning the treatment of his condition:

"Suppose you *could* take away the tics… What would be left? I consist of tics –there'd be nothing left." According to Lumsden (2013b, p. 172) "for him, a 'cure' would be death".[27]

This aligns with Nguyen's insights regarding the damage inflicted by torture. What is 'killed' in this form of treatment is the *particular way of desiring* that constitutes Sacks' patient as a self. While it may be the case that this way of being comes with tics, it nonetheless presents a certain connection with meaning that is just as valuable as anyone else's.

A problem here is that most persons naturally lose habits and change ideals, as Peter Strawson suggests. However, the way in which this normally occurs, being brought about by shifts in the relations and structures of care and attention that render a person's life intelligible to themselves, is quite different from a course of psychotherapy or medication. Specifically, these require someone to directly intervene on the person's behalf, in their life, in order to direct its flow towards a specific conclusion or set of outcomes.

The point I am making here is that nothing about self-formation should oblige persons to be rehabilitated in the way that Lumsden describes. The problem with the first of the two schemas of the limits of the self is that it appears insensitive to the fact that those who seem to lie closer to the boundary live lives imbued with the same fundamental meaning and value as those who enjoy rich, finely-textured being.

Note, though, that in the model I endorse, there is still a requirement that those who have had their capacity to desire destroyed be rehabilitated if possible. In a sense, these individuals have nothing left to lose but everything to gain. What is important is that once they have been rehabilitated, there is no imperative to continue to sculpt who they are into a more pleasing or 'better' self, although there may well be good reasons to do so independent of their basic existence as subjects of experience.

[27] See: Sacks (1986), cited in Lumsden (2013b, p. 172).

Susan Wolf's account of meaningful lives suggests something of a way forward from this. In so far as some lives are only minimally meaningful, imbued with either purely egoistical or impersonal sources of significance, it might be better to try and encourage those who live them into the intersubjective space in which more genuine meaning is generated. A weaker version of this claim might be that it is better to give those who live less meaningful lives every opportunity to live more meaningful ones.

Another facet of this issue is that the statements that Lumsden cites may not be indicative of every person who suffers from a pathology. For many, treatment is seen as a positive way out of a bad situation. Indeed, many persons who do not suffer from conditions like Tourette's Syndrome might have a similar desire to change who they are. They may not be pleased with the outcome of that change but at the time they chose to pursue it, becoming someone else was a vital personal ideal, just as important as pursing and cultivating existing desires is to other persons.[28]

My aim here is not to disparage the notion of rehabilitation or to provide a comprehensive account of when, where, and how it is permissible. Rather, I have sought to draw attention to the complex relationship between it and the boundaries of self-formation. The fact that in some cases, rehabilitation risks destroying a self or damaging parts of it beyond repair carries with it both positive and negative connotations. The important question that must be asked, where possible, is whether a person wants to change and whether they should coherently endorse altering who they are, given their relational connection to sources of meaning, value, and their desires.

Conclusion: Desire and value

[28] For a further discussion of these issues, see: Paul (2014).

I end this chapter with a brief summary and restatement of my position on the boundaries of self-formation and the reasons that they ought to be constructed at the point at which the capacity to desire is lost. Essentially, every position I have considered throughout my project has proposed some explanation of the relationship between lived experience and personal values. Generally, they have concluded that in order for me to be able to figure oneself as the subject of that experience, one must inhabit or bear some relevant connection to the personal values that arise within it.

For this reason, Lumsden's thread view is vitally important to accounts of self-formation. By noting how meaning and value can be dispersed far more disparately and flexibly throughout life than other narrative accounts permit, his position supports a much broader notion of the self, which is sustained the simple connection of some threads of meaning. A perennial issue for the various positions I have explored has been how to account for the changes between stages of life and the differences between persons that should be reasonably expected, a problem which Lumsden solves very effectively.

The appeal of the thread view is further reinforced, I believe, when contrasted with Nguyen's account of torture. It is a serious mistake to equate the lived experience of the disunified persons that Lumsden describes with the extremely traumatised. The crucial difference between these two cases is that by identifying threads of meaning, even disparate ones, the thread view describes a capacity for *desire* which is destroyed by extreme traumatisation.

For Nguyen's patients, both the capacity and the will to desire are gone, making it impossible to form any coherent connection between lived experience and personal values. Specifically, the kind of violation they have been subjected to makes it impossible for them to tolerate the necessary nuances or adopt the attitudes that enable them to make sense of their

lives in these evaluative terms. Their selves are fundamentally undermined by their estrangement from the processes by which they might recover, pursue, and cultivate values.

This invites concerns about whether some ways of life appear to lead persons dangerously close to torture and whether moral sainthood might actually constitute a form of torture. The crucial test is whether an agent's preferences and attitudes have been sufficiently modified and inhabited by the moral principles they would endorse as a moral saint before they commit themselves to the path it prescribes. The problematic instances of this or any other form of value fanaticism is, as I also noted in chapter three, that in order for values to genuinely describe and support self-formation, they must genuinely connect an agent to the world and broader contexts in which they live.

Contrasting elements of moral sainthood with the capacity for care that Katharine Wolfe's relational position makes use of further supports this point. Value fanaticism becomes problematic at the point where it surpasses or overloads the capacity to care and the relations that which inform it. Suddenly deciding to become a moral saint because, say, one has read Peter Singer might be a choice motivated by genuine moral concern for the wellbeing of the world and others but without the appropriate preparation and self-reflection, it is likely to be incredibly harmful.

Finally, Peter Strawson's remarks regarding the relationship between the coherence of the process of self-formation and ideals and values have further clarified what kind of account of rehabilitation of the self ought to be given. Particularly when contrasted with value fanaticism, the distinction drawn between personal ethical ideals and social morality helps establish what measure of coherence accounts of self-formation ought to appeal to.

Many of the problems that I have described over the course of this thesis, such as how the experiences of mentally ill subjects and children are best accounted for, are resolved by recognising that the capacity to desire and be open to care are essential to personal

development. This can be contrasted with the monolithic ‘moralised’ social injunctions that Strawson describes; as I have argued, approaches like narrative theory commit their gravest error when they mistake the ability to care about and respond to *this* as the limit on self-formation.

What this chapter has aimed to show is that accounts of self-formation must focus on how agents *personally* relate to their lived experience, comprised of both these further moral injunctions, their individual beliefs, desires, and intentions, and the way in which these all co-operatively interact on an intersubjective plane. For this reason, I take Peter Strawson’s framework regarding the relationship that subjects have with their own lived experience to vindicate Lumsden’s rejection of ‘whole’ lives and selves as the ideal for psychotherapeutic treatment.

Clearly, desire plays a central role in these positions and, from the very introduction of my thesis, an overarching question regarding self-formation and its limits has been: how does an agent coherently desire? Establishing a boundary on self-formation in terms of the capacity to desire is necessary for making sense of the relational view I endorse has made clear that at least when it comes to giving an account of the self, there is no *way* of desiring that is essentially better.

As Lumsden’s and Nguyen’s observations show, what is ultimately and fundamentally important for self-formation is the *capacity* to desire simpliciter, not the quality of one’s desires. With this said, as many of the positions I have described throughout my project attempt to show, once one has this capacity and establishes a self, further questions regarding the coherence of one’s exercise of it, such as those relating to hypocrisy, necessarily arise.

It is for this reason that I have placed such an emphasis on the notion of rehabilitation. For, if theorists take the role that desire plays as the ground for self-formation seriously, then they must also recognise the benefits, challenges, and dangers that come with trying to fix a

self. My aim here, it should be noted, has only been to raise these issues in light of the conclusions I have reached regarding the boundaries of self-formation, to suggest how the contours that an ethics of rehabilitation might be established rather than to argue for a specific approach myself.

Drawing this distinction leads me back full circle to the question and disagreement from which my inquiry regarding the self started. Having considered the further implications of the comments made by Locke and Hume and reviewed the current state of the debate surrounding this disputed notion, I believe I have established both a number of important limits on it and a number of critical points regarding it. Specifically, as this chapter has most clearly shown, I have attempted to do justice to how something like a forensic concept of the self enjoys a much easier relationship with beliefs, desires, and intentions than the typical Humean and neo-Humean picture suggests. Given this, I will conclude with a few short comments regarding the nature of the self's relationship with desire and the various roles it has played across my thesis.

Conclusion – Idealised Lives and Ideal Ways of Living

Narrative theories of the self are emblematic of both the strength of the connection between desires and the self and the problems that attend attempting to explain it. On the one hand, as MacIntyre so forcefully claimed, agents need a way of intelligibly representing their motivations to one-another over time. On the other, these explanations must coherently capture all of the salient aspects of a subject's lived experience or at least allow for divergent or new accounts that encompass them to be given.

Each of the claims that narrative theorists advanced attempted to represent the role desire plays in constituting the self. In the first claim, appropriate desires were both expressed by extended socio-historical contexts and actually responded to and taken up by the agents within them. According to the second, they were transmitted across and made live within these contexts by the framework of temporally-extended traditions and practices, which were limited and connected both to one-another and specific agents by the third and fourth claims regarding unity and teleology.

Although I disagreed with the kind of account of desires and their relationship with selves that narrative theorists advanced, I nonetheless accepted part of their core proposition. This was that desires and other self-constituting information must be minimally *representable* and that these representations need to leave a subject of experience *open* to the world.

These then motivated the critique of narrative theory I offered, which took specific issue with the limits narrative holism and teleology placed on subjects' capacity to construct a horizon through and against which they affectively interacted with the world. Rather than providing a clear interpretative lens, through which the properties of the self as an organisational matrix for important contents of a life, such as beliefs, desires, and intentions,

could be scrutinised, narrative theory was instead found to impose a distorting screen upon lived experience. This arbitrary barrier cut off a worrying proportion and population of subjects from the ranks of selves and also narrowed the scope of self-formation for those who it included.

Pursing this critique led me to search for an alternative account of openness and representation, which tracked the same aspects of lived experience that narrative theory identified as important but offered a more discerning schema for organising them. This was provided by Camp's character account, which, when fleshed out in terms of the question regarding the relationship between experience and desire around which I oriented my discussion, provided a stable basis for accounts of self-formation.

In order to reach this conclusion, more also needed to be said regarding the relationship between internal and external sources of motivation and the self. By considering the problematisation of the notion of 'character' by situationist scepticism, I moved towards a conceptual vocabulary which favoured an agent's *extendedness* rather than just their *characteristics* as constituent elements of their self.

Making this move then allowed me to expand on the basis I derived from the critique of narrative theory by questioning the nature of the extendedness of desire and its relationship with the subject of experience. Using Susan Wolf's account of meaningful lives as a paradigm, I first established an intersubjective notion of desire and how this and other motivational factors arise as significant for an agent within their life. Following this, I then considered two separate facets of this phenomenon, each connected to a different theory of relationality.

The clearest connection between the phenomenon of openness to desire and the constitution of the self that I drew from narrative theory arose in Katharine Wolfe's care-based relational account of self-formation. Providing a view on affect that incorporated not only an agent's internal resources but also connected them to the world of lived experience in the

intersubjective way that Susan Wolf and my response to situationism claimed was necessary, this view established the necessity of a thick concept of the self both for making sense of one's environment and one's particular responses to it.

A more subtle but equally important connection between the self and an openness to a horizon of care was established by an investigation of Classical Confucian relational theory. Here, the necessity of a limit on self-formation was most clearly expressed. Defining the elements of lived experience which require cultivation in order for an agent to not only meet their desires but to also appreciate them meant establishing a framework in which the subjectively and humanely possible were appropriately expressed.

An additional benefit to Confucian relational theories was the response they offered to the modernist/postmodernist divide in Western philosophy, which also addressed elements of the disagreement between Locke and Hume. As my earlier revision of the notion of character showed, some degree of the extension of the self into the world is necessary to make its desires coherent. By establishing patterns of behaviour, such as rituals, as central to self-formation, Confucianism goes some way to establishing this form of openness and providing a model for how it can be regulated so that agents can organise their desires coherently and interact with them intelligibly.

This led to my final move in attempting to fix the role that the representation of desire and openness to it and the world it permeates play regarding subjects of experience. If the self is to be measured by its relations and the motivations which these generate and cultivate, then the questions of how these might be lost and what their loss entails arise.

I answered these by first constructing a positive limit on openness to a horizon of care in terms of relational threads, which describe the minimal capacity to relate to their lived experience that subject must have in order to constitute a self from what they live. Having established this side of the boundary of self-formation, I then considered how extreme forms

of trauma, like torture, undermine even this basic ability to relate and neutralise the role that desire plays in making sense of lived experience.

The result of defining the limits of self-formation in these terms was that openness to care cannot be done away with without also doing away with a thick notion of the self. Even persons whose lives are dominated by extreme forms of care, such as the moral saint or Collector, still enjoy this capacity and, hence, engage in a process of self-formation. Identifying this further helped vindicate the problems I expressed with positions like narrative theory, in which particular, ideal *forms* of care were taken to be central, and to vindicate Camp's and Lumsden's views regarding the status of mentally ill persons as selves.

The concept I developed which captures this capacity is that of a 'normative injunction'. In Susan Wolf's work, a similar kind of notion was used to analyse and express the appropriate responses to meaningless lives. When confronted by someone like a value fanatic, persons typically feel the need to account for why they feel that life is being 'wasted' or misdirected. The reason for this is that in these cases there seem to be other, important sources of meaning that the agent in question is ignoring in pursuing and cultivating their particular kind of self-formation. That is, there is at least one other important thing that, from the observer's perspective, they ought to care about but fail to.

What my work establishing the extendedness and limits of self-formation has aimed to show is that a precondition of appreciating, let alone voicing these concerns is the presence of a thick self at the centre of the agent's life, which represents, organises, and expresses their own cares. Like the nexus of character traits Camp describes, this connects beliefs, desires, and intentions drawn from a subject's experiences and environment to allow them to project themselves into the world in terms of which of these elements of their life they are motivated to care about.

In this sense, while the notion of a normative injunction expresses the centrality of an openness to an appropriate horizon of care for agents, it does not by itself constitute that horizon. Rather, it is the relations that the injunction is constructed in response to and how *these* represent and interact with personal beliefs, desires, and intentions, which constitute the self who questions both their representation of own experiences and those of others.

This, I believe, captures Ricœur's concern with seeing 'one*self* as another,' in that a precondition of caring about and being open to the world in this way is the capacity to consider one's own relations and level such injunctions against oneself. In turn, this illuminates the problem that I began my project with: whether there is something that can be the subject of a forensic inquiry into an individual's behaviour.

This also captures and resolves precisely the worry that Chisholm expressed in rejecting the impression-based view of the self. What an interpreter of lived experience would find an impression of in a life can only make sense in terms of a relationship between a living subject and the world, a relationship that accesses a thick notion of care which is constitutive of a thick notion of the self. One need not go so far as to endorse a Cartesian ego at the heart of human life and subjectivity, however, despite Hume's misgivings, there does appear to be something, or more specifically some*one* actually living that life and experiencing it subjectively.

There does, then, seem to be a live, meaningful self, similar to what Locke described, at the heart of human life. At the very least, this plays a role like a centre of gravity in making experiences hang together coherently for their subject. However, for theories which attempt to fix a more substantive role for this locus of agency in human life, a refined account of how it relates to the world both as a care giver and receiver, who is open to both new experiences that broaden their horizon and harms which narrow or even risk destroying it, is needed.

What I hope to have demonstrated here is how precisely such a theory ought to be constituted. By identifying the untenable limits of narrative form, the problems that arise with

purely internal and external accounts of self-formation, and the centrality of relational threads of meaning to making sense of a human life, I have made a case for an intersubjective, care-based theory of the self. This position is supported by the actual relations that express beliefs, desires, and intentions, which obtain between subjects and not only makes sense of but describes lived experience in terms of the set of normative injunctions that can be plausible levelled against agents.

Ultimately, I believe that it is this which provides the best answer to the question of self-formation which I posed at the beginning of this project. If there is a *thing that lives* at the heart of human life – and there are good reasons for suspecting that there might be – then it must explain the lived experiences which constitute that life, or fail to do so in cases of extreme trauma, in terms of a relational self, which is constituted within a horizon of care.

Bibliography

Alfano, M 2013, *Character as moral fiction*, Cambridge University Press, Cambridge.

Aristotle, 2000, *The Nichomachean ethics. Book II*, trans. W D Ross, Raleigh, N.C., Alex Catalogue.

Camp, E 2011, ‘Wordsworth’s prelude, poetic autobiography, and narrative constructions of the self,’ *Nonsite*, viewed 19 February, 2021, https://nonsite.org/wordsworths-prelude-poetic-autobiography-and-narrative-constructions-of-the-self/

Camp, E 2009, ‘Two varieties of literary imagination,’ *Midwest Studies in Philosophy*, vol. 33, no. 1, pp. 107-130.

Camp, E 2013, ‘Slurring perspectives,’ *Analytic Philosophy*, vol. 54, no. 3, pp. 330-349.

Camp, E 2017, ‘Perspectives in imaginative engagement with fiction,’ *Philosophical Perspectives*, vol. 31, no. 1, pp. 73-102.

Chisholm, R M 2013, *Person and object: a metaphysical study* Routledge.

Darley J, and Batson C D 1973, ‘“From Jerusalem to Jericho”: a study of situational and dispositional variables in helping behaviour,’ *Journal of Personality and Social Psychology*, vol. 27, no. 1, pp. 100-108.

Dennett, D C 1992, 'The self as a center of narrative gravity,' in F S Kessel, P M Cole, and D L Johnson, (eds.) *Self and Consciousness: Multiple Perspectives*, Lawrence Erlbaum, pp. 103-115.

Fiorello, A 2019, 'Free will, values, and narrative selfhood,' *Philosophia*, vol. 44, no. 1, pp. 1-20.

Gallagher, S 2000, 'Philosophical conceptions of the self: implications for cognitive science,' *Trends in Cognitive Science*, vol. 4, no. 1, pp. 14-21.

Hardin, R 1995, *One for all: The logic of group conflict*, Princeton University Press, Princeton.

Harman, G 2009, 'Skepticism about character traits,' *The Journal of Ethics*, vol. 13, no. 2/3, pp. 235-242.

Harman, G 2000, 'The nonexistence of character traits,' *Proceedings of the Aristotelian Society*, vol. 100 no. 2, pp. 223-226.

Harman, G 1999, 'Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error,' *Proceedings of the Aristotelian Society*, vol. 99, pp. 315-331.

Howell, R 2016, 'Extended virtues and the boundaries of persons," *Journal of the American Philosophical Association*, vol. 2, no.1, pp. 146-163.

Hume, D 1739-1740/1878, *A treatise on human nature being an attempt to introduce the experimental method of reasoning into moral subjects and dialogues concerning natural religion*, Longmans Green, London.

Ivanhoe, P J & Van Norden, B W (eds.) *Readings in Classical Chinese Philosophy*. 2nd ed. Hackett Pub, Indianapolis, 2005.

Klein, S 2018, 'Remembering with and without memory: a theory of memory and aspects of mind that enable experience,' *Psychology and Consciousness: Theory, Practice and Research*, vol. 5, no. 1, pp. 117-130.

Li, C 2010, 'The Confucian concept of REN and the feminist ethics of care: a comparative study,' in Bell, D A (ed.) *Confucian Political Ethics*, Princeton University Press, Princeton, pp. 175-198.

Locke, J 1690/1997, *An Essay Concerning Human Understanding*, R. Woolhouse (ed.), Penguin Books, Harmondsworth.

Lumsden, D 2013a, 'Whole life narratives and the self,' *Philosophy, Psychiatry & Psychology*, vol. 20, no.1 pp. 1-10.

Lumsden, D 2013b, 'Narrative construction of the self,' *Te Reo*, 56/57, no.56, pp. 161-175.

MacIntyre, A 2007, *After virtue: a study in moral theory*, 3rd edn, University of Notre Dame Press.

Mao, X 2020, 'The three faces of vulnerability,' *Angelaki: The Journal of Theoretical Humanities*, vol. 25, no. 1-2, pp. 209-221.

Miller, S 2020, 'From vulnerability to precariousness: examining the moral foundations of care ethics,' *International Journal of Philosophical Studies; IJPS*, vol. 28, no. 5, pp. 644-661.

Mower, D 2013, 'Situationism and Confucian virtue ethics,' *Ethical Theory and Moral Practice*, vol. 16, no. 1, pp. 113-137.

Nguyen, L 2007, 'The question of survival: the death of desire and the weight of life,' *The American Journal of Psychoanalysis*, vol. 67, no. 1, pp. 53-67.

Paul, L A 2014, *Transformative experience*, Oxford University Press, Oxford.

Ricœur, P 1992, *Oneself as another*, trans. K. Blamey, University of Chicago Press, Chicago.

Ross, M 1989, 'Relation of implicit theories to the construction of personal histories,' *Psychological Review*, vol. 96, no. 2, pp. 341-357.

Rudd, A 2009, 'In defence of narrative,' *European Journal of Philosophy*, vol. 17, no. 1, pp. 60-75.

Sacks, O 1986, *The man who mistook his wife for a hat*, Picador, London.

Schechtman, M 2015, '"The size of the self": minimalist selves and narrative self-constitution,' in A Speight (ed.) *Narrative, philosophy, and life*, vol. 2, Springer Dordrecht, pp. 38-48.

Simola, S 2010, 'Anti-corporate anger as a form of care-based moral agency,' *Journal of Business Ethics*, vol. 94, no. S2, pp. 255-269.

Singer, P 2012, *Practical Ethics*, 3rd edn, Cambridge University Press, Cambridge.

Strawson, G 2004, 'Against narrativity,' *Ratio (Oxford)*, vol. 17, no. 4, pp. 428-452.

Strawson, G 2016, 'Narrative bypassing,' *Journal of Consciousness Studies*, vol. 23, no. 1-2, pp. 125-139.

Strawson, P 2008, 'Social morality and individual ideal,' in P Strawson, *Freedom and resentment and other essays*, 2nd edn, Routledge, pp. 29-49.

Thompson, K 2017, 'Relational self in classical Confucianism: lessons from Confucius' Analects,' *Philosophy East & West*, vol. 67, no. 3, pp. 887-907.

Velleman, D 1996, 'Self to self,' *The Philosophical Review*, vol. 105, no. 1, pp. 39-76.

Wolf, S 2014a, 'Moral saints,' in W Susan, *The variety of values: essays on morality, meaning, and love*, Oxford Scholarship Online, pp. 12-30.

Wolf, S 2014b, ‘The meanings of lives,’ in W Susan, *The variety of values: essays on morality, meaning, and love*, Oxford Scholarship Online, pp. 90-107.

Wolf, S 2014c, ‘Happiness and meaning,” in W Susan, *The variety of values: essays on morality, meaning, and love*, Oxford Scholarship Online, 2014, pp. 108-127.

Wolfe, K 2016, ‘Together in need: relational selfhood, vulnerability to harm, and enriching attachments,’ *The Southern Journal of Philosophy*, vol 54, no.1, pp. 129-148.

Wolfe, K 2017, ‘Are second person needs ‘burdened virtues’?: exploring the risks and rewards of caring,’ *Feminist Philosophy Quarterly*, vol. 3, no. 3, pp. 1-22.

Yi, H 2020, ‘Building narrative identity; episodic value and its identity-forming structure within personal and social contexts,’ *Human Affairs*, vol. 30, no. 2, pp. 281-292.