© Copyright by

Cristian Floring Mihut

2009

All rights reserved
DIACHRONIC AGENCY AND NARRATIVE UNDERSTANDING

Abstract

by

Cristian F. Mihut

In this dissertation I develop an account of diachronic agency that involves narrative understanding. This mode of practical reasoning emerges from the complex interaction between emotional cadences, relevant social contexts and the motivational scripts of agents who lay claims upon us. Narrative understanding is (a) essentially retrospective, (b) couched implicitly in the cadences of our emotional life, and (c) dependent on occupying stances of different dramatic personae. Narrative understanding generates distinctive reasons for action and is irreducible to practical rationalities that aim to integrate agency synchronically or diachronically. Intellectualist forms of practical reason, namely those involving universalizing, inductive generalizing or the creation of increasingly coherent desiderative profiles, do not explain how historical self-understanding engenders reasons for action, and presuppose its proper functioning.

Volition-based accounts of practical reason involve processes that structure agency diachronically. Some argue that the heart of agency is constituted by stable planning attitudes and related self-governing policies. But central cases of personal transformation over time do not sit well with this picture. Rather than the center of
agency, plans and policies should be understood as mere expressions of deeper commitments to persons. These commitments and the temporal transformation they entail are explained by narrative understanding in ways they could not be by reference to the policies that codify them. Other volition theorists hold that our deepest commitments are traced by our diachronic loves. I point out two main limitations for an influential view of volitional love. I suggest that paradigmatic features of narrative understanding are essential for overcoming these limitations.

Narrative understanding opens up the fruitful possibility of conceptualizing agency as a chorus of dramatic personae. While each persona can be conceived as a stable character that expresses an enduring emotional cadence, the agent sometimes simultaneously occupies different and even jarring narrative stances. Practical deliberation then implies inhabiting different dramatic personae in mutual dialogue. Furthermore, essentially retrospective phenomena such as personal forgiveness (and atonement) not only accommodate, but they seem to require a model of dramatic-narrative agency.
# CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ iv

CHAPTER 1 SYNCHRONIC AND DIACHRONIC AGENCY ............................................... 1
  1.1. The “Received Picture” ..................................................................................... 1
  1.2. Intellectualist Accounts of Synchronic Unity .................................................. 8
    1.2.1. Korsgaard, Reflection and Universality ..................................................... 8
    1.2.2. Smith and Desiderative Coherence ............................................................ 13
    1.2.3. Millgram and Practical Induction ............................................................... 18
  1.3. Resources for Narrative Understanding .......................................................... 26
    1.3.1. Narrative and Personal Identity ................................................................. 27
    1.3.2. Narratives and Moral Sensibility ............................................................... 28
    1.3.3. Narrative, Intelligibility, and Accountability ............................................... 29
    1.3.4. Narrative Explanation and Motivation ....................................................... 31
    1.3.5. Narrative and Moral Resistance ................................................................. 31
  1.4. A Quick Preview ............................................................................................... 32

CHAPTER 2 PLANS, COMMITMENTS AND SELF-UNDERSTANDING ..................... 36
  2.1. Michael Bratman and Agency ......................................................................... 36
  2.2. Refinements and Minor Criticisms ................................................................... 42
    2.2.1. Justificatory Regress Problem ................................................................... 42
    2.2.2. Lieberman and Commitment ..................................................................... 44
  2.3. Relational Commitments and Policies ............................................................. 46
  2.4. Self-understanding and Self-government ....................................................... 59
    2.4.1. Reflectiveness and Self-management ......................................................... 60
    2.4.2. Huck Finn and the Priority of Emotional Self-understanding .................. 69
  2.5. Retrospection and Self-Understanding ........................................................... 77
    2.5.1. Redemptive Events and Retrospection ....................................................... 78
    2.5.2. Emotions, Retrospection and Narrative ..................................................... 81
  2.6. Leonard: Planning Agent, Narrative Failure .................................................... 84

CHAPTER 3 LOVE AND DIACHRONIC SELFHOOD ................................................. 87
  3.1. Intuitive Constraints on Loving Agency ............................................................. 87
  3.2. Frankfurt and Volitional Love .......................................................................... 88
  3.3. The Diachronic Requirement and Volitional Love .......................................... 94
    3.3.1. Volitional Necessity and Limits of Agency ............................................... 94
      3.3.1.1. The Central, The Peripheral, and Transformation ........................... 97
      3.3.1.2. Restrictions on Structural Change .................................................. 99
      3.3.1.3. Fragmentation and Compartmentalization .................................. 100
    3.3.2. Caring and Loving in Diachronic Perspective .......................................... 106
      3.3.2.1. Caring as Diachronic Investment: The Prospective Mode ............. 110
      3.3.2.2. Caring as Diachronic Investment: The Retrospective Mode ......... 114
### CHAPTER 3 LOVE AS INVESTMENT AND VOLTIONAL SOLIPSISM

3.3.2.3. The Extended Frankfurt Account .......................................................... 117
3.3.2.4. A Final Reservation .............................................................................. 119
3.4. Love as Investment and Volitional Solipsism ................................................ 121
3.4.1. Iris Murdoch and Loving Attention .............................................................. 122
3.4.2. Characterizing Volitional Solipsism ............................................................. 126
3.4.3. Illustrating Volitional Solipsism ................................................................... 129
3.5. Concluding Unscientific Parable ................................................................. 134

### CHAPTER 4 NARRATIVE UNDERSTANDING AND THE DRAMATIC SELF

4.1. Preliminary Characterization ............................................................................ 137
4.1.1. Sub-personal narratives ................................................................................ 139
4.1.2. Narrative at the Personal Level ..................................................................... 146
4.2. MacIntyre and Intelligible Action ................................................................. 148
4.2.1. The First Stretch: Intention and Narrative .................................................... 150
4.2.2. First Objection: Of Life and Narrative ......................................................... 158
4.2.3. The Second Objection: Falsifying Life ........................................................ 161
4.3. David Velleman and Narrative Explanation ..................................................... 166
4.3.1. The Self and Its Guises ................................................................................. 166
4.3.2. Narrative and Causal Explanations ............................................................. 172
4.3.3. Cognitive and Emotional Coherence .......................................................... 178
4.3.4. Emotions as Quasi-perceptual Paradigm Scenarios ...................................... 182
4.4. Paradigm Scenarios And Dramatic Agency ..................................................... 186
4.4.1. Emotions as Interpretative Frameworks ....................................................... 186
4.4.2. Characterization of Dramatic-narrative agency ............................................ 187
4.4.3. Illustration of Dramatic-narrative Agency ................................................... 189
4.4.4. On Harmonies and Their Costs ................................................................. 192

### CHAPTER 5 PERSONAL FORGIVENESS AND DRAMATIC AGENCY

5.1. Map and Background Assumptions ............................................................... 197
5.1.1. Untidiness of Personal Forgiveness ......................................................... 199
5.1.2. The “Paradox” of Forgiveness ................................................................. 201
5.2. Empathetic Model of Aspirational Forgiveness .............................................. 202
5.3. Empathy, Opaqueness and Dramatic Agency ............................................... 208
5.3.1. Illustrating the Opaqueness Challenge ....................................................... 209
5.3.2. Answering the Opaqueness Challenge ....................................................... 213
5.4. Resentment and Dramatic Agency ................................................................. 217
5.4.1. The Depth and Breadth of Resentment ...................................................... 217
5.4.2. Butler’s Taxonomy ..................................................................................... 220
5.4.3. Amalgamation and Entrenchment ............................................................. 221
5.4.4. Resentment mixed with Pride ................................................................. 223
5.4.5. Resentment Entrenchment and Empathy ................................................... 225
5.4.6. Resentment mixed with Fear ..................................................................... 228
5.4.7. The Explanatory Power of Dramatic Model .............................................. 231
5.5. Prospects For Dramatic-narrative Agency .................................................... 237

### BIBLIOGRAPHY

.................................................................................................................................. 240
ACKNOWLEDGMENTS

For philosophical counsel and inspiration I’ve relied on many over the years, but especially Luke Potter, Tom Crisp, Thad Botham, J.R. Shrader, Christian Miller, and Donald Smith. Tim Gloege and Raully Donahue have sharpened my understanding of historical rationality, and this with grace, wit and humor. My colleagues Chad Meister, Tim Erdel, and Jim Stump have each in their own way been consistently supportive at all stages of this project. Brother Tim in particular has been a tremendous conversation partner, an insightful commentator on many drafts of my work, a resourceful guide to all things narrative, and a friend. I am thankful to John-Paul who keeps sharpening my ethical intuitions and cultivating my aesthetic sensibilities.

I am grateful to Karl Ameriks, Paul Weithman, and Cornelius Delaney for being gracious and helpful readers of this dissertation. In the latter stages of my work I have benefited enormously from Mike Rea’s careful reading, generous comments on every chapter, and unremitting encouragement. I owe a huge debt of gratitude to David Solomon, my advisor. I am thankful for his seminars that nurtured my thinking about the diachronic aspects of practical reasoning. He did not give up on me, nor did he despair even as the years of working on this project were piling up. His probing questions have always been more than helpful in refining my arguments. His philosophical insight has been highly valuable in chiseling my scrambled and amorphous intuitions.
I’ve relied on no one else more than on my wife Andrea. In so many different ways this work has only been possible because of her. She has been my most incisive and most gentle critic, her insights being present in every single chapter. As my practical outlook has been shaped by her wit, intelligence and humor in ways deeper and more enduring than I could ever credit, my work feels more like co-operation. And even though dissertation work was at times a demanding mistress, with patience, grace and humility Andrea has taught me much about diachronic love.
CHAPTER 1
SYNCHRONIC AND DIACHRONIC AGENCY

1.1. The “Received Picture”

In book IV of The Republic Plato has Socrates propose that worthwhile human action depends on bringing inner psychological forces into a certain unity or harmony:

And in truth, justice is, it seems, something of this sort. Yet it is not concerned with someone’s doing his job on the outside. On the contrary, it is concerned with what is inside; with himself, really, and the things that are his own. It means he does not allow the elements in him each to do the job of some other, or the three sorts of elements in his soul to meddle with one another. Instead, he regulates well what is really his own, rules himself, puts himself in order, becomes his own friend, and harmonizes the three elements together, just as if they were literally the three defining notes of an octave – lowest, highest, and middle – as well as any others that may be in between. He binds together all of these and from having been many, becomes entirely one, temperate and harmonious. Then and only then should he turn to action…In all these areas, he considers and calls just and fine the action that preserves this inner harmony and helps achieve it, and wisdom the knowledge that oversees such action; and he considers and calls unjust any action that destroys this harmony, and ignorance the belief that oversees it.¹

This Socratic (or Platonic) picture of agency seems assembled from a number of loosely associated fragments. One such fragment is the assumption that only forces and capacities internal to the psychological make-up of an agent are relevant to just and fine action. Another is that this psychological make-up can be analyzed into its synchronic components or capacities. Our psychological experiences can be atomized and itemized.

Another piece yet is the idea that the diverse forces or capacities must be brought into a kind of synchronic unity in the agent, analogous to the way musical dissonance must be turned into an arpeggio, if it is to sound good in the ears of an audience. An essential part of Plato’s picture, of course, is that wisdom (or practical reason) is and ought to be the capacity responsible for analyzing, rank-ordering, and integrating psychological attitudes or capacities. Finally, for our purposes, it is important to observe the conceptual connection between psychological integration on the one hand and desirable or just action on the other. In particular, it is the former that explains and justifies the latter.

Important parts of this picture, it seems to me, still frame many of the debates in the contemporary moral psychology. We might not be too far from truth to call this theoretical package the received view of practical agency. Some, even while disavowing Platonic moral psychology and theorizing, implicitly accept assumptions informed by this picture. Consider for instance Bernard Williams’ provocative and influential attack on the idea of external reasons\(^2\). According to Williams, whatever reasons an agent has for action they must be internal to her motivational set.\(^3\) The items populating a motivational set are “desires”, label applying democratically to dispositions of evaluation, patterns of emotional reaction, loyalties, as well as deep commitments. Williams claims that a subjective motivational set can be expanded, that is new desires can be created by

---


\(^3\) I realize that there are all sorts of ways to spell out the “internalist” position. See Robert Audi’s impressive taxonomy in his “Moral Judgment and Reasons for Action” in Cullity, Garrett (eds.) *Ethics and Practical Reasoning*, (Claredon Press, 1997). The kind of internalism that Christine Korsgaard and Michael Smith seem to agree on is a version of judgment internalism: if an agent judges that it is right (good) to act, then she is must be motivated to some extent to act, or else she is practically irrational. I do not mean to suggest that contemporary internalists self-consciously operate in the Platonic tradition. Rather, Plato’s manifest concern of grounding normative reasons in the psychology of agents seems still fashionable. And there might be good reasons for this trend.
practical reason in several of its modes: means-end reasoning, combining existing motivations, and the work of imagination. Williams seems to hold that in addition to these functions practical reason also plays a subtle role in unifying agency. Distinguishing himself from Kantians and Utilitarians to whom he attributes the belief that practical reason abstracts from particular persons and relations, Williams emphasizes character integrity understood as a diachronic loyalty to one’s identity-shaping ground projects. Although he rejects the aristocratic role of reason in attaining synchronic unity, and doubts the conceptual connection between synchronic unity and moral action, Williams still preserves the idea that integration around one’s basic desires and commitments is necessary for valuable or meaningful action.

Williams’ attenuated move is neither typical nor widely accepted. In fact it seems to me that influential contemporary theorists implicitly agree with Plato’s assumption that practical reasoning must essentially unify the agent synchronically, and that rational, desirable or moral action flows only from a suitably unified motivational profile. Synchronic unification comes today in different flavors. Some see the work of practical reason as a way of bringing inner consistency among different practical identities that we implicitly endorse when we act. On this view, the reflective distance from impulses or contingent identities results in synchronic unification, for it essentially allows the agent to endorse a more general conception of oneself as a human will, a conception falling under universal moral laws. For others practical reason fosters synchronic unity by way of

---


5 I take it that the canonical Plato exhibits what I called “the aristocratic” function given the control and governance reason has over the other psychological forces.

systematic justification of desires into larger, more coherent desiderative profiles. Some prefer still to speak of synchronically unified agency as the end result of practical induction, where analogous to the work of theoretical induction reason assembles from experience a grab-bag of practical judgments to be deployed in new circumstances. In a moment I turn to a brief examination of these options, all of which seem to have in common the predilection toward an intellectualist mode of practical reasoning.

Alternative accounts of integration avoid intellectualist forms of practical reasoning. Two paradigmatic attempts are my focus in chapters two and three respectively. Practical reasoning here seems less intellectualistic because agential unity depends mainly on a will whose integration is facilitated without recourse to reflective stances that imply universalizing, generalizing, or bringing about increasingly coherent desiderative profiles. On one such view volitional integration is the work of self-governing policies that involve practical planning; the other attributes synchronic and diachronic unity mainly to our enduring loving attitudes. Roughly, reason works toward diachronic unity either by filling out the details of an earlier plan, or by helping the agent accept the authority of her historical loves. Synchronically, this means that endorsing an action or identifying with a desire no longer has to involve strict theoretical consistency among self-conceptions or the expression of a will that falls under universal moral laws or following the advice of a more rationally integrated desiderative self. Rather, as we will see, these synchronic acts have subjective existential authority for the agent, for they fix and reveal what she really wants.

---

For the most part intellectualistic accounts of practical reason prioritize the synchronic unity of agency while volitional modes concern both synchronic and diachronic integration. My main aim in this project is to motivate the relevance and distinctiveness of narrative understanding for diachronic agency. A very quick way to broach the distinctiveness of narrative is to highlight the importance of phenomena involving retrospection: describing one’s college years, making amends for a wrongdoing, forgiving a past wrong, commemorations, changes of mind, repenting of past offenses, and conversions. If retrospection on one’s own idiosyncratic history essentially involves narrative understanding, and if the former can generate reasons for action, narrative understanding turns out to be a distinctive form of practical reasoning.

Narrative understanding is not equivalent to giving oneself the universal moral law, or to systematically justifying one’s desires by appeal to a rational advisor, or to planning or to volitional investment. In the process of telling, retelling or rehearsing particular personal stories or fragments thereof, narrative understanding can manifestly generate reasons for action. In addition such reflection is often taken for granted in the way particular social and psychological scripts underwrite our reasons for action. As I argue in the sequel, narrative understanding helps make diachronic sense of our lives, it explains plausibly how we can be committed to a variety of personal relations, and can ground decision and planning.

Contrary to one of the fragments from the received picture of agency, I hold that narrative understanding straddles the internal/external divide. Practical reason in this mode is already placed in a world external to the agent, for it not only responds to but is partially constituted by practical realities encoded in social stories and dialogue. At the
same time, narrative understanding allows agency to stretch outside the boundaries of one’s skin due to the existence of psychological states with appropriate functions. These functions explain the temporal continuity and stability of an agent’s intentions, the transformations agents undergo over time, and relevantly, the sensitivity to the extensional reality of other agents and to stories external to oneself. In chapter four I suggest that the attitudes responsible for the work of narrative understanding are the emotions.

Focusing on the relationship between narrative understanding and the emotions helps throw suspicion on the claims to exclusivity the other fragments from the Platonic picture seem to possess. In the previous paragraph I have suggested that we have reason to resist the internality assumption. But we should also emphasize that narrative promises a mode of understanding that moves in the opposite direction from atomization and itemization. As we shall see, at least some events or psychological states are intelligible when embedded in larger temporal or historical sequences in ways in which they would not be un-contextualized. Furthermore if narrative understanding is partially underwritten by emotions, we can expect practical reason to work in ways more subtle than those involving governance over capacities, desires or inclinations. A plausible outcome here is that an account of narrative understanding is consistent with rejecting an outmoded hierarchical faculty-psychology as well as any psychological model that implicitly assumes an essential combat between reason and emotions.9

9 I do not believe that all combat or internality has to be vanquished on my positive account. Rather, while some psychological items can and should be kept irrelevant to action, they are so in virtue of their contents, but not in virtue of the types of attitudes that they are. My thought here resists the suggestion that certain recalcitrant attitudes should be disqualified as worthless merely in light of belonging to certain faculties or merely because they tend to destabilize agency.
Finally, focusing on the connection between narrative understanding and the emotions helps question the central assumption that synchronic unity is necessary for meaningful human action.\(^{10}\) I aim to show not only that a fragmented agent could act meaningfully and justly, but that *pace* Plato sometimes meaningful action requires polyphonic and discordant agential stances. Narrative understanding can account for temporal persistence of some intentional sequences. But the agent must also be capable of resisting or questioning or distancing herself from any of these. I suggested above that one way she could do this is by listening to multiple, even jarring perspectives. The combined effect of the last three claims is that we need a model of agency that explains both the continuities and discontinuities inherent in meaningful action. In chapter four I propose a dramatic-narrative model that seems to accomplish this feat.

Before I move on to the intellectualistic views of synchronic unity, let me mention that I do not aim to dismantle the Platonic picture.\(^{11}\) I only mean to highlight valuable modes of practical reasoning can be traced outside its boundaries.

\(^{10}\) I use this umbrella term to cover moral, rational, desirable, and subjectively authoritative action.

\(^{11}\) Even though I call this the “Platonic Picture” I am inclined actually to think that Plato did not endorse it. One clue is that this account of psychological integration is the logical extension of a hypothetical concession to Glaucos in Book II of the *Republic*, where the latter is unsatisfied with the “city of pigs” Socrates had been describing. Another clue is that the model upon which Plato constructs psychological harmony is an immensely terrifying aristocracy. In the beautiful and noble city personal love is stunted, while eugenics, propaganda, political deception, censorship, and social engineering are musts. These absurd implications of perfect social justice together with a background that abounds with plentiful references to the jokes of Aristophanes have driven some to consider seriously the option the Republic is interlaced with jokes. One joke is of course that plausible folk beliefs have often absurd consequences. Another joke might be that philosophical positions that attempt to explicate folkish beliefs have absurd consequences. One does not have to hold that the whole of the Republic is an extended joke, though in the same way one shouldn’t underestimate the extent of Plato’s subtlety. Minimally, whatever Plato’s positive position, it catches shape after the series of jokes are seen to express *reductios* for false leads. It is hard to see how else to explain these and other perplexing features of the *Republic*. 
1.2. Intellectualist Accounts of Synchronic Unity

1.2.1. Korsgaard, Reflection and Universality

Responding to a remark by Williams, Christine Korsgaard explicitly claims her account of practical reason and agential unity is indebted to Plato’s manifest picture. Korsgaard holds in effect that without moral reflection that involves a claim to universality the agent has no chance at unification, at having a will. Reason possesses the prerogative to unify agency, to generate the will, due to its capacity for making universal laws:

Williams is exactly right when he says that he sees a Platonic inheritance in my view. For Plato too thinks that moral principles serve to hold the disparate parts of the human soul together, and in this way make the soul capable of unified and effective action. Moral principles, to put it in non-Platonic language, are what give the soul considered as a unified entity a will. The arguments first sketched in Republic I, to the effect that an unjust soul will fall into faction, and therefore will be incapable of acting effectively as a unit, say exactly this…With Plato, I believe that neither human souls nor human communities can be held together, can be unified, and so can really be unless they are (at least to some extent) Republics, submitting themselves to the rule of law. And that is why I think that freedom and autonomy require that we will in accordance with universal law.12

Placing aside the issue of the correct exegesis of Plato, it seems to me that Korsgaard makes at least three assumptions here. First, an unharmonious self is incapable of meaningful action. So, effective, meaningful human action presupposes a certain level of synchronic integration. This is a general assumption she shares with many others. Secondly, she holds more distinctively that moral principles must be responsible for the relevant integration of agency. The final assumption is that these principles must have the mark of universality. Thus, the relevant unity of agency, at

---

least with respect to moral action, is the one attained only when practical reason
generates a will that conforms to universal moral laws.

By the time these claims arrive, Korsgaard had already told us that in order to act
at all one has to act under a certain conception of practical identity (as a mother,
employee, parishioner, etc). But since any particular practical identity is contingent, the
only conception of oneself as a rational human self survives the fire of synchronic
reflection. The structure of practical rationality is such that once embarked in the process
of reflection on one’s desires or practical identities, one cannot stop short of endorsing
humanity as a formal universal conception for action. I cannot be rational and stop in the
process of reflection before implicitly endorsing my humanity, nor can I, after reflecting
on it, fail to endorse it. This is simply a matter of consistency.13

This is not a full-scale critical investigation of Korsgaard’s framework. Many
others have already done this. The question relevant to my project is how Korsgaard
accommodates the intuition that we are also temporally embedded agents, with changing
projects, commitments, self-conception, and desires. Does she have a story about
diachronic transformations? The best sense I can make of Korsgaard’s position is that
temporal transformations are intelligible only if the agent has an active will that survives

---

13 There are many questions here over which I simply have to fly. One might want to know for
instance why a reflective agent could neither stop the process of reflection before endorsing one’s universal
humanity as a reason for action, nor carry on the process of reflection on one’s humanity without endorsing
it. Korsgaard couldn’t mean that the rule constitutive of our human agential structure and driving us to
reflection is one characterize by logical necessity. Moreover, it seems that if it were a matter of
psychological constitution, there could be cases when such reflection would be pathological. So, surely
Korsgaard must mean the built-in rule of continuing reflection is normatively necessary for beings like us.
But if this is right, it is unclear why the process of reflection and endorsement of identities, has to stop with
the endorsing one’s humanity, understood as formal, universal condition of action. If self-reflection is
normative for our moral self-constitution, one could, after reflecting on one’s human identity, deliberately
fail to endorse the categorical imperative or to value one’s humanity. Some might, in light of certain
metaphysical assumptions about human nature, even have good reasons to make naked jest at the moral
law, be ironic toward it, or willfully transgress it. For this latter possibility see Raymond Geuss’s insightful
temptations or the flux of passive inclinations. The emphasis remains on the stability and 
endurance fixed by the synchronic moment of reflective endorsement. For it is this 
reflective moment in its generality and universality that generates an active will, a will 
allowing the agent to distinguish herself from undifferentiated, passive desires. But what 
is the justification for these claims?

First, Korsgaard explains that agency requires reflective distance from concrete 
impulses, and this in turn entails acting under the idea of freedom. But being free with 
respect to a particular desire means I could have acted differently on the occasion I am 
tempted to act on this desire. And this implies that reflective distance must conceive of 
this desire or of a particular practical identity in a more general way. Without the 
reflection that allows me to conceive of these ineluctable particulars under more general 
descriptions, I am a simple passive will, a pitiful stage for the ebb and flow of desires and 
impulses. This step can of course be resisted. Merely describing a particular desire as an 
instance of a more general type is insufficient to get me to resist the inclination on a new 
occation. At the same time, my will could be activated without conceiving of the 
tempting desire in a more general way. Suppose I am tempted to sacrifice working in 
favor of sleeping on a particular night. I do not have to think to myself “this is only a 
desire for sleep, you know you can resist it” in order to resist it. I am active in resisting it 
if I am moved by a very specific fear of leaving work that needs to be done, unfinished. 
Here, it seems another ineluctably particular situation gives me enough reflective distance 
to activate my will without the reflective outlook that requires me to scrutinize the desire 
for sleep from a general perspective.14

14 There is an additional worry which I do not develop here. It is not clear how Korsgaard’s discourse 
regarding the active will’s resisting the temptation of acting on impulses maps with respect to the earlier
The more significant worry is Korsgaard’s transitioning from generality to universality. The most her arguments show is that practical reason could activate a will with a claim to generality. But generality does not entail universality. And yet she claims:

If I give myself a law, if I am not merely the place where an impulse is operating, then what I do essentially involves a reference to other occasions when I might do otherwise – or, to pick up the point in the previous paragraph—to this occasion, regarded as possibly other, and so regarded in general terms. And that means that if I am to regard this act, the one I do now, as the act of my will, I must at least make a claim to universality, a claim that the reason for which I act now will be valid on other occasions, or on occasions of this type— including this one, conceived in a general way. Again, the form of the act of the will is general. The claim to generality, to universality, is essential to an act’s being and act of the will.15

On an uncharitable reading Korsgaard is simply conflating generality and universality. On a more charitable interpretation, Korsgaard is saying that under the auspices of freedom, when we reflect on any of the acts open to us, we can only consider action-types and not action-tokens. So, considered reflection on free acts open to us presupposes generality. But when we consider the reasons for acting, these necessarily involve universality. For reflection would have to consider explaining and justifying not only this particular act, rather acts of this general type, in these sorts of general circumstances. And presumably the only way to rationally justify categorically acting in this way rather than that, is to assimilate actions of that sort under universal principles. So, considered reflection on the reasons for free action presupposes not only generality, but universality.

parlance of endorsing humanity as a non-contingent identity in the face of contingent self-conceptions. Are these two ways of describing the same phenomenon? If not, I don’t see any reason why the temptation posed by an impulse be resisted from the perspective of contingent self-conception. If yes, a lot more needs to be said for the equivalences to be evident.

Let us suppose Korsgaard is right that an active will “is brought into existence by every moment of reflection”. Notice first that this moment of reflection would have to be rather complex. It would contain at least two “moments”: reflection on the action, implying an active general will and reflection on the reasons for acting, implying an active universal will. It might be fruitful to inquire whether Korsgaard is opened to the possibility of a scalar will the liveliness of which is in direct proportion to its degree of universality. Nonetheless, whatever one thinks about this, it remains dubious that the only way reason can justify an action is by assimilating it under universal laws. At any rate, we are now in a better position to understand why Korsgaard emphasizes the law-like nature of practical rationality and why she claims that the moment of reflective distance entitles the agent to claim generality and universality. I have started this section maintaining that Korsgaard agrees with the manifest Platonic picture. We can also point out some limitations.

As we have seen, Korsgaard’s project depends on many assumptions. First, meaningful action generally and moral action particularly can only flow from synchronically integrated agency. Second, an agent is synchronically unified only if her will is active as opposed to passively falling under the sway of contingent desires. Third, the will is quickened only by practical reflection at a suitable level of generality that formally entails universality. I think there are good reasons to challenge each of these assumptions. I do not however have the space to do this properly here. Still, it is almost a platitude that appreciating the temporal character of our practical experience implies valuing modes of thought that operate inside and through those contingent desires and practical identities Korsgaard appears to regard as irrelevant to practical reason. There is
substantial disagreement between us only if her view entails that the reflection generating a will that criticizes my impulses by conforming to universal laws is the only or most significant form of practical thought worth having.

1.2.2. Smith and Desiderative Coherence

Although Michael Smith does not directly reference the Platonic picture, his account implicitly inherits some of its fundamental assumptions. Smith agrees with Williams that practical reflection must be involved in the extension of our motivational sets through instrumental reasoning and imagination. In contradistinction from Williams, Smith thinks that reason’s most significant way of overseeing the functioning of the soul is by getting involved in systematically justifying desires. Although Smith does not hold that practical reflection generates a formally universal will in order to do this job, still the agent must occupy a “suitably idealized” perspective that could engender rational advice about which desires are worthy to be kept in one’s desiderative profile.

Practical reasons are analyzed starting from the platitude that “what is desirable for us to do is what we would desire to do, if we were fully rational.” At the very least, claims Smith, full rationality must entail that depressions, emotional distresses, and other

---


17 Normative reasons turn out upon analysis to be categorical imperatives. But these imperatives are not fashioned from the formal conditions of a universal will, rather from procedural considerations, that is they would be based “a single set of desires that all rational creatures would acknowledge to be systematically justifiable.” Smith, p. 177

18 The analysis of this platitude: “In other words – and now we are really turning the platitude into an analysis, for we are making explicit distinctions that are at best only implicit in the platitude – it tells us that what it is desirable for us to do in certain circumstances – let’s call these circumstances the ‘evaluated possible world’ – is what we, not as we actually are, but as we would be in a possible world in which we are fully rational – let’s call this the ‘evaluating possible world’ – would want ourselves to do in those circumstances. That is, it tells us that facts about the desirability of acting in certain ways in the evaluated world are constituted by facts about the desires we have about the evaluated world in the evaluating world.” Smith, p. 151
such maladies would have no constitutive function in the psychology of a fully rational agent. Reason must then be involved in keeping some and at the same time excluding other dispositions and desires. But this means that practical rationality essentially involves systematically justifying our current desires. The paradigmatic way we do this is analogous to the way John Rawls claims we aim to justify our evaluative beliefs through the procedure of reflective equilibrium. Systematic justification of desires involves creating general desires that explain and justify particular desires thereby securing more unified, and so more rational desiderative profiles:

Suppose we take a whole host of desires we have for specific and general things; desires which are not in fact derived from any desire that we have for something more general. We can ask ourselves whether we wouldn’t get a more systematically justifiable set of desires by adding to this whole host of specific and general desires another general desire, or a more general desire still, a desire that in turn, justifies and explains the more specific desires that we have. And the answer might be that we would. For in so far as the new set of desires – the set we imagine ourselves having if we add a more general desire to the more specific we in fact have – exhibits more in the way of, say, unity we may properly think that the new imaginary set of desires is rationally preferable to the old. For we may properly regard the unity of a set of desires as a virtue; a virtue that in turn makes for the rationality of the set as a whole. For exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of beliefs.19

Again notice the assumptions: (1) Rational agency is a function of the unity exhibited by a desiderative profile. (2) Desiderative profiles constituted by general desires explaining and justifying more specific desires are more unified than those lacking such general desires. (3) Desires are justified to the extent they can be integrated into whole desiderative profiles. The desires resisting integration are ad hoc, and thus are liable to be excluded. If these assumptions are taken seriously, synchronic integration seems to have some diachronic consequences. Practical reason in the mode of

19 Smith (1994) p. 160
systematically justifying desires could cause “wholesale shifts in our desiderative profile.” Under the continuous tutelage of justifying practical thought each change “will seem not just different from the old, but better; more rational. Indeed, it will seem better and more rational in exactly the same way, and for the same reasons, that our new corresponding evaluative beliefs will seem better and more rational than our old ones.”

One can plausibly object to each of three assumptions noted above. To illustrate just one such worry, increased generality is only one way to increase practical rationality. And we might conceive of situations where increased generality in some aspect of our desiderative profile makes it less suited for deliberation or action. Sometimes, especially in regards to temporal transformations, the more rational thing to do is to act on desires that cannot be integrated under more general desires. I desire cookies, I desire ice cream and I desire chocolate. On the one hand, it seems that synchronically I have a more integrated desiderative profile if I come to the realization that I desire sweet things. Suppose I probe around in there, and also find constitutive of my psychological profile the desire to avoid tooth decay. It seems that a more general natural desire that could explain and justify both desires could be something like a desire to enjoy sweets moderately while maintaining my health. This desire seems like the right kind of psychological item that integrates my desiderative profile on particular occasions when it could be driven asunder by the aforementioned competing desires. Thus, prima facie on any specific occasion I am more rational when the desire to avoid tooth decay is integrated under the desire to enjoy sweet things moderately while maintaining my health

---

20 Smith (1994) p. 161
than in alternative circumstances where the former usurps the justificatory force of the latter.

But now suppose I come to find out that I have been treating this more general desire as a way of rationalizing my preference for sweets. I am integrated, alright! But also my teeth have been steadily and gradually decaying. Might this be a situation in which I am more rational if I allow the force of my *ad hoc* desire to avoid tooth decay, or plumpness, or high blood sugar to usurp that of my more general rationalized desire for sweet things? My minimal point is that in light of self-discoveries, or new facts, or new circumstances it is sometimes desirable to accept the self-justificatory force of specific *ad hoc* desires precisely because they conflict with my more general wants.

More generally then, the systematic justification of desires should not depend exclusively on rational criteria reducible to synchronic coherence and generality. Sometimes holistic psychological and relational considerations, facts about the history and contents of desires are more pertinent to which desires should survive and which not. Narrative understanding allows for a form of practical rationality that can engage in systematic explanation of desires without necessarily claiming to integrate desires into more general desiderative profiles.

Yet even if it turns out that most desires become rational by being subsumed under more general desires, Smith’s model of practical reasoning might still require something like narrative understanding. The norms of rational subsuming under more general desires seem arbitrary unless guided by norms of coherence with the unfolding of my life story. For recall that my most general desires should be those that my idealized self would want. And furthermore in cases of conflict among my general desires, and in
cases when I have to identify the general desire under which a specific one gets
subsumed, I could only make a rational determination in light of my historical (narrative)
self-understanding. At the very least, a conception of my idealized self sketched in
abstraction from my history and relying only on structural considerations of occurrent
desires seems so severely impoverished as be implausible.21

Despite the dissimilarities in both content and methodology between Korsgaard’s
and Smith’s accounts, we have seen that the two make similar assumptions about rational
agency and synchronic unity. In light of psychological pressures seen as inimical to
rational agency, they both argue that synchronic unity is a value that fixes the shape of
rational agency. And furthermore synchronic integration (either in the will or in
desiderative profiles) obtains only when practical thought occupies reflective stances that
tend toward increased generality or universality. Even if these lofty functions of practical
reason are commonplaces for typical agents, does meaningful action have to result only
from synchronically integrated psychologies? If we are essentially temporally-extended
agents and if many of our reasons for action arise out of our past histories or plans for the
future, it is reasonable to inquire whether practical rationality might have modes that
engage diachronically. This does not entail dismissing the ideal of synchronic unity, but
(if indeed it is valuable to beings like us) it might mean refashioning it in light of the
repertoires of desires, values and normative beliefs grounded in our personal,
interpersonal and social histories.

21 I am indebted to Mike Rea for suggesting the line of thought in this paragraph.
1.2.3. Millgram and Practical Induction

A fascinating attempt to pay attention to the diachronic dimensions of practical rationality belongs to Elijah Millgram. In chapters 3 – 5 of his *Practical Induction* he defends the following three interrelated claims: (a) Without the exercise of practical induction actual human agents are not sufficiently unified for first-person deliberation; (b) It is impossible for human agents to be sufficiently unified for first-person deliberation without the use of practical induction (or, the capacity for practical induction is a necessary, irreplaceable feature of intelligent human agency); and (c) The explanation for (b) is mainly provided by presence of irreducible novelty in the world: there is no other way opened to intelligent agents for learning things that matter from experience except by practical induction. Consequently, the deepest fact about human nature relevant to practice in the presence of novelty is the capacity for practical induction.

Because his understanding of practical reason as practical induction brings together the ideal of synchronic unity and diachronic coherence I will deal primarily with (a) above. Here’s a reconstruction of what I take to be the main steps in support of (a)\(^\text{22}\).

(P1) Without a high degree of synchronic unity there is no agency or practical reasoning.

\(^{22}\) Millgram draws a heuristic distinction between synchronic and diachronic unity. I reconstruct in the text only the version that appeals to synchronic unity. As it turns out diachronic unification will require synchronic unity. Here’s a quick reconstruction of the diachronic version of the argument:

1. If my practical deliberation is to have a point, I should expect to act on the conclusion of a plan or a policy.
2. If I should be able to execute the plan, I need to be able to assume it won’t be bumped aside for no good reasons.
3. So, I need to possess the ability of discriminating between whimsical considerations and those that should be allowed to make adjustments to the original plan.
4. But (3) implies possessing the capacity for bringing potentially defeating considerations to bear on my deliberations.
5. And this just means that I must be synchronically unified.
(P2) An agent’s synchronic unity with respect to practical judgments (judgments about what is valuable) is strongly analogous to a mind’s synchronic unity with respect to beliefs.

(P3) A mind is synchronically unified to the extent to which a thought or mental item is prone to enter into more appropriate inferential relations with other thoughts that did not originally include it.

(P4) Therefore, by analogy, an agent is synchronically unified to the extent to which the agent’s practical judgments are prone to enter saliently into deliberation with other competing practical judgments. (From P2 and P3)

(P5) Bringing to bear one’s practical judgments saliently on deliberation consists in developing an intelligent sensitivity to the defeasibility conditions of one’s practical syllogisms.23

(P6) The sensitivity to the defeasibility conditions of one’s practical syllogisms could not consist mostly in comparing the relative strengths of desires. (Practical induction could not collapse into instrumentalism)

(P7) The sensitivity to defeasibility conditions involves mostly deploying practical judgments at an intermediate level of generality in a way that connects the competing interests expressed in the major premises of conflicting practical syllogisms. (The positive characterization of practical rationality)

(P8) Practical induction is responsible for furnishing an adequate enough stockpile of connecting practical judgments.

---

23 Millgram, p. 55
C: So, without practical induction, there’s no synchronous unity to deliberative agency.

From the views discussed so far, Millgram’s seems most resistant to some of the main fragments of the Platonic picture. Most significantly he seems to reject what I earlier referred to as the internality assumption. Practical reason is not only a matter of synchronizing inner elements, but it is also essentially involved in learning new things from experience. This being said, Millgram agrees with Plato, Smith and Korsgaard that synchronous unity is valuable, and that it is even more fundamental than diachronic unity. Furthermore, it is claimed that practical induction explains the unity that is necessary for deliberation.

But what is practical induction, and how does it function? Millgram warns he is merely sketching a practical analogue of theoretical induction, which unluckily for us, “is not very well understood.”\(^\text{24}\) Since experience is always of particulars, and since we will learn what matters practically from experience, practical induction is a form of “learning what is worth doing by generalizing from particular experiences.”\(^\text{25}\) Although a quick summary is unjust to the complexities and ingenious developments of Millgram’s account, we can compress his insights to notice that practical induction is divided among four main operations: i) it forms rock-bottom desirability judgments through the experience of pleasure and the experience of friendships ii) it stockpiles and structures these desirability judgments iii) it rationally deploys these desirability judgments as needed in particular circumstances such that they typically result in actions, and iv) it integrates synchronically the agent by connecting conflicting desirability judgments.

\(^{24}\) Millgram, p. 45

\(^{25}\) Millgram, p. 44
through intermediately general practical judgments (that is to say practical induction is involved in the work of any practical syllogism)

To illustrate how practical induction functions for Millgram consider Beth’s dilemma. Beth is a philosopher split between the policy of not going to talks by visiting speakers on the one hand, and the desire to have dinner with one such speaker whose published work is usually exciting. Millgram construes this as a tension between two practical judgments that are the work of practical induction: “Speakers are not worth hearing because they tend to be competently professional rather than deep or even clever; one dines with speakers because one expects philosophically exciting table conversation”\(^{26}\). The *intermediately general judgment* which connects the previous competing judgments, also fashioned by practical induction, might be: “this speaker’s work is for the most part philosophically exciting”\(^{27}\).

No quarrels so far. Nonetheless, while practical induction could indeed produce the above-mentioned practical judgments, it is not at all clear whether it is also the relevant process effective in making the intermediate judgment salient to solving the practical dilemma. The materials at hand are the work of practical induction, but the solution is “improvised”\(^{28}\). I can understand how *practical induction* can furnish the rough materials of deliberation, but it is less clear how precisely it is also involved in making salient the content of one judgment upon another. Rather, the inferential

\(^{26}\) Millgram, p. 63  
^{27} Ibid.  
^{28} Earlier, in the discussion of Alison, Millgram expresses the same thought by claiming that the solution to the practical problem is *ad hoc*. And on the face of it, that turns out to be an important and desirable thing given the claims made in chapter five.
connection among judgments seems either based on semantic connections or on fortuitous events, but not clearly dependent on further generalizations from experience.

Millgram might reply that the better we get at practical induction, the better we get at *improvisation*. Our ability to bring to bear the relevant bridging practical judgments upon our conflicting practical syllogisms depends on our having developed already a rich enough repertoire of practical judgments. This might be right, but notice that increasing our sensibility to defeating considerations and therefore increasing our stockpiles of practical judgments could also foster disintegration and instability of agency. This consideration should be particularly pertinent given Millgram’s central commitment to the claim that practical induction is essentially responsive to the increasing irreducible novelty and complexity of our practical world.29 To the extent that normal agents genuinely internalize the radical novelty of the world, to that extent they may treat any of their practical decisions with tentativeness and to that extent agential indeterminacy creeps in.

29 Interestingly, Millgram claims that in the reality of novelty lies an argument to the effect that practical rules couldn’t be universal. The argument can be reconstructed like this:
(1) Our world is full of irreducibly novel things and circumstances
(2) If novelty is an irreducible feature of our world, then intelligent agents cannot avoid inventing concepts necessary to capture this novelty.
(3) If agents cannot avoid inventing (coming up with) new concepts to mostly capture novelty, agents cannot avoid developing novel interests.
(4) If agents cannot avoid developing novel interests, novelty is essentially practical.
(5) So, our world is full of novelty that is relevant to action.
(6) If value-novelty is a basic feature of the world, no one will be able to anticipate the practical needs of a creature so designed as to be incapable of learning what matters from experience.
(7) If not even God can anticipate the practical needs of creature C, then the only way open to us to equip her is with ways of learning its interests from experience.
(8) If novelty is pervasive, then general rules may have unforeseen realistic exceptions.
(9) So, we couldn’t equip creature C with general exception-less rules.
(10) Thus, we should equip creature C with the ability to deploy defeasible rules of practical inference. But if novelty and complexity are irreducible, one wonders how helpful defeasible rules will be for integrating agency. The point is that practical induction needs to strike a perfect balance between agent’s stability and openness to novelty. And we would want to hear more from Millgram about just how that is accomplished.
Re-visit Beth. Suppose she has acquired through practical induction many somewhat general practical judgments, and that she’s really bent on improvising. This means only that she’s really good at finding ad hoc disconfirming instances of her practical judgments, disconfirmations that are also based in proper generalizations from her experiences. As soon as an intermediate practical judgment occurs to her, a judgment that would have otherwise solved her practical dilemma, another bridging judgment pops up. “But philosophers whose published work is exciting turn out to be conversational bores.” This judgment is in tension with Beth’s earlier bridging judgment. So, Beth would under normal circumstances decline the talk and the dinner. But remember this is a dramatically anomalous world, and Beth is really good at improvising. A new bridge: “When there’s the prospect that another’s philosophical creativity may inspire one’s own it is profitable to tolerate some boredom.” And on it goes. The general point that I’m trying to make here is that to the extent to which normal agents encounter the genuine novelty of the world and are apt at improvising, vacillation and indeterminacy could follow. If sensibility to defeating considerations is cultivated, practical induction may tend to undermine the presumed unity of action. If in Beth’s case we can see how processes that track better and tighter connections among practical judgments could continually delay decision and action. Judgmental unity understood as the disposition to create increasingly connected networks of intermediately general practical judgments

---

30 One might wonder whether the analogous problem holds for theoretical induction. If theoretical induction works fine, naturally, one explanation might be that from a pragmatic point of view we have no choice but to adopt a limited number of working hypotheses. It would be too costly to always reconsider our theoretical generalizations, to look for disconfirming instances even for the most entrenched frameworks. If one is attracted to the view that the main reason induction seems to work is due to socialization, one can consider the work of theoretical anarchists such as Paul Feyerabend. He seems to hold that something like distrust of any stable research program is a precondition of real openness to defeasibility conditions.
does not secure either emotional coherence or unity of action. On the flip side, if one values decisive action or emotional coherence, it is hard to see that practical induction is alone, or even primarily, responsible for the salience effective in attaining these.

From among conceptions emphasizing synchronic unity attained through generalizing judgments, Millgram’s view seems more amenable to accommodate narrative understanding. Part of the explanation for this is that practical induction involves diachronic processes shaping the perception of our practical world and the vehicles responsible for storing and later conveying that practical information. This being said, narrative understanding could offer distinctively corrective insights for Millgram’s position. As I argue in chapter four, narrative understanding does not offer haphazard perceptions of our practical reality, but those structured thorough and by the diachronic flow of emotions. As we will see, these perceptions have a dual, related intelligible context: one given by their social context and one summed up in a certain emotional cadence.

If this picture is correct, the perception, storage and transmission of practically relevant information should not be effected, pace Millgram, only through judgments generalized from experience. Rather, desirability trains of thought can be stored in complex social and psychological stories. Furthermore, the way these trains of thought can be brought to bear on decision depends not only on their narrative cadence but also on occupying distinctive empathetic positions that could move agents to re-conceive their practical reality. Phenomena such as personal forgiveness and atonement reveal that agency must involve forms of practical thought and subtle inference that allow for non-linear re-evaluations not only of practical judgments, but of more fundamental practical
and emotional outlooks. And these transformations in emotional sensibilities are better explained by my dramatic-narrative model than by practical induction.

I have looked at three main accounts of practical reason that emphasize in different ways the synchronic unity of agency. For Korsgaard the agent is unified in the moment she reflectively endorses her humanity as a reason for action. Smith maintains unity is generated by integrating specific desires in the more general desiderative profile of a suitably idealized rational advisor. For Millgram unity is a measure of the tendency of the whole network of practical judgments to enter in inferential relations with each other. For each thinker reason’s capacity for generalization is essentially involved in the process of unification. There are also striking differences. Whereas Korsgaard and Smith seem content to account for unity statically, in abstraction from the way desires, inclinations, or practical identities were formed, Millgram emphasizes that practical induction is essentially diachronic, external and historical. Consequently, his view of synchronic unity depends partially on generalizations that are fluid and better suited to

31 This claim needs some justification. It seems not only that there are non-judgmental modes of thought pertinent to an agent’s diachronic coherence. More significantly, practical judgments and the inferential connections between them cannot provide an exhaustive map of all the springs of meaningful human action. I argue in the sequel that some emotional outlooks and the linguistic practices supporting them are not expressible without residue in practical judgments, let alone in practical syllogisms. The situation is more like this: only fragments from our emotional lives ever make it explicitly into practical judgments.

To press this point let me refer to a main line of argument in Alice Crary’s new book Beyond Moral Judgment (Harvard University Press, 2007). She argues here for a broadening of our conception of moral thought beyond the narrow confines of moral judgment. The argument runs like this. If all forms of moral thought essential express individual moral outlooks, and if an individual’s moral outlook is shaped by sensibilities internal to all her linguistic capacities, “it follows that the view might be described as one which ethics is conceived as distinguished by a preoccupation not with judgments in one region of language but with a dimension of all language.” (Crary, p. 47) I agree with Crary that practical outlooks as they are expressed in emotional sensibilities cannot be restricted to practical syllogisms, but “they encompass her entire personality – her interests, fears, ambitions, her characteristic gestures and attitudes and her sense of what is humorous, what is offensive and what is profound.” (47)

Even if one thought synchronic unity valuable, it seems questionable that Millgram’s ideal of synchronic unity (a measure of the defeasibility of one’s practical syllogisms) is likely to bring about coherence to the whole “texture of a person’s being.” And to the extent that the latter is valuable to our practical lives and deliberations, practical induction remains mostly in a supporting role, and it has to share the power of shaping agency with other forms of practical thought.
deal with temporal transformations of agency. In the next section I merely list some of the main resources that can inform more generally a plausible account of narrative understanding.

1.3. Resources for Narrative Understanding

In the last several decades there has been much interest in the concept of narrative. Theorists in fields such as cognitive science, psychology, psychiatry, psychoanalysis, counseling, literary theory, social studies and theology have brought to the forefront the explanatory power of narrative.\(^\text{32}\) In order to exemplify what is so attractive to theorists from a variety of fields about the function of narrative, consider this extended paragraph from *Narrative and Consciousness*:

> As the various chapters in this book demonstrate, personal stories are jointly constructed through social relations, are shaped by our nature and culture, and are logically compatible with the naturalistic scientific method of investigation…Narratives are available for observation; they are public and shared, and as a result their examination does not of necessity result in subjectivism and relativism. Narratives are governed by culturally embedded formal rules – character roles, time and plot structure, and so forth – that define

whether they are recognized as coherent and cogent. Personal narration gives continuity to self and meaning to action as it locates and values present activity in the context of past experiences and projected outcomes. Narrative is also tied to the development of signs and language and as such is constrained by and consistent with developmental biology and human neurobiology as well as linguistic structures and narrative traditions.33

One can appreciate the contemporary interdisciplinary work on narrative while remaining cautious about how it is often invoked as a panacea for a number of theoretical and practical puzzles. Nonetheless, this cautionary attitude should not stop us from investigating whether or how narrative could inform contemporary philosophical conceptions of agency and practical reasoning. For this purpose I turn to briefly mentioning some past uses of this concept in philosophy, focusing particularly on moral psychology.

1.3.1. Narrative and Personal Identity

Marya Schechtman34 and Marc Slors35 have recently argued that the best way to account for personal identity over time is by reference to a narrative conception of self-constitution. Schechtman argues that the life of a human being has the form of a biographical narrative, since events such as the actions, beliefs, and intentions of the person cannot be understood apart from an intelligible story-line. Moreover, such events cannot get their meaning piecemeal, but rather from being contextualized in broader plot-like structures belonging ultimately to the life of the agent.

33 Fireman, McVay Jr., Flanagan, p. 5
As understood by Schechtman a significant feature of narrative involves an important modification of the Neo-Lockean model of psychological relations. She argues that individual beliefs, desires, values, and traits that make up a person’s life cannot be first “identified as isolated elements at a time and then connected by relations of similarity to psychological features at other times…but that the beliefs, values, and desires that make up these relations already deeply intermeshed when we first identify them.”36

Marc Slors agrees with Schechtman’s rejection of the atomistic individuation of mental contents. He argues that since the mind is essentially diachronic, the relations of continuity between temporally distinct psychological states cannot be simply a matter of causal connection or qualitative similarity, but that they are “connections of semantic interdependence between the contents of these mental states.”37 The connections of semantic interdependence contribute to the narrative continuity of a person, and both Slors and Schechtman agree that narrative continuity is presupposed by psychological continuity.

1.3.2. Narratives and Moral Sensibility

In a different mode from those stressing the narrative structure of our perceptual experiences, some argue that particular works of fiction (narratives) are necessary for sharpening moral sensibilities.38 Narrative understanding here refers to the very specific

37 Slors (2001) p. 4
ways in which good literature, films, or dramas develop our attentiveness to nuances and complexities of the moral life. Without the education afforded by compelling human paradigms encountered in such stories we would be less adequate human agents. The plausible view here is that good literature (or good stories) helps to educate the reader’s emotions and develop her moral imagination. Martha Nussbaum argues that the skills developed in the course of engaging great literature not only make one a more perceptive and emotionally attuned moral agent, but they also deliver concrete and precise moral knowledge and self-understanding.

1.3.3. Narrative, Intelligibility, and Accountability

Consistent with the above conception of narrative, Alasdair MacIntyre argues that since our actions are intelligible to ourselves and to others only in the context of the narratives in which they play a role, we are essentially story-telling animals. Moreover, he argues that the justification of practical action is a function of its intelligibility: “To identify an occurrence as an [intelligible] action is in paradigmatic instances to identify it under a type of description which enables us to see that occurrence as flowing intelligibly from a human agent’s intentions, motives, passions, and purposes. It is therefore to understand an action as something for which one is accountable.” MacIntyre seems to suggest here that actions are understood and justified in light of the narrative structures that they are part of. And furthermore, communication and dialogue show that these narrative structures are not merely phenomena internal to our consciousness but also

---

40 A. MacIntyre, *After Virtue* p. 209
external features of communities and social groups that partially fix the reference of an agent’s practical identity.

Consistent with MacIntyre’s conception of narrative as partially dependent on social contexts, Charles Taylor argues that we gain an understanding of our lives only in relation to some good, and only by integrating this relation in a broader framework that allows us to see how we have become what we are and the direction in which we are moving.41 Given that self-knowledge requires a story-like contextualization of the self with respect to the good, “we must inescapably understand our lives in narrative form, as a ‘quest’.”42

The self that we come to understand on Taylor’s story is not the punctual, monological, or individualistic self of the Cartesian provenance. Rather Taylor proposes a dialogical model of the self which crucially a) involves the mediation of the body as a source of encoding one’s temporal projects and b) involves conversation with other selves. The self finds one’s own articulated identity only as an “interlocutor” in social space.43

---

41 “It has often been remarked that making sense of one’s life as a story is also, like orientation to the good, not an optional extra; that our lives exist also in this space of questions, which only a coherent narrative can answer. In order to have a sense of who we are, we have to have a notion of how we have become, and of where we are going.” (Charles Taylor, Sources of the Self, (Harvard University Press, 1989) p. 47)

42 Ibid., p. 52

1.3.4. Narrative Explanation and Motivation

David Velleman has made recently an effort to map out the motivational roles that narrative self-conceptions might have.\(^ {44}\) One way in which he explores the possibilities and limits of narrative explanations for practical motivation is by focusing on ways in which fictive scenarios can serve as the story lines in light of which only certain actions makes sense as continuations of an agent’s own story. He uses the example of a smoker who could be motivated to stop smoking by identifying imaginatively with a non-smoker.\(^ {45}\) This imaginary script could generate reasons for action that are rational insofar as the agent becomes a more rational person. In addition congenial to my project, as I explore in much more detail in chapter four, Velleman holds that narrative explanation is significantly grounded in the life of the emotions.

1.3.5. Narrative and Moral Resistance

One significant role that narratives have played in contemporary moral philosophy is that of shaping and reshaping the moral identity of agents. Some have seen stories as essential ways of resisting formation of the self in terms of the oppressive stories that the dominant culture tells about the marginalized and the outcast. Others have seen in stories modes of repairing damaged done to identities, a narrative reframing of fragmented self.

Margaret Urban Walker, for instance, although resisting the whole-life narrative approach of Taylor and MacIntyre, agrees that local narratives are necessary for

\(^{44}\) See D. Velleman, “Self as Narrator” (available on his website); “Motivation by Ideal” *Philosophical Explorations* Volume V (2), May 2002 pp. 89 –103; “Narrative Explanation” (2003)

\(^{45}\) Velleman, D., “Motivation by Ideal”
intelligible actions.\textsuperscript{46} She develops an account of resisting the so called “necessary” identities that dominant culture thrusts upon the disposed and the marginalized. An alternative account of repair of such damages to one’s identity is provided in explicitly narrative terms by Hilde Lindemann Nelson.\textsuperscript{47} She argues that counter-stories are narrative ways of resisting the master plots which suppress or oppress moral agency.

Susan Brison’s insightful and gripping book, \textit{Aftermath}, offers evidence that narrative plays an essential role in the remaking of a self consequent to trauma. Compellingly she argues that the narrative memory of trauma survivor “is an act on the part of the narrator, a speech act that defuses traumatic memory, giving shape and temporal order to the events recalled, establishing more control over their recalling, and helping the survivor to remake a self.”\textsuperscript{48}

1.4. A Quick Preview

My positive account of dramatic-narrative agency is inspired by many of the previously mentioned original insights. I am greatly indebted to MacIntyre’s idea that narrative understanding is a distinctive mode of rationality that characterizes intentional action. In chapter four I respond to some of the main objections against this characterization and show its plausibility. I do not however embrace MacIntyre’s

\begin{flushleft}
\textsuperscript{46} Margaret Urban Walker, \textit{Moral Understandings: A Feminist Study in Ethics}, (Routledge,1998) p. 148
\end{flushleft}

\begin{flushleft}
\textsuperscript{47} She argues that “identities…are complex narrative constructions consisting of a fluid interaction of the many stories and fragments of stories surrounding the things that seem most important, from one’s point of view and the point of view of others, about a person over time.” (H. L. Nelson, \textit{Damaged Identities, Narrative Repair}, (Cornell University Press, 2001) p. 20)
\end{flushleft}

\begin{flushleft}
\end{flushleft}
conception of the narrative self. Whereas I think it is plausible to think of enduring trains of intention on the narrative model, when it comes to agency things get more complex.

The practical self is simultaneously heir to multiple stories, to many historically conditioned guises of self-understanding, only some of which find articulation on any particular occasion. Henceforth a more viable model for agency attempts to capture the continuities but also the interruptions and discontinuities that affect deliberation and action. My model conceives of the agent as a diachronically persisting chorus of personae, each persona expressing some scripted way of interpreting the world. Deliberation is then conceived implicitly as the dramatic interaction of these historically developed personae.

What gives this picture of agency plausibility is the important and subtle functioning of the emotions. I reference David Velleman’s masterful inventory of positions here, but draw on Peter Goldie’s and especially Ronald de Sousa’s view of the emotions.49 De Sousa argues, for instance, that emotions are diachronic states of mind not reducible to any other type of mental states. Since they permeate through beliefs, desires, and commitments, found at every level of our mental architecture, they turn out to be essential for self-understanding. Moreover, de Sousa argues that emotions originate in paradigm scenarios, which are loaded ways of construing the world, mediating between present and past, urging us to represent the objects of our present experience through the scripts essentially embedded in our past. He claims that “paradigm scenarios

are dramas – stories played out by actors of memory and phantasy, up for imitation by the flesh and blood actors of our life.”

I elaborate this picture of agency in the last chapter where I argue that some really plausible constraints on personal forgiveness make sense in light of my dramatic-narrative account. This is also the place to showcase how narrative understanding helps explain essentially retrospective action in ways irreducible to self-governing policies, volitional love, or practical induction.

I conclude this chapter by giving a quick synopsis of the next two chapters. I have argued here that intellectualist accounts of practical reason unduly emphasize the synchronic unity of agency. Chapter two and three deal with accounts of practical reason which are neither intellectualist nor exclusively synchronic. Michael Bratman claims that self-governing policies and plans reveal the seat of agency. Bratman’s view has initial plausibility as an account explaining the integration of the person through action. For plans seem to be the right kind of psychological structures that provide stability, reliability and consistency for our decisions over time. My main claim is that even though planning and its related expression through self-governing policies can rationally integrate the agent, this account also has limitations. These limitations are revealed when we consider ways in which an agent’s temporal self-understanding can generate reasons for action without relying on self-government or planning. Rather than revealing the heart of agency, Bratmanian self-governing policies should be understood as derived and defeasible expressions of our more fundamental commitments to persons. Looking

50 Ronald de Sousa, (1990) p. 438
ahead, I suggest that self-understanding and personal commitments distinctively involve narrative.

Volitional configurations distinct from planning make their way into discussions of diachronic agency in the work of Harry Frankfurt. Distinguishing from Bratman who locates our truest self in self-governing attitudes, Frankfurt holds that our loving investment integrates the agent diachronically and synchronically. I pursue two main reservations about Frankfurt’s account. The first is that conceptualizing love as an essential volitional limit to agency has some serious drawbacks in explaining plausible transformations of agency over time. The second is that volitional loving amounts to a kind of diachronic investment into others that neglects the extensional reality of the beloved. And that seems incompatible with our common understanding of love. Again, looking ahead I argue that narrative understanding can account better for changing practical attitudes over time. I also hold that narrative understanding involves essential modes of attentiveness to the reality of others, therefore presenting the possibility of grounding attitudes of love that avoid volitional solipsism.
CHAPTER 2
PLANS, COMMITMENTS AND SELF-UNDERSTANDING

2.1. Michael Bratman and Agency

My eventual aim is to show that the concept of narrative can be fruitful in contemporary moral psychology. But the road to narrative involves an elaborate story that has its point of departure in the contemporary theory of action. I begin by sketching Michael Bratman’s attempt to ground subjective normative authority in the functioning of complex planning attitudes. I isolate this view because Bratman is one of the few theorists working with the assumption that agency is essentially diachronic.

Moral psychology has been dominated in the last couple of decades by synchronic accounts of moral agency. These claim to generate accounts of practical reasoning and normative theories of human action simply from an examination of the agent at a time-slice. Kantians and Humeans almost uniformly agree that a complex evaluative scan of our current psychological states (beliefs and desires) can reveal the motivation for morally right/good action.¹ Running against this picture, Michael Bratman’s interesting and sophisticated account of moral agency considers the human experience of time an essential feature of human action; our temporal-extendedness is essential to our being moral and more generally, practical agents.

¹ See most recently Christine Korsgaard The Source of Normativity and Michael Smith The Moral Problem
This being said, Bratman’s account of diachronic agency is not without problems. I use the normative regress problem as a stepping-stone to highlight a more fundamental limitation for Bratman’s view. I aim to show that central commitments essential to one’s agency are not reducible to planning or self-governing policies. Bratman claims that an agent’s practical self is constituted by plans and self-governing policies due to our extensive need for self-management. I reply by arguing that the need for the type of self-management he envisions is exaggerated. Furthermore, a workable form of self-management can be attained in the absence of explicit self-governing policies.

Finally I hold that our need for self-understanding is deeper than and even illuminates of our need for self-management. I argue that relational commitments necessary for self-understanding are supported by retrospective diachronic states that cannot be plan-like. Unlike the reflection involved in treating self-governing policies as reasons for action, the type of reflection associated with retrospection and self-understanding involves subtle forms of emotional attentiveness and narrative understanding. By beginning to explore the connections between emotional attentiveness, narrative understanding and relational commitments, this chapter anticipates the positive account of narrative-dramatic agency developed more fully in chapter four.

Bratman’s model proposes several kinds of psychological structures that support, motivate and justify the action of temporally-extended agents. Let’s schematically review these. First, there are planning structures, which essentially infuse stability and reliability in the decisions of agents over time. A planning agent is a purposive agent

---

2 An early statement of planning agency occurs in Bratman *Intentions, Plans, and Practical Reasoning*, (Harvard University Press, 1987) view that is refined later in a number of articles.
because his desires and moral beliefs attain coherence by being integrated in planning structures. Consider Greg (a fictional character) who wants to give up smoking. He has just read Bratman, so he starts planning. Maybe he develops a daily strategy that ensures he does not have the daily pocket-change necessary for buying cigarettes, or maybe he plans to take a route to work which avoids passing by his customary tobacco shops. More discerningly, perhaps he begins to plan for an incremental redirection of his strong inclination toward smoking to chewing tobacco-flavored gum. So, he starts buying packs of gum containing various concentrations of tobacco flavoring, and as the days go by, Greg commits to only chew gum of the same or lower concentration than the day before. Relevantly, in the execution of these plans Greg’s identifies with his initial desire to give up smoking. It is no longer a passing, first-order desire, but it acquires a more robust agency-structuring status. And if Greg’s plans are stable enough across time he might just kick the habit.

A second main feature of Bratman’s account of agency is structural hierarchy. The idea of hierarchical psychological structures can be traced at least as far as Plato’s view of the soul. According to the canonical Plato, for things to go well and just contemplative rationality like an enlightened monarch controls the emotive and the desiderative functions of the agent. In Bratman’s view the seat of agency is not found in contemplative rationality but in a set of hierarchical plan-like volitional attitudes dubbed “higher-order policies.” Like contemplative reason however, these policies are normative in the sense that they regulate, guide, and can justify first-order desires and motivations. As we will soon see, policies can be understood as complexes of enduring desires that...

---

3 Illustrative is Plato’s Republic where the soul emerges as a hierarchical structure with the rational part in control of the spirited part and the appetitive part of the self.
have acquired their “high-order” status by being reflectively endorsed. Like Plato then, Bratman holds that an agency’s true self is essentially constituted by certain kinds of higher-order policies. At the same time agents can and do act on desires not endorsed by their policies and unconnected to any long-term plans. Sometimes our actions betray the commitments to our truest self.

To exemplify this phenomenon, consider St. Paul’s seemingly paradoxical statement from the Letter to the Romans “For what I want to do I do not do, but what I hate I do.” Bratman’s account can explain plausibly the apparent practical tension by appeal to psychological hierarchy: St. Paul is presumably reporting that occasionally he acts on desires that do not conform to his higher-order policies. When he acts on the first-order desires of the flesh, he thereby becomes alienated from his own true, authentic, spiritual self. Presumably, an integrated and harmonious agent is one whose lower-level psychological attitudes are in sync with one’s higher-order policies.

For Plato the ultimate justification for the normative power of contemplative rationality lies in its straightforward intuition of moral reality. Grounded directly in agent-independent reality, rationality’s normativity does not require monitoring or supervision by additional psychological attitudes. In this way Plato avoids the infinite regress looming for accounts of agency that establish normativity in psychological structures. The costs for Plato are well-known: a presumed implausible moral ontology, epistemic and motivational difficulties.

Unlike Plato, volitional hierarchy accounts do not appeal to the perception of an independent moral reality. So where does the normative force of higher-order policies come from? In fact, if the normative force emerges strictly by reflectively endorsing
desires from lower levels, hierarchical accounts of agency seem plagued by an infinite regress problem. As Gary Watson rightly points out, that the agent values X (or that X has normative force for the agent) cannot be explained entirely by the fact that the agent desires X at a particular psychological or volitional level. No matter how high the level at which the will is determined, it is intelligible and perhaps even desirable to go on asking why the agent should value the determination of her will there.

Bratman’s answer to this problem is intriguing. He suggests that when higher-order policies merge with planning structures, some of the newly emergent psychological attitudes could acquire a brand-new function. These new policies justify themselves by playing not just a motivational role but also an important justification-conferring role upon effective desires at lower psychological levels. Bratman labels these special reflexive and reason-providing policies “self-governing policies”.

To illustrate the functioning of such novel psychological structures consider the desire of an alcoholic to refrain from drinking. Perhaps his commitment to Alcoholics Anonymous and his fear of backsliding jointly reinforce a higher-order desire to refrain from drinking. Furthermore, suppose he is at a party where there are social pressures to drink. He may resist this temptation by deciding to treat the higher-order desire to refrain from drinking as reason-providing. The idea is that if the agent treats the desire to refrain from drinking as a kind of plan of action for the evening, it effectively becomes a policy.

---


6 Consider this quote: “These are policies concerning what weight, if any, is to be given to desired ends in motivationally effective deliberation. Following a discussion of related matters by Robert Nozick we can say that such policies involve a ‘commitment to make future decisions in accordance with the weights it establishes...’” (Michael Bratman, “Valuing and the Will” *Philosophical Perspectives, 14 Action and Freedom*, 2000)
that justifies itself in light of the governing functions it acquires. Perhaps throughout the night the agent thinks to himself, especially when tempted to reach for a drink “I have adopted this policy for the night, so come what may, I have got to stay true to it at least until tomorrow”. In these circumstances it seems this policy is effective in expressing the agent’s core values and in justifying his action. The agent thus treats this reflexive decision as a self-governing policy, allowing it to effectively shape and guide his motivations, and to give his audience a glimpse of what the agent values. Practically, by being effective in deliberation and shaping volitional profiles self-governing policies attain subjective normative force from their own contents.7

Bratman argues that self-governing policies furthermore get essentially involved in supporting and constituting Lockean psychological ties characteristic of the agent’s temporal existence.8 By capturing what is valuable in the cross-temporal organization of practical agency, self-governing policies partially constitute the moral identity of the agent over time. Being constituted by planning structures and higher-level policies, they make up the essential core of agency. According to the position, self-governing policies thus explain three main values typical agents care about: continuity over time, normative force, and authenticity.

In recent revisions Bratman adds that stronger forms of self-authenticity and self-autonomy would require the agent to know and to be satisfied with the content of his

---

7 A worry for Bratman here might be that the account seems to allow the possibility of generating self-governing policies for addictive or obsessive behavior. And these policies clearly shouldn’t be justified. But Bratman’s view requires that higher-order capacities reflectively endorse lower level desires. As such, it is hard to see how a being aware of his obsessive behavior would want to treat compulsive desires as reasoning providing in the future. So, Bratman would like claim that addicts and obsessives have diminished self-governing capacities in the relevant areas.

higher-order self-governing policies. It is not clear however whether these psychological states act as independent valuations of the content of the self-governing policies, or whether they are constitutive parts of these policies.

2.2. Refinements and Minor Criticisms

2.2.1. Justificatory Regress Problem

A first problem for Bratman’s view is that normative regress mentioned earlier may not be stopped after all. The worry regards the manner in which the self-governing policies receive their normative force. An unrepentant critic could ask about the value of deciding to treat one’s higher-levels desires as effective in deliberation. Deciding to treat a higher-order policy as reason providing for X may not necessarily translate in valuing X. Bratman’s appeal to knowledge of and satisfaction with particular self-justifying policies would miss the point of the challenge again. For it is possible to be psychologically satisfied with one’s self-governing policy and still not assess it as morally or practically authoritative. Our alcoholic friend might be satisfied with his decision to refrain from drinking due to exhaustion or to confusion or due to a pathological desire to avoid self-blame. And clearly the invocation of these attitudes does not explain the worth or value we attribute to agency in such moments of decision. On the flip side, some attitudes may be practically normative to our agency even if we fail to be fully aware of them or psychological satisfied with them.

In an attempt to beef up the intuition that self-governing policies can have a special kind of normative authority, Bratman invokes the idea of “reflexivity”:

---

9 See especially Bratman Michael, “Three Forms of Agential Commitment: Reply to Cullity and Gerrans”
The self-governing policies that are central to the model of autonomy that we are constructing will be in part about their own functioning. Such a policy will favor treating certain desires as reason-providing as a matter of this very policy. The idea is not that such reflexivity by itself establishes agential authority of the policy. Agential authority of such attitudes is, rather, primarily a matter of Lockean role and satisfaction. But in a context in which these conditions of authority are present, a further condition of reflexivity ensures, without vicious regress, the endorsement of self-governing policy that seems an element in full-blown self-governance.\(^\text{10}\)

Conative attitudes garner normative force only under several conditions. Bratman appears to say here that an agent normatively endorses a self-governing policy in case:

a) it is diachronically functional (Lockean condition)

b) The agent knows about and is satisfied with her self-governing policy (satisfaction)

c) it will be in part about its own functioning. (reflexivity)\(^\text{11}\)

The introduction of the reflexivity condition does not make it entirely obvious how the normative regress problem is avoided. Reflexivity either implies a brand new psychological state monitoring a self-governing policy (or a complex of such states) or it does not imply it (self-monitoring is already part of the content of each self-governing policy). If the first option is what Bratman has in mind, then we can go on inquiring whether the agent is satisfied with this brand new psychological state, and the regress is looming. If the latter option is true, that is, if self-monitoring is built into the content of each self-governing policy, it is hard to see why we need to introduce a new condition beyond psychological satisfaction with a certain volitional profile. As such, introducing


\(^{11}\)Bratman may suggest that this strong reading is not what he has in mind. Perhaps the reflexive content of the self-governing policies is fixed holistically (that is by appeal to other self-governing policies) rather than individualistically. I don’t think that this distinction would change essentially the argument that I am making here.
condition c) is either redundant or it continues to foster reservations pertaining to infinite regress. Either way, we have yet to bridge the gap between the psychological states of satisfaction and self-governance on one side and those of normative endorsement on the other.

It might be worth reflecting here for a moment on the metaphors and images that Bratman’s conceptual network employs in the effort to account for agency. The superficial worry is that concepts such as “planning,” “policies,” and “self-governing policies” more adequately describe and prescribe the functioning of institutions or corporations, rather than that of human persons.12 But focusing on surface grammar might provide reason to dismiss too facilely Bratman’s proposal. Instead my main evaluation focuses on whether the functional roles of psychological states described in institutional or corporate language can adequately reveal the core of human agency. My diagnosis is that the normative regress problem intimates deep-seated problems for planning agency. As I lay the framework for an alternative understanding of agency, I begin with a suggestion and a new metaphor from Marcel Lieberman’s interesting work at the interface between agency and moral realism.13

2.2.2. Lieberman and Commitment

Lieberman insightfully suggests that the complex psychological state of commitment is more fundamental for deliberation and persistence of agency than the

12 Plato is of course upfront with his reader about the establishment of a Utopian Republic. He tells us that he will first create the Utopia because justice in the aristocratic state would give us a picture writ large of the justice in the soul. Had Bratman first given us an account of a just or a righteous corporation as Plato does for the polis, his account of moral agency would have seemed more attractive.

13 See Marcel Lieberman, Commitment, Value, and Moral Realism, (Cambridge, 1998)
plan-like attitude of policy. Although the psychological functions of commitment overlap those of Bratmanian policy (it guides practical reason, integrates the agent diachronically and facilitates interpersonal coordination) Lieberman holds that it diverges on account of the central role self-understanding plays in the formation and maintenance of commitments:

The difference between commitments and policies might best be described with a spatial metaphor: substantive commitments lie nearer the center of who the agent is, while policies are found at the periphery. Although there will be occasions in which one must act on the basis of commitment or policy, or be forced to reconsider them, the failure to act on a commitment occasions a further reconsideration of one’s self-conception or a reassessment of one’s values, both of which further destabilize the commitment, while reconsideration of the policy ends with the policy itself – the reconsideration does not reverberate to the center.  

Let’s take stock for a moment. I have argued that Bratman’s interesting merger of volitional hierarchy and planning structures still falls prey to a version of the endorsement regress problem. Even if the agent becomes increasingly adept at supervising and governing his desires, beliefs, and policies, this does not guarantee that she endorses or values these attitudes or practices. Diachronic volitional supervision does not entail either moral normative force or authenticity. It is at this point that Lieberman’s analysis of commitment may prove especially fruitful. On the substantive commitment model, the agent endorses and values her actions because her cross-temporal motivational structures are filtered through self-understanding. My positive proposal of how self-understanding shapes deeply our agency will have to wait until chapters four

---

14 Lieberman, p. 83

15 Notice that one helpful advantage the concept of “commitment” has over “self-governing policy” regards tightening the connection between self-understanding and valuing. I can treat a certain desire as a self-governing policy, without being committed to it. I can be deeply committed to a desire or relationship without expressing it in an explicit self-governing policy.
and five. The main aim of the rest of this chapter is to support the modest thesis that some psychological attitudes essential to self-understanding cannot be subsumed under Bratmanian plans or self-governing policies.

In the first stage of my investigation I explicate how a Lieberman-inspired model puts pressure to refine or rethink a number of Bratmanian assumptions about agency. Thus, I investigate briefly a) the pressure to favor diachronic commitment over volitional hierarchy as a guide to what the agent values, b) the pressure to favor self-understanding over self-governing as a guide to self-valuing, and c) the pressure to favor an interpretive conception of self-understanding rather than one relying on strict causation.

In the second stage, I argue that commitments lie closer to the heart of agency than policies because they involve diachronic states that are essentially retrospective. On the one hand, retrospective states undermine the diachronic stability of self-governing policies, loosening their motivational and normative force. On the other hand, retrospective states are important to self-understanding partly because they involve chains of narrative (or semantic) relations and adopting certain dramatic postures that are jointly essential for meaningful action. Again, a more detailed theorizing of narrative understanding and dramatic posturing will have to wait until chapter four.

2.3. Relational Commitments and Policies

Oscar Schindler appears volitionally integrated. At least in Steven Spielberg’s portrayal from Schindler’s List his agency seems organized around self-governing

---

16 I was only superficially inspired to use this example by reading Nomy Arpaly and Timothy Schroeder’s “Praise, Blame and the Whole Self” (Philosophical Studies 93: 161 – 188, 1999). They use it and cases like it to argue against all versions of Frankfurt-inspired hierarchical views of agency. They argue that cases of reversed akrasia show that the practical self identifies with actions outside the perimeter
policies that involve the grotesque idea of profiting from the war by employing Jews from the Warsaw ghetto that otherwise may be doomed. It is not that the audience does not get glimpses of Schindler’s inner conflict or of the possibility that empathy for Jews has a genuine effect on his actions throughout the narrative. Nevertheless, his most consistently manifest motivations are those of a disenchanted war profiteer. According to Bratman’s position, Schindler’s most authentic self is expressed through these temporally engrained higher-order planning structures.

But the end of the movie reveals an intriguing phenomenon: Schindler makes known that his desire to save the Jews has been operative all along. He mourns not having saved more Jews when he could have easily done so by selling his car (for ten Jews) or his golden pin (for two Jews). The protagonist’s heart-felt lament and the audience’s reaction to it are puzzling on Bratman’s account. For as I have described the case, the altruistic desire to help the Jews has been diachronically kept largely external to Schindler’s core agency. Consequently, an altruistic expression seems like a betrayal of Schindler’s planning agency or perhaps an instance of weakness of will. This may be a natural and plausible interpretation of the story. I do not mean to cavalierly dismiss the possibility that the audiences unwittingly project their own altruism onto Schindler, or that he lacks relevant self-knowledge in his expressions of regret. Even more cynically, of the volitional core of agency. While I sympathize with their general intuition, I highlight the problem for the diachronic versions of hierarchical accounts of agency. Further, I hold that reversed akrasia reveals a particular problem for Bratman, but remain agnostic about whether a similar problem is posed for Frankfurt. For the latter, although volitional hierarchy is conducive to motivational effectiveness, it need not be constitutive of the agent’s deep moral or rational identity. By the same token, there might be features of one’s practical self (instances of strong evaluations) that bypass volitional hierarchy. But unlike Frankfurt, Bratman *aims* to ground moral authority in the core motivational structures of agency, so his plan-hierarchical hybrid account seems more exposed to the problem signaled by Arpaly and Schroeder. I argue in the next chapter, that if love has the diachronic normativity that Frankfurt says it does, then his view will be threatened by analogous cases.
with the Nazis defeated and his own survival at the mercy of the Jews from his factory, Schindler might be putting on a compelling show to save his skin.

Deflationary or cynical interpretations aside, there are also compelling reasons to believe Schindler’s latter-day altruism has unfeigned motivational force, more or less central to his self-conception. First, in the narrative it is the deep act of gratitude by the Jews that elicits Schindler’s expression. They offer him a freshly fashioned golden ring with an inscription in Hebrew “He who saves a life saves the world”, along with a paper signed by all the Jews from the factory advocating his amnesty in case he’s captured by the allied forces. Second, his spontaneous expression of altruism is reliable because it reveals how a certain instrumentalist aspect of his planning self has remained stable across time. Believably, his empathy is integrated with other attitudes revealing that implicitly Schindler still values maximizing outcomes. Relevantly, the relevant transformation occurs only in what Schindler now counts as worthy outcomes. His expression reveals that somehow he has implicitly moved from assigning ultimate value to money to that of saving human lives. Thus, if Schindler’s expression reveals a firmly planted altruistic desire, it is either because it has persisted in diachronic attitudes different from the original self-managing policies, or has been incognito present in them. Or, still yet because it is possible for an ephemeral desire to catch hold instantaneously in such a way as to unravel a fully integrated volitional profile. On any of these possibilities however, Schindler’s altruism cannot be explained fully by appealing to manifest expressions of self-governing policies.

A general trouble for the hierarchical-planning view thus is the possibility of such intelligible akratic agents. This agent’s action has an obvious moral, prudential, or
virtuous worth even though the motivations underwriting it have been effectively banished from the center of his agency. So are there any resources available for Bratman’s view to explain how an agent’s transforming self-conception informs or is informed by her manifest self-governing policies? Can it explain the phenomenon of Schindler’s mourning and regret?

As we have seen, for Bratman planning structures contain implicit valuations which structure agency over time. These valuations however, do not necessarily imply stronger notions of moral endorsement. Plans reveal stable and temporally coherent patterns of preferences. Recall also that a self-governing policy is a plan-like attitude supported by and supporting the Lockean continuity of the self. It is because of their ties with links constituting the persistence of agency over time, that self-governing policies allegedly have authority to be effective in deliberation and to guide action. Furthermore, the valuations expressed by higher-order planning structures are not just assembled as a sum-total of the valuations expressed by lower-level planning structures. Presumably these valuations are “low-level” because they are implicit in plans and intentions that have not been reflectively endorsed. Rather, the self-governing policies have special authority. Says Bratman, they “feed back to and help shape what such a creature can plausibly be said to value.”

The previous point is especially telling for at least two reasons. First, it emphasizes an authorial asymmetry holding between higher-level and lower-level conative attitudes. Higher-level policies have authority to shape the values at lower-levels in ways that the reverse does not hold. Second, there is a dynamic,

---

transformational aspect to the interaction between levels. First-order plans and preferences are being transformed and molded in light of higher-order policies. So, from the asymmetrical way in which the preferences at higher-order plan levels reflectively transform the considerate preferences implicit in lower planning structures, a picture emerges of a presumed stronger sense of valuing (endorsement).

If Bratman’s account is to accommodate Schindler’s expression, it needs to explain his transformation in terms of higher-level volitional attitudes. As I have argued above, if the transformation is explained by appeal to lower-level plans or by non-volitional attitudes, this would be utterly mysterious and unmotivated given authorial asymmetry. But why not believe that attitudes at any level could be supported by and be supporting of Lockeian ties of persistence over time? If Schindler’s case is a possibility sketching a compelling story of diachronic transformation, commitments essential to self-conceptions can take hold and survive independently of being expressed in higher-order policies. It seems the main criterion for authorial control involves persistence of agency over time, and this in turn need not to involve essentially explicit self-governing policies. Does Bratman have plausible reasons to hold that diachronic self-conceptions are shaped by commitments expressed primarily through such policies? He seems to formulate a response to the worry that self-understanding is prior to self-government, for in at least one place he states that the unique authority attaching to higher-level policies is grounded in the special perspective of the agent:

Begin by asking: what constitutes the perspective of the agent within which the self-governing policy is seen as having presumptive normative authority? Well, so long as we are talking about agents like us, the agent of the action is an agent who persists over time and whose agency is temporally extended. And this is a deep fact about the kind of agents we are; or so I have averred. So it is natural to understand the relevant, agential perspective as the perspective of the temporally
persisting agent whose agency is temporally extended. And that may well be what we are thinking when we say that within the agent’s perspective the self-governing policy has presumptive normative authority. This thought would be explained and justified by the tight connection between such self-governing policies and temporally extended agency. Since self-governing policies have, because of this tight connection, agential authority, they play a central role in constituting the very perspective that is at issue in such talk about presumptive normative authority. Since these policies in part constitute the relevant perspective it is no surprise that within this perspective they have a special authority.\textsuperscript{18}

Bratman’s argument can be reconstructed like this:

(1) Our temporal extendedness is a deep fact of agency

(2) The agential perspective is the perspective of the temporally extended agent.

(3) Self-governing policies partially constitute the relevant agential perspective

(4) So, self-governing policies are essentially involved in the perspective of the temporally extended agent.

(5) Whatever gets involved in the perspective of the temporally extended agent enjoys special normative authority (implied premise)

C: Therefore, self-governing policies enjoy special normative authority

An interesting feature of this argument is Bratman’s implicit concern to ground agential authority in deep features of agency, a concern congenial to my project. Another is Bratman’s recognition that diachronic self-conception is essential for agency. Premise (2) above suggests that the deepest features of our agency can be filtered through our diachronic self-understanding, our time-bound self-conception. This is another agreeable feature of the account, a feature making the conceptual distance between Lieberman’s view of commitment and Bratman’s understanding of policy seem not so vast.

Premise (3) however is problematic. It is ambiguous between two readings:

(3*) Self-governing policies partially constitute the relevant agential perspective in a way that is privileged over other attitudes that also partially constitute that perspective. (privileged constitution)

and

(3**) Self-governing policies partially constitute the relevant agential perspective in a way that is not privileged over other attitudes that also partially constitute that perspective. (non-privileged constitution)

If Bratman is committed to something like a (3*) reading of (3), this premise is too strong. Why should higher-order self-governing policies enjoy a special constitutive relation to the agent’s self-conception? The main problem here is that no argument is provided to exclude diachronic attitudes distinct from self-governing policies from getting involved in privileged relations of diachronic constitution. If the deep features of agency are filtered through one’s diachronic self-conception, it would take some argument to show that the psychological attitudes particularly advantageous in shaping one’s temporal self-understanding are plan-like, let alone hierarchical plan-like policies. Why should “high” and forward-looking psychological states be privileged articulations of agent’s “depth”? It is at least as plausible to suppose that commitments or psychological attitudes not couched in policy-like language can constitute a deep feature of agency.

A related problem here concerns an intuitively shallow sense of self-understanding that emerges from Bratman’s account on the assumption that (3*) is true. If self-understanding is diachronically fixed primarily by the asymmetrical interaction
between higher-order policies and lower-level plans, this agential perspective seems rather impoverished. Recall from the prequel that Lieberman prefers to ground self-conception in the psychological attitudes of substantive commitment rather than in policies partially because he thinks the latter hover toward the fringes of one’s identity whereas the former come closer to the center.19 But can we identify types of psychological states that fit Lieberman’s functional description of “substantial commitment”?20 And in what sense are substantial commitments closer to the heart of the agent’s diachronic perspective than self-governing policies?

Intuitively, some commitments seem to reach deep to the heart of one’s self-conception in a way that cannot be reduced to policy-like language. For instance consider relational commitments such as husband’s faithfulness to his wife. It seems to me that although this particular commitment could be expressed through any number of explicit self-governing policies, (say, “make time for getting to know wife”; “don’t give undue attention to other women”; etc), the complex emotional-normative content of such a diachronic attitude (or set of attitudes) reaches deeper than the set of policies supposed to codify them. Generally, it seems to me that policies are underdetermined by the relational commitments they are supposed to codify. If I am right about this underdetermination feature, it seems that the commitments intrinsic to deep, meaningful

---

19 Interestingly all sorts of ritualized behaviors might look like they express some higher-order policies that the agent wouldn’t endorse as part of her self-conception. Consider the consistent use of spanking as an educational tool in Eastern European block countries during Communism. A student might justifiably believe that his teacher uses spanking as a policy for cultivating in students the desire to learn. At the same time the teacher might not at all conceive of her-self as a “spanker”, nor value this practice. Similar observations could be made about addictive and obsessive behaviors.

20 Lieberman seems to think that there is no difference in kind between being committed to a friend, to a friendship, and to finishing a project. Each of these represents a form of substantial commitment for him. I am doubtful about this, but at this stage in the argument I am not concerned with drawing subtle distinctions among various forms of commitment. Suffices to say at this point that on my view commitments to persons are prior to and paradigmatic of any other form of commitment.
relationships are epistemologically prior to the self-governing policies that constitute an agent’s perspective.

Return for a moment to the Schindler case. Suppose that initially his agency is constituted by commitments to maximizing financial profit, but not explicitly to saving the Jews. Predictably, we would expect Schindler’s self-governing policies to express these commitments. But suppose that over time his commitments change initially to saving the Jews as a reliable way to increase revenue, and finally to saving Jews even at the cost of Schindler’s financial ruin or that of his personal safety. Such transformations are plausibly explained in the story by his acquaintance with the Jews and by developing deep, meaningful relationships that allow him to value their humanity. At the same time, Schindler’s deepening commitment to saving Jews is compatible with any number of apparent self-governing policies. For large intervals of his emotive transformation he can even retain the self-governing policy of a staunched industrialist. Indeed, Schindler shrewdly intuits that the only way he could act on the desire to save some Jews is to be explicitly committed to profiting from them.

I am not claiming that Schindler’s relational commitments are deeper only because they can be expressed simultaneously through apparently self-interested but implicitly other-regarding policies. More significantly, they enjoy a certain epistemic priority for they can explain how the change in his self-governing policies expresses the change in his commitments. For plausibly Schindler moves from policies expressing explicit and implicit self-interest in a first stage, to those expressing explicit self-interest but implicit other-interest in a second instrumentalist stage, to merely claiming self-
interest as a disguise for holding policies that express explicit and implicit other-interest in a third stage, to finally claiming explicit and implicit other-interest.\footnote{The plausibility of this interpretation depends crucially on what I called “the third stage”. A crucial episode in the narrative lends support for this view. Schindler decides that his factory will produce defective military equipment, and he entrusts this information only to main Jewish manager of the factory. At the same time he intensifies his campaign of buying as many Jewish workers as he can. This is not a just a risky business decision, but an unintelligible one on the assumption that he is still implicitly committed to increasing revenue, or ensuring his personal safety.}

Even if deep relational commitments do not always enjoy epistemic priority over self-governing policies in the constitution of diachronic self-understanding, from the standpoint of meaningful valuation they must be more fundamental. It seems that knowledge of which, if any, self-governing policy is salient in the constitution of agential perspective would have to depend on what is meaningful to the agent at that particular moment. A faithful husband is committed to his wife because he values her, and arguably it is this meaningful valuation that primarily shapes his self-conception as a husband. To see this point consider two possible scenarios, each describing a husband’s commitment to his wife.

In one case, certain first-order desires of faithfulness toward the wife become part of the husband’s self-governing policies by repeated acts of reflective endorsement. Over time, however he begins ever so slightly to assign more importance to his policies emerging from his self-conception as a husband, rather than to the needs and desires of his wife. More charitably perhaps, the wife’s concerns continue to enter the agential perspective of the husband but they do so \textit{primarily through the content of his self-governing policies}.

In a second case, let us suppose that it cannot be said that the husband reflectively endorses any explicit higher-order policies of faithfulness with respect to his wife.
Rather, he is concerned with her needs and affections, and overtime he becomes quite emotionally attuned to her deep concerns. One of her concerns is that he be faithful to her. His faithfulness to her is primarily constituted by his first-order diachronic attitude of valuing her and her concerns. Over time, as a result of maintaining this pattern of valuing, he quite naturally comes to regard himself as a faithful husband even in the absence of explicit higher-order policies. All other things being equal, the husband’s commitment in the second scenario seems to me at least as central to his self-conception as that of the husband in the first case. Furthermore, in the second scenario caring for the wife or for the relationship itself can easily be seen to effectively replace the function self-governing policies. All things being equal, comparing the two scenarios shows that relational commitments without policies are at times preferable to policies without relational commitments.22

If self-governing policies are needed in relational commitments, they should be better thought of as psychological structures that under certain conditions express the deep aspects of one’s self-conception. Alternatively, they could be thought of as pointers that other non-planning attitudes have already shaped the one self-conception. In either case, it would be the meaningful-relational element that bestows the needed justification and explanation for the authority of self-governing policies.

---

22 If this line of reasoning is generalized, it seems to have the consequence that commitments to persons are axiologically more basic than commitments to rules and principles, so clearly more basic than policies. If it turns out that commitments to persons enjoy a special normative force, they may possess the right features to avoid the justificatory regress worry that notoriously plagues hierarchical accounts of agency. Additionally, if this view turns out to be right, it predicts that higher-level commitments to rules or self-governing policies will be short-lived or easily undermined when not grounded in deep commitments to persons. Conversely, when grounded in deep relational commitments, self-governing policies may indeed be expected to be stable and be effective in explaining behavior. Consequently, the view has the capacity to explain plausibly and interestingly failures, successes and transformations of agency over time.
Let me illustrate the same point in a different way. Suppose that by having an affair an otherwise faithful husband breaks his commitment to his wife. Suppose furthermore that as a result he feels shame, regret and inner turmoil. According to the (3*) reading of Bratman’s account, the husband’s feelings of regret ultimately is explained by his failure to have given due preference to some policy or set of policies which partially constitute his agential perspective as a husband. But this reading reveals a rather superficial sense of the husband’s self-conception, and it does not seem to explain copiously the husband’s deep feelings of loss, regret and brokenness. In fact, it is not clear that the husband need be aware or even modally conscious of any broken self-governing policies, in order to compellingly explain his deep feelings of loss or regret. I suggest that a better explanation of the husband’s brokenness is his coming to perceive himself as a person capable of emotionally hurting his wife and/or as a person capable of performing actions that betray the trust of his wife.

Similarly, I gather, we would be inclined to regard more valuable the lament of a hypothetical Schindler whose self-conception contained caring for Jews in the absence of explicit policies to that effect. The lament of another hypothetical Schindler who bemoaned not forming the relevant self-governing policy in the absence of relational commitments, would be not only less meaningful, but largely unintelligible.23 Breaking some self-governing policies cannot plausibly constitute a more basic explanation for either the transformation of one’s self-conception or for the ensuing self-blame. If

23 Those inclined to object to this claim of relative meaningfulness do so on the basis of two faulty assumptions. 1) That reliable action follows only from endorsing some set of self-governing policies. 2) That emotional attitudes couldn’t ground reliable sets of self-governing policies. But both assumptions are suspect. Reliable action could follow from emotional sensibilities unmediated by engrained planning profiles (I argue for this in the next section). But even if this claim is questionable, planning profiles might only have intelligibility in light of an already stable emotional sensibility.
meaningful commitments are essential to our deep agential perspective, and if the former are not always filtered through explicit self-governing policies, we have reason to investigate diachronic psychological attitudes that contain and explain relational commitments.

If I am correct in my observations up to this point, then we ought to give up on a (3*) reading of (3) and adopt (3**) as a better interpretation. But if this is correct, then Bratman’s conclusion that self-governing policies have special normative authority is no longer so clear. Undoubtedly, it is plausible to hold that self-governing policies have a kind of authority over the agent, say motivational authority, or even a special practical justificatory priority (say over the psychological pull of immediate desires), but it is a significant step to transfer that kind of authority immediately over to the normative authority flowing from one’s self-conception. If diachronic depth is really what fixes agential authority and if depth is not always transparently codified in self-governing policies, then it is an open question as to which psychological states have constitutional priority in the formation of one’s self-conception. Again, this is not to deny the motivational authority of higher-order policies, only their claim to constitutional authority, to the kind of agential authority that emerges from one’s self-conception.

In this section I have argued that cases like Schindler’s show how commitments to persons are telling generally against prioritizing self-governing policies in the constitution of agency. I have argued that while we can agree with Bratman that some authority is derived from one’s diachronic self-conception, there is no reason to suppose that hierarchical volitional structures have a privileged role in the constitution of her temporal perspective of agency. I have indicated that in the case of relational
commitments, the language of policy and self-government is at least sometimes unhelpful in unveiling the nature of self-understanding.

2.4. Self-understanding and Self-government

Bratman could agree with most of my argument thus far, but add that it is unhelpful for a general account of agential endorsement. He might object that deep relational commitments are either too idiosyncratic to shape the understanding of all forms of commitment, or that they lack a signature of reflectiveness required by agency. Non-policy like relational commitments may very well constitute deeper self-understanding in a certain range of circumstances, but these situations may be too parochial to count centrally in shaping a general account of reflective agency. Bratman could further suggest that even if relational commitments may generally constitute a deeper perspective of agency independently of volitional structures, it is more desirable for such an account to require the determination of the will at higher levels so that this deeper agential perspective could be reflectively sanctioned. That is, provided that there are connections between self-governing policies and the diachronic attitudes that happened to have historically constituted deep self-understanding, the former must enjoy a certain reflective justificatory priority over the latter. Relevant diachronic attitudes supporting meaningful commitments are validated through higher-level volitional identification. In this way self-governing policies have an edge in the constitution of agential perspective that also explains how self-understanding could be reflectively normative. Bratman’s account may thus deliver a more general account of agency in which self-governing policies possess a unique presumptive normative authority.
My strategy in answering this dual objection goes as follows: I first attempt to formulate a general response to the reflectiveness problem supposedly affecting basic commitments to persons. I aim to show that the kind of reflection important in shaping agency need not invoke self-governing policies. In order to substantiate this claim I turn to the case of Huck and Jim where we see how subtle forms of reflection could be constitutive of relational commitments. Once we see the widespread function of these subtle forms of reflection, I suggest that the parochialism challenge is also plausibly answered.

2.4.1. Reflectiveness and Self-management

First, let us approach the reflectiveness problem. Recall that planning hierarchical structures are claimed to be essential for a relevant kind of self-knowledge. The agent’s reflective self-understanding is presumably shaped mainly through reflectively endorsed self-governing policies. Bratman however allows that there is a form of agential reflection effectively tied to the functioning of first-order plan-like commitments and policies. In an attempt to integrate diachronic attitudes such as plans in Harry Frankfurt’s hierarchical view of self-constitution, Bratman holds that a mode of reflection is already constitutive of first-order plans and policies, because these diachronic attitudes already involve valuing:

But now consider an alternative model of reflectiveness. This model highlights first-order policies about what to treat as a reason in one’s motivationally effective practical reasoning; and it says that such policies are reflectively held when they are appropriately tied to (even underdermined by) evaluative reflection. Here we have a central role for plan-type commitments concerning practical reasoning (to which we can extend our account of agential authority); and we have a kind of reflectiveness; but we do no yet have conative hierarchy.  

24 Michael Bratman, “Planning Agency, Autonomous Agency” pp. 48-49
If Bratman is correct here, it is unproblematic to look for forms of reflection characteristic of non-plan-like diachronic attitudes that could be essential involved in relational commitments. Crucially, subtle valuations occur in a number of diachronic attitudes distinct from planning: loving, caring, forgiving, repenting, encouraging, empathizing, etc. Evidently, in Bratman’s view the forms of reflection associated with such first-order attitudes and commitments may not be sufficiently robust to infuse full normative authority in the agent’s perspective. But why is that?

Bratman argues that in addition to first-order diachronic attitudes, (whether they are plans, policies, or deep commitments) a typical human psychology is populated by an assortment of wild motivational demands that do not grow from these basic or deep commitments, and actually aim to undermine of them. The upshot is that human psychology needs a self-managing system outfitted with a distinct mode of reflection to control these inauspicious attitudes. Importantly, the need for this self-management system grows out of our basic need to pursue basic commitments. Here is a reconstruction of Bratman’s argument:

1. It is reasonable to suppose that typical human agents pursuing basic commitments will know about adverse attitudes to them and about the need for self-management.

2. Thus, it is reasonable to suppose that typical, properly functioning human agents will endorse a form of functioning that serves this need: a self-managing system.
3. Thus, a typical agent’s basic commitments will contain as part of their content a commitment to manage first-order dissenting motivational pressures.

4. Thus, self-governing hierarchy is “part of a reasonable response to fundamental, pervasive, and (following Cooper’s Aristotle) permanent human needs for self-management in the effective pursuit of basic commitments.”

5. Thus, it is a reasonable and common human response for agents with relevant self-knowledge to develop conative hierarchical commitments in order to shape what has subjective normative authority for the agent.

This argument drives home the point that the pressure for conative hierarchy and higher-order reflection arises out of our pursuit of basic commitments coupled with a kind of unproblematic knowledge of all those first-order antithetical attitudes that threaten to undermine them. This “pervasive self-management problem” puts pressure on reflection at higher conative levels to perform a function which reflection intrinsic to first-order diachronic attitudes could not manage. Consequently, if one desires a worthwhile degree of self-knowledge, a basic priority attaches to developing and maintaining self-governing volitional structures:

This practical problem exerts pressure on those commitments to be higher-order. And once we recognize this point, we can go on to see such higher-order commitments as potential elements in a Frankfurtian project of self-constitution. If, in contrast, we were to model reflectiveness, internality, and self-government without appeal to conative hierarchy, we would be in danger of failing to take due account of this pervasive practical problem.

There are a number of worries plaguing the argument above. First, supposing that there is a pervasive practical problem, the transition from premise 1 to premise 2 is still

---

25 Bratman, Michael, “Planning Agency, Autonomous Agency” p. 50
26 Ibid, p. 50
questionable. Why need we endorse a whole self-managing system simply on the basis of our knowledge of recalcitrant psychological pressures facing our basic commitments? Couldn’t our dealings with these detrimental or undermining motivational pressures be equally effective in the absence of a whole system of management? Under some circumstances it is reasonable and perhaps desirable to approach self-management in a more piecemeal fashion, rather than being committed to an entire system.

Suppose there is pervasive novelty in the world. Practically this means we can reasonably expect the world to take our agential perspective by surprise in ways we could not have anticipated. If this is right, it seems that we can remain committed to the pursuit of our basic commitments, but there would be no rational pressure to strongly endorse a whole self-managing system. If agency is to become adaptive to the ever-increasing complexity and novelty of our social world, to our increasingly complex diachronic relationships, at most it seems desirable to hold only tentatively to a self-managing system, if one we must have. To resist endorsing a self-managing system implies viewing recalcitrant motivational pressures in our lives not always on the combative model. More positively, the alternative implies cultivating a form of attention subtle enough to make policy adjustments “on the fly,” when faced with complexity and novelty. Strongly endorsing a whole self-management system in a way as to confer it constitutional priority may over time render the agent disinclined to notice subtle changes in oneself, others and their interrelationships and may leave the agent ultimately handicapped in making timely adjustments. And this is hardly a desirable feature of agency.
Second, even if we suppose that some self-managing system is rationally desirable for agency, does this mean that there is rational pressure to adopt conative hierarchy? (This is to question the transition from premise 2 to premise 3) The pressing question here is whether Bratman’s self-governing model is superior to competitive models that do not rely on volitional hierarchy and seem equally compelling in securing the agent’s pursuit of basic commitments when facing psychological forces that tend to undermine them. Consider one such account, that of “silencing” originating with John McDowell.27 Unlike the merely continent, according to McDowell, the virtuous individual’s moral perception is so attuned to moral reality that she no longer explicitly rejects or overrides competing motivational forces. The relevant perception of what matters morally contains such a strong attention to the virtuous features of reality, that it silences all reprehensible concerns without a struggle.

Analogically, we could reply that a self-managing system could emerge from the pursuit of our basic commitments that might function on the “silencing model”. The agent might possess a mode of attention internal to her diachronic attitudes that already “silences” those concerns that are detrimental to his commitments. I am not arguing here that a McDowell-inspired model is more plausible than Bratman’s. My minimal point is that if a particular kind of attention to one’s commitments could plausibly function as a silencer of competing psychological forces, we can control our commitments, but without explicit self-governing hierarchy. So, pace Bratman, self-management systems could function in the absence of explicit policies.

27 John McDowell, “Virtue and Reason” and “Are Moral Requirements Hypothetical Imperatives?”
Let’s suppose that my first two objections either miss the mark or that Bratman has a satisfying response to them. There is still a third significant worry. Recall that Bratman is trying to argue that a form of reflection associated with higher-order policies is the natural response to the pervasive management problem, and that this is different in kind from the reflection inherent in basic, relational commitments. But even if this is true, higher-order reflection must always depend on a prior form of reflection which does not involve conative architecture. This may seem like a trivial and uninteresting point, but I think it has important consequences for the points made earlier.

From the outset Bratman assume that understanding which psychological forces undermine basic commitments is commonplace for agents like us. Clear cases of such motivational pressures, he claims, are “certain bodily appetites, and certain forms of anger, rage, humiliation, indignation, jealousy, resentment and grief.”28 Interestingly Bratman lists together indiscriminately types of emotions, moods, appetites, other-directed and self-directed attitudes, and psychological states that may be more or less associated with or constitutive of diachronic states. If there is a label that could characterize them all, it would be something like “recalcitrant psychological pressures”.

But this brings us to an important point. Knowing which psychological pressures are recalcitrant with respect to our basic diachronic commitments is an intricate and refined achievement. Reflection associated with self-government already assumes this sensible understanding, which is not always part of the content of policies themselves. Furthermore, I claim that sometimes it is desirable for this discriminating attention not to be swallowed up by the reflection associated with self-government. In fact, sometimes

---

higher-order reflection should depend on a prior form of reflection, one involving careful discrimination among those motivational pressures that are advantageous, ambiguous or undermining with respect to our basic commitments.

It may be objected that this form of attention could be part of the content of some self-governing policies. But if attention falls under the tutelage of higher-order policies employed to combat recalcitrant dispositions, it will tend to regiment these in psychological clusters that remain motivationally inert. This seems to follow from the combative model that assigns motivational and evaluative priority to higher-order policies. And my point is that attention shouldn’t always function like this! For sometimes motivational forces in our psychology do not come with ready-made labels neatly dividing the labor between the faithful and the unfaithful. Sometimes we need prolonged diachronic attention to figure out which psychological dispositions will turn out to be agency enhancing or agency undermining. So, reflection should not only keep the unruly at bay, it should be expected to perform the fundamental and complex function of discriminating between valuably disruptive dispositions and those manageable but unhelpful. Thus, this discriminating form of reflection seems to me prior to and more basic than the higher-order reflection involved in advancing self-managerial policies.

Let me only briefly illustrate some complex ways in which so-called recalcitrant attitudes can relate to our diachronic basic commitments. This underscores the necessity of acknowledging a mode of reflection internal to the agent’s perspective that takes into consideration the quality of her inner emotional life. In some cases attitudes labeled “recalcitrant” are actually supporting or enhancing of our long time commitments. For instance, grieving the loss of a dear one or expressing indignation at injustices can be
highly unmanageable attitudes threatening the stability of our basic commitments. But in central cases they are appropriate responses.

A parent’s mourning the loss of a child expresses a deep form of commitment and mutual love. Typically the absence of grieving in such circumstances can be problematic, underscoring either lack of self-knowledge, or callousness, or unwarranted stoicism. Furthermore, some recalcitrant attitudes indirectly give rise to projects and diachronic attitudes that are essentially tied up with our deep commitments. A politician’s retrospective indignation at her apathy with respect to the Rwandan genocide, for instance, could drive her to engage in projects that transform her into an individual essentially committed to social and international justice.

Some recalcitrant motivational pressures are thus not always antithetical to our ground projects, but they can nourish, support, and accentuate them. Others can indirectly trigger commitments that turn out to be basic to our diachronic practical identities. But in some situations certain recalcitrant psychological pressures (anger, bitterness, and resentment) should challenge or even reconfigure our basic commitments.

Consider situations in which the repertoire of moral concepts used by agents to express commitments is impoverished due to culturally or socially internalized practices of oppression or subordination.

As Diana Tietjens Meyers points out, employing only a repertoire of safe and manageable attitudes (she refers to this as “emotional vanilla”) could render moral perception “impervious to cultural or institutionally entrenched forms of injustice.” But if our emotional responses outstrip particular conceptualization schemes, then proper

---

sensitivity to the rancorous emotions can in certain circumstances be our only accurate pointers of injustices and oppression. Furthermore, Meyers points out that being attentive to and calibrating bitterness, anger and resentment can develop a state of the art capacity of *dissident moral perception*, which is the capacity to know when to move from “vanilla” to “vindalloo.”\(^{30}\)

It is less than clear that we ought to protect our basic commitments against a monolithic, clearly-defined set of recalcitrant psychological attitudes. As such, it becomes increasingly less obvious that Bratman’s pervasive practical problem is most fundamentally a self-management problem. Rather, more pressing seems to be the problem of self-knowledge and attention to our practical worlds. In a worse case scenario for Bratman’s view of self-management, if relevant controlling functions are built into the content of practical perception necessary for deep self-understanding, higher-order reflection and conative structures seem redundant. In more fortuitous circumstances self-managerial psychological structures may be necessary, but have to be perpetually calibrated in response to a detailed and subtle form of self-awareness.

Proper self-knowledge entails a trained, evaluative attention to a whole gambit of emotions, moods, desires, as well as basic agential commitments better construed on the dialogical model that I propose in chapter four. In contrast to the combative model seemingly proposed by Bratman, engaging in a reflective evaluation of a psychological attitude involves taking on a perspective different from that of this attitude. But this reflective distance need not entail volitional hierarchical structures. To adopt the dialogical model of attentive self-understanding is to reject the dichotomy that assigns

---

\(^{30}\) Ibid, p. 209.
default authority to some mental states over others. Surely, some basic projects and commitments come with a ready claim to significance. My claim is that if we start from the significance and meaningfulness that we attach to deep relational commitments and in the process employ self-reflection as attention to the voice of other psychological attitudes, our model will uncover central functions of agency more basic than, or at least non-reducible to, Bratmanian self-governing framework.

Let us take stock again. In this section I have argued that Bratman’s pressure to introduce self-governing policies in a prime justificatory-conferring role with respect to basic commitments is hardly compelling. I have argued that pursuing our basic commitments need not rationally pressure us to endorse whole self-managing systems. And even if it does, such systems need not involve conative hierarchy and higher-order reflection. Finally, I have argued that under closer scrutiny the motivation for introducing the planning hierarchy and higher-order reflectiveness is rather superficial. Perhaps it is too bold to claim that there is no self-management problem. A more modest and compelling thesis is that self-management depends on an exquisite form of self-understanding. The refined form of reflection that entails paying attention to our desires, emotions, and moods and their relation to our basic commitments, is not reducible to a set of self-governing policies and seems a precondition of proper self-managing.

2.4.2. Huck Finn and the Priority of Emotional Self-understanding

Let me now develop the intuition about the priority of self-understanding over self-government. I do so by considering a well-known case, commonly taken in the literature to illustrate an instance of reversed *akrasia*. This is the fictional, yet convincing
case of Huckleberry Finn. I start by critically considering the early interpretation of the case by Jonathan Bennett and the way in which Bratman might conceptualize Huck’s dilemma and his agency. Finally, I argue that an alternative interpretation of Huck’s psychological underpinnings consistent with prioritizing self-understanding has significant explanatory advantages.

The contours of the story are commonplace. Resisting the advice of his conscience, Huckleberry Finn befriends Jim the runaway slave.31 Having absorbed the social norms of 19th century Mississippi valley, Huck’s conscience demands to return Jim to his rightful owner. But when the time arrives for Huck to act decisively toward this end, he is first paralyzed by indecision. When he is psychologically capable of acting again, Huck facilitates Jim’s escape, contrary to the voice of his conscience. Jonathan Bennett sees in Huck’s psychological incapacity to act on moral principles a massive failure of rationality and determination of will. Bennett does not frame this as a failure resulting from rational conflict, because on his interpretation, Huck’s emotions are irrational. According to Bennett, Huck’s emotions cannot reveal any aspect of his self-awareness, thus they are unable to provide any reasons to counterbalance the force of “rational” demands:

…And when the moment for the final decision comes, Huck doesn’t weigh up pros and cons: he simply fails to do what he believes to be right – he isn’t strong enough, hasn’t “the spunk of a rabbit.” This passage in the novel is notable not just for its finely wrought irony, with Huck’s weakness of will leading him to do the right thing, but also for its masterly handling of the differences between general moral principles and particular unreasoned emotional pulls.32

31 Huck’s conscience could be taken as constituted by a set of moral beliefs and dispositions. Twain is masterful at ironically describing Huck’s own interrogation and mistrust of his conscience.

In a more plausible interpretation, Jennifer Rosner argues that Huck does not display weakness of will. She holds that early in the decision-making process Huck vacillates between determining his will with respect to his moral convictions and with respect to his desires to help Jim. But the fact that Huck fails to become unified in his volition with respect to morality does not entail that Huck fails to have any volitional determination. Revealingly, Rosner quotes Huck’s own account:

I was a-trembling, because I’d got to decide, forever, betwixt two things, and I knewed it. I studied a minute, sort of holding my breath, and then says to myself: “All right then, I’ll go to hell”…I would take up wickedness again, which was in my line, being brung up to it, and the other warn’t. And for a starter I would go to work and steal Jim out of slavery again; and if I could think anything worse, I would do that, too; because as long as I was in, and in for good, I might as well go the whole hog. (Twain, 1981 p. 206)33

For Rosner this passage indicates that Huck identifies wholeheartedly with the desire to help Jim, expressing thus a fundamental commitment inherent in a particular volitional structure. Because he is steadfast in his identification with the desire to help Jim, this “endorsement appears to be revelatory of Huck’s real self.”34 It is not clear whether Bratman would agree with Rosner’s interpretation. If Bratman sides with Bennett over Rosner, his view is problematic since the paragraph just cited reveals that Huck is a conflicted and reflective agent who aims to help Jim for understandable reasons. If Bratman endorses Rosner’s interpretation of Huck’s agency, his planning agency has to depend, I argue, on a subtle form of emotional self-understanding.35

34 Rosner, p. 110.
35 In chapter four, after I have introduced my positive account of narrative-dramatic agency, it should become clear why Huck’s emotional self-understanding is more basic than his plans or the determination of his will.
I suggest a different way yet of interpreting this story, a way that lends credibility to the model of agency that assigns priority to self-understanding over self-government. I agree with Bennett that Huck’s actions are caused by his emotions, but sustain that Huck’s emotions are meaningfully and rationally informed by his subtle self-awareness or self-discovery. I agree with Rosner that Huck is in some sense volitionally unified, but hold that this unification is filtered through a substantial sense of emotional self-understanding. The process by which Huck attains the determination of his will is complex, precisely because intricate and dissonant emotional factors play a larger role in shaping his self-understanding than is commonly noticed.

Against Bennett’s interpretation, it is not clear that the emotions which render Huck incapable of turning Jim in are irrational. It is true that Huck describes his action as a “weakening” and as a failure of his manhood, but this in itself is insufficient to sustain the charge of irrationality. Huck’s lack of spunk can reasonably be attributed to his strong compassion for Jim, so the critical issue is whether his concern for the runaway friend is rational. In my conception Huck is reflective, but his practical rationality resides neither in following abstract principles of a rule-based morality, nor does it merely express sophisticated self-governing policies marshaled for the purpose of realizing coherent plans. His is a reflective ethical action that expresses a subtle form of self-knowledge growing out of sophisticated emotional outlook. This emotional sensibility, although not articulated in classy moral concepts or judgments, belies a germane perception of moral facts.36

36 Notice too that the Huck and Jim narrative can also support the previous point that commitment to persons is more basic than commitments to principles, or self-governing policies.
Evidence for this claim can be extracted from a number of places in the story. Immediately preceding his running into the men who hunt for runaway slaves, Huck pretends to go scouting but he is actually planning to turn Jim in:

As I shoved off, [Jim] says: “Pooty soon I’ll be a — shout’n for joy, en I’ll say, it’s all on accounts o’ Huck I’s a free man…Jim won’t ever forgit you, Huck; you’s de bes’ fren’ Jim’s ever had; en you’s de only fren’ Jim’s got now.” I was paddling off, all in a sweat to tell on him; but he says this, it seemed to kind of take the tuck all out of me. I went along slow then, and I warn’t right down certain whether I was glad I started or whether I warn’t. When I was fifty yards off, Jim says: “Dah you goes, de ole true Huck; de on’y white genlman dat ever kep’ his promise to ole Jim.” Well, I just felt sick. But I says, I got to do it—I can’t get out of it.37

Notice first, when Jim reminds Huck of the way he conceives of this whole affair, Huck experiences characteristic psychosomatic reactions of sweat, emotional vacillation, and finally sickness. From Jim’s perspective Huck is a best friend, the only trustworthy white man Jim has met, a man of proven commitment and the one who will secure Jim’s freedom. It seems to me that the increasingly violent emotions that Huck is experiencing in response to Jim’s beckoning voice are physiological inscriptions that Huck is moved to gradually identify empathetically with Jim’s perspective. A further point here is to draw attention to Huck’s locution quoted in the previous citation, where he expresses his determination to help Jim: “All right then, I’ll go to hell!” This remark is illuminating especially since throughout their association, Jim has been emphasizing that he would go to hell for Huck. What this shows is a gradual but determinate empathetic identification that ends up in a deliberate endorsement of Jim’s perspective by Huck.

Second, notice that for Jim the force of principled rational morality is partly based on his empathetic identification with the perspective of others. In dialogue with his

---

37 Mark Twain, The Adventures of Huckleberry Finn, as quoted in Bennett, p. 29
conscience regarding his knowledge of the runaway, Huck blames himself not for violating some abstract moral rule, but for causing harm to Miss Watson:

Conscience says to me: “What had poor Miss Watson done to you, that you could see her nigger go off right under your eyes and never say one single word? What did that poor woman do to you, that you could treat her so mean?...I got to feeling so mean and miserable I most wished I was dead”.

Similar to the scenario described previously, Huck’s feelings give voice to his conscience, which here serve as a surrogate expressing the perspective of Miss Watson. If the reasons for which Huck is moved to any action lie deep in his basic empathetic responses, then pace Bennett, Huck’s conflict is not between some psychological states which are rational and some which are not. If the force of principled morality is grounded importantly for Huck in his empathetic identification with the perspective of slave-owners, and if his empathetic identification with Jim undermines the force of “moral concerns,” then if we want to settle the rationality of psychological conflict in Jim, we need to ask: under what conditions does empathetic identification with the perspective of others count as rational? And, how reflective can we expect Huck to be given his emotional formation?

In response to Rosner’s and possibly Bratman’s conceptualization of this case, I want to stress the essential role that Huck’s self-understanding plays in his becoming volitionally unified and the role that the emotions play in his self-understanding. I have already suggested that due to his empathetic identification with Jim’s perspective Huck comes to see himself as a person who would “go to hell” for his friend. But there is another aspect of Huck’s self-conception, aspect that emerges helpfully from Rosner’s quote. Huck perceives himself as a rebel. He points to his rebellious lineage, to his

---

38 Bennett, p. 28.
rebellious dispositions, and he becomes volitionally unified not only in helping Jim, but also in endorsing the lifestyle of a perpetual dissident with respect to the norms of his society.

It is hard to see whether Bratman’s account could accommodate our sense that Huck is functioning as an agent in this situation. In fact, if Bratman is correct, Huck’s agential perspective is shaped primarily by his commitments (which are plan-like) and the self-managing system yielding self-governing policies. But if that were true, Huck’s decision to save Jim would fall outside the perimeter of his enduring planning self. For as he rows away, we are finding out that Huck’s plans are to turn Jim in, that he has no genuine enduring plans of seeing Jim to freedom, and that he has at his disposal a self-managing system directed at squashing his natural empathy for Jim.

If Rosner is correct and Huck’s volitional unification reveals his true self, this is not due to any system of plans, policies, or self-managing attitudes that are already set in place. Rather, Huck’s endorsement follows the rather torturous path of emotional self-discovery. In the absence of any prior policy or plan, and yet on the basis of an attentive and progressive response to emotional cues, Huck seems to discover that his self-conception as a friend to Jim is basic to his identity. The realization of his basic deep commitment to Jim, accentuated perhaps by his self-narrative as a rebel allows him to silence additional competitive self-narratives that include that of an integrated citizen, a watchful neighbor, and perhaps even faithful believer. These “vanilla” psychological scripts will always be in the background of Huck’s self-understanding, perhaps waiting to be renegotiated, but for now Huck has decisively opted for “vindaloo.”
Bratman might of course agree with Rosner and accept a certain priority of self-understanding. He could add that Huck’s resolute endorsement of his desire to help Jim grows out of his basic commitment to Jim, especially when he is faced with the competing pressure which social-based morality imposes on him. On this gloss, “moral” principles might prove to be precisely those recalcitrant attitudes that Huck’s volitional determination has to manage and to beat. But volitional unification at this level could be only a sign of Huck’s deeper emotional sensibility that reinforces Huck’s deepest commitments. Moreover, Bratman’s network of plans, first-order policies, and self-government policies could not explain the transformation of Huck’s endorsement over time in the way that his empathetic self-understanding does.

I have aimed to illustrate in this section how a form of self-understanding might be prior to and more basic to than self-government. Using the case of Huck Finn I attempted to show that a form of emotional attentiveness could act as a management system upon one’s decisions. Huck’s finely attuned sensitivity to the plight of his friend, Jim, allows him to unsettle explicit self-governing policies that reflect the mores of his society. The reflection operative in Huck’s dilemma does not take the form of moral judgment. Nor is it a reflection that cords off attitudes deemed destabilizing of his plans. On contrary, it seems, Huck’s agency welcomes dissenting voices. His emotional reflection weighs the costs of remaining friends with Jim over against the costs of hurting Mrs. Watson and those of losing social and religious respectability. In the process of making fine discriminations among meaningful emotional trains, Huck comes to discover himself as a rebel, though he does not plan to be one.
2.5. Retrospection and Self-Understanding

In this section I aim to lay the groundwork for investigating some of the diachronic attitudes that explain the deep commitments to persons. Being committed to the current orthodoxy of event causation in the philosophy of action, Bratman seems to assume that any such diachronic states could be reduced to complicated relations of connectivity and continuity among events following strict deterministic causal laws. My suspicion is that a theory of human action aiming to explain agency wholly in terms of event causation has a hard time elucidating a most significant function of persons, namely the capacity for meaning-making. Distinctively human action is intelligible under descriptions of sequences of meaningful intentions. Thus, if Bratman’s account leaves out the capacity for meaningful action, it essentially ignores the capacity for self-understanding and self-interpretation. Self-government may be important, but in order to set up adequate policies, the agent needs to rely on a more fine-grained capacity which makes evaluative discriminations (articulations) on what is meaningful or significant to her.

We have seen in the previous section that self-understanding sometimes involves a mode of attentiveness to emotional states that function as pointers to meaningful events or relationships. The further question is whether we can identify some diachronic states that successfully characterize the functioning of such attentive reflection. I propose that retrospective or backward-looking diachronic states fit the bill as necessary for a practical

39 An interesting exception is Marcel Lieberman, who in his *Commitment, Value, and Moral Realism* gives an argument to this effect. He also points out in chapter 5 that even Davidson’s anomalous monism could be interpreted to resist the stronger formulations of the reductive event causation. A final comment, in recent years some theorists have begun to worry that broadly construed Lockean ties will not provide for continuity of agency across time. In the *The Diachronic Mind* Marc Slors argues in favor of semantic ties among events as constituting of personal identity

40 See the work of Charles Taylor (*Sources of the Self*) and Alasdair MacIntyre (*After Virtue*)
perception constitutive of an agent’s deep relational commitments. To see how deep relational commitments are shaped by retrospective diachronic states, consider the idea of redemptive event.

2.5.1. Redemptive Events and Retrospection

Thomas Kelly introduces the concept of “redemptive action” to refer to any present action which affects a series of past events by retrospectively endowing them with meaning. A redemptive event is any event (whether an action, perception, or occurrence) that triggers a retrospective reconfiguration of the agent’s past either by endowing it with meaning or by altering in relevant ways the meaning already present therein. Notice the meaning interdependence between the illuminated past sequence and the redemptive event. Were it not for the specific past sequence, the present event would hardly be meaningful. It may seem more like a whim or an arbitrary gesture. Similarly, were it not for the present event, the particular past sequence would not have as much relevance to the agent as it does. It would seem rather like a closed chapter in the life of the agent. It is done and put away. Alternatively, a redemptive event could simply alter the relevance of a past sequence in ways unanticipated or unforeseen initially. Even in its absence the initial sequence would have been meaningful to the agent, but due to the redemptive event the meaning of the sequence has altered significantly.

Consider Anne who gave up playing piano due to motherhood. On a lazy afternoon she fancies how much she enjoyed playing in the past, and she imagines what it would be like to start again despite her multiple current commitments. Later in the week,

41 Thomas Kelly, “Sunk Costs, Rationality, and Acting for the Sake of the Past” (forthcoming in Nous)
listening to a glorious piano concert resurrects in her a specific memory of playing piano for friends, and this memory quickly connects with other memories, creating in her the desire to take up piano-playing again. She decides to clear up her schedule for an hour every two days in order to start practicing again. She invests with meaning a series of past events connected with her being a piano-player; she acts redemptively.

Also, consider Huck’s realization that Jim has seen him as a most loyal friend throughout their voyage down the Mississippi. This realization is the redemptive event that invests with meaning Huck’s past actions and relationships, allowing him to act in the present in such a way as to continue to redeem those past events. Notice also, that while his friendship with Jim is “redeemed”, Huck’s relationship with Mrs. Watson may not be. For although there is a relevant sequence of past events that Huck could presently endow with meaning (he could act so as to return Jim to her), he will not. Also, notice that from the perspective of planning or utilitarian agency Huck’s redemptive action is personally hazardous. The moral seems to be that redemptive action can be subjectively valuable to the agent, even if it entails present risks or future losses.

Generally, in redemptive action a new intention results from the agent’s perception that past things, events, or relationships now acquire a new or different signification. This perception in turn can be triggered or enhanced by any number of events. Re-linking old friendships, giving up being resentful at those that have hurt us, forgiving or asking forgiveness, or atoning for an offense are all processes grounded in retrospective diachronic states. Interestingly, we are often thrown into retrospection by means of redemptive events that are neither planned nor subsisting in higher-order
policies. Typically, puzzlements, contingencies, or novelties of life toss us into such retrospective reflection.

Naturally, we may use plans to help us execute the redemptive intention, or in order to buttress ways of bring it about. But the relevant redemptive actions either retrieve, or revitalize, or transform sequences of past events in ways that these are now retrospectively coherent. Thus, if retrospective diachronic functions cannot be fully accounted for by plans and policies that Bratman considers basic to our self-constitution, and if the former are a significant feature of human agency, at the very least Bratman’s view needs to be supplemented by a model that can account for them.42

One might make a bolder move here. It might be held that retrospective diachronic states are not only irreducible to plans and policies, but they are more basic to agency then the latter. As we have seen, Bratman is committed to the view that the agent’s perspective is constituted by the deepest features supporting her temporal continuity. He innocently assumes that these deep features surface most clearly in basic commitments which are plan-like. But, what if the deepest features of agency become apparent not in times of agential stability (the calm pursuit of plans), but mainly in times of agential crisis and transformation? Naturally, the phenomenon of conversion, at one end of the transformational spectrum, attests to plausibility of assigning priority to self-transformation as a guide to self-valuation. A similar perception is effective in less dramatic occurrences, such as changing one’s mind about who to ask to prom. If we plausibly suppose that our deeper values come to surface in transformational or crisis times, we might be moved to an argument that assigns a basic priority to retrospective

42 Surely we can plan to be attentive to the past, but that assumes the working the capacities necessary for redemptive action and does not explain it.
diachronic states as opposed to plans.

2.5.2. Emotions, Retrospection and Narrative

To give more flesh to the contention that retrospective diachronic states are essential to self-understanding, I propose some desiderata that relevant such states must meet. I then provisionally identify the psychological states of emotions as fitting the functions of retrospective states relevant to self-understanding. A more fully articulated conception of the emotions, and of the narrative understanding, of which retrospection is but a part will have to wait till chapter four.

It seems to me that if retrospective states are to be effective in the formation of self-understanding they have to involve the following: (1) the capacity for reflexivity (they will be able to afford the agent significant self-referentially); (2) the ability to discriminate veridically and relevantly between sequences of past events; (3) the ability to articulate minimally the kind of person one wants to become; (4) the ability to evaluate and reinterpret both our past experiences and our self-conception; and (5) the ability to motivate and justify action based on capacities (1) – (4).

In chapter four I argue that this cluster of abilities is well accommodated by my model of dramatic-narrative agency. “Narrative” because human actions become intelligible, and thus meaningful only in the context of story-like plots.\(^{43}\) We come to understand ourselves and others against the backdrop of cultural, political, artistic, and religious narratives. Moreover, in the process of self-understanding we implicitly and sometimes explicitly evaluate the narrative sources of our self-conception. It is therefore

---

\(^{43}\) See especially MacIntyre *After Virtue* Chapter 15
through a number of relevant dramatic-evaluative postures that the agent can orient in the space of reasons for action and articulate a practical self-conception. This is eminently plausible case of Huck, for whom the role of a friend to Jim, unveiled by the latter’s narrative stance, makes its way to the forefront of Huck’s moral identity.

A lot more needs to be said about the model and functioning of dramatic-narrative agency. Since most of it will wait until chapter four, let me make some comments foreshadowing why I think that the emotions fill such an essential role in the constitution of retrospective states, and why they are essential to narrative understanding.

As I have said earlier, we would expect retrospective states and narrative understanding to be shaped not just by causal relations (event A, bumping into event B), but also by semantic or thematic relations. Can we say more than this? I believe that the relevant psychological states that couch such meaningful relations are the emotions. There are many reasons for this. First, there is serious empirical work showing that emotions are essentially diachronic, namely that they are constituted by the way in which they develop over time. Ronald de Sousa makes an even more intriguing claim, namely that emotions have essentially a narrative structure. They originate in what he calls

44 See especially the empirical work summarized by Craig DeLancey in his book *Passionate Engines*, and also the very nice description of it in J. David Velleman, “Narrative Explanation” *The Philosophy Review*, Vol. 112, No.1, (2003). Velleman recounts for instance that emotions: a) the first episode in the natural history of an emotion is its arousal by characteristic conditions (unlike desires); b) A series of physiological expressions (knot in the stomach, tingling, a lump in the throat) accompanied or followed by reflexive behavioral symptoms (smiling, scowling, cringing) also associated patterns of attention and interpretation: heightened awareness of danger, greater sensitivity to beauty. These are followed by c) motivational dispositions which could be either impulsive or be performed in the form of deliberate actions: attacking, caressing, gamboling. Finally d) “each emotion has a characteristic pattern of decay and extinction, involving conditions that characteristically dispel it and the mental states that characteristically remain in its wake.” (p. 13 – 14)
“paradigm scenarios” which like short dramas are “stories played out by actors of memory and phantasy, up for imitation by the flesh and blood actors of our life.”45

Second, emotions permeate all levels of experience from the pre-reflective and unreflective levels (feeling of dread, phobias) to the highly reflective level (feeling of resentment). Paying attention to the wide array of our emotional responses can then enrich our self-understanding, especially in the situations in which our moral vocabularies contain categorizations and concepts sanctioning relations of domination and oppression.

Third, the emotions and the experience of emotions are value-laden. Insofar as they help us relive or emotively remember things that matter to us, the emotions either point to values or they partly constitute them. Charles Taylor illustrates very nicely this point in talking about the experience shame. He says:

I am ashamed when I am shown up as contemptible or unworthy before others... A being can only be shamefully unworthy who is capable of a sense of shame, that is, who is sensitive to the standards which hold of a self-aware subject, and which demands that he show concern for his dignity and his good name, in short for his standing among men.46

If we take this point and the previous one seriously, we can provide another plausible reason for how retrospective diachronic states are more basic to our self-conception than plans are. If emotions outstrip conceptual categorization, and if emotions also are value-laden, the commitments formed on their basis will be more basic than either policies (Bratman) or commitments containing explicit judgments of values (Lieberman). Thus paying a discriminate attention to our emotional lives (and next

46 Charles Taylor, “The Person” in Carrithers, Collins, and Lukes The Category of the Person (Cambridge University), 1985 p. 264
chapters will develop this point) is in fact an exercise in retrieving and articulating those features of our identity that matter most for our self-conception.

Finally, the emotions enter essentially in the constitution of the moral virtues. Aristotle tells us that the virtues are rationality informed decisions which concern proper feelings with respect to oneself, others, and social goods. Perhaps it would not disturb Aristotle too much to think of the virtues as emotions infused with rationality.

2.6. Leonard: Planning Agent, Narrative Failure

Instead of summarizing this entire chapter let me end with a fictional illustration. My interpretation of it underscores the importance of narrative understanding, and the associated retrospective diachronic states as non-reducible to the planning theory. The main character of the movie *Memento*, Leonard, has a serious neurological condition that thwarts all his attempts to form new long-term memories. The condition reportedly started when his wife’s killer administered Lenny a blow to the head. Leonard can retain his life memories prior to this incident, but now he can only remember about 15 consecutive minutes of his life, after which these fresh memories fade away.

At a superficial level the plot reveals the distressing saga of Lenny’s vengeful search for the perpetrator responsible for his condition and for his wife’s death. In order to continue to pursue his target Leonard employs a complex system of Polaroid pictures and notes scribbled on his body as surrogates for his short-term memory. Despite Lenny’s plea that memory is unreliable and that only facts matter, it is soon evident that at least Teddy and Natalie (two other characters in the movie) have been feeding him incomplete or false information so they can use him for their own evil purposes. As the
story is filled in primarily from the perspective of Leonard’s subjective experience, contradictions emerge which convince the audience of the unreliability of the narrator.

At a deeper level however the movie raises significant epistemological, metaphysical, and moral questions regarding personhood, responsibility and agency. As the story is told backwardly, with the first scene of the movie figuring as the last event in the narrative, the audience is forced to experience the disorientation felt by the main protagonist. Even after having re-ordered the cause-effect structure of the reel, the audience is left to grapple with the meaning, the significance of the sequence. I think director Christopher Nolan surreptitiously wants us to see that just as the causal story of Lenny’s subjective experience can be given multiple interpretations, so the causal structure of the reel does not determine for the audience only one narrative interpretation.

So, here is one possible and plausible meaning-structure. Leonard’s failure as an agent is better explained by narrative understanding than by the planning agency. Despite his memory breakdown Leonard still has the capacity to plan for the future, and also the capacity for self-government. Albeit, these are truncated capacities, but it is evident that Lenny’s principal malfunction is not volitional. As one reviewer put it, Lenny is portrayed at once “utterly clueless and fanatically sure of himself”. Perhaps the plans drafted on the basis of Polaroid shots and practical axioms tattooed on his flesh “Find him and kill him” and “Memory is treachery” explain the minimal extent to which Lenny is an agent, but they do not explain his massive failure of agency.

Narrative understanding however does this. Leonard lacks the ability to form a narrative self-conception that is based on a veridical discrimination of meaningful events

---

from his past. What is evident is that Leonard is at each point in time a distinct time-slice agent who reinterprets his past to make it serve as an instrument for the present or some future Lenny-slice. Furthermore, he has no reliable external perspective (what I call in chapter four a dramatic persona) to act as an advisor on his present action or plans.

More fundamentally then, Lenny’s failure of agency is a moral one. Driven by the meta-narrative of revenge and distrust in a shapeless universe, he lacks the emotional understanding of the value of friendship, trust or compassion with respect to others and value of forgiveness and hope with respect to himself. It is not only the grip on his past that is loosened due to his inability to retain representational memories. Equally important is that Lenny has not emotional connection to any other reliable narrator. There are no external reliable narratives to give meaning and direction to his action. He appears to the other characters in the story as someone without a past. And this feature allows these characters to treat him merely as a hired gun to do their own dirty work. So, on my reading, this gloomy fictional world, populated by unsavory characters self-servingly reinterpreting the past in order to unleash the motives of greed and revenge, contains a moral: short-memories (vanishing of value frameworks) and strong wills bring about loss of personhood. Thankfully, this fictional world is nothing like ours.
3.1. Intuitive Constraints on Loving Agency

An account of practical human agency must meet two important constraints. The first rests on the Aristotelian intuition that the intelligibility of human action is essentially connected to our being characteristically equipped with the sense of time.\(^1\) Even if one rejects Aristotle’s strong teleology from *De Anima* or the *Nicomachean Ethics*, it remains plausible that our practical commitments are temporally dynamic. They are essentially involved in processes of formation, merger, maintenance, disruption, reinterpretation, and reintegration. Consequently, to the extent that loving commitments are an expression of an agent’s practical commitments, they too will tend to be dynamic. The second constraint is that accounts of agency should not lead to pernicious forms of practical solipsism. Consequently, a compelling analysis of love has to respect not just the intentional but the extensional reality of other persons. (I develop this point in the penultimate section of this chapter)

Here is a map of the argument in this chapter. First I examine critically Frankfurt’s canonical view of agency and the role of love in it. I then show that a plausible extension of this account can fulfill the first desideratum above. Subsequently,

---

\(^1\) This intuition is developed in a lot more detail in the following chapter where I argue that narrative understanding is an essential way of structuring agency.
I show that this plausible extension violates the second constraint. Concretely, a
Frankfurtian account of diachronic love is plagued by a pernicious version of volitional
solipsism. I suggest Frankfurt might have an answer to this worry only if his volitional
essentialism allows for psychological processes involving narrative attention to the reality
of others. Drawing on Iris Murdoch’s view on loving attention, I merely sketch here a
plausible view of narrative attention, a view that will find a fuller theoretical articulation
in the next chapter. I suggest that Frankfurt’s volitional essentialism cannot
accommodate narrative attention with ease.

3.2. Frankfurt and Volitional Love

At the center of Frankfurt’s work on practical agency is the important insight that
the distinguishing characteristic of a person is found not in her rational capacities or her
moral sense, but in her will.2 The heart of the person, Frankfurt consistently holds, is
found in hierarchical volitional structures that implicitly impose motivational or
valuational pressures on first-order desires. Evidenced in the humdrum phenomenon that
typical agents have relevant desires about their first-order desires, higher-order volitions
seem to be permanent fixtures of agency. Consider for instance a forlorn alcoholic who
desires to drink and also desires (say, more intensely but less effectively) that his desire
to drink not be so effective in moving him to action.

Even as early as his famous 1971 paper Frankfurt introduced the notion of
identification to help illuminate how higher-order volitions shape and reveal the agent’s

---

2 This does not have to imply that either set of capacities is not necessary for buttressing a person’s
volitional structures. Rather, just as a person’s rational deliberations seems to presuppose some structured
desires, or perhaps even volitional hierarchy, so it seems volitional identification with a desire can plausibly
be taken to involve a necessary reference to rational deliberation or reflection.
deepest dispositions. Considering the difference between a wanton and an unwilling addict, he focuses on the fruitful conflict the latter experiences between his occurrent first-order desires:

The unwilling addict identifies himself, however, through the formation of a second-order volition with one rather than with the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. ³

As some have observed, in explicating identification Frankfurt sometimes employs a set of concepts suggesting the active nature the will’s self-determination. We identify with what we truly desire through wholehearted decisions. At least in this original context Frankfurt seems to indicate that the active nature of identification provides a way of stopping the justificatory regress that plagues other conative hierarchical accounts:

As for a person’s second-order volitions themselves, it is impossible for him to be a passive bystander to them. They constitute his activity – i.e., his being active rather than passive – and the question of whether or not he identifies himself with them cannot arise. It makes no sense to ask whether someone identifies himself with his identification of himself, unless this is intended simply as asking whether his identification is wholehearted or complete. ⁴

Pursuing the same line of thought in a later paper, Frankfurt is forthcoming that identification most evidently involves deciding, and that this feature seems sufficient to bring wholeness, or harmony in the will, thereby conferring authorial justification to agency:

Since it is conspicuously by making a decision that a person identifies with some element of his psychic life, deciding plays an important role in the formation and maintenance of the self. It is difficult to articulate what the act of deciding


consists in – to make fully clear just what we do when we perform it. But while the nature of deciding is aggravatingly elusive, at least it is apparent that making a decision is something that we do to ourselves.\(^5\)

Contrasting the early focus on active identification, in his later work on identification and in his most recent work on love, Frankfurt seems to highlight the passive nature of the will. I do not think these distinct emphases pull Frankfurt in opposing directions. The reason for this is the plausible possibility that those things we cannot help willing are in central cases more fundamental to agency than those we decide to will. Another way of putting this is to say that decisive identifications depend asymmetrically on volitional necessities.\(^6\)

Volitional necessities, Frankfurt seems to hold, constitute “the heart of heart” of agency. They represent the very essence of the person. The main reason these volitional attitudes are characterized as necessities and as essential to personhood has to do with their substantial contents that delineating the territory of meaningful, free action. The gravitational center of the will (if this metaphor works) is filled in by our cares, loves and the category of the unthinkable. In important ways these are necessary first insofar as we do not have direct control over them, and second because they importantly shape and guide our decisions. Loving and caring foremost among volitional necessities set the

---

\(^5\) Harry Frankfurt, “Identification and Wholeheartedness”, in Frankfurt The Importance of What We Care About p. 172.

\(^6\) Here is a revealing paragraph in this direction. It appears in the Preface to The Importance of What We Care About in a context where Frankfurt explicitly reflects just on the presumed tension between the active and passive characterizations of the will: “It seems to me that discovering what we are is fundamentally, though perhaps only among other things, a matter of discovering what we must be. And it is thereby, to the extent that a person is defined by his will, a matter of discovering what we cannot avoid willing or cannot bring ourselves to will. The notion that necessity does not inevitably undermine autonomy is familiar and widely accepted. But necessity is not only compatible with autonomy; it is in certain respects essential to it. There must be limits to our freedom if we are to have sufficient personal reality to exercise genuine autonomy at all. What has no boundaries has no shape. By the same token, a person can have no essential nature or identity as an agent unless he is bound with respect to that very feature of himself – namely, the will – whose shape most closely coincides with and reveals what he is.” (ix)
boundaries of personhood. It is important to realize then that love in its substantive content and activity emerges as the most prolific structuring force of volition.\(^7\)

In the attempt to contextualize Frankfurt’s view of love it might help to briefly sketch some contemporary alternatives. According Eleonore Stump’s taxonomy, the “responsiveness account of love” refers to the family of views that analyze love as primarily a matter of responding to the intrinsic qualities of the beloved.\(^8\) The value or attractive force implicit in these characteristics elicits the love of the lover.\(^9\) By contrast, volitional accounts of love, of which Frankfurt’s is a paradigm, maintain that the value of the beloved is derived primarily from and depends upon the love of the lover. This account purportedly explains the constancy of our loves better than the responsiveness account, especially in circumstances when our beloveds undergo a change in their intrinsic qualities. Nonetheless, the implausible consequence for volitional accounts is that they make love largely independent of the qualities of the beloved. I show in the sequel that this objection is especially devastating for Frankfurt’s account, given that

---

\(^7\) Commenting on the substantive necessities of the will Frankfurt claims: “The will of a rational agent need not be, then, empty or devoid of substantial character. It is not necessarily altogether formal and contentless, having no inherent proclivities of its own. If a person’s will were a completely featureless instrument, with no capacity other than to transmute judgment about what to do into an effective expression of his active powers, then he would closely resemble the “bare person” to which Rawls says utilitarianism reduces the agent of rational choice. In fact, however, it is precisely in the particular content or specific character of his will – which may salubriously lead him to act against his judgment – that rationality of a person may in part reside. There is a mode of rationality that pertains to the will itself. Like the mode of rationality that is articulated in the necessary truth of logic, it has to do with the inviolability of certain limits. Logical necessities define what is impossible for us to conceive. The necessities of the will concern what we are unable to bring ourselves to do” (“Rationality and the Unthinkable” in *The Importance of What We Care About*, p. 190).


\(^9\) It is not my purpose to evaluate directly any of these options. Eleonore Stump mentions three main criticisms of the responsiveness account of love. First, implausibly any person possessing exactly the same intrinsic qualities as the beloved could thereby become an object of our love. Second, we can imagine very plausible cases where the intensity or the value of our love does not depend on the intrinsic characteristics of the beloved. (parents loving their children) Third, the account has purported implausible consequence that love disappears when the relevant intrinsic properties fade away.
loving constitutes for him the primary way of accounting for unity and consistency of agency over time.

Stump’s Thomistic account of love seems to cleverly combine the strengths of volitional account with those of the responsive account. On her view, love requires the presence of two interlaced desires. First, it requires a desire for the good of the beloved that does not have to be responsive to the qualities of the beloved. Second, a desire for union with the beloved that necessarily depends both on the intrinsic and relational characteristics of the beloved. While my purpose here is not to evaluate Stump’s account of love, I think that there is substantial agreement between her and Murdoch’s analysis of love, which I am inclined to accept. Concretely, Murdoch’s view of love as engrained diachronic attention to the intrinsic and extrinsic qualities of the beloved seems to me the right kind of psychological mechanism responsible for grounding Stump’s second desire of love.

But my main target is Frankfurt, so to him we return. As I have only briefly suggested, Frankfurt holds that love is essentially neither a whole nor a partial response to the intrinsic or relational qualities of the beloved. Contrary to more folksy views, love is also not essentially an emotional attitude, nor is it a relationship grounded directly in the history of the lovers. Instead, love originates and is dependent on volitional configurations that display the following essential features. First, love is disinterested.10

---

10 Frankfurt mentions that “selflessness” would also work to describe the relevant feature of love that he is after, but since his account of love includes illuminating self-love, for purely aesthetic he prefers “disinterested”. Can self-love be disinterested? Surely, says Frankfurt because here we have a clear case of the lover’s complete identification with the interests of the beloved. But isn’t disinterested self-love selfish? Here the answer is that it is truly in our best interest to selflessly love others: “The apparent conflict between selflessness and self-interest disappears once it is understood that what serves the self-interest of the lover, is precisely, his selflessness. The benefit of loving accrues to him only if he is genuinely selfless. He fulfills his own need only because in loving he forgets himself.” (“On Caring” in Necessity, Volition and Love p. 174)
That is, the lover’s loving investment must be shaped by a genuine concern for the flourishing of the beloved. Second, love is ineluctably personal, that is, its intentional objects are concrete individuals in all their historic and contextual particularity. Consequently, a lover need not purify all his interests by austere submission to impersonal, general moral principles. Against the tendency to make practical reason depend on abstraction and generalization, Frankfurt claims loving action overcomes selfishness by identifying with the contingent interests of the beloved.

Third, although love attaches to contingent particulars, the volitional configurations of love are both necessary and authoritative. Analogous to the kind of authority that reason or morality commands, the authority of our loves impresses itself necessarily upon us, shaping our agency, defining our volitional boundaries. The necessities of love, says Frankfurt, “mark our volitional limits, and thus they delineate our shapes as persons.” As I have already suggested, Frankfurt holds that volitional limits, fixed as they are by engrained higher-order desires, imply that we simply cannot willingly bring ourselves to do certain things. For instance, my volitional profile might be set so that I am incapable of betraying a childhood friend or of giving up my children for adoption. Consequently, volitional dispositions are partially constitutive of personhood. While not all volitional boundaries of a person are drawn by her loves, they do determine the deepest and most enduring such practical necessities. This shows that

---

11 There are also important dissimilarities. Whereas the reason and morality may impose formal or conceptual constraints on us as persons, the constraints laid out by love are substantial. In fact one could imagine that if Frankfurt is right and our volitional natures are essential to our personhood, when they come in conflict with the more formal necessities, each time the agent obeys the authority of the latter he essentially betrays himself.

love is reflexive. Loving essentially involves volitional attitudes toward the object of one’s love in such a way the lover is directly implicated. Our loves not only define who or what we are, they regard us. Thus although love must be disinterested, it is also essentially about the lover.

3.3. The Diachronic Requirement and Volitional Love

3.3.1. Volitional Necessity and Limits of Agency

Consider the following argument. (1) Loving is essentially constituted by volitional attitudes. (2) Volitional attitudes are higher-order configurations of desires that set the limits of personhood. This is to say that the things we cannot bring ourselves to do are explained largely by distinctive configurations of our will. (3) So, volitional loving sets the limits of personhood. My central aim is to show that the conception grounding loving agency in structural volitional configurations is a diminished and superficial. In this section I raise a number of related problems for premise (2) above. Certain kinds of temporal dispositional transformations seem to challenge Frankfurt’s volitional essentialism. Subsequent to my discussion in this section, my focus shifts to premise (1). I first explore a charitable way to construe Frankfurt’s view of love as the diachronic glue that binds agency together, and then show that loving understood in this way falls short of an important intuitive desideratum.

According to Frankfurt, a person remains the same if and only if her volitional limits stay unchanged. But depending on how these are characterized, volitional limits could change while practical agency remains intact. Intuitively, practical agency could also change while volitional limits would remain the same over time. Facing these
possibilities, my suggestion is that a more compelling account of diachronic transformation ought to involve concrete stories of causal change. In fact, I aim at a stronger conclusion, that structural volitional configurations have explanatory force only if supported by relevant diachronic practical attitudes that track these stories. If this is right, then Frankfurt implicitly relies on an understanding of love that plays this diachronic function. In the next section I argue that when charitably understood, Frankfurt’s view might be able furnish this alternative view of love, but that it is threatened by a version of practical solipsism.

Fred was brutal in high-school. You remember that in your junior year he beat you up once when you forgot to bring to school your Physics homework. You remember he was rather generous in handing out physical and psychological abuse, especially to the frail-framed guys in your class. He seemed to particularly enjoy shaming you before your mates. Let’s say you meet Fred seven years after graduation. You are now facing someone seemingly at peace with himself and at peace with you. He even seems to enjoy talking to you. He does not manifest one iota of the aggression that defined his high school interactions. He recounts only good memories of your high-school adventures. This thought enters your mind: “Fred is not the same person he used to be”.

According to Frankfurt’s volitional essentialism if his previous volitional limits have been essentially redrawn, “high-school” Fred ceases to exist.13 Perhaps we have no reason to distrust the commonplace utterance “He’s not the same person anymore”. For all we know, ordinary psychology and ordinary discourse are well equipped to detect changes in volitional dispositions. Interestingly, Frankfurt would characterize the

13 The example of Fred is my own. Frankfurt talks in more general terms about the reformation of a bully. I try here to provide a description of such a bully and his purported transformation, because my suspicion is that details matter in how we understand what volitional agency entails.
volitional change of a bully like Fred in terms of transformations in psychological
dispositions:

It might correctly be said of him that he is not the same person that he was,
meaning (among other things) that he is no longer inclined, and would no longer
be able even to bring himself, to do the dreadful things for which everyone once
feared and despised him. In virtue of his having reformed, the shape of his will
has altered; his volitional limits are different than they were. As we might quite
commonly say, he has become a different person. Perhaps there are some people
who do not know how he has changed, and who are therefore still wary of him. If
so, we might naturally tell them that their anxiety when he is around is misguided
because the dangerous brute about whom they worried no longer exists.14

According to Frankfurt’s volitional essentialism, when one’s volitional limits
have changed, her agency or practical personhood has also implicitly changed. Leaving
aside the reliability of folk discourse on this matter, my main worry is that if an agent’s
essence is understood as clusters of synchronic psychological dispositions, Frankfurt’s
model delivers only a superficial view of personal transformation through time. I argue
in the sequel that whether we should be anxious when facing Fred is not exclusively
dependent on the configuration of his current psychological dispositions. We must also
have a clear story of how his present configurations are tied to his past ones and how they
are likely to affect his future action. I suggest instead, that a combination of historical
processes and conditions turn out to support in relevant ways the synchronic structures of
the will. At the very least, Fred’s current dispositions and his endorsement of them
should have emerged in the right way from his previous volitional profile in order to
warrant the audience’s lack of anxiety in his presence.

14 “Reply to J. David Velleman” in Contours of Agency, p. 125.
3.3.1.1. The Central, The Peripheral, and Transformation

One worry derives from how to properly characterize volitional limits. If there is a distinction between central and peripheral volitional commitments, and Frankfurt seems to accept this, then the volitional limits of personhood might be delineated by peripheral and not only by central commitments. Suppose for instance that being a brute was not one of Fred’s central commitments in high school. Instead he merely acts like a brute, but this behavior is entailed or made probable by a central commitment to not losing face before his peers in conjunction with certain natural abilities he has in certain circumstances. He finds himself daily at recess in circumstances in which he intuits that a reliable way to ensure not losing face is by displaying his physical strength upon others. Redrawing the boundaries of the will such that he is now incapable of harming would not really mark an essential change in his agency. For imagine that over the years Fred maintains his commitment to not lose face before his peers. But now, his natural strength having dissipated, and having discovered his propensity for wit and sarcasm, he expresses his central commitment through clever self-deprecatory remarks in the appropriate circumstances. “Brutish Fred” and “Reformed Fred” possess the same central commitments while their manifest behavior is explained by changing peripheral commitments and the exercise of different abilities in different circumstances.

\[15\] Frankfurt acknowledges that even though our identifications or acceptances of certain configurations of desire set up the limits of our will, identification doesn’t always coincide with caring or love. Caring and loving are types of volitional necessities that structure the agency deeper than mere identifications or even moral endorsements. It is plausible then to suggest that even on Frankfurt’s own terms, identifying with one’s aggression is not a sufficient condition for fixing an essential feature of person, unless the person cares about that aggression.
This example spells trouble for two Frankfurtian assumptions. As the quote reveals, Frankfurt first assumes that a tendency to act brutishly reveals a central commitment to being brutish. Second, he seems to assume that if two persons (or person-stages) have a tendency to behave differently, this is explained by their possessing different central commitments. But, as we have seen, it is possible for a person to change his behavior without changing his central commitments. Furthermore, it is seems possible for a person’s manifest behavior to remain roughly the same and for his central commitments to change. The point is that stories of change from volitional possibility to volitional impossibility or vice versa do not explain by themselves which attitudes are essential to agency. Frankfurt holds that volitional configurations alone, understood as dispositions to behave in certain ways, set the boundaries of agency. I have suggested we can resist this.

But even if Frankfurt means that the transformation of agency depends on change in one’s central commitments, it is appropriate to inquire about the way central structural configurations come about. It seems the routes to dispositional change are directly relevant to our estimation of Fred’s transformation. And this is not just an epistemic issue. The concern is not only that the audience might be mistaken about whether Fred has undergone a relevant change. More pertinently, there are various possible routes through which Fred’s dispositional transformation can be causally enacted. Fred could come to have new dispositions as a result of an instantaneous conversion experience, as a

---

16 I am indebted to Mike Rea for helping me clarify the nature and expression of my argument in this section.

17 Imagine that Fred’s initial reason for acting brutishly is a desire to save face. Suppose that over time he “develops a taste” for brutalizing others even so as to become a torturer working for some secret agency. He might even get a kick from being a torturer. Nevertheless, his commitment to this way of life changes again when he discovers that the person he must torture is his long-lost high school sweetheart.
result of prolonged institutional detention, at the end of a six-month long seminar on self-improvement, or by pouring over Jane Austen’s novels. Let us suppose that in each such case the brutish tendencies of his high-school self are corded off so that they do not exert much if any motivational pressure on his current will.

My point here is simply intuitive. Whether redesigning the will’s boundaries happens directly or indirectly, instantaneously or gradually, voluntarily or non-voluntarily, the processes by which change happens directly affect which psychological dispositions become central, peripheral, or discarded with the passing of time. It is for this reason that an audience interested in evaluating the continuity of Fred’s agency might care about the concrete processes of transformation. Structural considerations alone do not seem to provide a clear criterion regarding continuity or termination of practical agency.

3.3.1.2. Restrictions on Structural Change

Are there any other restrictions that Frankfurt imposes on his account of agency transformation? Frankfurt suggests that the limits and necessities of the will (and love) are not necessarily laid out through deliberation or direct volitional control or even through ethical evaluation. In our example, the path from brutishness to reformation need not follow any of these trajectories. Yet, even though dissociating ourselves from former dispositions need not (and maybe usually does not) implicate our “higher natures,” it shouldn’t involve our “lower natures.” That is, repression, defensiveness, self-deception are excluded as acceptable forms of volitional transformation.18 Frankfurt

---

18 The reason why this matters diachronically is the same why it matters synchronically. Frankfurt says: “Resolving ambivalences by taking a decisive stand against certain feelings does not require (or even
also plausibly allows that past dispositions and desires that conflict with current ones could be (concurrently, consciously, or non-consciously) present in the overall psychic economy of the individual. Crucially relevant, however, is that the current higher-order volitional configurations have so been fashioned as to place divergent inclinations outside the borders of effective agency.

If the above conditions are the only restrictions volitional essentialism places on transformation of agency, they seem too weak to deliver the desired result. For suppose that Fred simply ignores or begins to care less and less about hurting others, not because of repression, but perhaps due to some other external or internal contingency. Suppose he gets a wearisome night job that drains him of the energy required to scheme harm against others. The desire to hurt them might still be present dispositionally, yet contingently it is overridden every day by the desire to get more rest. Fred’s volitional outlook is such that he wholly identifies with the desire for rest, though he still would want to inflict pain on others in its absence. Here the limits of the will have been altered non-repressively, and yet for all we know, if Fred takes a vacation, they may revert back to the original setting. The general point is that entrenched dispositions may be recalcitrant to change in surprising ways. They can be temporarily disabled non-repressively, while continuing to persist below the radar of reflective consciousness.

3.3.1.3. Fragmentation and Compartmentalization

Additionally, there is the problem that processes responsible for setting volitional limits may not establish correlative emotional limits. If the volitional is permit) that a person misrepresents those feelings or that he conceal them from himself.” (Reply to J. David Velleman, p. 126) The same point is made abundantly clear in “The Faintest Passion”
compartmentalized away from the emotional, it is not clear that personal transformation always tracks volitional transformation. Going back to Fred, suppose his volitional alteration occurs as a result of a dramatic religious experience. Fred acquires a new higher-order volitional outlook that renders him genuinely incapable of harming another person. Nonetheless, suppose he continues to consciously harbor feelings of animosity, resentment, even hatred. Fred’s incapacity to hurt someone does not show that he is a reformed person. For imagine that he still delights in the suffering of those whom he had previously abused. Each time he hears of their physical, psychological or economic afflictions he feels self-righteous contempt. “God’s vengeance released on them!” he thinks to himself, while also finding impossibly repellent the desire to smite them himself. In fact, Fred might experience this volitional impossibility in direct proportion to the delight experienced in what he interprets to be God’s vengeance on them. I suggest we should not call this person “reformed” in the relevant sense.

So far I have argued that focusing wholly or mostly on Fred’s synchronic volitional structures does not settle the issue of whether his agency has undergone an essential change over time. The widespread tenacious persistence of our entrenched dispositions together with the possibility that rational, moral, volitional, and emotive propensities get compartmentalized make it such that the synchronic moment of decision is an unreliable guide to telling where one’s essential commitments lie. Because the compartmentalization worry is more generally indicative of problems in Frankfurt’s account, let me dwell on it for a bit longer.

According to Frankfurt an agent could be volitionally integrated with respect to not harming another, while simultaneously rejoicing when the same person is harmed.
But the account allows not just for a division between the will and the emotions, surprisingly it also allows for the segmentation of the will itself. Volitional impossibility in one sub-region is fully compatible with volitional license in adjacent sub-regions in the will of the same person. Frankfurt himself makes this point in the discussion regarding a mother who cannot give up her child to an adoption agency, although she experiences the desire to do so on moral, rational, or prudential grounds:

The volitional necessity that is exhibited in her failure to bring herself to give up the child does not completely define the bounds of her volitional capacities. It only defines the bounds of her capacity to perform the act of giving up her child. It naturally does not limit her capacities to perform innumerable other actions, such as those involved in going to the adoption agency and completing all the work of giving up the child until the final act by which the adoption would have been decisively accomplished.19

Frankfurt’s point is that it is volitionally impossible for the mother to give up her child for adoption, while it is volitionally possible for to take steps in that direction. Now imagine a structurally similar possibility in Fred’s case. On Frankfurt’s own suggestion, reformed Fred would seem free to intend, to devise detailed plans of retaliation, even to set in motion a series of actions with the end result of harming others in ways indistinguishable from those of unreformed, brutish Fred. Nonetheless, provided he is incapable of carrying these intentions out (even at the last second), Fred counts as a new person. And this result just seems implausible. It seems arbitrary and unmotivated to hold that a brand new agency emerges by only a slight alteration of one’s volitional borders, while brutish Fred and reformed Fred share the wide majority of their dispositional traits in common.

19 Harry Frankfurt “Reply to Gary Watson” in Contours of Agency, p. 162
Perhaps Frankfurt could reply that a person is transformed to the extent to which more and more of his past volitional sub-regions get displaced and replaced by new ones. It is not the case that any one volitional limit determines continuity of personal agency, but perhaps coherent, significant clusters of sub-regions do.

A problem with this reply is that it falls prey to a version of the worry regarding the stubborn diachronic persistence of entrenched volitional attitudes. If one volitional limit can be neutralized temporarily by some non-repressive mechanism only to reemerge later, it seems the same could occur for clusters of such volitional limits. True enough, one may reply, but if enough volitional limits are neutralized temporally, a new person might in fact emerge. Imagine that Fred falls in love with a girl that returns his affections. Over time, due to the acceptance and care she gives him, and perhaps due to the change in perception by others, Fred’s brutishness is defused. Love does strange things to people. Mightn’t we correctly say that Fred is a new person? Let’s suppose however, the reply goes on, that after seven years of bliss the girl dumps Fred. Overwhelmed by self-doubt and self-defensiveness, Fred reverts to his old brutish self. It seems that one could plausibly say that the seven years of bliss constituted a temporal neutralization of large sections of his will. But one could equally hold that Fred underwent two personal transformations: from brutish Fred to blissful Fred, and then from blissful Fred back to brutish Fred. Whereas it is easy to see how peripheral commitments could be temporarily neutralized, if Frankfurt holds that personhood depends upon significant changes in central commitments, his view seems more plausible.20

20 Thanks to Mike Rea for pressing this objection
A first reply here is that even if Fred’s personal change is explained by transformations of his central commitments, it is hard to see how these in turn depend on structural conditions of the will alone. Redrawing Fred’s central commitments in this case is importantly explained by the presence and historical development of a loving relationship. In fact, it seems to me that the specific shape of that history and how Fred and the girl interpret it are the fundamental features crucial to whether Fred changes his central commitments.21

Second, this reply seems to underestimate the extent to which persons actually seem to get along quite well by compartmentalizing incompatible central commitments. Consider for instance the process of “doubling,” namely that of partitioning one’s volitional self into two or more semi-autonomous volitional wholes. Jeffrey Blustein suggests that Nazi doctors who tried to preserve their commitment to the Hippocratic Oath and to Hitler widely experienced the phenomenon of “doubling”:

From the standpoint of the original self [allegiance to Hippocratic oath prior to Hitler], it was another self who collaborated with the Nazi regime, and this splitting off of the Auschwitz self enabled the physician to continue to believe in his own integrity as a physician. From the standpoint of the Nazi doctor’s double, it was another self who would have qualms, for instance, about curing a patient only to get him well enough to be sent to the gas chamber. But there was a no standpoint from which the agent saw his incompatible commitments as commitments of a single self.22

---

21 Consider trying to fill in the second personal transformation. After Mary dumps him, looking back to the seven years spent together Fred might construe the history in different ways, construals that essentially alter the shape of his will. “It’s been a sham all along” he might say “she never really cared for me…All she wanted was to be in company of someone she felt could protect her.” Or “we’ve had some good years together, I wonder why she drifted away from me into the arms of John”. It is true that the personal history along with its interpretations does not entail that Fred will change his central commitments. At the same time, it seems to me his transformation is entirely arbitrary if it is independent of such retrospective analyses.

22 Jeffrey Blustein, *Care and Commitment* (Oxford University Press, 1991) p. 132
According to this interesting possibility, dispositions steeped in different causal histories could divide up the will into fragments mutually inaccessible from an evaluative point of view. As such, seemingly absolute limits of one volitional segment could be transgressed only by occupying the perspective of another such volitional segment. Suppose the character I called “Fred” was actually a high-school teacher in an Eastern-European country during the communist era. Standard educational practices allowed for physical and psychological punishments as motivational strategies. He was an excellent motivator by local standards, and that he cared deeply about promoting excellence. Nonetheless this central commitment of his together with his psychological make-up entailed being cruel to his students by universal norms. He simply couldn’t imagine being a good educator without punishing students for their lapses. Suppose that since you’ve graduated seven years ago he has become a father, and this has made him more compassionate to the plight of his students. When he looks at a reprobate student as someone’s child, he is incapable of carrying out the punishment. When he conceives of them only as failing students in his Physics class, he must motivate them. Now suppose you run into him, and as in the original description of the case, he shows no hint of aggression toward you. But that seems to be simply because he cannot conceive of you any longer as a reprobate student in his Physics class. Is Fred a reformed person?

Third, the reply does not deal with the more general worry about the division between will and emotions. As I have suggested in the previous chapter and as I defend in the next two, it is often the case that emotions give us a better clue to the continuation of our agency than explicitly endorsed central commitments.
My point thus far has been that the synchronic volitional necessity displayed in Fred’s incapacity of hurting others does not show that Fred is a reformed agent. And if this example generalizes well, structural features of agency cannot by themselves account for stories of personal transformation. The limits of the reflective personhood are not clearly set within the limits of synchronic volitional necessity, unless the latter also possess distinctive functions that integrate the agent’s will diachronically.

3.3.2. Caring and Loving in Diachronic Perspective

Does Frankfurt have a story to tell about how volitional necessities might play this diachronic role? He does indeed! He argues that among all volitional necessities, caring and its distinctive expression -- loving, are the most basic attitudes accounting for the continuity of human agency over time. If we cared about nothing, “we would be creatures with no active interest in establishing or sustaining any thematic continuity in our volitional lives.”

Frankfurt is well aware that caring is not and must not be reducible to synchronic configurations of the will, for he explicitly mentions the possibility that a human life is integrated volitionally moment to moment while also diachronically shapeless:

In other words, our capacity for higher-order desires and higher-order volitions might remain fully intact. Moreover, some of our higher-order desires and volitions might tend to endure and thus to provide a degree of volitional consistency or stability in our lives. From our point of view as agents, however, whatever coherence or unity might happen to come about in this way would be merely fortuitous and inadvertent. It would not be the result of any deliberate or guiding intent on our part. Desires and volitions of various hierarchical orders would come and go; and sometimes they might last for a while. But in the design and contrivance of their succession we ourselves would be playing no concerned

---

or defining role.²⁴

So, it is in response to the possibility of a synchronically unified but temporally incoherent life that Frankfurt introduces the concept of caring as having this distinctive diachronic role. Frankfurt points out that it is neither the intrinsic value of the objects of care, nor indeed their longevity that proves most basic for organizing diachronically human agency. Rather, caring has for us intrinsic value because “it is the indispensably foundational activity through which we provide continuity and coherence to our volitional lives.”²⁵ Two immediate upshots. First, caring is the glue that ties together sequences of agent-stages into a more-or-less coherent volitional diachronic whole. Second, caring is volitionally foundational, for by loving we ourselves account for our own agency’s continuity through time.

Can we specify more precisely how caring gives us continuity of agency? According to Frankfurt’s analysis, caring entails rejecting the extremes of motivational fatalism and of interventionism. A caring human agent could not participate merely pathologically in diachronic processes and attachments. When I say, “I care about X” there is a built-in expectation that the psychological forces responsible for my loving are not mostly external to my guidance or control. If I care about anything, I am not at the mercy of psychological processes over which I affect no control at all. So, caring excludes motivational fatalism.

Consider for a moment, at the other extreme, a character type of a radically different tenure, an agent who claims the freedom to reshape at any moment, any and all of her most basic concerns and cares. According to the romantic ideal of the free spirit,

²⁴ Ibid, p.162.
²⁵ Ibid, p.162.
as Raymond Guess puts it, “genuinely free and deep people are those who love masks and are deft at changing them.”  

Can the free spirit be a caring agent? On the one hand it seems that self-forging spontaneity does not have to repudiate loving or caring as defining features of agency, but only the permanence of the objects of such attitudes. On the other, Frankfurt would probably reject the free spirit ideal, but not because it is a necessary truth that our true loves are long-lived. Rather, his concern is to reject the possibility that we define the succession of our agency by direct control or fiat. “We are not fictitious characters, who have sovereign authors; nor are we gods, who can be authors of more than fiction. Therefore, we cannot be authors of ourselves.”

According to Frankfurt, to the extent that the free spirit’s love of changing masks is self-induced, to that extent she actually relinquishes the deepest loves in her life. So, paradoxically she would succeed only in performing successive acts of self-betrayal. It is for this reason that Frankfurt should enlist free-spirited self-creation among the enemies of agency, along with his other usual suspects: boredom, ambivalence, and wantonness.

If motivational fatalism and free-spirited self-creation are excluded, caring and loving must be constituted (at least partially) by diachronic volitional configurations falling only faintly, mediately, or indirectly under the control of agency. That being said, it remains somewhat unclear at this stage how this passive feature of loving fits with the other more active ways of integrating agency. For Frankfurt also requires that when the agent experiences conflict among desires, she must wholeheartedly decide on a course of action. Deciding, along with supporting attitudes such as endorsement, intuitively


requires more active control over one’s desires than merely falling under, or accepting the necessities of love. Perhaps an answer to this puzzle is that with respect to first-order desires deciding could be seen as active control, but from the perspective of one’s enduring loves, the decision is seen as a passive acceptance of a volitional necessity that brings integration to agency.

We have seen that Frankfurt is explicitly committed to structural/synchronic integrity. Furthermore, if caring/loving is most basic for agency continuity, then, as I argued above, Frankfurt must implicitly be committed to the claim that some volitional necessities are constituted partially by diachronic attitudes. It is also patently clear that at times synchronic integrity could come apart from diachronic integrity.28 So, supposing that wholeheartedness and loving have both synchronic and diachronic functions, we are pressed to understand the relationship between these functions.29 In the sequel, I first distill from Frankfurt’s work a model that throws some light on this relationship. The model centers on the insight that loving or caring is fundamentally a diachronic investment or a kind of temporally sustained ascription of significance. I develop this insight by extracting from Frankfurt two modalities of love: a prospective mode and a retrospective one, both of which reinforce the view of love as investment. I argue that while the interplay between these two modalities of loving gives some explanatory power

---

28 As I have argued above, the free spirit inhabits a diachronically incoherent life, although she can live wholeheartedly from one moment to another or even from one self-conception to another. The examples of diachronic integrity without synchronic integration are perhaps as easy to come by. We might imagine someone living by oracles, or an Old Testament prophet the shape of whose life is guided by a purported supernatural plan. Yet, this is also a person who from moment to moment is in the grip of ambivalence, self-doubt, and anxiety for although he has vision of his life-plan, he does not understand why he must act thus and so at particular moments in time.

29 I speak of both wholeheartedness and loving here, first because Frankfurt often suggests that being wholehearted is special form of love, namely self-love. Because of this, whatever the objects of our love, they must be such that we should love them wholeheartedly. (See the “Faintest Passion”)
to Frankfurt’s account, the model is also subject to some limitations. I develop the objections in the sequel.

3.3.2.1. Caring as Diachronic Investment: The Prospective Mode

There is much in Frankfurt’s writing to suggest the interpretation that loving is a forward-looking volitional endeavor ascribing new worth to objects or aiming to sustain an investment already made. Loving ascribes value to the beloved in the sense that it *creates* the ends of action and *directs* in specific ways the agent toward their implementation. Let me begin from certain things Frankfurt says regarding the way in which an agent’s decision can integrate her both diachronically and synchronically. In an early article in which he aims to show the importance of wholeheartedness to the constitution of agency, Frankfurt alludes to such a prospective mode of agential integration. Here he explores personal unification by comparing decisions regarding theoretical concerns with those regarding practical ones:

A person who decides what to believe provides himself with a criterion for other beliefs; namely, they must be coherent with the belief on which he has decided. And a person who makes a decision concerning what to do similarly adopts a rule for coordinating his activities to facilitate his eventual implementation of the decision he has made. It might be said, then, that a function of decision is to integrate the person both dynamically and statically. Dynamically insofar as it provides – in the way I have just mentioned – for coherence and unity and purpose over time; statically insofar as it establishes – in the way discussed earlier a reflexive or hierarchical structure by which the person’s identity may be in part constituted.\(^{30}\)

The idea here is that a decision is more like a *stable plan of action*. A plan requires a goal, a terminus. But it also implies a certain organization of the agent’s

---

\(^{30}\) Harry Frankfurt “Identification and Wholeheartedness” in *The Importance of What We Care About*, (Cambridge University Press) 1988, p. 175.
psychic items along both structural and temporal lines to ensure (or make probable) that the goal of action is reached. In deciding, the individual adopts a direction of agency formation through the particular reflexive arrangement imposed by orienting herself toward a target. It seems plausible that plan-like decisions operate at the volitional level as forward-looking rules that coordinate the agent’s actions over time by implicitly ordering the motivational configurations responsible for them.31

While some diachronic attitudes may be decision or plan-like, it is unclear that caring or loving fit neatly into this mold. As was mentioned at the end of last section, Frankfurt underscores that loving lacks the direct, voluntary control of decision, while also emphasizing its essential intentionality (the having of some object or end) and a corresponding structuring of this intentionality. Furthermore, since caring is constituted by purposeful, non-voluntary, and temporally enduring volitional attitudes, it seems better suited than voluntary decision to strongly define from within, as it were, the limits of volitional agency. Caring about an object, he says, consists

…in the fact the he [the agent] guides himself by reference to it. This entails that he purposefully direct his attention, attitudes and behavior in response to circumstances germane to the fortunes of the object about which he cares. A person who cares about something is, as it were, invested in it. By caring about it, he makes himself susceptible to benefits and vulnerable to losses depending upon whether what he cares about flourishes or is diminished. We may say that in this sense he identifies himself with what he cares about.32

The relevant point here is that if we abandon the direct, voluntary sense in which caring and loving are decision-like, there is an attenuated sense of caring-as-axiological investment. In loving, we invest ourselves into the objects of our loves (ideals, other

31 The work of Michael Bratman on planning and agency may be seen as one of the more promising attempts of unpacking with uncompromising rigor and novelty this aspect of Frankfurt’s insight.

persons, states of affairs). In this way our diachronic motivational structures acquire a meaningful direction, a trajectory, while our moment-to-moment volitional configurations are structured according to the importance or intensity of our loves. In caring, the lover either bestows worth upon the beloved or continues to maintain an investment already made. This endowment aspect of loving, especially in light of love’s reflexivity, deepens Frankfurt’s insight that loving others is not just consistent with but constitutive of self-love. To the extent we care about others, we are also likely to reinforce the reflexive frame through which we look at them, and become prone to incorporated it in our practical outlook. So, paradoxically, caring about other translates into care about oneself.

The passage quoted above might suggest that loving consists primarily in recognizing the intrinsically worthwhile properties of the beloved and in being guided by such a perception. One might even claim that investment is a reactive attitude grounded in the perception of such worth. But that would be to miss the mark of what Frankfurt has in mind. While in certain cases the perceived worth of an object elicits motivational responses, these responses merely simulate those of loving proper, or they involve peripheral cases of love. Paradigmatically, in loving the grounding relation runs the other way. Frankfurt claims that “what we love necessarily acquires value for us because we love it. The lover does invariably and necessarily perceive the beloved as valuable, but

33 Interestingly a number of other emotional investments can unify our will. As the next chapter makes clear resentment seems to operate analogous to what Frankfurt says about love here. In my view volitional invest follows a certain way of looking at the object of our attention. So, loving attention already implies a built-in sensibility to the needs and desires of the beloved couple with a desire to act for their good. It seems that Frankfurt could also make the desire for the good of the beloved constitutive of the act of volitional investment.
the value he sees it to possess is a value that derives from and that depends upon his love.”

The deep reason for prioritizing the importance of loving independently of the properties of the object of love can be seen by an analogy with the value of work. Frankfurt maintains that working is fundamentally and inherently valuable to us because it provides meaning and cohesion to our lives. At the same time it is not only the process of production that matters. The end results of toiling turn out to be the necessary conditions that support distinctive, meaningful action. By analogy, loving possesses inherent value, non-derivative from the properties of the object of love, because it provides coherence and orientation in axiological space. At the same time, the “beloved object is an essential and indispensible condition of the specific kind of value that loving possess.”

There is a straightforward sense in which Frankfurt’s suggestion is plausible. A young father’s love for his daughter leads him to make innumerable sacrifices and re-arrange priorities in his social and professional life. His loving bestows a particular kind of importance to her, and this independent of her cuteness, or her agility or any other such features. Clearly, his loving investment ends up redesigning the father’s former priorities and purposes, and inherently his previous volitional dispositions. Nonetheless, the daughter’s mere presence essentially elicits and directs distinctive patterns of paternal love, though again it is not her properties that ground it. If my interpretation of Frankfurt

35 Michelangelo sculpts David perhaps to avoid boredom or in order to channel his creative genius in ways beneficial to the Florentine people or to his own future career. But the fact that he works on David as opposed to the Sistine Chapel necessarily supports a distinctive investment of his valuable activity.
is on the right track, caring-as-investment of meaning explains very well the intrinsic, non-derivative value of loving, as well as the distinctive significance that the objects of love have for guiding and supporting the process of caring. The case of parental love illustrate prospective invest clearly enough.

The conclusion we’ve come to by now is that if loving is iterated investment of meaning in the objects of love, it is the most basic creator of ends of action. Frankfurt seems thus justified to claim that loving sets the deepest limits of what we can or cannot do. Deep volitional necessities mark the bounds of personhood because they express our most central attachments. That is, loving in its prospective investment mode lays out the coordinates of importance for agency. By projecting personal caring onto the map of human goals, love places the person in relation to those goals. And it is only by reference to this map that the person can come to understand her actions, and therefore her personhood. Consequently, if our volitional investments set the limits of our agency, we discover ourselves only by reference to them! Self-understanding always follows where loving leads.\textsuperscript{37} And this opens up the discussion regarding another modality of loving, namely its retrospective facet.

3.3.2.2. Caring as Diachronic Investment: The Retrospective Mode

The activity of loving not only colors the world with ends and purposes, it also reveals in retrospect, perhaps only indirectly and somewhat confusedly, the things we most care about. The nature of this disclosure and the way in which understanding must

\textsuperscript{37} In this context Kierkegaard’s words seem appropriate: "It is perfectly true, as philosophers say, that life must be understood backwards. But they forget the other proposition, that it must be lived forwards. And if one thinks over that proposition it becomes more and more evident that life can never really be understood in time simply because at no particular moment can I find the necessary resting place from which to understand it—backwards." (Journals…)
be submissive to the will is made evident in Frankfurt’s revisiting the case of the conflicted mother. To recall, while the mother seems initially divided between the desire to give up her child for adoption and the desire to keep her, in the end her volitional love is expressed by her \textit{incapacity} to give the child up. In a reply to a critic Frankfurt focuses on the way in which the \textit{phenomenology} of volitional necessity reveals whether the mother’s agency is integrated or disintegrated in the moment of decision. This paragraph is worth quoting at length:

For some mothers, in other words, it might be only in resignation and in disappointment with themselves that they accept as their own the will by which they are constrained. They would try to change this will if they believed it would be possible for them to succeed in changing it. Whether the mother experiences her submission to volitional necessity as a defeat, or whether she experiences it as a liberating release of her previously stifled or misunderstood real desires, is critically pertinent to deciding whether her volitional integrity is sustained or whether her agency is defeated. But how the mother experiences her submission to necessity is an additional feature of the situation, which cannot be inferred merely from the fact that she is unable to bring herself to give up her child.\textsuperscript{38}

According to Frankfurt the fact that we experience volitional incapacity by itself is insufficient to make it epistemically clear what are the deepest facts of agency. But the \textit{way that experience feels} just might. Depending on \textit{how} the mother experiences the submission to the necessity of love, she may be able to tell retrospectively whether her agency is integrated or not. The necessities of love together with the correlative emotions would sketch out the territory of the things she most cared about. The mother who experiences her inability to give up the child as \textit{liberation} \textit{discovers} both that she cares most about keeping the child, and the further fact that her loving has so structured her volitional investments over time that \textit{that} particular desire is what she truly cares about. She is diachronically integrated. The mother who experiences her inability to give up the

\textsuperscript{38} Harry Frankfurt “Reply To Watson” in Buss and Overton eds., \textit{Contours of Agency}, p. 164.
child as a defeat (maybe feelings of regret, guilt or shame) may thus signal a certain instability and incoherence of agency.

It is important to point out that Frankfurt does not claim that the emotions of satisfaction or disappointment are by themselves revealing of where our true loves lie. Rather they are value-markers only when supported by the relevant volitional configurations. Perhaps Frankfurt might want to add that such feelings are epiphenomenal on particular volitional organizations of the person. It is the moment of submission to love or the moment of loving investment that fixes that psychic organization of the whole will. The attendant emotions are therefore only reverberating movements of satisfaction or dissatisfaction with what is otherwise fixed at the volitional level. We may need some feelings to understand our volitional limits, but they are not, as it were, essential to constitute the attitude of loving or caring.

Here is the upshot of our discussion thus far. Both the prospective and the retrospective diachronic configurations of love seem to depend for Frankfurt on the present, that is, on the synchronous moment of caring investment. With respect to the future, loving devotion essentially imposes upon us purposeful ends and trajectories. With respect to the past, submission to love activates certain feelings that act as epistemic lights revealing the successive series of our most important past volitional investments. These feelings supervene on certain will configurations already laid out at the moment of decision or investment. At most, they could have only an epistemic but not a constitutive function in the economy of loving attitudes. Thus, one may say that the moment of investment is both something akin to a decision and something akin to a discovery.
3.3.2.3. The Extended Frankfurt Account

We can derive from the above discussion then two insights essential for Frankfurt’s view of love. I am going to call this the *Extended Frankfurt Account* (EFA). First, the synchronic moment of loving investment or acceptance structures both the forward-looking and the retrospective reach of our loving. Specifically, when past loves come in conflict with each other, through present investment the agent accepts one of these historical loves as having an implicit weight in the situation. Consequently, present tense loving imposes implicit orderings upon the histories of our loves. Equally important, in the forward-looking aspect of loving, anticipated (hoped for) aims impose corresponding explicit or implicit importance orderings on our present loves. This first condition, thus, expresses Frankfurt’s idea that present tense loving define diachronic agency by negotiating faithfully between our past and future cares. The second main feature of EFA does not require as much explanation, for it has already been articulated in the notion of investment. It is the idea that the importance of the objects of care is derived primarily (most basically) from the activity of loving and not from our perception of their intrinsically worthy qualities.

As an application of love’s diachronic role in self-constitution, we can give a more detailed explanation of what might be required for Fred’s conversion. According to the EFA, reformed Fred is a different person than brutish Fred in case a) reformed Fred does not care any longer about “motivating” his students through physical or

---

39 I use “acceptance” because this seems to be Frankfurt’s favorite term of late. See especially the replies in *Contours of Agency*.

40 I don’t mean to imply or to attribute to Frankfurt the view that loving is essentially revisionist. It is possible that Frankfurt’s account could imply this, yet I am trying to construe the relationship between present loving and its past loves as liberally as possible. What is an essential Frankfurtian theme is that present loving must avoid both the extreme of Free-Spirited loving, and that of historically fatalism.
psychological punishments; b) his present cares are causally efficacious in behaviors and intentions distinct from those of brutish Fred (Fred is described by others as a good father and an understanding teacher); c) Fred’s present satisfaction with his current cares has come about neither by fiat, nor self-deception nor by some other external compulsion. Importantly, Frankfurt’s account might have the resources to explain how Fred at some point has stopped investing in previous volitional profile, and instead began investing in another.

I do not mean to suggest that the notion of diachronic investment (EFA) solves all issues related to transformation of agency over time. When we consider the knotty ways of mature loves, their reckless instability or their stubborn entrenchment, the possibility that volitional prospective investment should eclipse the retrospective or that the retrospective investment should obscure the prospective does not seem at all far-fetched. Sometimes, people invest in future loves in order to escape their past, and sometimes they nurture the past, for fear of commitment to the future. Sometimes, of course, love strikes out on its own without any clear future purpose and without being tethered in any previous stable volitional disposition.

By raising these issues I want to point out that EFA provides only a limited explanation for agential transformation over time. Granted that the synchronic moment of investment bestows or transfers or intensifies importance onto the map of personhood, we still want to understand just how past investments are being balanced against future ones. Even if present loving investment would grow linearly from our preceding loves and move seamlessly on to our anticipated ones, we would still want to understand in more detail just how this happens. We would want a micro-structural, in-depth analysis
of the nature and function of the psychological forces that make possible this impressive diachronic act. Alas, human experiences of loving can hardly be claimed to follow linearly from past investment, or to move linearly toward anticipated returns. How loving brings to the foreground of our caring map some considerations and not others while avoiding fiat, self-deception, external compulsion and wantonness seems deeply mysterious.

3.3.2.4. A Final Reservation

Before turning to the solipsism worry, I express a final reservation to conclude this section. In the course of the argument thus far I have hinted several times that whenever emotional coherence comes apart from volitional integrity it is not clear where agency goes. The view of love as diachronic investment does little to allay this worry. According to Frankfurt, in extreme cases of irresolvable conflict between a person’s historically entrenched loves, agency can be destroyed. Consider Frankfurt’s description of Agamemnon’s inner conflict, which presumably leads to the dissolution of his agency:

Thus, Agamemnon at Aulius is destroyed by an inescapable conflict between two equally defining elements of his own nature: his love for his daughter and his love for the army he commands. His ideals for himself include both being a devoted father and being devoted to the welfare of his men. When he is forced to sacrifice one of these, he is thereby forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists. Hence, there can be no continuation of his story.41

In the story Agamemnon decides to sacrifice Iphigenia in order to facilitate his army’s taking off for Troy. On Frankfurt’s interpretation, this decision is destructive of

his agency because it constitutes an act of self-betrayal. Agamemnon is effectively forced to renounce an entire series of loving investments that hitherto have essentially constituted his agency. I am sympathetic to Frankfurt’s position that something gets irretrievably lost here. But I am not sure that it is Agamemnon’s agency, or personhood, or story. And I mean this in more than the trivial sense in which we read on that he survives to plunder Troy, and that he returns home to be murdered by Clytemnestra, his wife.

Agamemnon’s story and agency continue in a more robust sense. For once, his emotions of guilt and shame perhaps continue to testify that he is the father who killed his daughter. His subsequent emotional anguish would be unintelligible were it not for some kind of continuity. Suppose that Agamemnon does not express any feelings of shame and guilt, but instead the hero inside gets fired-up and he dedicates himself wholeheartedly to be the staunched leader of his men. We might say that his former self was destroyed and a new person born. But we might equally claim that Agamemnon is very adept at hiding or suppressing essential aspects of his identity in order to accomplish the task at hand – conquering Troy. The circumstances are such that he can only decidedly identify with the persona of the warrior by temporarily suppressing the role of the father.

Second, we can imagine all sorts of causally relevant ways in which Agamemnon’s loving investment in Iphigenia is maintained from a practical standpoint. He might build memorials, compose songs or poems honoring her, or pray daily to the gods about a gentle treatment in the Underworld. These actions would be intelligible in light of continued intentional attitudes that have Iphigenia as their loving focus. Finally,
the conception of agency that I introduce in the next chapter explains the drastic rupture, but also the relevant continuity in agency. On my view, one of Agamemnon’s dramatic personae is drastically altered, concretely his self-understanding as a father. But Agamemnon’s agency is constituted by the diachronic interplay among different dramatic personae, including that of being a heroic Greek leader, a husband, a brother, and relevantly that of someone who has implicitly internalized a set of social assumptions that makes it possible for the conflict between fatherhood and heroism to arise.

3.4. Love as Investment and Volitional Solipsism

Even if the reservations mentioned above are answered, there is another challenge striking at the heart of love-as-diachronic-investment. For lack of a better label, I refer to this as the worry of volitional solipsism. It seems plausible that for a range of human loves, caring-as-investment plays a crucial explanatory role. I have mentioned sympathetically the parent who invests himself in his child, regardless of whether the latter possesses intrinsically desirable qualities. Parental love aside, some human loves are such that they do not always require caring investments. There are even circumstances in which maintaining the investments Frankfurt’s view recommends can obscure kinds of love worth having.

Sometimes mature interpersonal loves require exquisite attention, appreciation, reinterpretation, and recalibration of the lover’s investment in light of real transformations sustained by the lover, the beloved and the relationship between them. In mature loves, diachronic emancipation from the necessities of love can be seen as desirable. For sometimes, it is precisely when we are released from patterns of volitional
investments that we come to see unsuspected characteristics of our beloved. Minimally, unless dynamic attitudes of openness toward the beloved are cultivated alongside of or are built into the act of investment, our diachronic loving may very well conceal the real aspects of the beloved rather than reveal them. Absent exquisite cognitive attention to the changes sustained by the beloved, iterated diachronic investment could lead the lover to love a simulacrum instead of the real thing. Effectively, this would mean volitional solipsism.

In order to highlight this problem, I first draw attention to Iris Murdoch’s view of love as diachronic attention to the reality of other persons, a view that inspires much of the current reservation for Frankfurt’s account of love. After a quick sketch of the Murdochian framework, I proceed to characterize more precisely volitional solipsism. Then I show that Frankfurt’s account of love is dangerously close to edges of volitional solipsism, if not merging into it.

3.4.1. Iris Murdoch and Loving Attention

Iris Murdoch agrees with Frankfurt that love takes as its object the ineluctably particular and historically shaped agent. However, Murdoch either downplays the traditional conception of the will, or makes it dependent on certain forms of moral perception. In contrast to theories of agency that emphasize the synchronic moment of decision, Murdoch focuses on the continuous function of evaluative awareness that informs and forms the outlook of agency between decisions. Because conscious awareness is always involved in processes of dynamic, intentional, and imaginative evaluation, the facts relevant to choice are largely fixed prior to the moment of action.
Responding to Stuart Hampshire’s view of agency Murdoch claims: “We evaluate not
only by intentions, decisions, choices [the events Hampshire describes], but also and
largely, by the constant quiet work of attention and imagination. The image here is not so
much that of a body moving [Hampshire’s image] but rather of a sort of seeping of
colour, or the setting of a magnetic field. When moments of decision arrive we see and
are attracted by the world we have already partly created.”42

A color spilling slowly over a surface alters it seamlessly, gradually, the effects of
the trickle being noticed only with the passing of time. By contrast, bodily movements
are typically direct, individually distinguishable, and their effects are immediately
obvious in the environment. Murdoch holds plausibly that practical attention to the
reality of other agents, analogous to a sluggish color spill, is essentially evaluative,
painfully gradual, and only diachronically intelligible. Decision, analogous to the
movement of an arm, is discrete, direct and immediately noticed. Minimally then,
practical attention depends upon a set of evaluative concepts different than those involved
in evaluating actions and decisions. In addition, Murdoch makes a more ambitious claim.
The complex evaluation evoked by attention is in some sense prior to, and more basic
than that attaching to intentions, decisions and actions. Loving action depends on loving
perception.

The freedom most desirable for moral or loving action is employed more subtly in
successive acts of continuous attention and imagination. More basic than merely acting
on one’s intentions, freedom is defined by Murdoch as the being of a virtuous agent,
which for her is either synonymous with or flows from the cultivation of specific modes

of practical attention. These modes of attention in turn are in dynamic and reciprocal interaction with an agent’s ready supply of evaluative concepts.43 That is, the agent’s conceptual repertoire relevant to practical action is always in the process of being stretched and refined. This happens partly because the public meaning of evaluative concepts is too crude to capture the complexities and nuances that come through in practical attention. And loving attention is complex partly because its proper objects are not impersonal states of affairs, but persons in all their idiosyncrasies, intricacies and obscurities. According to Murdoch then, loving is the sustained, continuous attention toward grasping the reality of other persons.

On Murdoch’s view loving is a very demanding cognitive task, for at least two reasons. On the upside, lovingly attending to another (at the very least) seems to demand becoming progressively competent in learning the beloved’s peculiar idiom.44 This thought rests on two plausible assumptions: that there is an immense variety of personal histories, and that individuals tend to privatize the meaning of public concepts as these are brought to bear on their own idiosyncratic, private experiences. One important consequence of Murdoch’s thought here is that genuine love tends to transform or to

---

43 “Moral freedom, if it is to be defined at all, cannot, it seems to me, be defined without reference to virtue. A mediocre man who achieves what he intends is not the ideal of a free man. To be free is something like this: to exist sanely without fear and to perceive what is real. I would be prepared to imply that one who perceives what is real will also act rightly. If the magnetic field is right our movements within it will tend to be right.” (Existentialists and Mystics p. 201)

44 In “The Idea of Perfection” for instance Murdoch gives the example of how a user takes the public concept of “repentance” into his own privacy. “The active reassessing and redefining which is the main characteristic of live personality often suggests and demands a checking procedure which is a function of an individual history. Repentance may mean something different to an individual at different times in his life, and what it fully means is part of this life and cannot be understood except in context.” (Existentialists and Mystics, p. 320)
refine the lover’s perception of the beloved in light of the latter’s history.\textsuperscript{45} Another consequence of this radical openness to the beloved is marked by some very specific attendant emotions felt by the lover in the presence of the beloved. The characteristic emotions responsive to the “vast and varied” reality of the beloved are initially those of terror or awe, followed by humble delight when apprehending the difference of others.

On the downside, loving attention is demanding for a more sinister reason. Our fallen human nature, argues Murdoch, corrupts and infects the natural functioning of imagination and attention. Though not a theist (at least not to my knowledge), Murdoch thinks moral psychology has to recover what she calls “ambitious pessimism.” Drawing on Freud she suggests that relying on an unambiguous view of motive (desire) and will is highly suspect.\textsuperscript{46} Of course, one need not rely on Freud to spot the deviousness and deceptiveness of human desire. Yet if desire, motive, and will are fallen, so is imagination. Murdoch makes no bones about this, and it is essential to her view that the functioning of attention and imagination is always tainted with fantasies and ego-affirming illusions.

Let us say that our natural disposition toward fantasy and selfish illusions can be properly characterized as a tendency toward affective solipsism. An affective solipsist is one who treats himself as the main object of his affections. It is easy to see now how Murdoch believes that we can make actual progress in overcoming our affective

\textsuperscript{45} Later she expresses the difficulty of opening oneself to the reality of others on the comparison with learning Russian: “My work is a progressive revelation of something alien to me, something which my consciousness cannot take over, swallow up, deny or make unreal.” (\textit{Existentialists and Mystics}, 373)

\textsuperscript{46} In “God” and “Good” she says about Freud: “He sees the psyche as an egocentric system of quasi-mechanical energy, largely determined by its own individual history, whose natural attachments are sexual, ambiguous and hard for the subject to understand and control. Introspection reveals only the deep tissue of ambivalent motive, and fantasy is a stronger force than reason. Objectivity and unselfishness are not natural to human beings.” (\textit{Existentialists and Mystics}, 341)
solipsism. The transformation lies in developing sensibilities that make us aware of the reality of other people. The difficult path of learning to really pay attention to others is *the only way* to escape the tyranny of the self. She says:

> Human beings are far more complicated and enigmatic and ambiguous than languages or mathematical concepts and selfishness operates in a much more devious and frenzied manner in our relations with them…(and the later)…The love which brings the right answer is an exercise of justice and realism and really looking. The difficulty is to keep the attention fixed upon the real situation and to prevent it from returning surreptitiously to the self with consolations of self-pity, resentment, fantasy and despair. (Sovereignty of the Good, p.375)

This is not the place to assess the merits of Murdoch’s conception of transformational love. Instead, I find therein an inspiration for some natural reservations regarding Frankfurt’s view of love, and this independently of her thesis regarding the moral ambiguity of the will.

3.4.2. Characterizing Volitional Solipsism

I turn now to characterizing volitional solipsism. Roughly, a volitional solipsist is someone who by the operations of his own will, say through iterated acts of volitional investment, forswears meaningful contact with the practical agency of others. I aim to show that loving as diachronic investment comes close to a dangerous form of volitional solipsism.

Let me start with a couple of quick points of clarification. First, I aim to characterize solipsism pertaining to practical attitudes, and not its epistemic variant. According to the textbook view of the problem of other minds, I have privileged, direct access only to the contents of my own mind. At best, I have only derivative, inferential access to the thoughts of others. Thus other people’s thoughts, pains and desires cannot
directly impact me. By analogy, if my agency is constituted essentially by volitional commitments and if these are impervious to being unsettled by the central commitments of other persons, I am a volitionally solipsistic agent. If Frankfurt’s view implies that loving investment separates an agent from others in this sense, this should count *prima facie* against his model of diachronic agency. The standard I employ here is no different than that claiming that the problem of other minds should count *prima facie* against a view of the mind as an insulated, self-sufficient, representational machine.

Second, we must avoid the danger of defining practical solipsism too broadly. I will simply assume that being directed toward intentional objects does not suffice to make a practical attitude solipsistic. If anyone would reject this very plausible assumption, then the solipsism under consideration is benign. The consequence of this restriction is that affective, sensual or volitional loving is not solipsistic simply in virtue of its directedness. If my practical attitudes are oriented toward the pertinent features of the beloved, this does not entail a dangerous form of practical solipsism.

Let me now bring together the mix of assumptions that reveal how Frankfurtian love paradoxically suggests a kind of separation from the agency of others. If Frankfurt’s conception is on the right track, loving action implicitly expresses the value of wholehearted integration (whether synchronic or diachronic). Consequently, diachronic loving naturally tends to eliminate, suppress, sideline or compartmentalize considerations that tend to disrupt agential unity. And if this is right, it is hard to see how Frankfurtian love can essentially involve recalcitrant attitudes that would destabilize synchronic or diachronic unification.
But if the Murdochean picture is on the right track, it is in the nature of love to do just that! For mature beloveds are naturally free to pursue their own good in complex and distinctive ways, especially as they themselves and their projects change over time under the force of considerations often independent and unanticipated by their lovers. So, if love entails a certain mutual fine-tuning to the changing of the other, lovers naturally change plans, expectations, desires, and even entrenched dispositions in response to the varying good of the beloved. Loving attention to others entails at least the possibility if not the cultivation of attitudes that disrupt the lover’s volitional harmony.

The importance of the mature beloved could not reside merely or mostly in our loving investment in them. If loving were essentially constituted by diachronic investment that ignores the complex ways in which the beloved’s intrinsic and relational qualities are temporally dynamic, the beloved’s own volitional essence would seem irrelevant to love. If diachronic loving is constituted primarily by self-reflexive attitudes of recurrent investment in others and if these do not essentially involve feedback mechanisms of responsiveness to their changing practical commitments, then the possibility of a pernicious form of volitional solipsism looms large.

Even if my beloved’s will is partially responsive to my own investment in her, I could not be so presumptuous to assume that the most pertinent aspects of her volitional reality are constituted through my serial investments. And even supposing that my concerns and cares get dispersed throughout her entire will, she will also extend, reconfigure or recalibrate these investments in ways unanticipated by me at the moment of loving action or decision. She is free to absorb my investment in ways outside my
control, since her volition is also at the same time in a dynamic traffic with all sorts of pressures, concerns, and cares independent of my own.

On Frankfurt’s view the most enduring and valuable aspects of my agency are constituted by my loves. These essentially are patterns of central volitional commitments established and maintained by my iterated investment in others. Withholding or foreswearing investment thus endangers the stability of my central volitional commitments and thus threatens to destroy me as an agent. The presupposition at work here is that loving commitment to another could never unsettle or override any of those substantial commitments that constitute my volitional essence. But this is precisely where the presupposition seems wrongheaded. For my beloveds have central volitional commitments of their own which should unsettle or disrupt my own substantial investments precisely on account of the love I have for them. Frankfurt’s view then seems solipsistic in this sense: it either presupposes that our beloveds have no volitional investments of their own, or that their commitments are insulated from undermining our pre-existing central commitments, or that our commitments which do get unsettled are compartmentalized away from other central commitments we have, such that agency is not dissolved. In effect, the practical agency of the people we care about is rendered causally inert.47

3.4.3. Illustrating Volitional Solipsism

Let me briefly illustrate the worry of volitional solipsism. I do so by turning to an example used by Rae Langton in her analysis of Proust’s hero from *Remembrance of

---

47 I am indebted to Mike Rea for comments that helped me express more clearly the worry of volitional solipsism at this stage of the argument.
Things Past, example that plausibly illustrates sensual solipsism. I modify the original example in order to bring out the important analogy between the two kinds of practical solipsism. As she evaluates Marcel’s love for Albertine, Langton finds the former blamable for having made Albertine’s real existence irrelevant to his practical life:

Albertine exists merely as a screen on which emotions can be conveniently projected. Consequently, the important thing is not the worth of the woman. Consequently, her worth is simply instrumental. If Albertine has no more inner life than a piece of furniture, a chair, a table, a display case, then she can have only the value of a tool. Then indeed she proves useful, as an engine for producing those feelings of love and grief that are so interesting to the literary narcissist.48

According to Langton, Marcel’s love does not necessarily reduce Albertine’s personhood to a mere sexualized body.49 On the contrary, Marcel endows her with a range of properties, but crucially, the emerging picture is crafted (whether consciously or unconsciously) in order to induce the relevant emotional reactions from the lover. Although the case of Frankfurtian volitional investment is different, there are relevant structural similarities. Instead of projecting onto the beloved the characteristics essential to the lover’s own satisfaction, the lover invests into what he perceives to be in the interest of the beloved.

One of the keys for seeing the problem here is that for Frankfurt a form of paternal love is the paradigmatic way to conceptualize caring. Paternal love has in common with Marcel’s romantic love a certain kind of propensity toward expansionism. Whereas Marcel invests in constructing an image of Albertine on the basis of predictions about his own expected satisfaction, the parent invests in his child on the basis of his


49 On Kantian story she weaves, Langton thinks reducing the object of love to only her sexual qualities is another form of love solipsism, but she is not interested in analyzing its features in this particular paper.
predictions about the child’s satisfactions. In both cases there is hardly any expected feedback from the object of love. Thus even though self-interest is clearly different from other-interest, but they both can be expansionistic.\(^\text{50}\)

Undoubtedly, Marcel’s love seems repugnant and dangerous to us because it is self-interested. It treats the extensional object of love merely as an instrument for projecting Marcel’s own fantasies. But an additional problem is evident when we come to appreciate the difference between Marcel’s conception of Albertine and the latter’s own practical self-conception. The problem is not only that Marcel constructs an ersatz Albertine as an intentional object for his own gratification. Additionally, her own self-conception and practical commitments are practically irrelevant to Marcel’s fantasies. For we can imagine sad possible worlds in which Albertine freely comes to see herself approvingly as Marcel does, as called to the vocation of satisfying his sensual wants. In this situation it is not only that her self-conception now largely overlaps Marcel’s ersatz Albertine.\(^\text{51}\) More important, her newly developed practical commitments could very well reinforce and intensify Marcel’s own desires. If this happens, Marcel is no longer a sensual solipsist, though he remains selfish and the relationship might still be a failure of some norm of romantic love.

Volitional solipsism does not have to entail possessive desire, or its extreme expression in desiderative solipsism. But it still remains problematic for mature loves. Imagine that Proust oddly enough tells the story of Marcel’s love for Albertine a bit

---

\(^\text{50}\) I don’t mean to suggest here that all forms of expansive desire are unjustified. It strikes as obvious that the expansiveness in parental love is justified. But is not clear to me however, that this model applies very well to mature loves or friendships.

\(^\text{51}\) The point I am trying to make here is independent of a particular causal model that secures this direction of fit. Albertine could come to find out about Marcel’s fantasy of her, and because she likes him independently she may try to secure his affections by molding herself into ersatz Albertine. She could also gradually take on properties of Marcel’s projection, without any direct or indirect influence from the latter.
differently. Imagine that Marcel proceeds to construct an *ersatz* Albertine not from the desire to possess her, but mostly from the honorable and altruistic intention to secure her own good. He has a high regard for her worth, and acts consistently to promote and maintain what he perceives to be her flourishing. Still, assume that Marcel’s desire for Albertine’s good depends on a conception of Albertine that is out of touch with what she truly values or how she perceives herself. If there are no realistic causal pathways by which the Marcel’s loving investment is fine-tuned or reconfigured in light Albertine’s practical self-conception, she still remains irrelevant to Marcel’s practical outlook. The concrete, extensional individual functions only as the condition, the trigger for Marcel’s emotions and desires, altruistic though they are. Marcel remains an altruistic, practical solipsist.

Frankfurt suggests that love has utmost value for us precisely because of the iterated patterns of volitional investment, and not because of the value we find in the intrinsic or extrinsic qualities of the beloved. Frankfurt does not completely eliminate the value of perceiving accurately the beloved. But seemingly the perception relevant to caring always rides the teleological horse of volitional investment. Consider this early statement that captures nicely Frankfurt’s unchanged stance: “The person does not care about the object because its worthiness commands that he do so. On the other hand, the worthiness of the activity of caring commands that he choose an object which he will be able to care about.”

Whatever attention the beloved receives, it is always guided and shaped by the lover’s prospective volitional investment. If Murdochian forms of exquisite attention

---

52 Frankfurt, Harry “The Importance of What We Care About” in *The Importance of What We Care About*, (Cambridge University Press, 1998) p. 94
always follow volitional investments that are iterated diachronically, they likely become disconnected from the particular and idiosyncratic idiom of the beloved. On the one hand, features the beloved considers eminently lovable become practically irrelevant to the lover’s action. Alternatively, the lover becomes invested in features that the beloved simply does not care about.

To correct the predicament of volitional solipsism, what is needed is that occasionally attention takes the lead and guides loving investment. And it seems that typical agents can and do experience such phenomena. Consider situations in which unexpected or unanticipated qualities of the beloved actually awaken or intensify caring. Furthermore, consider having cared about those we now perceive to be unworthy of our love. “How could I have been so blind!” is the anguished diagnosis when facing the betrayal of a friend. If surface grammar is at all reliable in these circumstances, we seem to hold our cognition and not our will responsible for having failed to pick up on the relevant cues. I do not want to exaggerate the extent to which attention is prior to or independent of volitional investment. The point I am trying to make is that the form of attention which always depends on iterated volition investment is too thin and too naïve to sustain the kind of caring relations we value in mature loves and friendships.

Unless loving agency is equipped with robust processes that diachronically calibrate the intensity and tone of loving investment in light of the beloved’s own investments, the lover’s action is maintained and motivated by ersatz beloveds. This is not to say that the Frankfurtian agent could not secure himself a subjectively meaningful life. The above only shows that in the absence of a dynamic calibration of loving investment, and to the extent that the lover’s conception of the beloved is systematically
disconnected from the beloved’s own practical self-conception, a caring agent can attain a subjectively meaningful life and still remain a volitional solipsist. That Frankfurt’s paradigmatic conception of caring should have this consequence may constitute a significant reason to reject or revise it.

3.5. Concluding Unscientific Parable

In the Hebrew Bible, the story of the prophet Jonah provides an enlightening caricature of volitional solipsism. Through the narrative we discover a prophet so invested in his parochial conception of Yahweh that he becomes desensitized to the latter’s real concerns. The Hebrew writer goes out of his/her way to convey that Jonah cares more about maintaining his integrity, by protecting his traditional understanding of the God of the Jews, than about attending to the features of the God he meets in the story. Despite God’s varied ways of trying to direct Jonah’s attention to his expansionistic compassion and to the pitiable reality of Nineveh, the belly of a ship, the belly of the whale, the shadowy booth, are all images that betray Jonah’s persistent blindness. Wrapped in self-pity and self-defensiveness, Jonah secludes himself from God and others. Under one description we could perhaps read a diachronically integrated prophet really caring for the purity of his religion or of his God. Under a different description, we see a comically sad man who has invested so much in his provincial conception of God, that he is systematically incapable of interacting with the real object of his loyalty.

There is a further delicious irony in the story. Paradoxically, the outsiders come to see things that the insiders should have intuited automatically. Those without a history of loyalty or loving investment in God come to know him, while Jonah in his
wholehearted commitment does not. The pagan mariners from the beginning of the story not only recognize that there must be a supernatural force behind the causal natural order, not only are they kind and compassionate to Jonah, but they come to gradually shift from the worship of their parochial gods to worshipping and fearing the God of Jonah. We would have expected Jonah to be alert, vigilant, theologically engaged and sensitive to Yahweh’s call of compassion toward those excluded by the circles of ethnicity, race and religion. But what we get in this story is the reverse. The outsiders are vigilant, responding to an even primitive knowledge of Yahweh, compassionate toward the insiders, while the insiders are callous, self-absorbed, and at peril of losing their own identity.

Finally, it is worth noticing that throughout the entire story God is thoroughly engaged, attending to the minutest details of the situation. God pays attention to the outsider and the insider in concrete, revealing ways. The story ends dramatically with God leaving the despondent Jonah to chew on a rhetorical question. Jonah had come to Nineveh expecting the fire and brimstone of God’s wrath on the pagans. Instead, God shows mercy to the repentant, leaving his elect prophet in the throws of despair. God explains himself: “And should I not pity Nineveh, that great city, in which there are more than a hundred thousand persons who do not know their right hand from their left, and also much cattle?” On the one hand, this shows the nature of Yahweh’s thorough compassionate vision, including not only pagans but also their animals in the range of his kindness. On the other hand, and more subtly, God invites Jonah one more time to share, to join into his own perspective. Subtly, God reminds Jonah that he himself did not know his left hand from his right hand when he chose to disobey and go in the opposite
direction from God’s call. If in his blindness Jonah can be the recipient of God’s attentive care, then so can the Ninevites.

Jonah’s story reinforces the Murdochian insight that loving as exquisite attention to the beloved is sometimes more valuable than love as committed diachronic investment. If this story is any guide, withholding investment sometimes can open up a way to be surprised by the characteristics or the action of the beloved. And this seems desirable even if such openness entails the risk that the interpretive frames through which we’ve been internalizing the necessities of love will be challenged and unsettled. Furthermore, cultivating forms of Murdochian attention seems valuable precisely because of such risks. For Jonah’s story also reveals the extent to which we tend to be paternalistic in our loves. If in our desire to maintain integrity and volitional consistency we do not shy away from loving ersatz gods, is there a reason to think we are more scrupulous in our loving people?
4.1. Preliminary Characterization

In the previous two chapters I argued that the retrospective aspect of narrative understanding cannot be reduced to self-reflexive planning or to volitional love. I have suggested that insofar as the latter are necessary for diachronic and synchronic unity of agency, they must depend on and be shaped by the former. That is, the cultivated ability to understand my past and take responsibility for it seems epistemically and axiologically basic for forming sensible plans and for maintaining volitional integrity.

Narrative attentiveness also seems to suggest original and valuable forms of responsiveness to second-personal reasons. This is partly a conceptual and partly a phenomenological claim. Attunement to our own personal histories and dispositions seems inseparable from attunement to the reality of those who have laid claims upon us. Unlike practical rationalities that focus on the first-person stance or the third-person perspective, narrative approaches emphasize second-person perspective (interlocutor) thus sketching fruitful, under-explored alternatives in the contemporary debates. The distinctiveness of the narrative approach resides in grounding forms of practical
responsiveness to others, richer and subtler than those based in the duty-bound concepts such as obligation or respect or accountability.¹

In this chapter I aim to offer an explication of narrative understanding. I begin by locating the concept of “narrative” at the appropriate explanatory level. My subsequent discussion revolves around two central arguments inspired from Alasdair MacIntyre’s work. The main upshot of my discussion of the first argument is that narrative understanding is a distinctive form of practical reason that straddles the internal/external divide of agency. Additionally, I argue that narratives are internally grounded in the rich life of the emotions, which I conceive on the model of quasi-cognitive interpretive frameworks. The main upshot of my discussion of MacIntyre’s second argument is that narrative does not properly apply to our concept of selfhood. Rather, when an agent (as opposed to an automaton) acts, she identifies with or expresses something like a dramatic persona already in dialogue with other such dramatic personae constitutive of her diachronic agency. Practical deliberation implies inhabiting different dramatic personae in mutual dialogue. Each persona in turn can be fruitfully conceived as a character with a history that expresses enduring emotional sequences. Thus the concept of dramatic agency builds upon the conclusions regarding narrative understanding.

¹ This comment regards Stephen Darwall’s newly developed account. While I applaud Darwall’s shifting the scope of practical reasoning from first-person the second-person perspective, the concept of mutual recognition derived from emphasizing the concept of blame and accountability seems to me at once too narrow and too broad. There are ways of being responsible or blamable for transgressing certain social or moral norms, without at the same time being moved by the reality of any concrete individuals. On the flip side, and more relevant for this project, not all modes of awareness of the concreteness and individuality of others necessarily ensue into strict moral accountability or responsibility to them. Our responsiveness to others is as fine-grained and richly textured as the conceptual richness of thick-valued judgments that undergird it.
To begin with, it is instructive to recognize some of the levels at which the concept of narrative can be invoked in the discourse of contemporary moral psychology and agency. It seems the concept of narrative can play a distinct explanatory role for events at levels as different as the sub-personal, personal, social and world-historical (we might call this level the supra-personal). Depending on the level at which the theorist chooses to emphasize the main work of narrative, one gets very different accounts of practical agency.

4.1.1. Sub-personal narratives

Let us call an austere sub-personal narrativist one holding that the events at the personal, social and supra-personal level are structured according to myths, whereas only processes occurring sub-personally, below the conscious level, have true intelligibility. Only the stories at sub-personal, sub-conscious level have clear, world-representing directionality, unlike the stories that characterize sequences of individual human action, socio-political events, or natural happenings. Generally, sub-personal stories denote any type of structures supposedly operating on events and processes not directly or manifestly accessible to human consciousness at, or immediately preceding, the moment of action. Stories involving explanations by appeal to genetic processes, Schopenhauer’s Will, Nietzsche’s Dionysian, or Freud’s death drive would exemplify the stock of sub-personal stories.

Arguably, at least in one of his explorations into the topology of the human psyche, Nietzsche could be described as an austere sub-personal narrativist. Drawing on the mythological underpinnings of Greek tragedy, the early Nietzsche famously invoked
the Dionysian and the Apollonian as the pre-reflective drives that configure psychic energies.² Whatever reasons for action might be cited by agents as diverse as the saint, the natural scientist, or the advocate of democracy, the true springs of their action are essentially constituted by the interplay of these pre-reflective drives.³

If one gets spooked by the prospect of explaining human action as an expression of Greek tragic drama, the alternatives for the sub-personal narrativist abound. Witness the enduring appeal of explanations as diverse as those proceeding from sociobiology or from psychoanalysis. That terms like “the selfish gene” or “Oedipal complex” are not conversation stoppers suggests that ordinary folks continue to find attractive both the idea of a storied sub-personal, and the idea that it explains the events unfolding at the personal and the social levels.

Interestingly, even if compelled by the reality of such sub-personal stories, those inclined to emphasize the independence of the various levels of explanation for human action will reject austere sub-personal narrativism. The possibility of level independence reveals the hefty burden of proof that an austere sub-personal narrativist must carry. First, he must tell a compelling story of the Nietzschean, Socio-biological, or Psychoanalytic (or other such) variety. Second, he must establish the merits of one particular sub-personal story relative to the competing stories at the same level. Finally, this theorist owes us a story of how the ostensibly intelligible explanations of personal or social events bottom out without residue in the veiled, sub-personal story.

² See Nietzsche, The Birth of Tragedy (Dover Publications, 1995)
³ Evidently, only retrospection can reliably uncover the true springs of human action. For the early Nietzsche, it seems that typical agents unwittingly internalize culturally available powerful images, myths or icons, which in turn determine the causal antecedents of their action. We have here the interesting possibility that social (supra-personal) stories can be assimilated into the sub-personal without the reflective control of the personal.
In his later years Nietzsche himself seems to have given up on the austerity of his youth. While Nietzsche seems now to be a pluralist about forces that shape his self-understanding—it is the playground for both observable and hidden forces—he still retains the possibility of mapping the hidden drives of human action through a genealogical investigation. At the same time he also becomes aware that the methods of investigating the manifest and the latent levels of human psychology belong to different orders. For instance, in *Beyond Good and Evil* Nietzsche likens uncovering the hidden springs of action to a “big-game hunt” into a virginal forest. The moral psychology which surveys the jungle of the human self by traditional epistemic means (dogs and beaters) appears not only ridiculous but also dangerous. The reason for this is that traditional, Greek-inspired scientific methods and virtues cease to function precisely where self-knowledge becomes interesting:

The drawback in sending scholars out into new and dangerous hunting-grounds where courage, prudence, subtlety in every sense are needed is that they cease to be of any use precisely where the ‘big hunt’, but also the big danger, begins—precisely there do they lose keenness of eye and keenness of nose.

Whatever we may think of his subtle attack on Platonic psychology here with its too sanguine epistemological assumptions, Nietzsche is not only concerned to point out the limits of scientific, objective methods of investigating the self. More positively, he seems to subscribe to the view that tracing the contours of moral agency is possible only if the historically situated knower tries to understand (and identify with) the linguistic, perceptual, emotive stance of concrete others who instantiate paradigmatic human

---

4 Consider for instance this passage from Frederick Nietzsche, *Beyond Good and Evil*, (Penguin Books, 1990) p. 74: “The human soul and its frontiers, the compass of human inner experience in general attained hitherto, the heights, depths and distances of this experience, the entire history of the soul hitherto and its still unexhausted possibilities: this is the predestined hunting-ground for a born psychologist and lover of the ‘big-game hunt.’”

5 Nietzsche, *Beyond Good and Evil*, p. 74.
experiences. To illustrate this, notice his honest ambivalence, his simultaneous attraction and repugnance toward the prospect of being a scholar of religious conscience:

To divine and establish, for example, what sort of history the problem of knowledge and conscience has had in the soul of homines religiosi one would oneself perhaps have to be as profound, as wounded, as monstrous as Pascal’s intellectual conscience was – and then there would still be needed that broad heaven of bright, malicious spirituality capable of looking down on this turmoil of dangerous and painful experiences, surveying and ordering them and forcing them into formulas. – But who could do me this service!...In the end one has to do everything oneself if one is to know a few things oneself: that is to say, one has much to do!6

The last couple of paragraphs may seem like distractions from our main argument. If so, they are important distractions. It seems that as Nietzsche fleshes out his genealogical perspectivism, his own theoretical understanding of human psychology takes an intensely personal turn. If one’s theoretical and practical stances are constituted by historically laden states of mind, then the psychologist can make sense of the practical stance of others only by some kind of empathetic, simulationist posturing. Nietzsche’s own example (he is happy to call himself a psychologist) is in marked contrast with much contemporary cognitive science, animated as it is by a blissfully reductive spirit. Mocker of the dogs and beaters of the moral psychology trendy in his days, Nietzsche would undoubtedly resist our contemporary attempts to reduce religious belief (or altruism, say) to the proper (or improper) functioning of cognitive sub-modules selected by hidden evolutionary mechanism.

One might be inclined to reject any particular account, be it Nietzschean or Darwinian, while still retaining a general worry about the explanatory power of reductive sub-personal stories. I offer here a suitably general reason to help attenuate this worry. If

---

6 Frederick Nietzsche, Beyond Good and Evil, p. 74.
narratives configure events only at the sub-personal level, then the stories that seem to organize personal and social events are either fictional or they are wholly reducible to sub-personal narratives. But if narratives are fictional at higher levels, might we not be tempted to hold that sub-personal stories are equally fictional? Suppose that personal narratives are helpful myths we cook up, pressured as we all are to make sense of our lives. Meanwhile, the real engines of our actions are those all-too-real sub-personal narratives, to which unfortunately we have no epistemic access. We could worry then that if these sub-personal stories are intelligible at all, though only by some genealogist of morals (or psychiatrist or social biologist), they too might be the fictional projection of other agents or intentional forces, albeit less manifest ones. And so we’re off on an infinite explanatory regress.

Now if narratives we employ as explanations at higher-level are not fictional, they could only be real in virtue of boiling down to nothing but sub-personal stories. But _prima facie_, there is no in principle reason to disqualify higher-level narratives from being conducive to truth. So, the austere sub-personal theorist must draw a compelling distinction between real, veiled stories on the one hand, and false, manifest ones on the other. And as he does this, the _explanatory_ gap between the levels can only widen. For the austere sub-personal narrativist would have to provide an account of how all the stories at the personal and social levels would be explained without residue in terms of sub-personal patterns. Theoretical reduction is understood nowadays either in terms of bridge-laws connecting distinct levels, or in terms of identity relations, or by appeal to

---

7 Fictional is to be distinguished from stylized stories. The latter have more reliable representational qualities than the former.
functional equivalence. But with respect to the first two models, it is very hard to see what explanatory regularities could hold between the higher and lower narrative levels. And what of the concepts entering such regularities? Would they be hybrid concepts, part sub-personal and part personal? Perhaps then the functional equivalence could be made to work. Perhaps we have a clear idea of what it is for the concept of “gene” to be functionally equivalent to the concept of the “DNA-molecule”. Yet even if one concept can reference an underlying causal processes that explains another concept, this hardly seems applicable to stories, which consist of complicated inter-dependent strings of concepts.

Consider Anscombe’s famous example about the difficulty of characterizing the intentions of a person you see pumping water. Imagine giving a functional description that tells a complex causal story involving neurons in interaction with the arm muscles, then contact with an iron handle and so on. This description would be compatible with all of the following stories but none of them are explanatorily reducible to it, nor are they reducible to one another: “The person is pumping the water”; “The person is poisoning the city’s water supply”; “The person is taking revenge on her enemies for having destroyed her home”.

Perhaps I am not creative enough to imagine functional reductions holding between stories. But even if narrative explanatory reductions could be carried out, it is still hard to see how they could be successful without importing explanatory concepts from higher levels. For instance, if the concept of “selfishness” applies primitively to

---


144
persons, and only derivatively to genes, Richard Dawkins’ popular explanation of human motivations in terms of genes is successful, but only by a sleight of hand.

Before we go on looking at some of the ways in which the concept of “narrative” can be invoked at what I called the personal level, a couple of caveats. First, I want to register the conviction that personal narratives are not reducible to supra-personal (cultural) narratives, neither are the latter reducible to the former. A natural extension of the argument thus far would conclude that **austere personal narrativism**\(^9\) and **austere supra-personal narrativism**\(^10\) are equally misguided.

Second, I recognize the possibility of mutual **shaping** and **influencing** among narratives from all levels.\(^11\) I leave these relationships unanalyzed and vague, since I care about the partial dependence of one level on another, and consequently about a certain fluidity of meanings between the levels. All the same, I want to affirm the independence of each of level.\(^12\)

---

\(^9\) It is hard to give examples of people falling into this category without caricaturing them. Those trusting the power and shape of their own experiences, while discounting both covert (psychoanalytic, Nietzschean, socio-biologic) and structural explanations for their action (Marxist, Hegelian, Foucauldian) seem to fit the bill.

\(^10\) All **real** narrative explanations of human action are structural explanations. Here again, one has options: Hegel, Marx, and Foucault could serve as inspirations for an austere supra-personal narrativist. Also, it bears saying here that a narrative at this level, if unconsciously internalized, could function as a motivational story at the sub-personal level.

\(^11\) I think it is important to emphasize that the notion of “shaping” or “influencing” already situates us into a conceptual network distinct from the strict causality of determinism or probabilism. How problematic is the assumption that something like **thematic** connections stretch throughout all the levels of human psychology? It seems to me that this issue is problematic only for those who are already committed to render all acts of interpretive agency in terms of Humean causation. My view here is consistent with David Velleman’s view on narrative. In his important article “Narrative Explanation” he argues that even though narrative plots track causal networks, this does not account sufficiently for the explanatory power of narratives. The main part of his argument there is devoted to showing how “something other than causality or probability serves the function of differentiating narratives from other genres and endowing it with its peculiar explanatory force.” (David Velleman, “Narrative Explanation” *Philosophical Review*, **112** (1), January 2003, p. 4)

\(^12\) This point is an important one in the argument of the second main section. There I make the point that although biological/Freudian stories make sense of emotional reactions at a certain explanatory level
4.1.2. Narrative at the Personal Level

Typically, philosophical accounts of psychological continuity rely on 

*psychological atomism.* That is, they assume that in order to account for the continuity of 

the self, the contents of human psychology should first be identified piecemeal, in 

isolation from one another at a time, and only then connected by causal relations 

(Lockean ties) of continuity or similarity to other discrete psychological events at 

different times. Furthermore, the assumption of *psychological atomism* has not only 

framed the personal identity debates, but has also made attractive and trendy a mode of 

explaining agency and motivation in terms of belief/desire complexes.

Yet some theorists working on personal identity either reject or qualify the 

assumption of atomism. For instance, Marya Schechtman\(^{13}\) and Marc Slors propose that 

narrative connections explain better the continuity of the self than the Lockean causal 

links common to standard versions of Psychological Continuity theory. According to 

Schechtman rejecting atomism implies affirming “that the beliefs, values, and desires that 

make up these connections are already deeply intermeshed when we first identify 

them.”\(^{14}\) A psychological state comes attached with intricate historical and semantic 

strings (not all of them causal) that give it its meaning and intelligibility relative to a 

structured context of other such mental states. As such, any given causal relation among 

distinct psychological items makes sense only in light of a prior understanding of 

---

\(^{13}\) See Schechtman’s book *The Constitution of Selves,* (Cornell University Press, 1996) where she 

argues that the narrative structure of a subject’s experiences and memories is a necessary condition (though 

not sufficient) for personal identity.


Explorations Nr. 2,* 2001 p. 100.
historical, social, or broader frameworks that situate this relation. Consequently, a
sensible atomism carries a commitment to some kind of holism.

Some, like Marc Slors, commit to a more modest thesis, namely that the standard
(Lockean-type) causal connections by themselves are insufficient to explain the kind of
connectedness necessary for personal identity. Psychological atomism might explain all
sorts of mental happenings, yet we need some extra-causal relations between the content
of mental states in order to account for personal identity over time:

Rather than causality, what binds together the mental states of such sequences are
relations of mutual dependence between the contents of these states. These are
relations in virtue of which the consecutive occurrence of mental states with the
contents they have make sense relative to each other, not relations by which these
consecutive occurrences are scientifically explained.  

Slors argues extensively that these relations of narrative-continuity (what he calls
N-continuity) can characterize phenomena as diverse as sheer perceptions, reflective
conversations, or considered accounts of the history of a friendship. Slors explicitly aims
to avoid commitment to “narrative” understood as particular literary genre, or to any
connotations of constructivism. Narrative connections must be essentially person-
relative, since the property of “making sense” necessarily references a possible or actual
audience. But this is a far cry from holding that the agent is the sole or even the primary
constructor of one’s identity. Furthermore, Slors’ narrative-continuity only commits him
to a minimal teleology: “For whereas in cases of robust teleology we can explain a
system’s present state by referring to the future purpose this state serves, in the case of

---

15 Marc Slors The Diachronic Mind: An Essay on Personal Identity, Psychological Continuity and the
psychological connectedness of the kind I am describing mental states often can only be rendered intelligible by later mental states in retrospect.”

In rejecting psychological atomism, a narrativist can hold more generally that at least some mental events are essentially defined by reference to the intelligible sequence of contents that they are part of. If she refers to particular mental states (feelings, thoughts, or perceptions) in separation from the sequence of contents (or process of events) that bestow intelligibility on them, the narrativist might be referring to abstractions helpful for elucidating some range of mental phenomena.

Even though my task involves “narrative” as a form of practical rationality and not persistence of the self, I can cull some insights even from those who use this concept in relation to personal identity. Generalizing the points made by Schechtman and Slors, we can take some first steps toward a provisional characterization of “narrative structure” at the personal level. First, an event is narratively structured only if it gains meaning by reference to other events in an experienced sequence. Second, the concrete sequence of events must be capable of being articulated (describable) by some agent. Third, paradigmatically the form of these psychological processes or sequences of lived events is socially decipherable. That is, there is the recognition that psychological holism might have to depend on a broader social context of shared stories.

4.2. MacIntyre and Intelligible Action

Since my main focus is not on personal continuity, I use the above insights only heuristically and insofar as they help elucidate the role of narrative for practical

---

16 Ibid, p. 88
rationality. A central discussion devoted to the relevance of narrative for practical reason unfolds in chapter 15 of Alasdair MacIntyre’s *After Virtue*.

MacIntyre draws here a penetrating analysis that aims to show that “narrative” belongs to a conceptual network that includes “intention”, “intelligibility”, and “human life”. The first leg of the argument maintains that the notions of human action, intelligibility and narrative are conceptually related:

1. Human action is intelligible if only if it is appropriately situated in truthful, enacted narrative sequences.

2. Human action is intelligible.

3. So, human lives and human actions are narratively structured.

The second leg of the argument extends the application of “narrative” from the concept of human action to that of the self, the conscious subject of human experiences. The argument here can be interpreted as structurally similar to the previous one:

4. A human self is intelligible if and only if the agent inhabits the unity of a character in her own actual or possible history.

5. Human selves are essentially intelligible.

6. Selves are essentially narrative.

I return to this second argument momentarily. For the remainder of this section, I focus on the first leg of the argument, considering several objections to it and relevant MacIntyrian replies. In section 4.2.1 I defend MacIntyre’s conception of narrative practical reason by interpreting it as being constituted by sequences of intentional processes that straddle the internal/external divide. In sections 4.2.2 and 4.2.3 I sketch out and further support MacIntyre’s replies to two fundamental objections to his view:
that narrative is essentially different from life, and that narrative falsifies life. Section 4.3 constitutes a more critical assessment of the second leg in MacIntyre’s argument. I argue that whereas the notion of narrative applies primarily to intentional action, it applies at best derivatively to the self. In particular, I argue that David Velleman is correct to hold the human agent is intelligible under different guises, some of which may not be narrative. Drawing on a recent conception of the emotions as quasi-cognitive interpretive frames, I suggest that these suitably can play the role of the inner psychological vehicles supporting MacIntyre’s conception of narrative practical rationality. At the same time, the emotions also furnish fruitful resources to view agency as cast of self-interpretive guises in dialogue. This leads me to end this chapter by proposing a two-tier account of dramatic agency as an alternative to MacIntyre’s narrative self, and to other competing conceptions of diachronic agency.

4.2.1. The First Stretch: Intention and Narrative

Starting from an analysis of human action, Alasdair MacIntyre argues that chains of intentions are intelligible to the subject and to others only as belonging to larger historical social contexts and practices. MacIntyre suggests that human action can be rendered intelligible only by making reference to descriptions in two different, but related contexts, namely the intentional context and the institutional-social context. In intentional contexts, an action is explained by reference to the agent’s intentions, which being of variable duration would involve descriptions reflecting the appropriate diachronic connections and the appropriate level at which they are occurring. Furthermore, integrating a human action in its appropriate intentional train (chain?)
requires both an (implicit) understanding of which of her belief/desire pairs are causally efficacious at the time and the temporal ordering that allows short-term intentions to be embedded in long-term ones.

Following MacIntyre’s example, consider the question: “What is he doing?” In response to this question my action can be described as falling under different intentional descriptions which are not only mutually compatible, but also ordered in some way: “Moving his fingers”; “Typing words on the computer”; “Completing a chapter in his dissertation”; “Trying to get a tenure-track job”; “Making his wife happy”. Notice that intending to complete this chapter seems to presuppose at least that I have (implicitly) the belief that continuing to type strings of appropriate words will serve this purpose and that I desire to do so. Additionally, it would seem that the demand of intelligibility, if indeed a recurring and constant feature of human psychology, encourages the integration of short-term intentions in larger and more stable long-term chains. Developing this example, my intending to type is incorporated in the longer-term intention of completing the dissertation, and this in turn is embedded in the larger and diachronically more complex intention of getting a tenure-track job. Finally this latter intention is steadily nested inside an even more complex and unwavering (alas, perhaps Quixotic) intention of making the wife happy.

It’s apparent from the above that MacIntyre believes intelligibility demands that psychological characterizations of intentions need to be simultaneously causal and temporal. But this thought raises at least two further questions. First, how precisely do iterated, embedded intentions in larger intentional contexts display the historicity or narrativity of action? Second, what is the relationship between the causal description
and the historical characterization of intentional sequences? The first question has a manifest and pretty compelling answer. Answering the second question turns out to be more complicated.

In response to the first question, MacIntyre seems to say that historicity and thus narrative enters the explanation of action both from the agent’s psychological context and from her larger social setting. The problem is that one might be tempted to think that the phenomenon of embedded intentions presupposes only some sort of teleological explanation, but not necessarily a narrative one. Possibly my intentions are structured a-temporally, synchronically, like Russian nested dolls: I intend to type for the sake of completing the dissertation, I intend to complete it for the sake of getting a job, etc. If this is a genuine possibility, the demand of intelligibility would seem to nudge toward teleology, but not necessarily historicity. So, whence historicity?

Part of the reply is that historicity enters the explanation due to the fact that the social settings of human action are essentially historical, and intelligibility demands that our psychological characterizations involve essentially our social settings. To illustrate, the settings, practices, or institutions that contextualize my intention to complete the dissertation have themselves evolved (and are evolving) historically. They have complex histories such as those exemplified by the historically contingent practice of requiring the completion of a Ph.D. dissertation in order to be seriously considered for an academic job in the American Universities. Or consider the way in which the institution of marriage has changed only in the last 30 years in the Western world, ways that ineluctably place expectations on the relationship between spouses, expectations considerably different from those 50, or 80, or 100 years ago.
Even more significantly, MacIntyre holds that history shapes not only indirectly or contextually human action. Rather, if human action is essentially characterized by nested intentionality that unfolds causally and temporally in a concrete historical setting(s), then it contributes *causally* to the agent’s own history *and* to the history of her settings:

We place the agent’s intentions, I have suggested, in causal and temporal order with reference to their role in his or her history; and we also place them with reference to their role in the history of the settings or settings to which they belong. In doing this, in determining what causal efficacy the agent’s intentions had in one or more directions, and how his short-term intentions succeeded or failed to be constituted of long-term intentions, we ourselves write a further part of these histories. Narrative history of a certain kind turns out to be the basic and essential genre for the characterization of human actions.17

It is important to note here a possible radical implication of MacIntyre’s position. One way to interpret the dependence of intentions on history and settings is hold that the semantic contents of intentional states depend essentially on historical (external) practices and institutions.18 But we can also take MacIntyre to espouse the more ambitious thesis that the very cognitive and conative structures and practical processes that support the function of human intentions are also partly, but essentially historical. To explain what I mean here, I borrow some seminal ideas from Mark Rowlands’s account of vehicle externalism in philosophy of mind.19

According to Rowlands, in contradistinction from *content externalism*, *vehicle externalism* claims that not only mental contents are individuated by items in the external

---

18 Though there are many ways to express content externalism, an intuitive one is that that the individuation of some mental contents depends on items that exist outside the heads of the subjects of those mental states.
environment, but so are the cognitive vehicles (mental architecture and processes) responsible for mental operations. Rowlands sums up this framework in four main theses:

1. The world is an external store of information relevant to cognitive processes such as perceiving, remembering, reasoning and so on.

2. A process such as perceiving is essentially hybrid – it straddles both internal and external forms of information processing.

3. The external processes involve manipulation, exploitation and transformation of environmental structures that carry information relevant to accomplishing the perceptual task at hand.

4. At least some of the internal processes are ones concerned with supplying the cognizing organism with the ability to use appropriately relevant structures in its environment.20

How might an analogous framework be developed for human action, instead of human cognition? I want to stress the “analogous” aspect, first because I am merely trying to explicate at more length MacIntyre’s central thesis (premise (1) above) that action is essentially narrative. Second, because if Rowlands’ vehicle externalism is correct, then if intentional processes depend on cognitive ones, perhaps it is more likely than not that a relevant version of practical vehicle externalism is also true. But neither the motivation for nor the truth of the practical framework needs to depend on the theoretical framework. In keeping with MacIntyre’s Aristotelian and Wittgensteinian

20 Rowlands, p. 189
persuasions, and remembering that for him the information necessary for action is
codified in terms of historical narratives, consider the following practical version.

1. The social world (setting) is an external store of codified information
   (narratives) relevant to intentional processes such as valuing, deciding,
   practical reasoning, emoting, etc.

2. A process such as practical reasoning is essential hybrid – it straddles
   both internal and external forms of intentional processes.

3. The external processes involve the manipulation, exploitation, and
   transformation of (external) historical narratives relevant to the piece of
   practical reasoning at hand.

4. At least some of the internal intentional processes are ones concerned
   with supplying the agent with the ability to use the relevant external
   processes in their historical setting.

I’m trying to decide whether to allow my five-year old daughter to play soccer in
a recreational league this fall. The framework stipulates that there is already scripted
information in the world pertinent to my decision. Scripts specifying the behavior of
soccer moms and dads every Saturday, scripts outlining the expectations placed on
college teachers on any given weekend, scripts regarding the socialization of little girls
through sports, and scripts regarding the physical talents, dispositions and desires of a
certain five-year old girl.21 As an agent attentive to my social world, I already possess
the raw materials for deciding, and they are located (perhaps most of them implicitly and
dispositionally) in my social setting. Furthermore, in light of wise counsel and

---

21 There are, naturally, many other scripts that I implicitly acknowledge with respect to this decision. The scripts above, if anything, tell about how much we have to simplify an actual case in order to be able to describe it.
conversations that also reference this scripted world, and in light of my own historical dispositions I will in the end come to enact a script (or a combination of scripts) that terminates in a decision.

Interestingly, MacIntyre explicitly acknowledges that the widespread phenomenon of conversation is a natural and paradigmatic locus of narrative practical rationality, since it is one of the most often used vehicles for accessing our social environments. Conversation presupposes implicitly agreed upon external scripts, the possibility of connecting internal intentional worlds with external descriptions, and the possibility of improvisation, that is, of causally co-contributing to existing narratives.

Furthermore, MacIntyre compellingly holds that if our external social world is constituted by an ever-increasing stockpile of narratives, then self-understanding and other-understanding can only come about by careful attention to the rich dramatic repertoires interwoven into its fabric. “Deprive children of stories and you leave them unscripted, anxious stutterers in their actions as in their words”, MacIntyre says. And this is not just a witticism about the development of moral vocabularies in children. Rather, in context this seems a general thesis about the proper contours of practical reasoning and human action. This dramatic-external setting along with appropriate forms codifying it, expressing it and responding to it, are actually constitutive of practical reasoning. A human agent “is not essentially, but becomes through his history, a teller of stories that aspire to truth.” And consequently practical rationality resides primarily

---

22 “We allocate conversations to genres, just as we do literary narratives. Indeed a dramatic work, even if a very short one, in which the participants are not only actors, but also the joint authors, working out in agreement or disagreement the mode of their production. For it is not just that conversations belong to genres in just the way that plays and novels do; but they have beginnings, middles and endings just as do literary works. They embody reversals and recognitions; they move towards and away from climaxes.” *(After Virtue, p. 211)*
neither in fashioning agency out of fresh cloth, nor in merely following abstract, timeless principles. Its main task is representing oneself on the map of this complexly structured practical terrain. I can only answer the question ‘What am I to do?’ if I can answer the prior question ‘Of what story or stories do I find myself a part?’”23

In the previous paragraphs I have suggested that MacIntyre has a ready answer to what I called the first question, namely, ‘In virtue of what do sequences of intentions count as narrative?’ I have also suggested that he should be taken as replying plausibly that narrative practical reasoning “straddles” the internal/external divide. An action is intelligible only if embedded in long-term intentions, which are the partial product of processes that reference external, historical narratives (or their relevant segments). Notice that in providing this answer we have gone a long way toward giving support to MacIntyre’s first premise in the first leg of the argument. Nonetheless, I haven’t said much yet about what I called the second question, ‘How should we understand the relationship between causal descriptions and temporal characterizations of actions?’ I postpone the detailed discussion of this question until the next main session. Here I will just limit myself to some quick, programmatic remarks.

MacIntyre holds that human action is intelligible. Yet given that intelligibility depends on both causal descriptions and temporal descriptions, it seems appropriate to ask “How do causes enter historical processes?” MacIntyre does not provide a clear answer here. Do causal descriptions pertain to a context wholly different and incommensurable with the temporal/historical context? If yes, then the account of agency will have to be fragmented at least between the two different modes of

---

23 MacIntyre, After Virtue, p. 216
explanation (see my discussion of Velleman’s view position in the next section). If however, the two contexts are commensurable, then a more detailed account of how causes track historical networks is needed. For it is a trite point that causal sequences do not always tell stories, and that narrative connections almost always involve something additional to causal connections among events.\(^{24}\) And if MacIntyre takes the narrative connection to provide a more basic explanation of human action and self than causal connections, then where and how do causal relations enter this story?

So far I have been discussing some reasons for believing the main premise in the first leg of MacIntyre’s argument. It’s time now to deal with two main objections to it. I’m going to call the first objection *narrative is different from human life*, and the second *narrative is falsifying of human life*. We can plausibly construe the first objection as an objection to the truth of premise (1). The second objection grants the truth of the premise but still rejects the conclusion.

4.2.2. First Objection: Of Life and Narrative

‘We live our lives forward and only understand them retrospectively’ goes the platitude. The first worry seems to be that since “[l]ife has no beginnings, middles, or ends” as Louis O. Mink says, then if narrative enters anywhere in the explanation of human action, it organizes mental events in ways essentially different from lived events. The problem seems to be that MacIntyre’s first premise is false because of an essential

\(^{24}\) This point will be emphasized in much more detail below in our discussion of the role of emotions for narrative understanding. For a preview recall E.M. Forster’s often cited quip: “The king died and then the queen died is a story. The king died and the queen died of grief is a plot. The queen died and no one knew why until they discovered it was of grief is a mystery, a form capable of high development.” Forster presents in the first sentence simply a sequence of events that does not constitute strictly speaking a narrative, a story. We get the qualities of narrative in the second and third sentence, even though, or rather, *precisely because* you can’t count on grief to cause death in a sufferer.
equivocation. It mistakenly presumes that properties belonging exclusively to artistic forms of representation also make a cozy home in lived events. But, the objection continues, intelligibility and living belong to incommensurable domains. See Mink again as quoted by MacIntyre: “There are hopes, plans, battles and ideas, but only in retrospective stories are hopes unfulfilled, plans miscarried, battles decisive, and ideas seminal.”

In response MacIntyre argues that retrospective intelligibility does not entail that narrative explanation belongs exclusively to artistic forms of expression. Even if narrative represents one way of ascribing meaning retrospectively to events, this is not only a feature of art. In fact, one may suppose that the widespread experience of losing someone dear essentially and naturally involves reference to “unfulfilled hopes” and “final partings”. MacIntyre makes a subtle point here. If lamenting the death of a loved one is a universal human experience, the story is lived first, and retrospectively recounted second, and perhaps represented artistically third.

And to someone who says that in life there are no endings, or that the final partings take place only in stories, one is tempted to reply, “But have you never heard of death?” Homer did not have to tell the tale of Hector before Andromache could lament unfulfilled hope and final parting. There are countless Hectors and countless Andromaches whose lives embodied the form of their Homeric namesakes, but who never came to the attention of any poet.

That we retrospectively characterize our hopes as “unfulfilled” and our ideas as “fertile” may very well indicate that a present mental attitude imposes a narrative ordering on a sequence of discrete past lived events. But equally this ascription could also express the intensification, or reversal, or diminution of lived stories or storied lives,

---

25 Alasdair MacIntyre, After Virtue, p. 212.
26 Ibid, p. 212.
whose content gets enhanced or transformed through retrospective recounting or artistic expression. Ascribing meaning through retrospective understanding is thus compatible with lived events already possessing the qualities of narrative. Furthermore, from an agent’s intentional stance it looks that her action in the world makes causal contributions to her existing stock of stories. At least in some areas of human endeavor, the agent’s intentions are not only explained by the pressure of narrative coherence (the intended action is intelligible as a continuation of some personal narrative strand) but also motivated by it. If living is motivated by specific or general desires to understand one’s life retrospectively, as it often is the case, the sharp distinction between intelligibility and living seems arbitrary and unsustainable.

The above point does not depend on the presence of an overly intellectualistic form of self-reflection. If it did, the Minkean objection would still have some force. It’s true that some intelligent agents unable to make sense of their lives report motivational failures. But likewise do those who do make sense of their lives, but who for whatever reason cannot care about their personal history any longer. They are empathetically dislocated from their past. And in extreme cases, some become stymied, apathetic, passively letting life happen to them, merely surviving and not living through and by the means of their creative agency. On the flip side, some agents are in the overwhelming grip of ideological or personal narratives that render them a bit too quick to act. The apathetic and the revolutionary testify to the causal powers of lived stories.

27 I find myself recalling Jordan (my five-year old daughter) expressing her desire to play soccer. I also remember seeing her kick the ball with her sisters and with the neighborhood kids on our street. I remember noticing those incipient skills that could easily be enhanced and developed by a good coach. So I find myself not only wishing or desiring to have Jordan play soccer in a fall league. I intend to do something about it, and my intention causally contributes to this particular strand of Jordan’s storied life and thereby to one of my own.
4.2.3. The Second Objection: Falsifying Life

The second challenge to face MacIntyre’s first leg of the above argument, plainly inspired by Sartre, represents a blunt rejection of the conclusion. Differently from Mink who holds only that narrative doesn’t apply to life, Sartre seems to attribute to his *The Nausea* character the belief that “to present human life in the form of a narrative is always to falsify it.”\(^{28}\) Since the storyteller imposes a meaning on series of discrete events, meaning which these could not have had at the live moment of action, to narrate is to baldly falsify. We may summarize the falsification argument as follows:

1. Human action is intelligible only if is appropriately situated in truthful, enacted narrative sequences.
2. But there are no and there could not be any truthful narratives.
3. So all human actions are unintelligible events.

As MacIntyre points out, there is no disagreement between him and Sartre on the truth of (1*). The trouble is with (2*). MacIntyre hints that the intuition supporting this Sartrean premise is analogous to the one that motivates Mackie’s error theory.\(^{29}\) The thought explicitly voiced in support of (2*) is that “[h]uman life is composed of discrete actions which lead nowhere, which have no order; the story-teller imposes on human events retrospectively an order which they did not have while they were being lived.”\(^{30}\) Patently, just as moral discourse is shot through with projective error, so is narrative discourse and practice. Furthermore, perhaps the same basic psychological mechanisms

---

\(^{28}\) Alasdair MacIntyre, *After Virtue*, p. 214.

\(^{29}\) Here is perhaps a fictionalist argument regarding moral judgments. If moral discourse is non-fictional, then moral concepts must refer. But there is nothing in the world that answers to them. So, moral discourse is fictional.

\(^{30}\) Ibid, p. 214.
are responsible for both forms of projection (moral and narrative), and perhaps their mutual interdependence is a complex expression of our basic need to enchant our otherwise cold and drab world.

I can’t deal properly with this worry here. An adequate treatment will have to make reference to the subtleties emerging from the resurgence of fictionalism in contemporary meta-ethics. However, some remarks seem in order. First, what is precisely the projective error in the case of narrative discourse and practice? Well, the error could be broken down in at least four different aspects: ontological, semantic, psychological and epistemic. The ontological claim says that the world presents itself to human consciousness in discrete meaningless, unprocessed chunks. So, if the narrative discourse is partly representational (the semantic presupposition), that is, if it functions as if stories (meaningful sequences) correctly signified reality, and if the ontological claim is correct, then the world fails to display the right kind of teleology that answers to this representational content. The psychological consequence follows: agents represent story-like properties due to psychological processes that have first projected these properties onto the world. The epistemic consequence is clear: if agents understand their actions by pretending that their actions are part of coherent stories, their narrative discourse (or mode of self-understanding) is strictly speaking false.

MacIntyre’s reply, as far as I understand it, is that the practical demand of intelligibility makes it reasonable to believe that the ontological claim and its consequences are false. He first invites us to imagine how we could explain human action if only representational processes functioned properly, in the absence of any “falsifying” narrative projection. If somehow projective error were eliminated, the
unfiltered representational content of action sequences would be more like that of disconnected events (maybe a chronicle of disjointed happenings) than those of narratives.31 And again, if we focused on any chronicled action from such a disjointed sequence, the action would be intelligible only if characterized as “an episode in a possible history.”32 So, even as conscious agents approach atomized events, they cannot extricate themselves from the psychological necessity to employ the devices of narrative. As plausible as this reply is, by itself is not sufficient to answer the Sartrean worry.

MacIntyre supplements this first reply with a direct assault on what I called above the ontological claim. He holds that from a phenomenological perspective, the world presents itself to the human consciousness already replete with meanings, already containing interconnected sequences of processed events at various levels of complexity. That is, characters in the social world are never isolated projectors of meaningfulness, rather “they plunge in media res, the beginnings of their story already made for them by what and who has gone before.”33

Another way to put MacIntyre’s point (although he does not do it in this way) is to imagine the kind of social world in which every character would be able to project his own private meaning onto the sequence of lived experiences. Practically this would be an incoherent world. If meaningful action were mostly a matter of personal retrospective projection, there would be no other constraints on acting meaningfully except biological

31 He quotes a very revealing passage taken from Dr. Johnson’s diary recording his travels in France: “‘There we waited on the ladies—Morville’s. – Spain. Country towns all beggars. At Dijon he could not find the way to Orleans. – Cross roads of France very bad. – Five soldiers.—Women. – Soldiers escaped.—The Colonel would not lose five men for the sake of one woman. – The magistrate cannot seize a soldier but by the Colonel’s permission, etc., etc.,’” (After Virtue, p. 215)
or existential. And if this were true, our personal and social worlds would lapse into unintelligibility.

MacIntyre argues that agents must live in social worlds that are jointly intelligible, and that this “shared intelligibility” makes reference to “narrative” inescapable. This necessary appeal to narrative falls out of observable formal and substantive constraints that must characterize human action in social contexts. First, the formal constraints are represented by the visible presence of other agents in our worlds. Characters are “constrained by the actions of others and by the social settings presupposed by his and their actions.” The formal constraints emanating from the action of others are both unpredictability – we must be open to be surprised by the action of others and by our responses to theirs-- and teleology – we must act “in light of certain conceptions of a shared future.” Since temporality is interwoven into our actions in such a way that sharing of certain past and openness to the future are prerequisites of mutual understanding, we must take ourselves to be moving toward a set of goals, and we do so without full, transparent knowledge of what will happen as we move.

But agents act in a world of shared meanings not only under the pressure of the formal constraints. As we have already noticed above, culture outfits budding agents with a repertoire of stories, varying from those highly literary to the mere humdrum

36 It’s important to point out that MacIntyre is not the only one noticing that these formal features of mutual understanding seem connected to narrative. In her “Poetics of Psychological Explanation” Deborah Knight seems to make this a central feature of psychological explanations generally. Manifestly inspired by Jerome Bruner’s claim that narrative is distinctive among text types due to its ability to deal “simultaneously with canonicality and exceptionality”, Knight goes on to apply this insight to psychological explanation: “Psychological explanation exploits the two fundamental features of narrative form in general: the capacity to represent both the canonical and the exceptional, and the capacity to represent the temporality of action and experience.” (p. 76)
scripts. These stories and scripts in effect encode standard (paradigmatic) intentional responses in such a way that we learn our roles in the social world only by reference to them, and become “anxious stutterers” in their absence. The upshot of this reply then, is that in light of the formal and substantial constraints on agency, human awareness encounters a world partly constituted by non-discrete interconnected strings of meanings, and some of the vehicles necessary for negotiating in this structured traffic. So, against the ontological claim, the world does not present itself to human consciousness in atomistic, arbitrary bits, but as already containing interconnected partly processed chunks of meaning.

But what if the social world possessed shared structured meanings only in virtue of their having being projected repeatedly by our ancestors? What if we, as a species, have been systematically engaging in such fictionalizing? If this were the case, then we would require a pretty compelling story to explain the pressure on our species to have developed such amazingly complex pretense processes. Furthermore, this story, whether of an evolutionary bent or not, would have to account for many things. First it would have to explain the causal efficacy of such systematic fictionalizing. Why is it that make-believe processes have been insuring cooperation, survival and culture in ways that truth-seeking practical processes could not have? Second, if the psychological modules aimed at survival systematically misrepresent the cold truth, what are the chances that those aimed at theoretical reliability are not also the result of iterated pretense? To the extent

37 MacIntyre again: “It is through hearing stories about wicked stepmothers, lost children, good but misguided kings, wolves that suckle twin boys, youngest sons who receive no inheritance but must make their own way into the world and eldest sons who waste their inheritance on riotous living and go into exile to live with the swine, 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 been born and what the ways of the world are.” (After Virtue, p. 216)
that we found the story about diachronic projective practices compelling, to that extent we might have to doubt its truth.

4.3. David Velleman and Narrative Explanation

4.3.1. The Self and Its Guises

   In the previous section I endorsed, developed and further defended against objections MacIntyre’s conception of narrative practical reason. I interpreted him to say that sequences of intentions are at least partially constituted by processes that straddle the internal/external divide. In this next section I take a more critical look at the prospect of unequivocally moving from narrative as a feature of the sequences of actions to narrative as a property of selves. I end up suggesting that agents do not require strong narrative unity in order to make sense of their lives or of the lives of others. In fact, both supporters and critics of “narrative” might be guilty of equivocation when they move rather freely from talk of actions as enacted narratives, to that of selves as narrators. Drawing on the work of David Velleman, I begin to argue that emotions are the internal cognitive vehicles that can be taken to properly ground practical narrative reasoning. I understand the emotions as modes of attention to the value-laden world of persons and stories. The emotions provide a helpful way of understanding how human action enacts truthful narratives. In fact, I argue that emotions can be seen to play for self-understanding a role analogous to the one played by conversations in interpersonal relations. By leading and tracking internal conversations among different self-interpretative stances, the emotions open up a way of understanding unity of agency on the model of a cast of characters performing a drama. This culminates in developing a
framework for dramatic agency that incorporates narrative at the intentional level, but also claims that the interaction among various forms of interpretive guises of the practical self is more improvisational than strictly narrative.

Recall the second stretch of our argument:

(4) A human self is intelligible if and only if as an agent she inhabits the unity of a character in her own actual or possible history.

(5) Human selves are essentially intelligible. 38

(6) So, selves are essentially narrative.

In support of (4) MacIntyre invokes considerations analogous to those involved in defending the narrativity of human action. Just like an action is merely an episode in some possible story, so the concept of a person in separation from her history is an abstraction. The self is not constituted by a series of person time-slices glued together by relations of psychological continuity. Since concrete historical development provides the shape of psychological continuity, and only teleological relations can illuminate causal ones, “the character” is conceptually prior to “the self”. 39 The self is only as intelligible as a concrete episode in the life of a character living his own life story.

Unsurprisingly, the justification for this stunning claim turns not on a metaphysical thesis, but on a phenomenological analysis of widespread practices and patterns of reason-giving among persons. From a practical stance, character is prior to self because I can expect to be held accountable for my past actions and experiences by

---

38 It might be objected here that selves are not essentially intelligible. For we might imagine selves so fragmented that they are beyond intelligibility. In response it might be replied that essentially here refers not all any conceivable situation, but to that which is of the essence of paradigmatic human selves. Perhaps, an Aristotelian framework is taken for granted.

39 The sense of “character” here is primarily that of a participant, an actor in a drama or a story; the moral connotations are not intended. Although for MacIntyre the rest of the argument will be moving from the dramatic to the moral usage, nothing in my argument relies on this development.
others who take me as a character experiencing a unique and meaningful human life. Also, because others can expect to be held accountable by me for their actions and experiences grasped as parts of their unique lives (or fragments of lives) moving toward some goal. To those that know me I’m answerable as a subject in my own story; those that I know answer to my questions from the settled perspective of characters in stories for which they are partially responsible. Depending on the extent and the mode in which we ask and respond to each other’s questioning, our lives and stories are interconnected. This notion of accountability further sediments the insight from the previous section that practical reason for MacIntyre is simultaneously narrative, personal and straddling of the internal/external, subjective/objective divide.

We may also expect challenges analogous to the ones raised against the first stretch. One challenge might be called the narrative self is a nice fiction but there aren’t any in real life. Analogous to Sartre’s worry for intentional sequences, one may object: the narrative self is either an unnecessary or a pernicious projection, while our true selves are episodic. The first may very well be espoused by Daniel Dennett, and

---

40 Paying attention to the wide ranging forms of human questioning and response can reveal many insights into understanding ourselves and others, and thus the degree of participation or involvement in each others’ narratives. Consider the difference between questioning in different modes: query and grilling, quizzing and interrogation, probing and suspecting, reservation and doubt, playful kidding and mistrust. Interestingly, the mode of our questioning as well as replying to others is only made intelligible in light of a storied background from which some scripts gain ascendancy over others. Furthermore, notice that perhaps much of our questioning of others is non-discursive and encoded in physiological-emotional responses. There is a good reason for this, as I will suggest below. For now it suffices to notice that even our embodied responses can take the form of questions and answers, are scripted and intelligible in light of narrative sequences. “He came in to ask forgiveness for betraying her. His hands were fidgeting, his mouth gasping for the right words. His visible sorrow met only by her implacable face, her scorching gaze. After two excruciatingly long minutes she raised her left eye brow, a mocking smile forming on her lips. He noticed. He turned. He exited the room in silence, his eyes aimless scanning the ground.”

41 See his view of the “self as a center of narrative gravity” developed in several papers: “The Origins of Selves” Cogito 3 (1989); “The reality of Selves” in Consciousness Explained (Boston: Little, Brown and Company, 1991) Ch., 13; and “The Self as a Center of Narrative Gravity” on line at http://ase.tufts.edu/cogstud/papers/selfctr.htm
second is most clearly evident in the recent work of Galen Strawson\textsuperscript{42}. While the discussion of Dennet and Strawson will have to wait for a different occasion, I turn now to David Velleman’s most interesting challenge to what I called the second leg of MacIntyre’s argument. Velleman holds that narrative has real explanatory power for real, non-fictional selves, and that it is one of the necessary modes of self-understanding. At the same time, he suggests, the agent can understand herself under different guises, in which case, there are gaps or discontinuities in her self-interpretive narratives that throw doubt on the prospect of unifying her agency around a single narrative. It bears mentioning here that the attenuated role that narrative plays in Velleman’s account could contain a plausible response to the challenges posed by Dennet and Strawson.

As an alternative to MacIntyre’s account of the narrative self and practical rationality, let us then turn to David Velleman’s recently developed account. This is important in the present context for some important reasons. First, Velleman agrees with MacIntyre that personal narratives provide an essential characterization of human intentions and actions. He also agrees that the narrative integration of intentions is a natural consequence of our drive toward intelligibility or self-awareness.\textsuperscript{43} Second, although he also agrees with MacIntyre that self-narration involves a fundamentally

\textsuperscript{42} See his “Against Narrativity” in Ratio XVII, No. 4, 2004 and “Episodic Ethics” in Daniel Hutto ed., \textit{Narrative and Understanding Persons}

\textsuperscript{43} Consider several piece of evidence for this. First, in his “The Centered Self” he characterizes his own view of practical reasoning as a “variation” of that of Elizabeth Anscombe. He seemingly agrees that our actions are guided by our conceptions of them, and that we cannot escape the drive to characterize them in increasingly integrative meaningful sequences. He says: “For we can attain integrative knowledge of what we are doing simply by framing and fulfilling integrative conceptions of our behavior, conceptions formulated in terms of the dispositions and circumstances that help to explain it.” (David Velleman, \textit{Self to Self}, Cambridge, 2005 p. 264) Further in his “From Self Psychology to Moral Philosophy” Velleman argues that what agents typically refer to as reasons for actions are nothing other than the material from their narrative context that is then marshaled to draw up an integrative act description: “A self-narrative can thus provide the meaningful act-descriptions that enable the agent to understand what he’s doing. When he instantiates one of these narrative act descriptions, he performs an action that ‘elaborated, fulfils and closes’ and episode in his self-narrative, so that his behavior is intelligible as part of the story.” (\textit{Self to Self}, p. 247)
different form of explanation than causal, Velleman holds that practical rationality is split irreconcilably between these modes of self-understanding. His pessimism regarding the prospect of integrating these two forms of practical reason represents a distinctive (different from Mink and Sartre) and an illuminating challenge to the sufficiency condition in premise (1) of MacIntyre’s first argument. At the same time, his view represents an explicit rejection of premise (4) in MacIntyre’s argument. In the latest refinement of his account Velleman argues that a self is intelligible only through its various self-reflexive guises, and that each of these guises depend further still on distinctive (and it turns out, incompatible) modes of self-understanding. Thus, narrative cannot be the only way of providing intelligibility to the self.

For a clear understanding of Velleman’s position, it might help to distinguish from the get-go between guises of the self and modes of self-understanding. In the very beginning of his book Self to Self he distances himself from those who take “the self” to refer to a proper part of a person’s psychology. Rather, the concept of “the self” expresses in different contexts different reflexive guises under which we come to understand ourselves. Since “[a] reflexive mode of presentation is a way of thinking that directs an activity or mental state at its own subject conceived as such”, talk of the self always is directed at the subject under a particular mode of description and internal to that conception.44

Velleman distinguishes and elaborates at least three different such reflexive guises. One guise denotes a person’s self-image, whereby the agent conceives of himself as one person among others. Under this guise I become the target of some reflexive

44 David Velleman, Self to Self, p. 3
attitudes (perhaps emotions such as shame, guilt, self-esteem and pride) by which I compare myself with others. When I feel good about myself, implicitly I might view myself from the perspective of others who show me admiration; similarly when I hate myself, I regard myself as a person that others might detest. Another intrinsically reflexive guise, according to Velleman, is that by which I pick out my past and future selves. The attitudes that constitute this guise represent from a first personal perspective some person-stage of mine that is accessible through experiential memory or anticipation. As Velleman puts it, the referents of these thoughts “are simply the past and future persons whom the subject can represent as the ‘I’ of a memory or the ‘I’ of a plan – persons of whom he can think reflexively, as ‘me’.”

There is another guise that Velleman spends a long time elucidating. Before we deal with it, consider the already complex interactions between the two guises introduced above. It’s true that the two can and do come apart. I remember harming squirrels in my boyhood, but presently I display no feeling of shame or indignation. (reflexive memory, but no relevant self-image) I currently feel shame at the mere thought that I could harm a little critter, but I don’t remember that in my boyhood I did such things (self-esteem, but no reflexive memory).

But also notice the complex ways in which such guises can and do interact. I remember harming squirrels in my boyhood and that causes me presently feel shame and regret for my past self, and indignation at the mere thought of doing something like it today (reflexive memory cause both past-directed shame and present shame). Or, I remember the humiliation I felt when my parents discovered I was harming squirrels. Or

---

a current sense of humiliation retrieves a memory of my past doing, but because I’ve changed so much since, I don’t feel much by way of empathetic identification with the subject of my squirrel-harming memories. I may even mutter to myself “shame on him”, and strikingly the content of that utterance is evoked by my current shameful feelings via experiential memory. Furthermore, this complex interaction between the two guises (nesting, causality) seems to depend on both causal and narrative of grasping events. We will return to this.

4.3.2. Narrative and Causal Explanations

The third main guise that Velleman discusses is that of the self as the source of autonomous action. He argues that this reflexive guise represents the practical reasoning by which the subject makes sense to himself of his autonomous action. It is helpful to distinguish three historical stages in Velleman’s conceptualization of the way understanding gets involved in representing autonomous action. In his earliest writings, reasons for actions are rendered intelligible entirely in the terms of folk-psychological causal notions, such as belief/desire complexes. In an intermediate stage Velleman seems to think that in understanding her autonomous action, the agent must rely on self-narration, and that this form of explanation is merely a re-description of what goes on at the causal level. In the latest refinement of his view, Velleman gives reasons to believe that narrative understanding and causal-psychological understanding constitute

---

46 See his Practical Reflection and The Possibility of Practical Reasoning
47 See “Self as Narrator”; “Willing the Law” and “Motivation by Ideal” all in Self to Self
“independent and potentially competing modes of practical reasoning.” Furthermore these two irreconcilable modes of self-understanding inform the three different guises of the self.

Practical reasoning is disjointed Velleman now believes mostly because an agent’s search for self-understanding aims at fulfilling radically different needs (or wants). Roughly speaking, narrative understanding responds to our overwhelming need (want) to make emotional sense of our lives, whereas causal understanding answers the call to trace scientifically the source of actions, maybe even to treat ourselves as scientists searching for reliable explanations for our behavior.

Velleman’s splitting practical rationality in these two largely irreconcilable modes depends mainly on his analysis of the distinctiveness of narrative explanation. In broad strokes, the distinctiveness of narrative is connected to ordering events in such a way as to initiate, develop, and conclude familiar emotional arcs or cadences in an audience, even when causal explanations for why the events occurred are absent. Narration engages the affective responses of an audience in such a way that it “knows how to feel about the narrated events as they unfold, and how it feels about them at their conclusion, having arrived at a stable emotion in retrospect.”

So, how should we understand the way in which narratives enter the life of the emotions? According to one model, the basic structure of a story – the narrative plot (a beginning, a middle, and an end) is to be analyzed in terms of our affectionate natures. According to a second model, the relationship is reversed: the emotions ought to be

---


49 Velleman, “Reply” p. 285
analyzed in terms of basic narrative plots. According to the first option, narrative structures are entirely reducible to the ebb, flow, and dissolution of a natural drive or process.\textsuperscript{50} The second model is inspired mainly by the work of Ronald de Sousa on the rationality of the emotions. According to de Sousa our emotions should be mainly analyzed in terms of \textit{paradigm scenarios}, which are, roughly, ways of construing or interpreting the world.\textsuperscript{51} Emotions are accordingly pervasive, developmental, non-linearly progressive interpretive frames through which past emotions and current or expected experiences are cast.

In a clever synthesis of the above two models, Velleman holds on the one hand that the emotions can be \textit{usefully described} as diachronic interpretations of events or objects at which they aim. On the other, he claims that these distinctive modes of interpretation reside largely in biologically programmed patterns. While he avoids restricting himself to any one available biological story (Freudian or Darwinian) Velleman seems to decidedly endorse a general functional (empirical) analysis of emotions as physiological plots. For him, the emotions are biologically realized events characteristically constituted by beginnings, middles, and ends: arousal followed by physiological symptoms, followed by sequence of behavioral symptoms, followed by reflexive patterns of attention and interpretation, followed by motivational dispositions, followed patterns of decay and extinction. Consider the following functional characterization of a paradigm case of fear. I hear a suspicious noise in the middle of the night. My pulse speeds up, my hands sweat, and I get out of the bed to check it out. My

\textsuperscript{50} Velleman references Peter Brooks’ version of this option, although he dismisses it as being too restrictive and reductive. Brooks tells a Freudian/Biological story where the beginning of a narrative is like an itch that needs scratching, and where the scratching provides the resolution of the desire.

\textsuperscript{51} See especially chapter 7 of Ronald de Sousa \textit{The Rationality of Emotion}, (MIT Press), 1987
sweating palms pick up the broomstick on my way down the stairs. I notice a bulge dancing in one of the window drapes. It looks like an intruder hiding, and my arms unleash repeated, frantic blows almost involuntarily. But I’m beating the air. Someone left a window open, and the wind was playing a convincing trick on me. I go back to bed relieved, my heart pumping from the exercise.

Not only individual affections are to be understood as biological narratives that typically display a characteristic beginning, middle, and end. What’s more, strings of emotions can be usefully seen as making sense of series of events by framing them in such a way as to (typically) bring the audience to a stable feeling toward the whole series. Velleman illustrates vividly the beginning, unfolding, and conclusion of an emotional cadence.

In the story of Oedipus, for example, worry arouses curiosity, which gives way to foreboding, which is swept away by shock, and then by horror, which finally comes to rest in grief (and, for the audience, in Aristotelian pity). The audience doesn’t know why Oedipus and his father crossed paths to begin with, and so it cannot explain the event that triggered the tragic developments of the plot; but it does recognize the emotional arc of those events, and it knows how to feel about the overall course of events at the end.52

The complex act of emotional grasping at the end of an emotional sequence has at least three important features, which we can easily identify as unification, retrospective subsumption and biological grounding. First, modeled roughly on an intellectual grasping whereby a sequence of events is given a shape and directionality by being catalogued under summary concepts such as “mutiny” or “treasure-hunt”, emotional understanding assimilates and unifies disjointed events and catalogues them in the “ticks and tocks” of biological rhythms. Second, this unification tends to subsume the

52 Velleman, “Reply to Catriona MacKenze” p. 285
previously felt emotions in the sequence, such that the end of the cadence marks a stable attitude not just about the ending, but about the whole story. Third, and most significantly for the “splitting” argument, Velleman argues that since the biological rhythms grounding emotions are stockpiled in physical architectures and processes (vehicles) different from those that house discursive patterns, emotional grasping is of a different kind from intellectual grasping. So it is that we come to the most basic explanation for why narrative understanding is essentially different from causal understanding:

A story therefore, enables its audience to assimilate events, not to familiar patterns to how things happen, but rather to familiar patterns of how things feel. These patterns are not themselves stored in discursive form, as scenario or stories: they are stored rather in experiential proprioceptive, and kinesthetic memory – as we might say, in the muscle-memory of the heart. Although the audience may have no discursive memory of events such as those of the story, it nevertheless has an experience of dija senti, because its emotional sensibility naturally follows the ups and downs of the story, just as a muscle naturally follows the cycle of tension and release.53

Events can be assimilated both in discursive patterns and emotive patterns. Grasping events by way of causal understanding involves deploying concepts that are in the business of unifying phenomena. Clearly then Velleman must hold that causal understanding itself filters lived events through some kind of conceptual cadence. But whereas the physical processes housing causal patterns aim at an objective understanding of the world, those grounding narrative patterns aim merely at a kind of subjective (emotional) resolution. One mode, aiming at wider explanatory coherence, asks why events happened in the way they did; the other seeks to bring the experienced events into

53 Velleman, “Narrative Explanation” p. 19
the familiar fold of emotional sense. Thus, according to Velleman, diachronic agents are biologically inclined to cast events into different and irreconcilable intelligible blueprints.

Granting that these distinctive blueprints can and do overlap, Velleman’s crucial claim is that agents are biologically inclined to treat the coherence of narratives as replacement for causal explanations. Our fundamental problem is that we are hardwired to fill the causal gaps in our practical understanding of the world with stories that respond primarily to the biological need for emotional resolution. We thus treat emotional closure as a substitute for causal closure. This is the real problem of projective error for narrative understanding. It is not that stories project a narrative structure onto an unprocessed reality, for even causal explanations presumably impose some conceptual filter upon events. Rather, the problem is that audiences (agents in search of self-understanding in our case) are prone to mix-up emotional coherence for explanatory coherence. As Velleman puts it:

Having made subjective sense of historical events, by arriving at a stable attitude toward them, the audience is liable to feel that it has made objective sense of them, by understanding how they came about. Having sorted out its feelings toward events, the audience mistakenly feels that it has sorted out the events themselves: it mistakes emotional closure for intellectual closure.54

A slick response to Velleman’s argument would be to challenge the biological story he weaves. One worry here is that his argument for the essential split of practical rationality relies on an outdated “faculty” psychology. It is as if emotions constituted, if not the direct foe of the intellect, at least a subtle competitor inclined to overshadow it. Beside this, many cognitive scientists recognize that the brain does not store information along independent emotional and intellectual tracks. Rather, there is good empirical

54 Velleman, “Narrative Explanation” p. 20
evidence to suggest that the emotional processes are essentially involved in all acts of traditional cognition.\textsuperscript{55}

4.3.3. Cognitive and Emotional Coherence

Now even if Velleman’s story regarding the physiological grounding of emotions is mostly correct, it does not entail that discursive processes could not affect non-discursive ones and vice-versa. In fact, it seems that an audience’s appropriate emotional responses to a story often depend on its having grasped discursively at some point in its career the salient causal features of the world. Velleman argues convincingly that even those finding Oedipus’ story absurdly implausible still resonate with it, because we all “know the feeling of being undone by our own efforts.”\textsuperscript{56} That is, it is the emotional closure that explains this resonance, even while intellectual closure is lacking. But it is worth pointing out that feelings of this kind are formed partly because some of us have been undone by our own efforts in absurdly implausible ways. In fact, one could say that Oedipus’s story either evokes or expresses writ large aspects of the absurd events we’ve encountered or of our inexplicable personal tragedies.

We respond appropriately with horror or fear to Oedipus’ plight, because as an audience we inhabit at least a double perspective. As Velleman would agree, the first perspective is that of an empathetic stance that identifies with the main character’s condition. In our case, we (the audience) identify with those engrained dispositions that lead Oedipus to act as he does. The second however is a stance external to that of the


\textsuperscript{56} Velleman, “Narrative Explanation” p. 21
main character that affords only us, the audience, pertinent knowledge of the plot and other characters, knowledge we know Oedipus doesn’t possess. It is the dramatic distance between our caring about Oedipus and our knowing what he does not that creates in us the relevant emotional cadence. Thus, the two stances are jointly necessary to induce in an audience the appropriate emotional sequence.

As I illustrate below, the interplay between the internal and external perspective explains much of our emotional responses not only to literary drama, but also to events and persons. An agent grasps her stories retrospectively, and she does so from at least two perspectives. The internal stance is that of having experienced (or imagined) a sequence of events as an actor engaged in unsuspecting ordinary behavior. One external stance is the retrospective perspective of the narrator possessing now knowledge simply unavailable to the actor at the moment of action. Another external stance is simply the perspective of an interlocutor, who does not have to possess any new information of the lived (or imagined) events, but could still frame them differently. We might say that switching between internal and external perspectives constitutes the emotions’ unique way of tracking cadence and thus a certain kind of meaning over time. By allowing their contents to be examined and enacted from different perspectives the emotions are thus essentially performative. Due to this feature the emotions have the added benefit of explaining how various self-reflexive guises might be internalized over time. I hurt a friend, and imagining how he now views me, I am driven by the consequent shame and guilt to ask forgiveness. Knowing that I am prone to making a fool of myself in public, I turn down the invitation to go a party by imaginatively taking the advice of a remorseful and disappointed future self. If my hurt friend is the one issuing the invitation, I have to
take yet the perspective of another imagined advisor who judges whether reconciliation is preferable now in spite of my likely social blunder.

Furthermore, the interplay between the actor (internal) and the narrator’s or the interlocutor’s perspective can fruitfully show that the relationship between emotional and cognitive closure is more complex than Velleman acknowledges. I use the example below to show that emotional closure is not always a projection onto causal networks.

Returning from her grandfather’s funeral my wife Andrea brought home a lovely bouquet of flowers. Nobody was surprised to see the beloved family kitten, Geneva, play happily in the flowers. Little did we know that among the flowers Geneva found most exciting there were some Asiatic Lilies, which interestingly are deadly poisonous for kittens. This information was news even to the florist. Geneva died of renal failure seven days after grandpa’s funeral, and we couldn’t help seeing a causal connection between the two events. At the same time different members of our family felt sadness mixed with responsibility. “Why did I have to bring home that flower bouquet?” asked Andrea. “Why wasn’t I more scrupulous in investigating things noxious to little kittens?” I asked. “Why didn’t I have the power to see that the lilies are poisonous?” asked Alexandra, one of our seven-year old twins.

The death of a kitten is undoubtedly one of the smaller tragedies of life, but my point is that in these types of situations agents typically attempt to attain intellectual closure by way of emotional resonance and vice-versa. If my question, or Andrea’s, or Alexandra’s are at all intelligible, they provide some evidence that as emotional agents we are inclined to track causal networks. Coming to know that chemical X is found in Lilies and that it is deadly to cats reveals that Alexandra shouldn’t blame herself for not
being aware of the relevant causal connections, and so interestingly can provide emotional closure for her. Small and large tragedies do not reveal unambiguously that narrative closure is just a poignant substitute for intellectual explanation.

Audiences wish Oedipus had enhanced knowledge of his situation, and retrospectively we wish we had wider knowledge of causal tracks. Sometimes intellectual closure can bring naturally audiences and agents to emotional closure. And sometimes uncovering causal explanations can only intensify our lack of emotional closure. And sometimes, causal explanations are not at all sufficient to explain why a sequence of events happens to us in the way it does. What is clear is that we all have to live only with a smattering of causal knowledge, that the world of practical action must be perpetually experienced as causally open. So, stories may very well encode our lived understanding of the limits of causal explanations and our emotional response to this situation.

I’ve argued thus far that neither the biological story Velleman tells, nor the case of smaller tragedy sufficiently shows that narrative explanation is based in emotive-physiological patterns at the expense of discursive ones. I’ve also suggested that the interplay between first personal (actor) perspective and external (narrator, interlocutor) perspectives creates quite complex ways in which discursive patterns affect emotive ones and vice-versa. But can we give a better characterization of the emotions role in narrative explanation? Is there a theory of the emotions that better accommodates the complex way in which narrative explanations seem to combine causal connections, first-personal

———

57 Imagine coming to find out that the florist is terribly ill-disposed toward the feline family. She’s been deliberately planting Asiatic Lilies in all of her flower bouquets on the chance that some cats may ingest the poison. This piece of knowledge while contributing relevantly to the causal explanation may dislodge the emotional equilibrium achieved from attributing Geneva’s death only to a series of unfortunate events.
experiences and retrospective attribution of meaningfulness?

4.3.4. Emotions as Quasi-perceptual Paradigm Scenarios

More progress can be made here if we emphasize the second model that enters into Velleman’s synthesis. As I have indicated above, Velleman seems to prioritize the biological explanation, and consequently the interpretive grids through which emotions codify events turn out to be mostly analyzable in terms of biological rhythms. But what if we emphasized instead the perceptual content of emotions, and let the biological chips fall where they may. The alternative model draws its inspiration primarily from Ronald de Sousa’s work on the rationality of emotions. According to him, the emotions are quasi-perceptual interpretative frameworks, or ways of seeing the world that possess both a representational and expressive-motivational content.

According to de Sousa the emotions are reducible to neither beliefs, nor desires, nor to simple internalized somatic responses. Rather, the emotions are complex perceptions that track patterns of salience in our world. They frame our world in so far as it is relevant to action, and aim to understand it axiologically. The salience of the emotions is explained both developmentally and structurally. de Sousa argues that we learn both our emotional vocabulary and our standard emotional responses through paradigm scenarios which are selected first “from our daily life as small children and later reinforced by the stories, art, and culture to which we are exposed.”

As we mature, we may thus expect that our earliest paradigm scenarios may be supplemented, supplanted, refined and developed by a variety of socio-cultural meanings.

de Sousa claims that the dual aspect of paradigm scenarios should enable us to understand both the intentionality of emotions and their expressiveness. “Paradigm scenarios involve two aspects: first, a situation type providing the characteristic objects of the specific emotion type…and second, a set of characteristic or “normal” responses to the situation, where normality is first a biological matter and then very quickly becomes a cultural one.”59 For example, a pet’s death is saddening because it characteristically references the loss of something dear. So, if Alexandra saliently perceives the event of Geneva’s death, then, by virtue of grasping the object of her sadness (losing Geneva) she will also be motivated to respond appropriately.60 And if the child is astute, or lucky to have been the recipient of a sound emotional education, she might also be capable to articulate or communicate intelligently about the way she feels.61

The connection between linguistic proficiency and emotional competence suggests that without a certain level of linguistic sophistication some very real features of paradigm scenarios will remain obscure to some agents. Without a finely tuned capacity to describe experiences, subtle nuances and differences among traits, events, or actions will simply melt away, and so some agents will lack the capacity to tune in emotionally

59 Ibid, p. 182

60 I am assuming here that a standard, culturally shaped response to the death of a beloved pet is something like distress, or grief.

61 If articulacy is a measure of the depth of emotion and hence of its appropriateness, we noticed a significant difference between Alexandra (age seven) and Ethan (age four). Whereas both reported sadness, Ethan could only say that he missed Geneva, and pretty soon learned to exploit this paradigm scenario. For even about two months after Geneva’s death, Ethan invariably claimed he was crying because he missed Geneva even in cases where that was evidently not the case. Alexandra on the other hand, did not extend this paradigm scenario unjustifiably. Furthermore, the explanations for her sadness were compelling, complex, and dependent on a nuanced grasp of causal relations. Recalling in heartwrenching detail that unique and playful path that Geneva carved out in her world, the articulation of her sadness revealed the following: a) Replacing Geneva is impossible for no other kitten would to exactly what she did. b) Even if a pet resurrection is causally possible, that will have to wait for the future; it doesn’t fix the problem now. c) Celebrating the life Geneva did live doesn’t help, for remembering her only accentuates the current sadness.
to the relevant features of specific paradigm scenarios. One plausible consequence of this is that a deliberate cultivation of insightful conversation, good literature and honest friendships would in effect refine and widen the external stances through which we can fine-tune emotionally to our world. As I have adumbrated in an earlier chapter, Jim’s description of their voyage down the Mississippi river allows Huck to see himself as Jim’s friend and liberator. Interestingly, it is the illiterate Jim who has a keen moral perception of the situation. When Jim calls Huck his only friend, his best friend, and the only white gentleman that ever kept his promise to him, these words amount to a coherent, and compelling paradigm scenario through which Huck can re-assess his emotional investments and his actions. Dramatically, Jim’s perspective places such an emotional brake on Huck that he can’t carry through with his intention to turn Jim in.

I have argued thus far that emotions are value-laden ways of perceiving the world that display essential intentionality and expressivity. Emotional range, appropriateness and accuracy depend both on early inculcation and on the mastery of a sufficiently complex linguistic repertoire. More directly relevant to our discussion regarding the relation between narrative and causal explanation, it seems that emotions are also involved in directing toward the world in characteristic ways the cognitive states or processes responsible for causal explanations. That is, the salient patterns that frame our emotional perception of the world also direct and guide in their train beliefs and desires. Consequently, it would seem that emotions set the contours of practical rationality. A nice example again proceeds from Mark Twain’s characterization of Huck’s inner struggle. Notice how his moral reasoning resulting in the intention to return Jim back is guided by feelings of pity toward the “poor Mrs. Watson”:
Conscience says to me: “What had poor Miss Watson done to you, that you could see her nigger go off right under your eyes and never say one single word? What did that poor woman do to you, that you could treat her so mean?...I got to feeling so mean and miserable I most wished I was dead.”

But how do paradigm scenarios guide practical reasoning? How do emotions direct beliefs and desires? De Sousa is again helpful here. He says that “emotions set the agenda for beliefs and desires: we might say they ask the questions that judgment answers with beliefs and evaluate the prospects to which desire may or may not respond.” Accordingly, emotions self-effacingly set us in pursuit of answers to our practical puzzles; they set up the perimeter where causal explanations are deployed. Yet though they display this type of directionality does not imply that the emotions are arbitrary or merely projective. Comparing them to Kuhnian paradigms that shape contours of normal science, de Sousa observes the emotions’ fruitfulness in opening new ways of carving up the practical world. Even if new paradigm scenarios may significantly reconfigure previous ways of perceiving the world, they still make contact with it:

If the world is real, it will look different as we move around. When a paradigm scenario suggest itself as an interpretation of a current situation, it arranges or rearranges our perceptual, cognitive, and inferential dispositions in terms of some real configuration of human experience.

We are protected both against projection and against radical relativity of perception by the built-in patterns of salience derived from paradigm scenarios. An external perspective from a friend may induce me to retract some of my earlier...

---

62 Quoted in Bennett, p. 28

63 de Sousa, *The Rationality of the Emotions*, p. 200

64 “Like scientific paradigms, in the sense of Thomas Kuhn, emotions are better at stimulating research in certain directions than at finding compelling and fair reasons for their own adoption.” (*Rationality of Emotion*, p. 198)

perceptions of myself as a quite broadminded husband. It may instead cultivate the suspicion that I’m overprotective and jealous, and so it may help reconfigure significantly beliefs, desires, and practical inferences. No matter, if the resulting emotional pattern is more fine-tuned to my current situation than a previous pattern or perception, there is no reason to think it constitutes an illegitimate projection either onto the world, or onto my previous causal explanations. In fact, it seems that the emotional refinement has led me to discard or filter previously unhelpful causal processes, and I’m in a better position now to understand myself, my wife, and my friend.

4.4. Paradigm Scenarios And Dramatic Agency
4.4.1. Emotions as Interpretative Frameworks

This view of the emotions as constituted by inter-subjective modes of practical attention has many advantages. First, it explains Velleman’s compelling insight regarding the self-reflexive guises through which agents engage their practical worlds. If intentional consciousness is situated at the confluence of different and perhaps occasionally discordant paradigm scenarios, then our self-presenting guises are filtered through or constituted partially by patterns of emotional attention. Second, as I have already suggested, this view of the emotions also provides an alternative to Velleman’s position regarding the different and incommensurable orientations of narrative and causal explanations. If emotions are grounded in paradigm scenarios, then causal processes are represented as directed by narrative processes of varying degrees of complexity and sophistication. The sophistication will be a measure of the accuracy, precision and detail
present in the articulation of a sequence of emotions or several such concomitant trains.

4.4.2. Characterization of Dramatic-narrative agency

Third, this view allows us to develop a compelling model of diachronic agency as the locus of characteristic dramatic exchanges among different but relatable paradigm scenarios. At one level, narrative is a real feature of agency for it shapes individual emotional frameworks and the larger social scripts that partially constitute these. In the previous section I argued that emotional trains direct intentional processes toward different features of the world. Agency is thus found at the intersection of varied interpretive perspectives. But here it seems that the concept of “narrative” is no longer sufficient to explain the interaction among these perspectives. I propose that the interaction be understood on the model of an inner dialogue, conversation or dramatic performance. Practical reasoning can thus be seen to involve alternate identifying with different interpretive perspectives, analogous to the interaction of different dramatic personae in a drama. While the personae or narrative voices have clear histories and settings, the dialogue among the personae also has a genuine causal power over how the drama goes on: which voices are silenced, which gain ascendency, and how or whether new characters can enter the stage.

The benefits of this proposal can be seen more clearly in situations of practical reasoning. If the emotions reveal relevant features of our practical world by fine-tuning and broadening our ways of seeing it, the agent can examine a situation from different axiologically charged perspectives. Given that these perspectives have antecedents in the emotional history of the agent, reflecting her self-understanding modes, practical reason
consists in bringing to bear upon specific situations historically salient ways of seeing the world. Agency can thus be characterized as the locus of multiple dramatic personas cashed out in terms of multiple and mutually accessible emotional postures.

Some intentions as well as decisions could be thus seen as outcomes of an internal “conversation” taking places anyhow among various interpretative perspectives. Typically, the fundamental paradigm scenarios to have a voice in deliberation are diachronically stable, and one measure of such stability is given by the paradigms’ tendency to constitute the agent’s self-reflexive guises. When all things go well we should expect that the ability an agent has to be internally conversant with his emotions translates well into the ability to converse with people and vice versa. This being said, natural continuities should not be taken as entailments or even probabilities. It is conceivable that people highly attuned to their own emotions may not be able to articulate them to others, or that those attuned to others may display flagrant failures in self-knowledge.

Let me hurry to add that being conversant with one’s emotional repertoire does not have to lead to emotional coherence. In fact, if we cultivate our emotional lives through the external stances of friends, experts, experiences, good art and fiction, we may develop a sharper vision of our specific circumstances, but internal coherence would seem elusive. Analogous to the way conversations between people can be stopped, interrupted, then picked up again, the dramatic model of agency allows for a chorus of voices, for dissonance, for interruptions, and for silence. This seems to me a more realistic model than those aiming at emotional, judgmental, or volitional harmony. Multiplying the outlooks of our emotional lives seems preferable to both
compartmentalization and fictive integration. The dramatic model of the self thus affirms the value of assorted narrative perspectives upon our lives and the value of keeping these in conversation with one another.

4.4.3. Illustration of Dramatic-narrative Agency

Let me illustrate how multiplying the emotional outlooks can in fact function profitably in enlarging our moral sensibility to others. The example below also shows how narrative-dramatic agency facilitates a creative exchange between what we called the internal and the external perspective while avoiding the Scylla of mere personal projection and the Charybdis of mere inculcation of accepted social norms. Revisit Huck as he keeps on rowing away with the intention to turn Jim in, only several moments subsequent to the latter’s heartfelt expression of generosity and friendship. We might expect to find at least four different emotionally charged interpretive frameworks that simultaneously cast different lights on Huck’s practical dilemma. We might say there are four different characters or voices engaged in a dramatic performance.

The first voice expresses Huck’s developed, steady association with Jim during their flight, but which Huck wouldn’t characterize as friendship, or not anyway as the type of friendship which would sacrifice for the sake of the friend. Maybe this is the perspective of the casual adventurer who perceives the alliance with Jim as an occasion to skip out on the banality and restrictions of the provincial life. The second paradigm scenario expresses Huck’s feelings of pity for Mrs. Watson and related feelings of shame and guilt, in light of the perception that she is the rightful owner from whom he has wrongfully stolen. Notice that this interpretative framework is highly cultural, and Mark 189
Twain masterfully and ironically dramatizes the extent to which Huck has internalized a culturally dependent emotional perspective, which the reader can (hopefully) see as deeply flawed.  

The third dramatic voice Huck hears reinforces but is still distinct from the second. It is associated with the religious overtones that would condemn to hell any outlaw that violated established social conventions. As he is breaking the law by continuing to hide Jim, Huck has good cultural reasons to fear the proverbial hell, which is now firmly in sight. Finally, as I suggested above, Huck also sees the point of Jim’s plea. And this external perspective, resolutely different from both his own framing of the adventurous association and from his internalized cultural scripts, makes its way steadily into Huck’s emotional sensibility.

There are several moments in the narrative when these four competing voices contend for Huck’s attention, and Huck manages to experience a healthy kind of emotional discord by alternating perspectives. As he allows each of these different self-reflexive modes a voice, Huck likely gets a better grasp of the entire situation. First, he widens and deepens his view of the friendship with Jim by steadily adopting the emotionally and dramatically charged story as seen from the latter’s perspective. In Jim’s story, of course, Huck functions as the hero, the courageous savior who risks big to help his friend out. As he listens to Jim’s account, Huck now comes to see himself through the former’s eyes; he is not so much a fickle adventurer as he is Jim’s devoted rescuer and friend.

---

66 This is not say that all of Huck’s inferential capacities or belief/desires complexes are also flawed. Notice that some such practical inferences are quite adequate: “You should return what you unjustifiably took. I took something unjustifiably. So, I should return it.” The flaw is in wide variety of standard culturally dependent paradigm scenarios that educate citizens to look at people as possessions.
Second, to the extent that he internalizes Jim’s perspective, Huck also finds some resolution to social interpretative grids, without bringing them decisively upon his final decision. He continues to feel regret that he robbed Watson, and is now prepared to go to hell in order to rescue Jim. One paradigm scenario grounded in the pity he initially felt for Miss Watson terminates in regret, the others, grounded in the fears of hell and of social opprobrium, find resolution in Huck’s embracing these apprehensions. Interestingly, the diachronic shape of the respective emotional sequences seems effected by his budding feelings of friendships for Jim. Now if my description of this case is psychologically accurate, can we say more about how Huck can come to assign deliberative priority to one emotional paradigm over another?

A plausible suggestion here is that if Jim’s own take on the lived events illuminates Huck’s own perception of the adventure story, Huck might also imaginatively identify with Jim’s interpretative stance regarding other events or situations. For instance, Huck might think “How would Jim perceive my feelings of regret for having stolen him from Watson?” and “How would Jim perceive my fear of going to hell for helping him?” Or suppose for a moment that in the grip of fear he might reflect on Jim’s perspective: “I won’t go to hell, so I’m turning him in. It makes me sick to my stomach to betray a friend.”\(^67\) My suggestion here is that the iterated play of such alternative emotive perspectives allows Huck in the end to settle upon a course of action. I am not suggesting that Huck consciously explores all of these different perspectives, let alone that he forms these particular practical judgments. The thought here is that Huck

---

\(^67\) How would Huck’s stories look from the perspective of a relationship of genuine friendship? How would a story of genuine friendship look from the perspective of a story of fear?
effectively identifies with these emotional perspectives in the course of whatever thought processes he goes through.\textsuperscript{68}

To develop the above perspective, Huck’s reflective action takes place only after the implicit interplay of emotional perspectives, and only after a certain acceptable reverberation has been achieved as a result of the dialogue among the various dramatic personae. At the very least, it seems that Jim’s paradigm scenario finds the right kind of resonance in Huck’s own emotional economy, timbre which blunts the force of culturally-induced emotional patterns. By his emotive interaction with Jim, Huck comes to interpret himself simultaneously as a loyal friend, courageous white man, outlaw, and eternally damned. And with this cast of characters he is prepared to live.

4.4.4. On Harmonies and Their Costs

By way of concluding this chapter I address an obvious objection. As I have hinted above, the dramatic-narrative model does not aim explicitly at harmony or internal coherence. The worry then is that this model commits the agent to experiencing systematic ambivalence, irresoluteness and motivational indeterminacy. So, is there any form of psychological harmony available to the emotionally enriched agent? Is there a way to prevent the different voices allowed to express interpretative frameworks from collapsing into a cacophony?

In response, I share with Wayne Booth the assumption that the conception of the self as a plurality of historically developed dramatic voices (or characters) is morally

\textsuperscript{68} Thanks to Mike Rea for helping me clarify this thought.
preferable to several conceptions of psychological integration. In a turn of phrase intended to subvert the idea of authenticity as “purity of heart” Booth claims:

Our true authenticity, in this view, is not what we find when we try to peel away influences in search of a monolithic, distinctive identity. Rather it is the one we find when we celebrate addition of self to self, in an act of self-fashioning that culminates not in an in-dividual at all but in – and here we have to choose whatever metaphor seems best to rival Mill’s bumps and grinds of atomized units – a kind of society; a field of forces; a colony; a chorus of not necessarily harmonious voices; a manifold project; a polyglossia that is as much in us as in the world outside us.69

The type of psychological integration that Booth targets here is that of “being true to one’s self”. The main idea is that there is a core self and if we are to have genuine self-knowledge or authentic action, the impurities must be discarded. Diachronically, this form of existential integration can be understood as being committed to one’s volitional loves (Frankfurt), or carrying out one’s plans (Bratman), or even as loyalty to one’s fundamental ground projects (Williams). The main costs for this type of integration involve susceptibility to practical solipsism, and explanatory deficiency for psychological transformation in phenomena such as conversion, repentance, and forgiveness.

As an alternative to the existentialist, integration can be conceived as a way of functionally arranging the parts with respect to the whole. This conception of integration-as-functional harmony has a reputable history, going at least back to Plato’s view of the just soul in the Republic. Diachronically, some may be tempted to read MacIntyre’s view of narrative agency as displaying this type of integrity. MacIntyre suggests that the unity of agency is to be rendered in terms of the unity of a life in search for an end destination, a telos. “The unity of a human life is the unity of narrative quest”

69 Wayne Booth, “Individualism and the Mystery of the Social Self; or Does Amnesty Have a Leg to Stand On?” in Barbara Johnson (ed.) Freedom and Interpretation, (BasicBooks, 1993) p. 89
he famously states. Some interpreters take this to mean that the agent is unified to the extent to which she manages to integrate all the events in her life in a single coherent narrative. But it seems to me that MacIntyre explicitly rejects the static framework that assumes fitting events in a life is analogous to filling in a puzzle. Thus, narrative unity for MacIntyre does not entail the integration of all the different psychological parts of a self (desires, commitments, or evaluations) into a coherent single life story. And by extension, it does not seem to entail either the integration of all expressions of practical judgments into a coherent, consistent evaluative set. Rather, the very notion of a quest involves a constant sizing up, redirection and re-evaluation both of the goal(s) of the search and with respect to one’s own place on the map relative to the goal(s).

My view of agency is in tension with both versions of psychological integrity described above. Existential integrity in all its forms as well as integration conceived as fitting parts into a whole, seem to exclude important dissonant voices or to domesticate them in the process of psychological integration. And I have maintained that these voices are often essential to dramatic transformations of agency, so it seems we need a more liberal conception of integrity.

Furthermore, a large measure of integration is achieved biologically and socially, as our emotional outlooks are formed in particular communities. Additionally, as I argue in the next chapter, from a psychological point of view, certain paradigm scenarios such as fear and resentment lend themselves naturally to function as motivational gravitational

---

70 MacIntyre says: “It is in the course of the quest and only through encountering and coping with various particular harms, dangers, temptations and distractions which provide any quest with its episodes and incidents that the goal of the quest is finally to be understood. A quest is always an education both as to the character of that which is sought and in self-knowledge.” (After Virtue, p. 219)
centers. If this is accurate, it is in our interest to adopt psychological models that help disintegrate us.

Finally, the psychological costs of perpetually expanding the “company we keep” might be really high, as we have to live with inner tensions, unresolved questions, and ruthless attention to the emotive stance of others. Nonetheless, the moral benefits of becoming members of each other, may trump the psychological costs. Booth claims that our selves are literally constituted by the selves and life stories of others: parents, playmates from infancy, schoolmates, friends, foes, and even favorite literary characters. I will not go as far as Booth to claim that I am constituted literally by the selves of others. Mine is a more attenuated claim that my self-conception is essentially informed by self-conceptions of many others with whom I dialogue. If one further agrees that understanding another is essential for valuing them, then I can also agree with Booth’s estimation:

My value consists largely in the values or “voices” I have absorbed, and in the continuation of the dialogue among them – among my present selves and further selves that I/we hope to encounter. Whatever differences in value one finds among lives or moments in life are thus insignificant when compared to the universally shared value of enacting a dramatic story line. Every prisoner, every murderer, every torturer shares this potentiality for dramatic change and growth into the future. 71

I have introduced a two-tier view of agency. When we characterize intentional actions it is appropriate to think of enacting narratives, and to aim at integrating intentional sequences into coherent wholes. However, when thinking about the unity of agency, I think we should replace the discourse of narrative with a related one of dramatic performance. Whereas different intentional strands can appropriately be

---

71 Booth, p. 92.
characterized as personal narratives that inform and form our lived lives, the characters that enact them need not synchronize all these stories along a single axis. Rather, as we are accountable to others and as we hold others accountable, our interpretative frameworks would fall under the diverse and sometimes incommensurable relational pressures. And consequently agency would be constituted from the interchange of voices that improvise around some central themes.
CHAPTER 5
PERSONAL FORGIVENESS AND DRAMATIC AGENCY

5.1. Map and Background Assumptions

In the current chapter I explore some connections between personal forgiveness, narrative understanding and the dramatic self. Although forgiving seems to require expressing, embodying or being disposed to embody some type of affectively transformational story\(^1\), influential accounts of diachronic practical agency seem badly equipped to explain this phenomenon. If agency is constituted essentially by self-managing plans or by projective loves that structure volition, forgiveness seems either a recalcitrant, subversive or otherwise peripheral feature of practical rationality. But what if forgiving and being forgiven is more central to our self-understanding? My main claim is that narrative understanding is conducive to forgiveness only if it is supported by an appropriate conception of the dramatic agency.\(^2\)

\(^1\) See Griswold, Charles Forgiveness: A Philosophical Exploration (Cambridge, 2007); Margaret Urban Walker, Moral Repair (Cambridge, 2006); Cheshire Calhoun “Changing One’s Heart” Ethics, 103 No. 1 (Oct. 1992); David Noviz “Forgiveness and Resentment” Philosophy and Phenomenological Research, 1998 LVIII June, 1998 to mention a few recent accounts.

\(^2\) I do not have the requisite space to show how my agency model handles the case of divine forgiveness. In the Christian tradition the Incarnation could plausibly be seen as God inhabiting a dramatic persona that allows him to understand the waywardness and violence of humanity. If the story humanity’s fall is construed as the primordial act of violence against God, the story of Christ’s life, suffering, and death can be seen as God’s exposing, absorbing and undoing the initial affront and the derivative violence. Crucial to this interpretation is Christ’s exclamation directed toward God Father in the height of his suffering on the cross. “Father, forgive them for they don’t know what they are doing.” This is obviously a severe simplification, but still we might interpret this as an exchange between two dramatic personae of
In the sequel I presuppose that forgiveness is essentially a retrospective phenomenon and that any model of practical rationality worth its salt must at least be consistent with it. The main task of this chapter is to argue that the model of agency developed in the previous chapter and a desirable form of interpersonal forgiveness mutually support each other. On the one hand, I show that the analysis of agency as a cast of thematically connected characters illuminates puzzling aspects of forgiveness. On the flip side, the phenomenon of robust interpersonal forgiveness deepens and elucidates further the account of dramatic-narrative agency.

First I describe some of my background assumptions regarding forgiveness, and introduce the paradox of forgiveness. Second, I explain Cheshire Calhoun’s empathetic model, which sets out to answer this paradox. Drawing a distinction between aspirational\(^3\) forgiveness and minimalist forgiveness, Calhoun argues that the former is possible given the psychological possibility of telling aspirational stories about our perpetrators. In the third section of the chapter I raise the opaqueness problem for Calhoun’s psychological model, and argue that my dramatic-narrative model provides a helpful diagnosis and solution. In the fourth section I raise the resentment entrenchment problem, and argue that the dramatic-narrative model provides a compelling explanation for this phenomenon, and points to fruitful ways of upholding the ideal of aspirational forgiveness. I end with promissory notes about a wider application of the dramatic-narrative model.

---

God, one having identified empathetically and narratively with the plight of humanity, the other focused on the gravity of the offense.

\(^3\) Throughout I use “aspirational” interchangeably with “robust” and “radical”
5.1.1. Untidiness of Personal Forgiveness

My focus here is interpersonal forgiveness. Generally, it must involve a number of complex and distinctive attitudes enacted by a person who has been injured, hurt or victimized by another, who is the object of these attitudes. I will not tackle here the complex challenges facing analyses of international, political, or civic forgiveness, nor will I substantiate my suspicion that these, if appropriate, would be modeled on interpersonal forgiveness. Furthermore, I am not very optimistic that an analysis of forgiveness (even in the simplified interpersonal contexts) will yield anything resembling necessary and sufficient conditions. Rather, expressions of interpersonal forgiveness seem to me to refer to an untidy and amorphous cluster of phenomena.

There are at least three reasons for this pessimism. First, the same expression of forgiveness can pick out any stage in the process of forgiving. It could refer to an event that merely signals the intention to forgive with the same ease as to an event that marks the achievement of reconciliation. By picking out different aspects of a rather complex and fluid process, expressions of forgiveness are thus inherently resistant to unqualified unification. Second, as some theorists have noticed, forgiveness implies relinquishing certain forms of resentment, and so it presupposes a measure of affective re-orientation. But the emotions themselves are rather notoriously hard to analyze. Generally, emotions amalgamate effortlessly with other mental states. They naturally blend in with and bleed through moods, desires, intentions, beliefs and plans. From among all emotions,

---

4 The intuition here is pretty clear, I hope. The idea is that in inter-communal victimization contexts (e.g., an ethnic group inflicts violence on another) or hybrid contexts (e.g., an individual is victimized by an ethnic group, or political organization; or a institution, corporation, community is harmed by a single individual), it makes sense to speak of forgiveness as a possible coherent attitude only if the features of the parties involved in the harming context can be modeled on relevant features at play in interpersonal forgiveness.
resentment is particularly adept at permeating through practical mental contents and processes. So, if forgiveness involves at least partially disowning resentment and if resentment is dispersive and multi-faceted, forgiveness must follow suit in relevant ways.

Third, theorizing about forgiveness seems difficult due to its dependence on contextual cultural, social and religious practices. For example, it should come as no surprise that forgiveness has profound religious connotations. The theorist analyzing forgiveness without paying attention to the religious symbols and images that have given its conceptual and emotive contour either operates in an overly reified theoretical space, or merely reads her cultural or personal biases into the analysis. On the other hand, those reducing forgiveness merely to a religious phenomenon are likely to unwittingly theorize in the proverbial grip of a picture. If we want to avoid clenches of all sorts in our theorizing about forgiveness, Avishai Margalit’s methodological quasi-humanism points in a helpful direction. According to Margalit, humanism in ethics is committed to two claims. One is that “human beings are the only source of justification for ethics and morality” and the other is “that humans are a sufficient source for justification of ethics and morality.”5 Margalit accepts the first claim, but denies the second. Methodologically then, the theorist of forgiveness is free to draw on religious imagery and practices, as long as the distilled results are applicable in a broader human context. By extension then, the theorist of forgiveness should be free to draw on various cultural and historical conceptions, provided that the resulting conceptions of forgiveness are fruitful beyond those conceptual borders.

5.1.2. The “Paradox” of Forgiveness

The *locus classicus* of most contemporary discussions of forgiveness is Aurel Kolnai’s paper on the so-called paradox of forgiveness.⁶ Here is a summary: When we forgive, we either uphold the tie between the perpetrator and his wrongdoing or we do not. If the link between the agent (perpetrator) and his wrongdoing is broken, then it is morally inappropriate to evaluate the agent apart from the act, so forgiveness is unnecessary. If the link between agent and wrongdoing is maintained, then forgiveness collapses into condoning. So, forgiveness is either redundant or it is a mere case of condoning. So, forgiveness is impossible.

Some of the early responses to this paradox conceived of the possibility of forgiveness on the judicial model: the victim is judge, jury and executioner. Other responses developed civil court models: the victim is the plaintiff who is free to either offer or withhold forgiveness. Lately the focus has shifted from the judicial and the civil, to the empathetic models: the victim is supposed to make sense of the wrongdoing by contextualizing the harm within the in biography of the perpetrator.⁷ Since my main aim is to showcase the merits of my dramatic-narrative view of agency, I will not evaluate here all these responses. Rather, I remain narrowly confined to one specific model of empathetic forgiveness, albeit, an intriguing and paradigmatic one.

---

5.2. Empathetic Model of Aspirational Forgiveness

Cheshire Calhoun’s model of robust or “aspirational” forgiveness aims to avoid the two risks already foreshadowed in Kolnai’s original paper. One desideratum is to not blunt the connection between the original offense and the character of the offender at the time of the offense. If the victim allows this connection to fade due to rationalizations, tiredness, fatigue, lack of attention, guilt, or the desire to move on, then it is not forgiveness, but appeasement that is offered. If the connection is dulled due to the existence of really good reasons (e.g., evidence that “the perpetrator” is fundamentally good but this time he acted out of character, or that he was coerced by features outside of his control, or that he has reliably repented of his action), then the type of forgiveness offered is “cheap”, significantly lessened in value, or unnecessary. Indeed, if relevant features of our interpersonal reality show how our resentments were unjustified to begin with or are no longer warranted, this is a plausible response. In these types of situations, extending forgiveness seems less of a gift we make to our offenders, but something morally demanded by the situation we initially misconstrued and is retrospectively seen in better, sharper light. So, this is only a minimal type of forgiveness, if at all.

The second main desideratum is that robust forgiveness be given for intelligible reasons. That is, we expect normal people to act on understandable grounds that are open to public deliberation and in good standing with the shared norms making practical life possible in particular communities. Forgiveness offered without any justification would
seem psychologically impossible for beings so shaped by resentments, and morally repugnant to boot due to its ostensible arbitrariness. If we withhold moral judgment with our perpetrator’s violence in clear view, are we not morally blameworthy? Are we not arbitrary or perverse when we forgive robustly?

One of Calhoun’s tactics is to challenge the psychological model that makes only minimalist but not aspirational forgiveness possible. The dilemma (rather trilemma above) has force only due to a psychological model presupposing that to count as persons individuals must in some sense deserve forgiveness. An individual’s harmful action or behavior is either morally intelligible or it is not. If there are available justifications and excuses for the wrongdoing, the offender is forgiven. If however such cannot be provided, the offender falls under the unmediated demand to repent and reform. Not that forgiveness is inconsistent with the demand to repent, but it seems that the desert model implausibly makes this demand a necessary condition for extending forgiveness. If, for a variety of reasons (inattentiveness, malignancy, or slothfulness), the agent however fails to repent and reform, she counts as a diminutive person, morally stunted or underdeveloped. So the desert model operates under the supposition that normal agents either must deserve forgiveness due to some available justification. Alternatively, in the absence of justification and if the agent fails to reform, her personhood is significantly diminished.

Calhoun’s first step in developing a positive account of aspirational forgiveness involves abandoning or amending the desert model. As we will shortly see, some hold that resentment functions as a person’s primary alert system; it signals that some moral boundary of the victim has been violated. Nonetheless, Calhoun argues that making a
person’s reactive attitudes responsive only to moral demands seems exceedingly reductive. Her proposal in effect is to conceive ways of framing a victim’s reactive attitudes to injury in such a way that these are not all subsumed under resentment. Could the victim respond adequately to the offender from a standpoint distinct from that of moral desert? Is it psychologically feasible for the victim to undergo a change of heart rooted in dispositions distinct from those demanding that the perpetrator be morally accountable for her action, or that she repent, or otherwise take reliable steps in that direction? And even if this psychological model is possible for resentment-riddled beings such as us, is aspirational forgiveness a desirable moral outcome? Calhoun intends to argue that this is so.

It seems to me that Calhoun is right to appeal to narrative understanding as a first step in the direction of developing a more helpful psychological model of forgiveness. She points out that typical stories told by the victims about their perpetrators fall into two main categories. They either tell stories that make it possible to imagine the perpetrators as the continued recipients of their goodwill, or stories that frame them in a light in which they no longer are the appropriate objects of reactive attitudes. In this latter case, the perpetrators are sketched or recounted as diminished persons. They are seen as having in some sense disqualified themselves as desirable candidates for human relationships, due to psychological or moral defects, or on account of a vicious choice, or due to their systematic indifference to their victims. But this attitude, argues Calhoun, presupposes the belief that only those who can make moral sense of their actions should count as persons. And if victimizers fail to account for their actions before some moral tribunal, “they must be psychologically stunted, malignant, or indifferent,” and such unfit for
being part of our moral community. They don’t deserve our forgiveness.

But this is precisely the presupposition making only minimalist or desert-forgiveness possible. And this presupposition Calhoun’s model challenges. For normal persons do not respond to moral demands only, but also to biographical demands. Insofar as we are temporal beings, insofar as we successively encounter different arrangements of physical events, and often the surprising, unexpected or even irrational claims of others, we need to make narrative sense of our lives. Calhoun argues that this pressure on normal agents to maintain a cohesive self-understanding through time is an essential feature of the psychological model that helps make aspirational forgiveness possible. To the extent that wrongdoing is intelligible in the context of the offender’s life, to that extend the victim could tell stories that make it possible for her to imagine the offender as the recipient of her good will. 10 And to this extent, aspirational forgiveness remains a possibility. 11

So, aspirational forgiveness unfolds in circumstances in which even though an injurious action does not make moral sense, it can still make biographical sense in the context of the offender’s life. For even the unrepentant offender cannot extricate himself from the mandate of leading a narratively integrated life. Aspirational forgiveness transcends minimalist forgiveness by requiring the victim to “deprioritize” the moral

---

10 The suggestion here seems to be that the aspirational story minimally should provide some way for the victim to contextualize (or understand the genealogy of) the dispositions from which the wrongdoing proceeds. For this will enable the victim to see that the offender himself acts under complex self-understanding pressures.

11 Calhoun’s response is a specific manifestation of a more general strategy for defusing the paradox of forgiveness. The strategy, suggested to me by Mike Rea, envisions the possibility of drawing the distinction between holding someone responsible for a wrongdoing and blaming him for it. In forgiveness, we could refrain from blaming the victimizer while we also holding him responsible for the wrongdoing. Notice how Calhoun’s response exemplifies nicely an instance of this general approach. Understanding the narrative context of the offender’s action presents one way to reduce or neutralize our blaming him. At the same time, we can continue to hold him moral responsible for his wrongdoing.
standpoint in order to “hunt” for more profound explanations of how wrongness ensues from and is interwoven into the complicated patterns of the offender’s life. While we continue to regard the harm done to us as morally unintelligible, we also come to see that “wrongdoing is less likely to be a blow directly aimed at us than simply shrapnel from something else more complicated and more interesting in the person’s life.”

The Hebrew Bible tells the fascinating story of Joseph the dreamer and the victimization he suffers at the hand of his older brothers. Surely they seem morally culpable for selling their visionary brother into slavery. And as Joseph reflects on his past from some filthy Egyptian prison, his feelings of moral resentment seem justified. But he might begin to imagine what life would have been like for a Judah or a Simeon. He might come to see that their resentment and anger make sense given Jacob’s uneven attention toward his sons, or the strictures that nomadic culture places upon the older males of the clan, or maybe even in light of Joseph’s own self-absorbed visions. This does not mean that Joseph excuses their action, but only that he is now open to regard his brothers in a more humanizing way, aware that the pressures on their own lives, the ongoing demands of their father and the harshness of desert life, partially explain (but not justify morally) how they became mortally annoyed with him. Thus, if Joseph can empathetically enter his brothers’ shoes, and make sense of their horrific action, while he can still condemn unequivocally their wrongdoing, he is well on the way to embodying aspirational forgiveness.

Calhoun’s psychological model with its focus on the empathetic understanding of the offender’s diachronic dispositions is not idiosyncratic. In his paper “Forgiveness and

---

12 Calhoun, “Changing One’s Heart” p. 93
Self-Respect,” David Novitz argues that the attempt to empathetically understand the offender is a necessary condition for robust forgiveness. On Novitz’s view the ability to empathize must be cultivated to the point where it overcomes the reactive moral resentment of the victim. While Calhoun begins from the idea that moral resentment must be kept, as it were, in healthy tension with the attitude of making sense out of the wrongdoing in the context of the perpetrator’s life, Novitz’s more ambitious move is to make narrative empathetic acquaintance a necessary causal force responsible for beating moral resentment. By contrast, Calhoun’s aspirational forgiveness requires only that the victim adopt an empathetic stance from which she will prioritize telling an aspirational story about the perpetrator rather than a desert story.

While the turn toward empathetic understanding of the aggressor’s life context seems to advance in important respects our understanding of forgiveness, this condition, I argue, is in need of a finer theoretical development. I propose to do just that by showing how my dramatic-narrative agency model explains successful aspects of empathetic understanding while correcting some of its limitations. I will focus on two limitations affecting Calhoun’s aspirational model of forgiveness. A first is that the desire to understand the offender’s narrative can undermine the possibility of aspirational forgiveness in the absence of a relevant narrative perspective that grounds empathy. (I call this the opaqueness worry). The second limitation emphasizes that empathetic understanding must resist being overrun by resentments. Over time these tend to

---


14 Interestingly, Calhoun’s view doesn’t seem to require that the empathetic narrative fight off the feeling of moral resentment. Rather, and this seems more amenable to my own position, aspirational forgiveness requires that the victim place herself in a position, or be shaped in such a way as to be able to prioritize contextualizing narratively the harm over expressing moral disapproval.
entrench their grip of self-reflexive modes by disseminating in self-effacing ways throughout our emotional histories (I call this the resentment entrenchment worry). I argue that where opaqueness and resentment entrenchment show the limitations of the psychological model deemed to make aspirational forgiveness possible, the dramatic-narrative agency explains the force of these objections and answers them compellingly.

5.3. Empathy, Opaqueness and Dramatic Agency

First, let’s consider the opaqueness worry about the intelligibility requirement. The (Calhoun-Noviz) empathetic model assumes that in paradigmatic cases the victims would have or could imagine having access to relevant contextual features of the offender’s life that render the latter’s wrongdoing intelligible. But in some cases the victims simply lack access to such facts. Some perpetrators happen to be strangers or outsiders; others are acquaintances that deliberately hide their past. Sometimes the springs of action are hidden or impenetrable to agents themselves at the moment of action. And sometimes the offenders and victims alike trace perhaps too readily the cause of wrongdoing to a historical or external condition. In circumstances like these the victim may hardly have any hard facts to support a reliable understanding of the offender’s motives or life context. For these reasons, in some circumstance it would be hard to imagine telling the aspirational story. If motivational opaqueness reaches deep in psychology of the victims and offenders alike, we would tell helpful fictions or myths that help us move on, while the wrongdoing lays unexplained and un-contextualized.

More significant to my argument, even when the victim does have access to relevant aspects of the offender’s narrative that contextualize the wrongdoing, this fact
may not help trigger the right kind of empathetic response necessary for aspirational forgiveness. Sometimes, the more the victim learns about her perpetrator’s life story, the more bitter and resentful she becomes. The extended example I discuss below seems to shows that, pace Calhoun, narrative understanding can undermine aspirational forgiveness. Although this example is persuasive, it doesn’t force me to give up either on empathetic understanding or on robust forgiveness. Instead, my model of dramatic-narrative agency explains why understanding the offender doesn’t work in this case, thus alleviating the apparent tension between narrative on the one hand, empathy and forgiveness on the other.

5.3.1. Illustrating the Opaqueness Challenge

In J. M. Coetzee’s novel Waiting for the Barbarians the representatives of a “civilized” Empire attempt to control the border by engaging in cruel expeditions against those labeled “barbarians” or “enemies” but who are mere nomadic people. The main drama is set in motion when the local governor of the Empire crosses over into the desert wilderness in order to help take a barbarian girl (for whom he has inexplicably fallen) back to her tribe. Upon his return to the provincial border town, he loses his functionary job to a younger and more ruthless officer from the capital. He is arrested, and through a Kafkaesque turn of events he is beaten severely with sticks and slugged with a hammer for the sake of extracting a confession of betrayal. As he is hauled away in shame to the place of the execution, the ex-functionary tries in vain to make sense of the motivations

---

The example below is supposed to be challenge the following conditional that Calhoun seems committed to: “To the extent that wrongdoing is intelligible in the context of the offender’s life, to that extend the victim could tell stories that makes it possible for her to imagine the offender as the recipient of her good will”
that drive his torturer. It is remarkable to notice how his effort to humanize, to
understand his offender, to contextualize the harm done against him, is met initially by
the perception of a well-functioning cog in the machinery of the Empire. His torturer is a
well-trained biological automaton highly versed in the art or rather the science of torture.
(Notice the comparison with surgery)

I look into his clear blue eyes, as clear as if there were crystal lenses slipped over
his eyeballs. He looks back at me. I have no idea what he sees. Thinking of him, I
have said the words torture... torturer to myself, but they are strange words, and
the more I repeat them the more strange they grow, till they lie like stones on my
tongue. Perhaps this man, and the man he brings along to help him with his work,
and their Colonel, are torturers, perhaps that is their designation on the three cards
in a pay-office somewhere in the capital, though it is more likely that the cards call
them security officers. But when I look at him I see simply the clear blue eyes, the
rather rigid good looks, the teeth slightly too long where the gums are receding. He
deals with my soul: every day he folds the flesh aside and exposes my soul to the
light; he has probably seen many souls in the course of his working life; but the
care of souls seems to have left no more mark on him than the care of hearts leaves
on the surgeon.  

In the next scene the ex-functionary beckons his torturer to give him an
explanation, some reasons why he devotes himself to this line of work. And why is he
the target of the torturer’s violence? The only response the torturer offers to this plea is
illuminating:

Do you see this hand?” he says. He holds his hand an inch from my face. “When I
was younger” – he flexes his fingers – “I used to be able to poke this finger” – he
holds up the index finger – “through a pumpkin-shell.” He puts the tip of his finger
against my forehead and presses. I take a step backwards.  

Coetzee’s descriptions perhaps dramatize the truth that sometimes no imaginable
stories the victim tells about his perpetrator could make the wrongdoing intelligible. In
the first quoted paragraph the victim’s empathetic imagination crashes, as it were, against

17 Ibid, p. 118.
the towers of systemic oppression and functional routine. The clear blue eyes, as crystal lenses covering the eyeballs, might signal the impenetrable nature of the violence hiding comfortably in the human flesh of the perpetrator. Perhaps imperialistic systems desensitize flesh-and-blood humans to become torturers, surgeons of souls. Perhaps it makes sense for lives committed to administering violence to be depersonalizing to such an extent that relevant empathetic capacities are significantly diminished. Yet this explanation goes only so far.¹⁸ For what remains (and should remain) opaque to the victim are the reasons why the Empire administers torture to begin with, and more importantly, how it is that this particular officer has become a torturer, how it is that he has assumed and internalized this particular role in the economy of the Empire.

Crucially, the victim wants to understand whether the torturer carries out his characteristic work under a certain mode of self-understanding. And surely, given the presence of this desire, the victim seems disposed to tell an aspirational story about his perpetrator. Surprisingly the ex-functionary resists his initial impulse of treating the torturer as a mere objectified gear in the torture machinery of the Empire. He aspires to occupy a different standpoint from which he can regard the officer as the object of some proper reactive attitudes, and not just that of blind rage or resentment.¹⁹ We may speculate that the victim is searching for an external perspective from which he can continue to view himself as human, especially in the aftermath of the horrific desecration

¹⁸ I am not claiming that there is no explanation here. In fact, I think relations of power can explain a lot about fragmentation and compartmentalization of empathy. A prison guard at Guantanamo could be unremittingly gentle with his children and unflinchingly cruel to the inmates. While this phenomenon is explained well by the guard’s belonging to different “power systems”, I venture that this explanation is not very satisfying to those experiencing the rough end of the stick.

¹⁹ Interestingly, the strangeness of the words “torture”, “torturer” might indicate that Coetzee’s functionary experiences a natural psychological limit to his resentment. The search for a deeper explanation of the wrongdoing in this story shows at least just how unsatisfying un-contextualized resentment can feel to the victim.
of his personal boundaries. If the victim can make sense of the humanity of his
perpetrator, there is at least the chance that the victim counts as a person in the latter’s
eyes, and so a chance that he might aspire to integrate his own self, shattered by the
seemingly absurd, un-intelligible events of torture. So it is that the functionary imagines the officer as engaged in distinctively human activities (playing cards, friendships), these musings tempting him to dig for a deeper explanation of the disposition to torture in the broader context of the officer’s life.

Yet when the oppressor does provide a personal narrative, it is of the wrong kind. The pumpkin-pressing anecdote no doubt reveals a relevant sense in which the offender has been engaged in meaning making throughout his life, but the narrative is empathetically opaque, emotionally unintelligible to the victim. Consequently it is hard to imagine how this story could stimulate in the victim reactive attitudes that will make the officer the recipient of his good will. And this feature defeats the expectations set up by Calhoun’s account. This feature needs explanation by a more refined psychological model.

The pumpkin anecdote explains perhaps the officer’s qualifications for becoming part of a depersonalizing, violent system. But if the system is senseless to begin with, the person who makes such a cozy fit in it could hardly be more intelligible to the victims of Empire. More importantly, the anecdote intimates that the tortured appears from the officer’s perspective less as a patient to a surgeon than as a vegetable to by-passer. The self-reflexive script the tormentor invokes to explain his action evidently diminishes or obscures the personhood of the victim. The latter is perceived akin to an arbitrary inanimate object, serving only to showcase the perpetrator’s brute strength. The
perpetrator’s narrative thus adds the proverbial insult to injury. In Coetzee’s story aspirational forgiveness would thus seem psychological impossible, and this is neither for lack of aspiration nor for lack of a narrative. Even though the victim can see the offender involved in meaning making in his own life, that very activity does not provide the victim with right kind of external perspective to see how he matters *qua* person in the story of the offender. Consequently, there seems to be no way for the victim to understand the humanity of the offender, and thus no way at the moment for him to reclaim his shattered self.

5.3.2. Answering the Opaqueness Challenge

Interestingly, my account of dramatic-narrative agency can explain successfully (I argue) the collapse of empathetic understanding in this case. Recall that on my view dramatic-narrative agency has a two-tier structure. At one level I hold that individual diachronic intentions are explained and guided by rich emotional lives. I conceive of the emotions as biologically and socially grounded interpretive frameworks, displaying more-or-less integrated narrative cadences. Furthermore, any intentional narrative strand or overlapping such strands that are diachronically stable can function reliably as a distinct mode of the agent’s self-understanding. Thus, the sum-total narrative strands effective in (brought to bear upon) any decision or action constitute different modes of the agent’s self-understanding. At another level, I characterized the diachronic interaction of self-understanding modes on the model of a cast of different characters participating in the same drama. This model seems to me to explain well the experience of continuity and discontinuity of reasons brought to bear on action, the dynamic
interplay among the agent’s different reflexive postures, as well as the ability to respond “on the fly” and directly to the demands other agents place on us.

In the fragment from Coetzee’s story, the shape of the functionary’s diachronic experiences can be sketched as a complex emotional arc commencing with the physical trauma. This consequently opens up a number of possibilities for attaining emotional closure all of which are squelched eventually in a fresh emotive-psychological trauma. The first stage of this emotional cadence arises in response to the arbitrary torture occasioning puzzlement and resentment. In the second stage, analogous to the way in which theoretical crises in the sciences create the need for new scientific paradigms, these emotions create the need for empathetic projections and associated reflexive narrative postures that explain the physical trauma, the motivations of offender, and the possibilities of reshaping the broken agency of the victim. In the third and final stage, the victim’s empathetic response seems to terminate in repulsion and disgust.

Furthermore, it seems to me that the arc of this particular emotional cadence is explained by the dynamic, one might even say improvisational, interaction among various modes of self-understanding. Until the event of physical abuse the ex-functionary is described as oscillating between at least two self-reflexive modes. Most of the early events in the novel are framed through the interplay of these two first-personal perspectives. Through a first self-reflexive posture the protagonist understands himself as the apathetic and disillusioned provincial governor with a sustained interest in archeology. Through the second he sees himself as someone inexplicably attracted to the

---

20 I don’t mean to suggest that there is no emotional or psychological trauma co-present in the acts of torture, or that what I take to be the resulting psychological event in the series couldn’t also be co-instantiated in a new even of physical trauma. In fact, the functionary’s stepping back seems to support precisely this point. My point here is only that these two are distinct events setting the boundaries of a rather complex emotional narrative unfolding before our eyes.
exotic Barbarian woman that was once his prisoner. But as serial torture disrupts or shatters these modes of self-understanding, the threat is that the emerging mode of reflexivity that silences all others is one filtered through resentment. The real danger to the integrity of the functionary’s self-conception is that the description “victim of torture” is now internalized either as the dominant self-reflexive mode, or as the mode that henceforth permeates throughout his other central self-reflexive modes. So, in opposing the impulse to view the officer only as his torturer, the functionary in effect resists reducing his central self-understanding modes to just that stunted and derivative one.21 By trying to tell an aspirational story regarding his torturer, he is effectively trying to tell an aspirational story about himself.

Inviting the perpetrator to speak might also signal an invitation for the offender to view himself from the perspective of the victim, and thus to come to recognize the extent to which he has traumatized the victim. Seen from this light, the invitation itself is an empathetic act both because it may facilitate the perpetrator’s self-discovery and because by exposing the violence, it may be necessary step in inching toward forgiveness. This empathetic invitation then seems to be a reaching out for some story that will provide a reliable reflexive mode through which the victim could frame the violence done against him and an act of charity toward the violator, as he gets to speak, to explain his action. But when the officer (the captain) opens his mouth, it is clear that the story we get is not one that could serve to diminish resentment. Rather, the officer’s narrative only reinforces and even deepens the ex-functionary’s fear, namely that he can only regard

21 Torture stunts people because it shrinks the importance of the real past practical identities people have. Torture creates derivative identities because the tortured experiences oneself repeatedly and traumatically at the cruel mercy of the torturer. She experiences herself literally and figuratively in the hands of the torturer.
himself as a victim of torture. The anecdote conveys that the officer has self-consciously appropriated the events the victim describes as “torture” under the innocuous guise of “gardening”. It’s not that the victims do not enter his personal narrative. They do! It’s not that the victims don’t matter. They do! But they matter at best only as automata. They enter his self-understanding only as objects deprived of the capacity for self-reflexivity, as frail shells of pumpkins.

It is no surprise that this revelation makes the ex-functionary recoil in horror. The narrative voice he has allowed for the purposes of qualifying or ameliorating his own resentment actually exacerbates it. The officer’s mode of self-understanding reveals a moral monster, an opaque human being. And that diminishes the prospect that the victim’s own personhood be reintegrated by appealing to the offender’s perspective. Forgiveness at the moment seems psychologically impossible. As the psychological trauma of this revelation can only deepen the felt resentment22 (or desensitize the victim to the personhood of the violator), if forgiveness were to be possible again, what would be needed is a new empathetic story, a story that explains all of the previous puzzling facts: the physical torture, the humanity of the victim, the humanity of the torturer, and how the latter has come to internalize this reductive and awful mode of self-understanding. This complex narrative posture, if at all available, will have to retrospectively recast these events while also generating complex emotional responses that loosen the secure hold of resentment or the freshly landed disgust.

22 I don’t know if “resentment” or “callous revulsion” better describes what the victim feels in light of this revelation. I am trying to suggest here that that the emotion felt is retrospective, that it subsumes under a more pernicious guise the initial felt resentment. Initially, we may presume, the physical harm was processed under the guise of “random shrapnel”. But now the victim sees clearly that his torture is not accidental to the identity of the torturer but essential to his self-understanding. Consequently the nature and intensity of the attending emotion will attune to this new understanding, and likely will retrospectively reframe even the initial merely reactive resentment.
From this case that sets a limit to aspirational forgiveness we can generalize with caution. My account of agency predicts that aspirational forgiveness is psychologically possible only if a) there is a narrative posture that humanizes the offender and explains his hurtful action and b) this narrative perspective outfits the victim with a reflexive mode such that when set in dialogue with the choir of competing reflexive modes demanding moral accountability or worse, the net result will be to make resentment dissonant, or sidelined, or reduced in intensity. 23

5.4. Resentment and Dramatic Agency

5.4.1. The Depth and Breadth of Resentment

Related to the opaqueness worry there is perhaps an even more serious problem, one that proceeds from our affective natures. Even if there are biographical stories about offenders that contextualize or mitigate their wrongdoing, even if the victims have cognitive access to them and can successfully bring them in conversation with their anger, there is still the risk of emotional entrenchment and projection. In the aftermath of a sustained hurt and under the ensuing pressure of fear and anger, motivational and desiderative structures are likely to change dramatically. Over time, as the harm perpetrated is internalized and rehearsed in standard emotional routines, the interpretations of our own past and of the past of those who harm us, are likely filtered through and guided by complex yet subtle forms of fear and resentment. At the same time, as typical agents we also feel the practical pressure to move on with our lives, and when we do, we often do so without accurately assessing our past. Consequently, in

23 By this weak formulation I aim to indicate the possibility of establishing a narrative perspective in the mix that doesn’t get subsumed under resentment. Furthermore, this approach is compatible with various ways of making resentment irrelevant to decision and action.
order to become reliable narrators of the wreaked harm, victims need to simultaneously
avoid the psychological pressure to magnify resentment and the pragmatic temptation to
shrink it.  

Calhoun’s psychological model seems to me insufficiently fine-tuned to deal with
the problem of resentment entrenchment. In fact, if I am correct about the breadth, the
depth and subtle infiltration of resentment throughout the emotions relevant to practical
attitudes, telling generic aspirational stories does not seem conducive by itself to
developing aspirational forgiveness. These stories are fraught with difficulties *internal* to
the very dispositions deemed responsible for answering the self-understanding demands
of victims and offenders alike. Something about the very nature of self-understanding
and something essential to resentment make it such that appealing to generic biographical
demands alone is unlikely to sideline resentment. The dramatic-narrative model by
contrast has the potential to answer successfully the resentment entrenchment worries,
while also hinting at helpful ways to safeguard a measure of self-understanding necessary
for aspirational forgiveness.

In his wise book *The Vehement Passions*, Philip Fisher argues convincingly that our
emotions are dynamic diachronic forces that chart out from within, as it were, the
contours of agency. He says:

> The passions arise from, marking for our notice and the notice of others, a militant
> state of the will, patrolling its own borders, or what it imagines those borders to be.
The intensity of fear and anger, along with the tendency of formal descriptions of
> the passions to begin with either fear or anger, makes clear this negotiation of the
> radius of the will in which the self remains in a state of alert to any injury to the
> claims that the will makes on the surrounding world.  

---

24 It seems to me that often victims engage in both tendencies at the same time: manifestly they claim
to have moved on and are unwilling to face the hurt and resentment, while secretly the resentment and even
hatred is often amplified.

In imagery less flamboyant than border patrol, but no less compelling, Peter Strawson had also already noticed in his famous 1962 paper that certain types of resentment are “reactive attitudes” naturally stirred by violations or by envisioned violations to the perceived boundaries of personhood. Resentments often can be seen to contain markers that point to the rights and duties properly due to oneself.26 A number of contemporary discussions follow Strawson here, emphasizing that forgiving too readily may show a form of disrespect toward oneself, a kind of lack of self-regard. Resentment thus seems naturally fitted to remind a victim or potential victim that she is worthy of moral respect, that she is a proper person. Although Strawson emphasizes the reactive nature of resentment (its personhood-protecting aspect) and Fisher its more belligerent forms (its personhood-invading aspect), these descriptions are compatible. Resentment can have both a reactive and a dynamic aspect. In fact, as our subsequent discussion reveals, there is a wide variety of psychological phenomena that could properly be labeled “resentment”, ranging from merely passive, biological reactions all to the way to highly reflective, higher-order cultivated attitudes.

To get an appreciation for the breadth and depth of resentment it seems appropriate to consider Joseph Butler’s interesting and subtle discussion. His examination of the breadth of resentment casts a significant light also on its multifarious depth. I use Butler’s fruitful investigation as a springboard to my claim that this passion naturally attaches itself to a variety of emotions and attitudes, thereby infiltrating deep into our central desiderative and self-reflexive structures. We might even say that for a host of reasons connected to self-concern and self-preservation, resentment is naturally and

deeply promiscuous. I will simply assume that psychological models that explain and face up to the resentment entrenchment problem are necessary for individuals who aim to tell reliable aspirational stories of forgiveness. And I argue that my dramatic-narrative model is superior to Calhoun’s in this respect.

5.4.2. Butler’s Taxonomy

Butler distinguishes sudden anger and its various abuses on the one hand from settled (or deliberate) resentment along with its various abuses, on the other. Sudden anger is for him primarily a biological and not a moral passion. It arises strictly as a natural, self-defensive mechanism in response to the prospect that body boundaries may be violated. Thus the proper object of sudden anger is harm, to be distinguished from injury, which although possibly including harm never describes just a biological happening but contains implicitly moral valuations and is thereby the proper object of settled or deliberate resentment. The latter can be understood as a feeling of moral

27 There are several important distinctions here that Butler might not have cared to characterize more precisely. One is between a reflective disposition and an unreflective one. The other set is between an ingrained disposition and a fleeting one. Finally there is the distinction between dispositions actively directed toward distinct objects of attention and those non-intentional, or reactionary dispositions. On the one hand, Butler seems to assume that biological harm (when occurring in separation from moral injury) tends to trigger reactionary, fleeting, unreflective sorts of resentment. On the other, he assumes that moral injury tends to trigger reflective, engrained, and intentional emotional responses. While reflective, engrained and intentional dispositions come apart in all sorts of interesting ways, perhaps Butler might argue that there are special reasons for which they “go together” in the case of resentment.


29 Butler says: “The reason and end, for which man was made thus liable to this passion, is, that he might be better qualified to prevent, and likewise (or perhaps chiefly) to resist and defeat, sudden force, violence, and opposition, considered merely as such, and without regard to the fault or demerit of him who is the author of them”. And later, in the same paragraph: “But, considered as distinct from settled anger, it stands in our nature for self-defence, and not for the administration of justice.” (Fifteen Sermons, p. 95)

30 Clearly the two responses can coexist in the same person in all sorts of ways. The same event can be an instance of harm and moral injury, thus triggering both sudden anger and deliberate resentment. In Butler’s own words, the difference lies in that deliberate anger “is never occasioned by harm, distinct from
indignation that arises quite naturally in response to cruelty and injustice. This feeling is
as naturally intense as it is socially valuable when we witness the plight of others, but it
is especially reverberating and less fleeting when we ourselves are injured. In this case,
Butler holds, resentment has significant psychological utility, for it can be used as a
weapon against those threatening the boundaries of our moral self, or our personhood.

Given its conduciveness to social harmony and personal integrity, Butler seems
rather optimistic about the role of settled resentment in interpersonal relationships. Even
so, he takes the time to reveal some of its “abuses”. Among these he mentions our
tendency to exaggerate the injury done (“out of partiality to ourselves”); the tendency to
conjure up injury where none was committed; the tendency to construe only harm as
moral injury and thus respond with settled resentment when we ought respond strictly
with sudden anger; the tendency to respond with disproportionate resentment to a real
injury committed; and the tendency to maintain resentment and even take revenge simply
out of the desire to gratify the initial resentment, which, let us assume was formed
appropriately.

5.4.3. Amalgamation and Entrenchment

Butler’s optimism wanes a little in the very next paragraph after he bemoans the
futile task to taxonomies the endless deviations of resentment. He makes at this point the
following interesting observation:

injury; and its natural proper end is to remedy or prevent only that harm, which implies or is supposed to
imply, injury or moral wrong.” (Butler, p. 97)

31 Famously Butler claims that resentment arises against “vice and wickedness: it is one of the common
bonds, by which society is held together; a fellow feeling, which each individual has in behalf of the whole
species, as well as of himself.” (96)

32 I disregard here his concerns about the abuses of sudden anger.
But there is one thing, which so generally belongs to and accompanies all excesses and abuses of it [resentment], as to require being mentioned: a certain determination, and resolute bent of mind, not be convinced or set right; though it be ever so plain, that there is no reason for the displeasure, that it was raised merely by error or misunderstanding. In this there is doubtless a great mixture of pride; but there is somewhat more, which I cannot otherwise express, than that resentment has taken possession of the temper and the mind, and will not quit its hold.33

One way to express Butler’s thought here is that over time resentment tends to amalgamate with other emotions (he mentions pride) in such a way as to become an entrenched self-reflexive mode. Interestingly, Butler claims that resentment’s natural propensity toward entrenchment is evident in its abuses, that is, in circumstances where agents intentionally or unintentionally characterize either harm or excusable injury in terms of culpable injury. My suspicion is that in genuine cases of culpable injury resentment’s propensity toward amalgamation and entrenchment becomes even more robust, if not always more evident. At the very least, there is no reason to suppose that resentment takes possession of mind and temperament more moderately when the injury is real and culpable than when it is fictional or excusable.

It might be helpful to understand “taking possession” as a scalar property of a complex (perhaps non-linear) tendency a) to subsume sudden biological reactions under entrenched intentional dispositions and b) to subsume intentional dispositions under diachronically stable higher-order outlooks or reflexive modes.34 If the emotions are patterns of salience that direct practical reasoning, and if resentment is a co-opting (or disseminating) emotion working against the background of typical memory routines, we

33 Butler, *Fifteen Sermons*, p. 99

34 One ought not forget that here I build on the arguments from chapter 3 and 4. There I argued that reflexive modes are often grounded in the complex life of the emotions. And that one important aspect of cognitive emotions is their directionality. This refers to their capacity to direct beliefs, desires and practical reasoning.
can expect resentment to generate stable and yet improvisational interpretive scripts, scripts that in one way or another rehearse and keep fresh the harm or injury. We may thus say that resentment is deep (entrenched) as a measure of its tendency to infiltrate or commandeer the most diachronically stable and motivationally central self-reflexive guises of the person.

5.4.4. Resentment mixed with Pride

Let me revisit the story of Joseph and his brothers in order to illustrate concretely the diachronic amalgamation of resentment and pride. Several decades have now passed since his brothers sold Joseph into slavery. Threatened by starvation, the brothers travel to Egypt to buy grains so their families may survive. Still under the pressures of nomadic life, still under the roughshod hand of Jacob, ironically they are now vulnerable to loss. By contrast, Joseph has now become, partly through his ability to interpret dreams (ironically, the same ability that got him into trouble with his brothers), Pharaoh’s mighty treasurer. The meeting scene: “When Joseph recognized his brothers, although they did not recognize him, he was reminded of the dreams he had about them. He said to them: “You are spies. You have come to see the nakedness of the land.”35 Joseph’s early dreams had not been those of your typical teenager. In them, he had been the center of attention, worshipped by his brothers, mother and father. And now, it seems, the dreams had finally actualized.36

This memory seems to explain Joseph’s initial pretense, the half-feigned harshness

---

35 New American Bible, Genesis 42:8,9
36 I am not suggesting this is the only way to read this story, or even the primary way. Nonetheless, only I need is that it is a possible interpretation of this story.
at play in locking his brothers up for three days, and the harshness of keeping Simeon (his older brother) in custody until the others trek back to Palestine to fetch Benjamin, Joseph’s youngest brother. Whatever virtues he developed along the way, if Joseph frames the current meeting as a vindication of his teenage visions, his actions betray enduring resentment. Furthermore, it is a peculiar resentment, one filtered through a kind of smugness or self-importance. Now that the tables have turned, he will show them how the “dreamer boy” can play with their lives, and he will take his time doing it. If this interpretation is possible, it illustrates how over time resentment can infiltrate and endure couched in the self-reflexive guise of smugness. It shows that the self-descriptive script “I am your brother whom you sold into slavery” could give voice to a complex resentment that over time has been fed by and has in turn fed the smugness articulated in the self-description “I am the dreamer who predicted you will revere me”.

Nonetheless, while this example works well to show a certain degree of entrenchment through mutual mingling, the resentment seems perhaps moderate (and perhaps local). For when Benjamin is brought to him, Joseph reveals his true identity and embraces his brothers. Perhaps he judges that they have atoned sufficiently for they wrongdoing, or perhaps he offers forgiveness for reasons totally independent of their atoning. Below I give another example of a form of resentment filtered primarily through

37 It may be replied that Joseph is trying to teach his brothers a moral lesson or test their moral progress. Will they return for Simeon? Or will they leave him to rot in the Egyptian prison, just like they disposed of Joseph in the past? Several quick responses. First, even if this is Joseph’s primary motivation, it can still be interlaced with both resentment and smugness. Second, the smugness is particularly evident, if Joseph is described as entitled to track the moral progress of his brothers. Finally, strictly speaking Joseph presents the brothers with a devastating moral dilemma: either abandon Simeon, save Benjamin but run the risk of starvation, or bring Benjamin to Joseph, thus run the risk of hurting Jacob again if anything were to happen to him. That is to say, Joseph is not quite making his brothers’ lives easy.

38 One intriguing possibility is that for Joseph resentment can be kept alive only by the self-reflexive guise of smugness. On this possibility, to the extent that Joseph dissociates himself from the self-conception of a visionary, he will also let go of resentment.
fear, which seems to me more pervasive and more entrenched.

5.4.5. Resentment Entrenchment and Empathy

But before I move on, I want to point out why I perceive that the resentment entrenchment is problematic for Calhoun, and more generally for empathetic models of robust forgiveness. I began from Butler’s rather upbeat discussion of various forms of resentment and its social and personal utility. By extension I argued that as a harmful or injurious event is taken up into memory, the resentment felt typically evolves from a mere biological reaction to a complexly engrained interpretive script that is likely to amalgamate with other emotions. And when it disseminates among other self-interpretive paradigms such as fear and arrogance, resentment is likely to become a dominant, though sometimes self-effacing, higher-order mode of self-understanding.

Before I move on to explain the virtues of my model relative to those of Calhoun’s, let me make a brief methodological point. The theory that I offer is neither a priori nor is it a hypothesis probabilistically developed from a multitude of empirical examples. Rather, I take it that my model has two fundamental advantages. First, it is offered as a corrective to the empathetic models of forgiveness that avoid the reductive trends of judicial and civil models. Second, and more importantly, the psychological model I propose seems to explain a wide variety of psychological phenomena related to forgiveness and recalcitrant emotion of resentment. Also, as I suggest in the last section the model is fruitful for explaining psychological phenomena associated with atoning for one’s wrongdoing. Thus, its theoretical goodness can be assessed analogous to the way scientific hypothesis are assessed on the basis of criteria revealing an inference to the best
explanation. Other models might make fewer assumptions or have greater elegance. I hold that the narrative-dramatic model has great explanatory power with respect to forgiveness and has fruitful applicability to retrospective phenomena beyond forgiveness.39

As Calhoun tells it, her psychological model not only allows but requires the unrepentant offender to remain under the condemnation of unreduced moral indignation.40 In fact this is a key aspect that distinguishes aspirational forgiveness from minimalist forgiveness. Now if I am right in my Butler-inspired comments, keeping the offense and the offender fresh before one’s victimized eye tends to disseminate and entrench resentment, and even amplify it beyond the limits of the moral.41 So, if she must keep the unrepentant offender the target of unreduced indignation, it is hard to see how the victim could without self-deception deprioritize the moral standpoint. But this deprioritizing is the natural consequence of a changed heart, the upshot of the relevant psychological transformation that makes room for telling aspirational stories. Therefore, unless a more fine-grained model of empathetic psychology is provided, Calhoun’s desideratum of telling aspirational stories runs strikingly against the desideratum of keeping the unrepentant offender fresh before one’s eye. The more proficient we get at the latter, the more we are likely to fail at the former.

39 Thanks to Mike Rea for drawing my attention to this methodological point.

40 See Calhoun’s insistence that resentment must be kept clearly in our sight. “What distinguishes a minimalist forgiving of me from aspirational forgiving is the story you tell about me. Minimalist forgiving tells a story that places me under some description that makes me an inappropriate object of resentment (or of as much resentment). In aspirational forgiveness, I, as object of forgiveness, remain under an intentional description of culpability that warrants continued unreduced resentment.” (“Changing One’s Heat”, p. 81)

41 If the story I weaved thus far is on the right track, in the wake of an injury the hypertrophy of fear or the hyperactivity of pride can easily merge with resentment, enabling the victim to internalize at one end the extreme fright and numbness and at the other, vengefulness.
Perhaps Calhoun would reply that the undeniable mark of undergoing a genuine change of heart just is *deprioritizing the moral while continuing to see the injury in all its hideousness*. She might add, plausibly, that this balancing act is very hard. This is why aspirational forgiveness is, well, *aspirational*. While robust forgiveness is indeed difficult to attain practically, a generic empathetic-psychological model does not explain at right level of structural detail how aspirational forgiveness is possible. I propose that what is needed is a diachronic psychological model that explains how typical agents can a) feel an appropriate form of anger at the remembered initial injury and simultaneously can b) take up dramatic perspectives allowing them to tell aspirational stories, but aspirational stories whose content is partially shaped by c) their awareness of the miscellaneous entrenchments of resentment.

Notice that on this model the possibility of telling aspirational stories depends explicitly on proper self-assessment and other-assessment. If iterating in memory the injury likely enhances fear or pride or the desire to see the offender suffer, aspirational stories of forgiveness must include multiple reflexive stories that counter entrenched resentments. Predicated on the possibility of attaining a workable measure of self-knowledge, these stories will aim to diminish fear, or pride, or the desire for revenge. Furthermore, stories of resistance essentially must *recognize* the natural rhetorical power of disseminating resentment and associates, and *aim to undermine* it. So, reliable stories of resistance depend on cultivating the ability to occupy simultaneously multiple dramatic perspectives about our past injuries and about our diachronic responses to them. For, if resentment is a multivalent swaying passion, then sidelining it involves resisting the tendency to allow past injuries slip into the gravitational center of our self-conception.
And if resentment, pace Butler (and Calhoun), is more subtle and insidious than acknowledged, we need to learn to effectively recast our dramatic circumstances, sometimes even by listening to voices from the periphery of our self-conception.\textsuperscript{42}

Practically this means allowing personal memory, friends, family, imaginative art, and even foes to furnish empathetic interpretative perspectives that interrogate or shock our settled emotional routines.\textsuperscript{43} Clearly what I have just outlined is one way in which the dramatic-narrative model responds to the resentment entrenchment problem.

5.4.6. Resentment mixed with Fear

A compelling statement of the way in which fear deepens and fortifies resentment over time is found in Margaret Atwood’s novel \textit{Cat’s Eye}.\textsuperscript{44} Here she tells the fascinating story of Elaine Risley who returning to her childhood stomping grounds (Toronto) for a retrospective show is compelled to piece together the broken fragments of her traumatic past. Atwood describes how commonplace harms even in the life of nine-year-old girls can over time amplify into lifelong traumas and can continue to shape the course of protagonist’s memories, emotional reactions, and even guide her creative

\textsuperscript{42} Here is perhaps a helpful image to balance Butler’s view of resentment as a self-defensive weapon. For some (no doubt) idiosyncratic reason I began thinking about resentment as a fainter version of Plato’s tyrant. In Book IX of the \textit{Republic} the tyrant is described as an omnipresent force incapable of making friends, teetering between the ingratiating and the vengeful, doing all it can to maintain control of the constitution. If this analogy is at all adequate, resisting resentment would be like a guerilla warfare against a very persuasive but self-effacing tyrant. I think Butler is correct in this: resentment is one of the main psychological glues tending to integrate agents over time. And it seems to be doing this by telling and retelling coherent stories that naturally dense up around the value of self-defense and security. But if I am right, the kind of psychological unity achieved through defensiveness is not worth much, partially because it tends to fictionalize the past, to promote poverty in emotional lives and to stunt self-conceptions.

\textsuperscript{43} This is clearly consistent with the practical version of vehicle externalism I outlined in the previous chapter.

\textsuperscript{44} Margaret Atwood, \textit{Cat’s Eye}. (Anchor House: Random House,1998)
work.  Through her memory Elaine sympathetically reframes her understanding of Cordelia, her best childhood friend who through her wild and arbitrary behavior had shaped Elaine’s earliest experiences of fear and loneliness. Elaine also sympathetically recalls her mother’s failure to protect her sufficiently from the abusiveness of her friends Grace, Carol and Cordelia. In my discussion I want to focus on Elaine’s recounting of Grace’s mom, Mrs. Smeath.

Mrs. Smeath is initially depicted as a stern woman suffering from a “bad heart” on account of which the friends must “tiptoe, walk quietly, stifle our laughter, do what Grace says.” Her severity manifests in manifold ways, often finding in Elaine the object of her disgust: she treats the un-churched Elaine with smug charity, she is unbending about table manners, she demands that Elaine memorize Scripture and then despises her for her excellent performance. The fear of losing Grace’s friendship makes Elaine keep returning to the Smeaths, although she feels relentlessly judged, scrutinized, and watched. Particularly evocative in this respect are Elaine’s church visits in the company of the Smeaths. During her first visit, Elaine is particularly impressed with the huge stained-glassed windows. The largest one portrays a “Jesus in white”, white bird circling his head, with a thick black inscription underneath that reads: THE . KINGDOM . OF . GOD . IS . WITHIN . YOU. Another has a woman, dressed in blue with a white cover partially hiding her face, reaching out to a man “with what looks like a bandage wound around his

45 This is not something incidental to her project, but seems essentially connected to her view of memory as including and deepening prior events. The third paragraph from the book has Elaine contemplate on the nature of her remembering: “But I began then to think of time as having a shape, something you could see, like a series of liquid transparencies, one laid on top of another. You don’t look back along time but down through it, like water. Sometimes this comes to the surface, sometimes that, sometimes nothing. Nothing goes away.” (Cat’s Eye, p. 3)

The inscription underneath: THE.GREATEST.OF.THESE.IS.CHARITY.

On White Gift Sunday Elaine brings to church Habitant pea soup and Spam for the poor. Even though the white gifts, “bleached of their identity”, look sinister to Elaine, she still worries that the Smeaths would perceive hers as inadequate. Worse, she feels judged for having to read the words to the songs from the screen, as Grace, who knows them by heart, is intently watching her. She reflects: “I want to shine like a candle. I want to be good, to follow instructions, to do what Jesus bids. I want to believe you should love your neighbors as yourself and the Kingdom of God is within you. But all this seems less and less possible.”

Fast forward a couple of decades. Elaine now recalls her accomplishments as a progressive young artist in search of recognition, in the context of the broader feminist movement of the 60s and 70s. At an all-women artist show she displays all her paintings of Mrs. Smeath. In one painting entitled “Leprosy” she appears with half the face peeled off, but the main attraction is “White Gift” which is in four panels:

In the first one, Mrs. Smeath is wrapped up in white tissue paper like a can of Spam or a mummy, with just her head sticking out, her face wearing its closed half-smile. In the next three she’s progressively unwrapped: in her print dress and bib apron, in her back-of-the-catalogue Eaton’s flesh-colored foundation garment – although I don’t expect she possessed one – and finally in her saggy-legged cotton underpants, her one large breast sectioned to show her heart. Her heart is the heart of a dying turtle: reptilian, dark-red, diseased. Across the bottom of this panel is stenciled: THE . KINGDOM . OF . GOD . IS . WITHIN . YOU. It’s still a mystery to me why I hate her so much.

With the passing of years Elaine’s subdued childhood resentment seems to have escalated to hate. Filtered through her artistic genius the hatred is precise, progressing

---

48 Ibid, p. 137.
and dissecting. Some of the ordinary objects and events of childhood (Spam, half-face, bad heart, church inscription) are taken up and given new ominous (even irreverent) meanings as they are recycled in Elaine’s gravitational field of hatred for Mrs. Smeath. From the retrospective perspective of the narrator, the seething resentment of the young artist is grounded in the deep intentional objects of her memories. Though motivationally efficacious and intentionally highly structured, the artist’s resentment remains a mystery for Elaine’s synchronic consciousness. This invites the hypothesis that the intentionality of resentment had been at least partially guided by the intentionality of Elaine’s diachronic fears. This hypothesis is confirmed as we pick up the narrative thread. For during the exhibition, just as Elaine looks away from Mrs. Smeath (the paintings) she seems to see a flesh-and-bone Mrs. Smeath moving toward her: “It’s as if she’s stepped down off the wall, the walls: the same round raw potato face, the hulky big-bone frame, the glittering spectacles and hairpin crown. My gut clenches in fear; then there’s that rancid hate, flashing up in an instant.”

5.4.7. The Explanatory Power of Dramatic Model

Elaine’s diachronically internalized fears sustain, deepen, and sharpen her resentment. It heightens her attention, her natural and cultivated alarm system to those features and properties she initially found threatening. It moves Elaine to associate (sometimes irrationally) clusters of her real, historical fears around the object of her primal fear (Mrs. Smeath), even to the extent that the latter seems to absorb the former. If this is right, fear functions as a self-defensive mechanism that commandeers (or at least

guides) even highly reflective processes that go into creating works of art. On this interpretation, we can see the explanatory power of the dramatic-narrative model. Fear and resentment are interpenetrating paradigm scenarios that frame diachronically Elaine’s increasingly hateful conception of Mrs. Smeath. Growing around this weighty intentional object, they have been configuring the stories through which Elaine understands herself ever since childhood. Thus, they infiltrate lastingly throughout Elaine’s central reflexive modes, including her self-conception as an artist.

The dramatic-narrative model also predicts that if forgiveness is possible here, Elaine would have to occupy some dramatic-empathetic perspective that allows her to recast her conception of Mrs. Smeath and to assign new meanings to the intentional objects that have been pulled in the gravitational center of hatred and fear. If there are aspirational stories to be told here, some have to be resentment resistance stories, some of which will target Elaine’s fears, and some her rancid hatred.

Fast forward again. Elaine is now an acclaimed elderly artist and herself the mother of two young women. She is beckoned back home, to Toronto, to the place of her childhood traumas for a retrospective show in her honor. Arriving early at the gallery, she reflects on her life’s work as it is chronologically arranged\(^{51}\), devoting the longest time to a re-examination of “Mrs. Smeath”. A first step in Elaine’s retrospection is a direct and reflective acknowledgement of the early fright that had led her to scorn and desecrate Mrs. Smeath in her artistic representations, scorn which perhaps unintentionally also exaggerated the fear:

Next to them is Mrs. Smeath; many of her. Mrs. Smeath sitting, standing, lying

\(^{51}\) Interestingly Atwood puts in Elaine’s mouth the enigmatic: “Chronology won out after all”, perhaps signaling the perspective of a narrator who has already achieved some measure of emotional closure to her life’s drama, if we take that drama to be neatly represented in the stages of her development as a painter.
down with her holy rubber plant, flying with Mr. Smeath stuck to her back, being screwed like a beetle; Mrs. Smeath in the dark-blue bloomers of Miss Lumley, who somehow combines with her in a frightening symbiosis. Mrs. Smeath unwrapped from white tissue paper, layer by layer. Mrs. Smeath bigger than life, bigger than she ever was. Blotting out God.  

In a second step, acknowledgement of the exaggerated fear leads to an unexpected confession. What in the mind of the budding artist was merely a mysterious hatred, in the hindsight of the seasoned introspective narrator is deliberate, carefully choreographed malevolence: “I put a lot of work into that imagined body, white as burdock root, flabby as pork fat. Hairy as the inside of an ear. I labored on it, with, I now see, considerable malice.” One might even say that as Elaine occupies a dramatic perspective external to her fear, she is enabled to recognize the grip resentment had on her life and art. The ruthlessly honest self-assessment is concurrent with another perception and self-perception. In this third moment of her retrospection, unrestricted by fear or self-deception, Elaine is free to recognize that some of Mrs. Smeath’s features were taken up in her memory and represented artistically without malice, realistically, even empathetically:

It’s the eyes I look at now. I used to think if these were self-righteous eyes, piggy, and smug inside their wire frames; and they are. But they are also defeated eyes, uncertain and melancholy, heavy with unloved duty. The eyes of someone for whom God was a sadistic old man; the eyes of a small town threadbare decency. Mrs. Smeath was a transplant to the city, from somewhere a lot smaller. A displaced person; as I was.

Notice several features about this description. First, Elaine’s empathetic stance is possible because it is grounded in a reliable self-awareness and a reliable offender-

52 Ibid, p. 443.
53 Ibid, p. 443.
54 Ibid, p. 443.
awareness. It is only because she continues to hold Mrs. Smeath morally condemnable (her smugness), and because she has confessed her malicious exaggerations that Elaine can now reliably empathize with her. Second, now that the self-interpretive scripts of fear and malice no longer blind Elaine, her empathy can be seen as a form of visual attention responsive to some real features of the world that hitherto had gone unnoticed.\(^{55}\)

In my preferred language of the emotions, we are witnessing the formation (or activation) of a brand new interpretative paradigm that has the potential to correctively recast Elaine’s memories and associations.\(^{56}\) Significantly, Elaine’s empathy for Mrs. Smeath is not grounded in a pre-existing disposition or desire, but in a perception that itself has historical precedents.\(^{57}\)

Third, once the empathetic posture gets a hold, it becomes a growing self-interpretive script that continues to loosen the holds of fear and resentment. As it moves from the periphery to the center of Elaine’s self-interpretive stories, this perception reconfigures the relations holding among the intentional objects conditioned historically by her fear and resentment. Elaine’s aspirational stories thus essentially shock and destabilize the network of meanings that had been historically attached to her perception of the monstrous Mrs. Smeath. The recognition that Mrs. Smeath is herself a displaced

---

\(^{55}\) It is interesting that empathy itself is born out of the reflection on the pictorial representation of Mrs. Smeath’s eyes. Also, reflecting on her past representation of Mrs. Smeath Elaine reflects: “I have not done it justice, or rather mercy. I went for vengeance. An eye for an eye only leads to more blindness.” (Atwood, 443)

\(^{56}\) It is essential to notice that the formation or activation of this paradigm scenario cannot be explained by appealing Bratmanian plans, Frankfurtian loves, or Milgramian supplies of practical judgments. Rather it arises in response to the real features of the world that now get noticed, given a much sharper perception of the background that has led to Elaine’s blindness to this feature.

\(^{57}\) Elaine has occasionally had to turn the paintings facing the walls (she couldn’t bear looking at those eyes). Also, on the occasioned of her first feminist exhibition, Elaine’s artist friend Jody exclaims regarding the numerous depictions of Mrs. Smeaths: “It’s woman as anticheesecake…Why should it always be young, beautiful women? It’s good to see the aging female body treated with compassion, for a change.” (Atwood, 380)
person effectively shows she is no longer a gravitational center for Elaine’s swirling self-conceptions. Or if she still is, she is no longer now seen as an overpowering divinity with a reptilian heart. The corrected picture of Mrs. Smeath as a “transplant” and herself the victim of an internalized fear of God shows her as a mirror image to Elaine, a doubling of Elaine’s own need to be loved, protected, welcomed home. And this new picture also makes it possible for Elaine to occupy yet one more dramatic perspective, to view herself empathetically through Mrs. Smeath’s eyes. To imagine a dramatic perspective that allows her to understand how she functions as a person in Mrs. Smeath own story:

Now I can see myself, through these painted eyes of Mrs. Smeath: a frazzle-headed ragamuffin from heaven knows where, a gypsy practically, with a heathen father and feckless mother who traipsed around in slacks and gathered weeds. I am unbaptized, a nest for demons: how could she know what germs of blasphemy and unfaith were breeding in me? And yet she took me in.  

When set against the context of Mrs. Smeath’s hard life, Elaine can finally see herself as the object of Mrs. Smeath’s charity, something that shouldn’t be mocked. There is smugness there to be sure, but it is smugness partly explained by the latter’s displacement, as we have seen. The upshot is that embedded in Elaine’s empathetic stance toward Mrs. Smeath is an empathetic stance she can take toward herself. She is now capable of seeing herself not only as a victim, but also as the recipient of Mrs. Smeath’s empathetic attention. Someone who, albeit in her own misguided and displaced ways, is at least attempting to make a home for her.

Elaine is lucky to have available the relevant external emotional stances from which to launch a corrective re-interpretation of her past. Her paintings awaken a chorus of dissonant voices that enlarge her conception of Mrs. Smeath, and of the many objects and

58 Atwood, *Cat’s Eye*, p. 443
events whose perceptions had been stunted by entrenched resentment. By listening to these voices her own self-conception is enlarged, not only generally in the sense that she is now the person who is capable of forgiving Mrs. Smeath. More specifically, she does not to see herself merely as the victim. Her paintings thus serve for her as a reliable narrator and advisor and interlocutor.

As Elaine, typical agents are beset by hurts and injuries going all the way back to their childhood. As Elaine, typical agents are not unambiguously the victims; sometimes they slip naturally into patterns of revenge or even victimization of others. Typical agents are perhaps not as lucky as Elaine to have right before their eyes relevant external frames that provide corrective takes on their past. If robust forgiveness demands reliable retrospective attention, and if this involves telling not only aspirational stories about our offenders but also enlarging our emotional self-understanding, we need all sorts of external perspectives. We all need reliable advisors and interlocutors. These often need to be sought out and cultivated through the sustained company of friends, family, fictional characters of good literature, and art.

In this extended example I have tried to show at least three things. First, I have argued that fear and related emotions can mingle with resentment in ways that make the latter an entrenched feature of self-understanding. Second, I argued that if aspirational stories of forgiveness are possible, they essentially have to reference stories that unravel the hypertrophy of fear and resentment. Finally, I have shown how the narrative-dramatic model explains successfully how such corrective stories must be told from different and changing dramatic perspectives.
5.5. Prospects For Dramatic-narrative Agency

Fear and pride are only two of the emotions providing breeding grounds for resentment. My hunch is that the deeper we articulate the various forms of fear and pride, the more confirmation we will find for the entrenchment of resentment. But these specific investigations will have to wait for a different time.

In this chapter I have focused on the empathetic model of forgiveness. I have claimed that narrative-dramatic agency helps elucidate it and solves some of its problems. But I think it can also cast a very fruitful light on the process and conditions of atonement. Atoning could be conceived as a characteristic pattern of action undertaken by an offender to the end of redressing or remedying a wrong perpetrated against the victim, pattern that essentially involves appropriate emotional responses. If remorse and repentance are emotive dispositions necessary for atonement, it seems that the offender has to develop an appropriate narrative understanding of the perpetrated harm in the context of the victim’s life.\footnote{Interestingly, this claim has important applications in Christian theology. It’s perhaps not surprising that believers have devoted so much attention to the life, passion and death of Christ. If, as I suggested earlier Christ is the innocent victim of human violence, and if believers are now called to be “crucified with Christ”, effectively this might mean adopting that dramatic stance, being drawn into a narrative that makes even forgiveness of the radical other, of the worst enemy, possible. Interestingly for Christians, the drama of the suffering and death of Christ provides both the possibility of being forgiven, but also as they atone for their violence by taking on the perspective of the crucified God, believers learn to make atonement, that is, to forgive others in ways they couldn’t have done otherwise. I thank Mike Rea for suggesting this line of thought.} In fact one might speculate that for atonement to be attainable, external emotional stances similar to those employed in forgiveness must be effective. So, the possibility of “aspirational” atonement would be more plausible on dramatic-narrative conception of agency.

An interesting Homeric story serves to illustrate my point. As the Phaeacian poet Demodocus sings about Troy’s fall, Odysseus enjoys incognito the praise of his
cleverness and courage. Yet, at a crucial moment in the recounting, moments after Demodocus sang about how Odysseus devastated the fallen city “like the War God Himself”, an unsung memory creeps upon him:

A woman wails as she throws herself upon
Her husband’s body. He has fallen in battle
Before the town walls, fighting to the last
To defend his city and protect his children.
As she sees him dying and gasping for breath
She clings to him and shrieks, while behind her
Soldiers their spears into her shoulders and back
And as they lead her away into slavery
Her tear-drenched face is a mask of pain

So too Odysseus, pitiful in his grief. 60

Surprisingly even though the narration provides only the external perspective of the victor, experiential memory leads Odysseus to see himself through the eyes of the conquered. As the perpetrator Odysseus identifies with the plight of the victim, this external perspective on his own actions (presumably causally responsible for the woman’s plight) provides him the opportunity for remorse and atonement. I won’t pursue this point much farther here, except to claim that emotions such as grief and remorse are the appropriate forms of empathy conducive to telling an aspirational story of atonement, and insensitivity the central unruly emotion. In atoning the perpetrator’s insensitivity toward the victim can be seen to play a role analogous to that played by the victim’s resentment toward the offender in the case of aspirational forgiveness.

Since we are often victims and often offenders, and often victims and offenders with respect to the same individual, we navigate complex emotional waters by alternating

---

different but related external perspectives on our action. In this area, I suspect, the narrative-dramatic model has a lot of explanatory power.
BIBLIOGRAPHY


______. Actual Minds, Possible Worlds (Harvard University Press, 1986).


______. *Necessity, Volition and Love* (Cambridge University Press, 1999).


______. The Importance of What We Care About: Philosophical Essays (Cambridge University Press, 1988).


______. “Rationality and the Unthinkable,” in The Importance of What We Care About: 177-190.


______. “Freedom of the Will and the Concept of a Person,” Journal of Philosophy, 68.1 (January 1971); reprinted in The Importance of What We Care About: 11-25.

Frei, Hans W. The Eclipse of Biblical Narrative (Yale University Press, 1974).


Goldie, Peter. The Emotions: A Philosophical Exploration (Claredon Press, 2000).


____. *Subjection and Subjectivity: Psychoanalytic Feminism and Moral Philosophy* (Routledge Press, 1994).


____. *Practical Induction* (Harvard University Press, 1997).


______. *Damaged Identities, Narrative Repair* (Cornell University Press, 2001).


______. *The Birth of Tragedy,* tr. Clifton Fadiman (Dover Publications, 1995).


Sacks, Oliver. *The Man Who Mistook His Wife for a Hat and Other Clinical Tales* (Simon and Schuster, 1985).


______. *Ethics and the Limits of Philosophy* (Fontana, 1985).

______. *Shame and Necessity* (University of California Press, 1993).
