

Tony Soprano, compartmentalized: An Achilles' heel in the deep self view

Ke Zhang

This is the accepted manuscript of a paper forthcoming in Ethical Theory and Moral Practice.

Please cite the published version when available.

Abstract

Standard deep self views of moral responsibility contend that we are responsible for what we do if and only if our actions issue from our deep selves. This suggests that responsible agency is unified in a crucial sense. In this paper, I identify a tension between a special phenomenon of compartmentalization and unified responsible agency, and elaborate a novel criticism of the deep self view. Given the deep self view's requirement for unified responsible agency, a severely compartmentalized evildoer is at least less responsible. However, I argue that the reason such an agent is deemed less responsible within the deep self picture commits deep self theorists to judging that all compartmentalized wrongdoers are likewise less responsible. This significantly undermines the plausibility of their views.

Keywords

Compartmentalization, deep self, responsibility, unity of the virtues

1. Introduction

Because of the extremity of evildoing, it is sometimes easy to stereotype evildoers as one-dimensional—they wake up, destroy people's lives, and go to bed. However, at least some of them are more complicated than that. Indeed, a cruel murderer on the downtown streets can be a loving partner at home. They can be a good partner because they care about maintaining a good relationship. Outside of their family life, however, they can become the worst nightmare for anyone who gets in their way. This phenomenon can be captured by a special kind of compartmentalization, where, in some domains of life, a compartmentalized agent exhibits patterned and systematic

behaviors that align with some practical concerns, while in other domains, they exhibit behaviors that are contrary to the same concerns.

Although the responsibility literature has not paid much attention to it—aside from Michael McKenna and Chad van Schoelandt’s (2015) discussion of motivational and behavioral incongruence across different areas of activity (more on this in Section 6)—the phenomenon of compartmentalization has facilitated extensive discussions about the unity of the virtues and virtuous agency (Badhwar 1996; Sreenivasan 2008, 2009; Schaubert 2014; Stichter 2021). The debate there reveals a tension between compartmentalization and unified virtuous agency. When considered with more clarity, I suggest a similar tension arises in responsibility theories that require unified responsible agency. In this paper, I focus on standard deep self views of moral responsibility.

My focus on deep self views relies on a crucial feature of these views: they play a prominent role in advancing the idea that responsibility requires unified responsible agency. According to such views, being responsible for what one does requires coherence between elements of one’s psychological constitution and one’s behaviors (Frankfurt 1971; Watson 1975, 1987; Bratman 1997, 2004, 2005; Shoemaker 2015; Sripada 2016; Matheson 2019; Gorman 2019, 2022). This coherence means that our responsible agency is unified in a crucial sense. The requirement for unified responsible agency will lead deep self theorists to claim that a compartmentalized evildoer is at least less responsible for their actions. As I will argue, reaching this judgment comes at the expense of concluding that all compartmentalized wrongdoers are less responsible, which significantly undermines the plausibility of these views.

In what follows, I will first examine a compartmentalized agent, Tony Soprano from the TV series *The Sopranos*, and identify a special phenomenon of compartmentalization that occurs in his life (Section 2). After discussing different ways to interpret deep self views, I will propose the most plausible understanding of the requirement for unified responsible agency in the deep self picture

(Section 3). Drawing inspiration from discussions about the unity of the virtues and virtuous agency, I will argue that this requirement for unified responsible agency motivates deep self theorists to judge a compartmentalized agent like Tony as less responsible (Section 4). However, the reason they would use to make their case will commit them to judging that all compartmentalized wrongdoers are less responsible (Section 5), which undermines the plausibility of their views (Section 6).

2. *The Phenomenon of Compartmentalization*

Tony Soprano¹ is trying his best. He cares about his family and tries to protect them as best he can, though admittedly far from perfectly. In the wider world, however, he is a committed criminal—cruel, greedy, and most importantly, a taker of people’s lives. Given his commitment to both things, there is a patterned and systematic incongruity in how he comports himself in interpersonal relationships across these domains of life. “Why can’t he carry even a molecule of his kindness and respect toward his family over to his enemies and their families?” asks the viewer. Indeed, some viewers, including myself, are deeply perplexed by this. We are intellectually and emotionally invested in complicated characters like Tony at least partly because they present interesting cases of compartmentalized practical agency. This phenomenon of compartmentalization challenges us to form apt judgments about the character of evildoers like Tony and the extent to which these evildoers are responsible agents.

In fact, this phenomenon has already been extensively discussed by virtue ethical theorists in the debate about the unity of the virtues and virtuous agency. An important element in that debate stems from a perceived tension between compartmentalization and a unified conception of virtuous agency (Badhwar 1996; Sreenivasan 2008, 2009; Schaubert 2014; Stichter 2021). In this paper, I argue that a similar tension arises in understanding and theorizing about responsibility, particularly

¹ *The Sopranos* is a TV series about the New Jersey mob boss Tony Soprano, who struggles with *doing the right thing* in both his family and criminal lives. He suffers from mental breakdowns due to his lifestyle and career choices, and experiences identity crises as a committed criminal with a family.

in deep self views that commit to a unified conception of responsible agency. But first, let me more precisely articulate the phenomenon as follows:

Compartmentalization An agent is compartmentalized with regard to a practical concern across a given set of domains of life when they issue patterned and systematic practical concern-conforming behaviors in some of those domains, but issue patterned and systematic practical concern-nonconforming behaviors in some others of the domains.

This formulation suggests that domains of life must be individuated in a certain way for compartmentalization to occur. But what is a domain, and how are they to be individuated with respect to one another? Consider first Neera Badhwar's (1996) notion of a domain. According to Badhwar, a domain is an area of practical concern. Our practical concerns are expressed through certain normative attitudes and psychological features regarding what we judge ought to be done. Because we have a variety of practical concerns in life, and because some of these concerns cluster in normative ways, at least some domains are distinguishable from others (316-320). For example, an aspiration we have in our personal life to, say, cultivate an appreciation of fine art requires us to set goals, pursue values, and monitor our motivational and emotional states differently from how, in our professional life, we aim to develop skills that improve work productivity. This creates a distinction between our personal and professional lives. Similarly, we sometimes separate our professional life from our social life because the codes of conduct differ. In certain cultures, traits acceptable in social settings are deemed inappropriate in professional settings due to different normative and cultural expectations. Adopting this understanding of domains and domain-individuation reveals an important sense in which many of us are compartmentalized agents across at least some domains of life.

However, the compartmentalization that occurs in Tony's life is of a special kind, and we can build on Badhwar's account to flesh it out. Characterizing a domain of life in terms of practical concerns, at least sometimes, gives rise to an asymmetry between the individuation of different domains and the individuation of practical concerns relevant to those domains.² For example, even though we distinguish between our personal life and our professional life, the practical concerns in these two domains are not always completely isolated from each other. Indeed, we can have the same practical concerns, such as caring for others in interpersonal relationships, in both domains of life. One may care for their partner and their colleagues in different ways, and the degree to which one cares for them may vary. For example, some may express care more reservedly with colleagues, while certain family histories may make it harder for others to care for family than for friends. Nevertheless, caring for others remains an overarching practical concern from which caring for different people across these domains derives. However, if one's care is so narrowly focused that it extends only to their family, and their attitudes and behaviors differ drastically between domains such that they become hostile toward those outside their family, then they seem genuinely compartmentalized in their care for others across these domains. This helps explain the compartmentalization between Tony's family and criminal lives: it results from a drastic divergence in how (and indeed even whether) the practical concern about how he treats other people is realized across these two domains. Moreover, the incongruity in his attitudes and behaviors across these domains is patterned and systematic—he never stops making attempts to protect his family, nor does he stop striving to be the worst nightmare of his enemies.

To be sure, there are alternative ways to interpret Tony's attitudes and behaviors that do not render him compartmentalized as I have formulated it. For example, perhaps he does not have an overarching practical concern of caring for others, but rather two narrower practical concerns: loyalty to his family and cruelty toward anyone who gets in his way. Indeed, these two concerns

² One way to approach this is by distinguishing between trait-relevant situations as "nominally convenient" domains and as "psychologically meaningful" clusters. See Duckworth & Tsukayama (2015: 397-398).

might cohere with each other—being cruel toward those who get in his way is one way in which he remains loyal to his family.³ In these cases, he would not be compartmentalized in the same sense that I suggest he is.

However, interpreting Tony in these ways leads to a further consequence. Consider his cousin, Junior Soprano—a primary family rival—who repeatedly does things that create obstacles for Tony in protecting his other family members. As a result, Tony cannot remain loyal to Junior and instead acts in a systematic way: he stays loyal to the other family members, but sabotages Junior whenever he can. In this case, we might say that he commits to an even narrower practical concern—loyalty to family members who do not interfere with his ability to remain loyal to others. The implication of detailing Tony’s situation goes beyond offering a compelling interpretation of a TV character. We can imagine other cases in which we continue to add contextual details about the circumstances under which an agent acts systematically, each corresponding to increasingly narrow practical concerns to which they are committed. But if we understand a practical concern this narrowly, motivational and behavioral incongruity would disappear from human life.

Nevertheless, one might think that acknowledging motivational and behavioral incongruity is one thing, while acknowledging the kind of compartmentalization I have identified is another. It could be argued that when we act incongruously with regard to a given practical concern across domains, it typically falls into one of three scenarios.

First, we may be divided in our pursuits or feel conflicted in our commitments regarding a given practical concern across domains of life because the conditions necessary to achieve these things simultaneously are not met. For instance, there may be a clash of schedules, a conflict of interests, or certain interpersonal constraints. In this case, once these constraints are removed, the compartmentalization disappears. Second, our dividedness and inner conflict regarding a given practical concern across domains may express ambivalence. Here, we are torn between two

³ I thank Dong An and Neal A. Tognazzini for suggestion here.

competing courses of action that both express and constitute our core identities, yet we cannot wholeheartedly choose one without compromising crucial aspects of who we are (cf. Frankfurt 1992; Coates 2017). In this scenario, we are compartmentalized when we cannot fully commit to either of the two competing courses of action, even though both are essential to our core self. Third, we might simply be suffering from self-deception, believing we are divided and conflicted while our behaviors suggest otherwise. In this case, no real compartmentalization occurs.

However, acknowledging the above cases does not preclude the existence of the special kind of compartmentalization I have identified. This is a form of compartmentalization that cannot be readily resolved by addressing certain contingent conditions. Nor does it involve an agent who is ambivalent about realizing a given practical concern across different domains. Instead, it involves someone who is wholeheartedly incongruous in their realization of the concern. While this compartmentalization may accompany self-deception, it does not arise from it.

Zeroing in on this specific kind of compartmentalization is not a new endeavor. We tend to see repeated incongruous behaviors as an indicator in evaluating someone's character. In virtue ethics, there has been extensive discussion about the extent to which the presence of incongruous behaviors relevant to a given virtue or virtues across domains indicates a lack of those virtues. If so, then ascribing the relevant virtues to the agent would be illegitimate. A prime example of this claim can be found in discussions about the situationist challenge to virtues. According to this challenge, social experiments have shown that people's character exhibition varies in response to situational variations. For example, rather than exhibiting stable traits of benevolence and compassion across situations, people are more likely to engage in helping behaviors when in a better mood, especially when that mood is induced by random rewarding incidents (Isen & Levin 1972). Alternatively, the presence of another person who does not show a tendency to help can influence whether one chooses to assist someone in distress (Latané & Darley 1970). These phenomena are in tension with the virtue ethical claim that to be truly virtuous, one's virtue-conforming behaviors should be

consistent across situations, further supporting the idea that when virtue-conforming behaviors are incongruous across domains, they disqualify the agent from possessing the relevant virtue.⁴

The phenomenon of compartmentalized virtuous agency, manifested in incongruous virtue-conforming behaviors, can be directly applied to the concept of unified virtuous agency.

Specifically, if the thought is that our virtuous agency is unified in the sense that possessing one virtue requires a virtuous agent to exhibit patterned and systematic virtue-conforming behaviors across domains,⁵ then the phenomenon of compartmentalization invites us to reconsider this understanding of virtuous agency—perhaps virtuous agency is less unified than previously believed, or fewer people are genuinely virtuous. This reveals a tension between compartmentalization and a unified conception of virtuous agency, which, in turn, invites us to reexamine the conditions for the ascription of virtues: When an agent's exercise of a virtue is incongruous, resulting in a disunified virtuous agency, they are not virtuous regarding that virtue.

In the following sections, I will argue that a similar tension arises in understanding and theorizing about responsibility, particularly for those responsibility theories that require unified responsible agency. Specifically, I will examine the position of deep self views of moral responsibility on this topic and argue that, given their requirement for unified responsible agency, they cannot offer a diagnosis of Tony's responsibility without undermining the plausibility of these views.

3. The Requirement for Unified Responsible Agency

Let me first offer a general formulation that covers the various positions within standard deep self views.

⁴ For a comprehensive survey of what situationism is in social psychology and the development of various challenges from it, see Miller (2014: ch. 4).

⁵ This is one implication of some versions of the thesis of the unity of the virtues, which I consider most relevant to my discussion here. For a survey of different versions of the thesis of the unity of the virtues, see Wilson (2021).

Generic An agent is responsible for what they do and all that they bring about if and only if their actions issue from, and so express, their deep self, as characterized by certain elements in their psychological constitution.

Deep self theories like Harry Frankfurt's (1971), Gary Watson's (1975, 1987), and Michael Bratman's (1997, 2004, 2005) are captured by this generic formulation. Frankfurt argues for an account of free will in terms of an agent's hierarchical desires. According to him, an agent's deep self is characterized by their second-order volition about their first-order motivations; that is, that which they desire to desire for themselves expresses who they truly are. So, they act freely and are morally responsible for what they do if and only if they act from a first-order desire that is licensed by a second-order desire. In comparison, Watson's account of free agency suggests that an agent's valuational system is an important source of their motivation, and they act freely and responsibly if and only if their actions issue from, and so express, what they value. Bratman offers an alternative picture on which an agent's planning for self-governance over time is constitutive of their deep self; the plans and the self-governing policies that they carry out over time make them who they are. So, an agent acts freely and responsibly if and only if they act in accordance with their plans and self-governing policies.

There are also hybrid theories inspired by the above views, which too are captured by *Generic*. Deep self theorists such as David Shoemaker (2015) and Chandra Sripada (2016) contend that the deep self should be understood in terms of a disjunction of the motivational and valuational elements of an agent's psychological constitution; namely, their caring and evaluative commitments. Thus, an agent acts freely and responsibly if and only if their actions issue from, and express, either elements of their motivational make-up, or of their evaluative commitments, or of both. More recent developments of the deep self view contend that the deep self is identified, through narrative explanation or upon reflection, with elements of an agent's psychological

constitution. Examples of this position can be found in Benjamin Matheson's (2019) and August Gorman's (2019, 2022) accounts.

Before I proceed, let me state an important qualification. In this paper, I focus on deep self views of moral responsibility in which an agent is *accountable* for their behaviors in the sense that renders them liable to blame or praise.⁶ However, not all deep self theorists claim to argue for accountability. While traditional deep self theorists are more concerned with the purchase of an agent's deep self on the grounding of accountability (Frankfurt 1971; Watson 1975, 1987; Bratman 1997, 2004, 2005), more recent developments of the deep self view tend to focus solely on responsibility in the *attributability* sense (Shoemaker 2015; Sripada 2016; Gorman 2019, 2022). According to these views, an agent is attributionally responsible for their behaviors just in case they are properly attributable to their deep self, which requires standing in the right relation with it.

This shift of focus from accountability to attributability is well motivated. Traditional deep self views have been widely criticized through counterexamples: surely, one can be held responsible for actions that do not express who they truly are, and acting from one's deep self may not render one accountable if, for instance, one was covertly manipulated. Nevertheless, in my discussion, I am exclusively concerned with responsibility in the accountability sense because it is less obvious, and thus more puzzling, whether a compartmentalized agent can be held responsible for what they do for which they are attributionally responsible, particularly when their incongruous behaviors align with different elements of their psychological profile. Indeed, according to deep self views of attributability, we can attribute incongruous behaviors to compartmentalized agents as long as these behaviors align with the psychological elements that arise from within who they truly are. However, whether or to what degree they can be held accountable given such compartmentalization remains unsettled. Therefore, if my challenge to *Generic* in the accountability sense based on the problem of compartmentalization is successful, it will offer a new criticism of such views.

⁶ Although this paper will focus only on accountability for wrongdoing, and thus on the sense in which it renders blameworthiness, I will leave it as an unfinished philosophical project to further explore whether the same implications can be drawn for praiseworthiness.

Generic suggests that responsible agency requires that 1) an agent has a deep self, and 2) their behaviors cohere with their deep self in a certain way. Although it is not always clear when they are presented, I suggest that deep self views should *not* be taken as claiming that a responsible agent must have a *singular* deep self, and that every time they act freely and responsibly, they must act from that deep self *as a whole*. Few would be responsible for much if that were the case. This is because, in reality, our mental life is multifaceted with respect to our diverse and disparate commitments, projects, and relationships, and it is often in flux, lacking perfect coherence. Indeed, deep self theorists should allow the deep self to be domain-specific. For example, someone may have a deep self in the domain of health but not one in the realm of aesthetics. They may have practical concerns about health matters, which are expressed through the psychological elements defined by deep self theorists, but none related to the arts. Within the domain of health, they may focus solely on physical health, excluding mental health. In this case, they have a deep self regarding physical wellness.

Things can get more specific. Suppose our agent, who presumably cares about physical health, is a hopeless vaper and has resolved to quit vaping, but nonetheless let go of their resolution every time they are at a social event. Given this detail, we might say that they have a deep self regarding physical health *when they are alone* but lacks such a deep self in that regard *when they are at a social event*.

Things can get even more specific. Suppose that, due to their resolution to quit vaping, our vaper got rid of all their vaping kits. When they vape at social events, they borrow vapes from others. However, they do not just grab any stranger's vape because according to them, "even *I* do not sink that low!" For health and hygienic reasons, they only share vapes with friends. Considering these more fine-grained elements of their psychology and the contexts in which they act, we might say that although they lack a deep self regarding physical health *when vaping in general social settings*, they do seem to have a deep self *when vaping in social functions with their friends*. These

details help capture their various practical concerns, which correspond to different deep selves manifesting in their patterned behaviors. However, if we zoom in even further, we might discover more fine-grained psychological elements, potentially leading to increasingly narrow practical concerns regarding physical health, and particularly vaping, and an infinite number of deep selves that are expressed by them. This further results in a paradox that they both have a deep self with regard to vaping and do not have one. But if someone's deep self regarding a practical matter is allowed to be so fragmented in the deep self picture that such a paradox arises, it is reasonable to question whether the concept of the deep self is still useful for understanding them and their behaviors.

To avoid this paradox, we can understand a given deep self—focused on certain practical concerns—as unified through its manifestation across relevant domains. But why must this unification be realized across domains? This can be supported by the fact that a practical concern is often constituted by normative constraints embedded in a particular set of domains and is carried out through various dimensions related to those domains. Acting in accordance with a practical concern, therefore, implies acting within those constraints and through those dimensions, all of which are tied to specific domains.

Take vaping as an example again: it is a leisure activity with both personal and social dimensions. A vaper may vape privately or in public. They purchase supplies from others and often connect with fellow vapers. Thus, when a responsible agent has a practical concern regarding physical wellness—particularly in relation to quitting vaping—they act in response to the norms and expectations of both their personal and social lives. By choosing to quit and successfully quitting, the agent not only stops vaping on a personal level but also discontinues certain social activities associated with it. A deep self in relation to quitting vaping is therefore unified through an agent's actions that express with this concern across relevant domains. This unification is realized in patterned and systematic behaviors, while allowing for occasional lapses. Indeed, sometimes,

divergent behaviors are one-off occurrences, and deep self views should not be so demanding as to disqualify an agent's responsibility due to one-off ill-performances. So, if our vaper consistently finds themselves using their friends' vapes at social events, despite resolving to quit and successfully abstaining when they are alone, they do not satisfy the unification between a given practical concern and practical-concern conforming behaviors across relevant domains.

So, to reasonably interpret the deep self views, we should grant that there be a certain level of specificity to a given deep self for the views to retain their core. However, the deep self must not be subject to unlimited fragmentation. That is, an agent may have different deep selves in different domains of life and regarding different practical matters, and the relevant practical concerns could be specified by more fine-grained psychological dispositions and normative constraints within or across relevant domains. Nevertheless, given the coherence requirement in deep self views, there must be a level of unification between any given practical concern and practical-concern conforming behaviors across relevant domains.

Let me now state that the most plausible understanding of deep self views is to recognize their commitment to a more refined conception of unified responsible agency. According to it, a responsible agent must act in accordance with a given deep self, which is expressed by certain practical concerns, and do so in a patterned and systematic way across relevant domains of life, allowing for occasional, one-off non-conforming behaviors.

Understanding deep self views as committing to this refined conception of unified responsible agency has its merits. It is less demanding than understanding the deep self as a singular whole, and it avoids cases where the deep self is seen as subject to unlimited fragmentation. Additionally, it aligns with the idea that to a significant extent, we *are* what we *do*, especially what we *do a lot*: our character and who we truly are emerges from systematic and patterned behaviors across various domains of life.

4. *The Diagnosis of a Compartmentalized Tony Soprano*

Recall Tony's compartmentalization: he cares about the welfare of his family, but he could not care less about the pain or danger experienced by others (caused by him) simply because they are his enemies. This shows that a practical concern about caring about others is present in his psychological constitution. However, his target is too narrow—it includes his family but excludes his enemies and their families. The manifestation of this concern is restricted only to his family life and does not extend to his criminal life. According to deep self views, which commit to a unified conception of responsible agency, this compartmentalization results in Tony having a disunified responsible agency, and thus at least diminishes his responsibility for what he does.⁷

To illustrate this point, consider Gopal Sreenivasan's (2008) discussion of *holes* in virtuous agency (609-611). Sreenivasan suggests that the phenomenon of compartmentalization regarding virtues can be best characterized by an agent's having holes in their virtuous agency. Specifically, if an agent exhibits patterned and systematic non-virtuous behaviors regarding a particular virtue in some domains of life, but demonstrates virtuous behaviors concerning the same virtue in other domains, then we say there is a hole in their virtuous agency.

Depending on the moral status of the virtue-nonconforming behaviors, there can be two kinds of holes in virtuous agency. If an agent exhibits patterned and systematic virtue-nonconforming but *non-egregious* behaviors regarding a particular virtue, then they have a *blindspot* in their virtuous agency with respect to that virtue. For example, an agent might show compassion toward a friend going through a tough time but show no compassion toward a colleague

⁷ For instance, Benjamin Matheson (2019) proposes that an agent is morally responsible only if they act from certain psychological elements, grounded in a coherent psychological set that they *own*, wherein such coherence is provided by an ideal narrator's narrative explanations of aspects of their life. Given this view, a compartmentalized Tony Soprano may still be responsible, albeit *less* so, since the ideal narrator may be able to offer only a partially intelligible story for his compartmentalized behaviors across relevant domains. Similarly, other deep self accounts may appeal to a concept of responsibility that admits of degrees in order to explain Tony's diminished responsibility. Might a deep self theorist judge that Tony's responsibility is *undermined* such that he is *not at all* responsible? I think so. Therefore, to clarify and for reference, when I say that deep self theorists judge a compartmentalized agent to be *less responsible*, I also mean to include cases where some of them hold that such an agent's responsibility is undermined to the point that they are *not responsible at all*. However, for simplicity, I will set this aside in the main text. I am grateful to an anonymous reviewer for suggestion here.

in similar circumstances. In this case, they have a blindspot in their compassion. Conversely, if an agent displays patterned and systematic virtue-nonconforming behaviors that are *egregious*, they have a *blackspot* in their virtuous agency regarding that virtue. For instance, if the same agent not only fails to show compassion but actively expresses animosity toward their distressed colleague, they have a blackspot in their compassion. Sreenivasan then argues that having a blindspot regarding a virtue does not disqualify an agent from possessing that virtue—they remain virtuous, albeit defectively. However, having a blackspot regarding a virtue does disqualify them from possessing that virtue—they are not virtuous in that respect at all.

To be sure, virtuous agency and responsible agency are not the same. Although one may act both responsibly and virtuously, acting virtuously is not required for responsibility, and one may act virtuously without being responsible. For instance, someone might protect a stranger solely for monetary reward—an action for which they are responsible—without acting from the virtue of courage. Or consider someone coerced into protecting a friend from greater harm by hitting them in the face; they might exhibit the virtue of courage without being held responsible for the harmful behavior.

Within the deep self framework, a deep self need not be virtuous: there may be practical concerns—expressive of certain deep selves—that are not connected to any virtue. In such cases, an agent may act from a deep self without acting virtuously, or act from a virtue without acting from any deep self, and thus would not be responsible according to the deep self view. Therefore, having unified virtuous agency does not imply having unified responsible agency, just as disunified virtuous agency does not imply disunified responsible agency.

Nevertheless, the distinction between egregious and non-egregious behaviors in virtue ethical discussions is helpful for understanding Tony's actions: as a compartmentalized evildoer, he commits egregious acts against his enemies in his criminal life, while displaying caring attitudes and behaviors in (most of) his family life. In turn, the distinction between blindspots and blackspots

can shed light on his responsible agency. We tend to sense that something is *more seriously wrong* with an agent who does something *really, really bad* compared to one who does something *bad, but not really, really bad*. The idea is that whatever explains an agent's exceptionally wrong behaviors might significantly diminish or even undermine their responsibility. Consider Susan Wolf's (1987) infamous Jojo, who grew up idolizing his dictator father, an evildoer, and ultimately became an evildoer himself. Jojo's upbringing was not responsive to objective moral standards, leaving him without the basic capacities for moral responsibility. Arguably, he is not a suitable candidate for responsible agency. Or consider Neil Levy's (2007) Billy, who developed into a cruel criminal partly due to a brain tumor that interferes with his basic moral responsibility-conferring capacities, and for that reason does not satisfy the threshold for responsible agency.

To further illustrate, applying Sreenivasan's distinction between blackspots and blindspots, a compartmentalized Tony exhibits a blackspot in his responsible agency. This blackspot is evident in his patterned and systematic egregious behaviors concerning certain practical concerns in some domains of life but not in others, which also shows that a unification between his behaviors and certain practical concerns across relevant domains is not satisfied. Given the deep self views' requirement for unified responsible agency, I suggest that their diagnosis of Tony is as follows:

- (1) A responsible agent must have a unified responsible agency in the sense that if they have a deep self that is expressed by certain practical concerns, then they must act in accordance with them across relevant domains of life in a patterned and systematic way.
- (2) Tony Soprano issues patterned and systematic practical concern-nonconforming behaviors that are egregious in some domains of life but not in others, and thus exhibits a blackspot in his responsible agency.

- (3) Having a blackspot in his responsible agency means that he does not meet the required level of unification between some of his practical concerns and systematic, patterned behaviors across relevant domains, and thus he lacks unified responsible agency regarding those practical concerns.
- (4) Therefore, his responsible agency regarding those practical concerns is at least diminished. That is, he is less responsible for actions regarding those practical concerns across relevant domains.

So, our deep self theorists are committed to judging that when Tony protects his family from his enemies' threats, he is less responsible for his actions. Similarly, when he threatens his enemies to harm their families, he is less responsible for that either. This is due to the fact that, in his compartmentalization, his responsible agency is disunified, and thus diminished.

There are different ways to challenge these verdicts. For one, they seem highly counterintuitive—surely, he can be fully responsible for some of his behaviors, albeit not for others, despite the compartmentalization. Or he can be fully responsible for all his behaviors regarding his treatment of his family and his enemies. In what follows, I elaborate one novel criticism of the deep self views, and argue that their diagnosis is problematic for the following reason: although their reasoning is valid, the diagnosis comes at the cost of undermining the plausibility of such views. Specifically, concluding that Tony's responsibility for his evildoing is diminished extends to concluding that *all* compartmentalized wrongdoers are less responsible for their wrongdoing. This conclusion is a hard bullet to bite, especially for deep self theorists.

5. *The Diagnosis of All Compartmentalized Wrongdoers*

The key move in the diagnosis of Tony is the idea that patterned and systematic *egregious* behaviors diminish his responsibility in relevant domains of life. However, this move may be less available to

deep self theorists than it appears. I argue that the reason Tony is less responsible in the deep self picture lies not in the egregious nature of his actions but in something else. This reason will turn out to support the claim that all compartmentalized wrongdoers—whether as extreme as Tony or less so—are less responsible for their wrongdoing.

To proceed, let us reconsider the different implications of having blindspots versus blackspots in one's virtuous agency. Sreenivasan claims that having blackspots disqualifies someone from possessing the relevant virtue, and I have suggested that this facilitates deep self theorists' diagnosis of Tony. Nevertheless, there is a relevant challenge to Sreenivasan's approach that should not be overlooked. Specifically, the distinction between having a blindspot and having a blackspot may not translate into different implications for virtuous agency; both may imply the disqualification of a compartmentalized agent from possessing the relevant virtue or virtues. Nancy Schauber (2014) advances such a challenge. Her argument aims to preserve the Aristotelian picture of virtuous agency in the face of the phenomenon of compartmentalized virtuous agency. On the Aristotelian picture, she says, virtuous agency requires that a virtuous agent act from a certain psychological constitution—they must have a grasp of what the virtuous thing to do in a situation is and be properly motivated to act accordingly. Schauber argues that neither having blackspots nor having blindspots satisfies this criterion. This is because the occurrence of patterned and systematic virtue-nonconforming behaviors indicates that the agent does not act from the necessary psychological constitution, and so whether or not the behaviors are egregious is beside the point.

This is insightful because deep self views have a requirement for responsible agency that is of a similar spirit with Aristotelian virtuous agency—acting freely and responsibly requires that an agent have a particular psychological constitution and act from it in a certain way. More specifically, as I proposed, deep self theorists are committed to a unified conception of responsible agency, wherein a responsible agent must act from their deep self regarding certain practical concerns across domains of life in a patterned and systematic way. The core of this requirement is

structural, and it is manifested in behavioral patterns. So, in the deep self picture, Tony's compartmentalization indicates that he does not meet the necessary condition for unified responsible agency because his compartmentalization is structurally divergent from unified responsible agency, and this is *not* because he has a blackspot in his responsible agency. The issuing of patterned and systematic *egregious* behaviors is one feature of his diagnosis, not part of its explanation.

Now, if we accept that Tony's diminished responsible agency in the deep self picture is due to a structural divergence that is manifested by his compartmentalization, the same reasoning can be applied to other compartmentalized wrongdoers who are not as bad as Tony; they too would be deemed less responsible for their wrongdoing because their compartmentalization indicates that they do not meet the necessary condition for unified responsible agency either. So, given the deep self views' requirement for structurally unified responsible agency, the assessment of Tony's diminished responsibility commits deep self theorists to making the same judgment about any wrongdoer who is compartmentalized in a crucial sense that Tony is.

6. *An Achilles' Heel in the Deep Self View*

Two approaches can be advanced to challenge the conclusion that both compartmentalized Tony and those who are not as bad as him are less responsible for their behaviors. Regarding the claim that there is no relevant difference between them, one can certainly defend the sentiment that there is a meaningful difference between a wrongdoer who commits egregious acts and one who does not, and that this difference has significant implications for their responsible agency. For instance, endorsing a historical dimension in the necessary condition for responsibility could support this sentiment. Arguably, compartmentalization that accompanies egregious wrongdoing might indicate a personal history the existence of which diminishes an evildoer's responsibility, whereas compartmentalization that accompanies ordinary wrongdoing does not imply the same kind of

history. Specifically, if, during an agent's personal history, factors beyond their control prevented them from being given the opportunity to develop or exercise the necessary responsibility-conferring abilities, they would not meet the requirements for responsible agency.⁸

If we have good reasons to believe that an evildoer like Tony has experienced something in his personal history (or failed to experience something) that prevents him from accessing opportunities to develop or exercise certain abilities necessary for responsibility, for example, the ability to do otherwise, or the ability to respond to what is morally right—thereby explaining his evildoing—then we could argue that he is at least less responsible for what he does. On the other hand, if we have good reasons to think that an ordinary wrongdoer typically does not have a personal history in which they were deprived of such opportunities, then there is a principled difference between the evildoer and the less severe wrongdoer. In this case, the latter would be responsible, or more responsible, than the former.

However, standard deep self theories are famously not history-sensitive in two key senses. First, in a number of deep self views (Frankfurt 1971; Watson 1975; Sripada 2016; Gorman 2019), what matters for moral responsibility is whether an agent identifies with certain elements in their psychological constitution, and such identification is a-historical. According to those views, were an agent to undergo covert manipulation at a certain point in their personal history, they would still act freely and be morally responsible for their actions, as long as they identify with the newly adopted set of psychological elements as a result of manipulation, and act in accordance with them. Thus, whether compartmentalization that accompanies egregious behaviors indicates a personal history that differs from one indicated by compartmentalization that accompanies ordinary wrongdoing does not create a relevance difference between these two groups of wrongdoers. Consequently, if Tony is less responsible for his behavior due to his compartmentalization within the deep self picture, then an ordinary wrongdoer is likewise less responsible.

⁸ For discussions on the relevance of personal history to responsibility and how certain personal histories might diminish or undermine responsibility, see Mele (1995), Haji (1998), Watson (2004), McKenna (2012), Cyr (2023), and Zhang (2024).

Second, while there are deep self views that account for an agent's diachronic identity in assessing responsibility (Bratman 1997, 2004, 2005; Matheson 2019) and can explain why a manipulated agent is at least less responsible than one who is not manipulated, these views do not require that an agent must have been afforded, at some point in their personal history, certain unimpeded opportunities to develop or exercise the abilities necessary for responsibility. This means that only certain features of an agent's personal history are relevant to their responsible agency, such as the absence of manipulation, and thus, these features create a distinction between an evildoer and an ordinary wrongdoer. In this case, there is not a relevant difference between a non-manipulated Tony and a non-manipulated wrongdoer who is less severe. Then, if a non-manipulated Tony is less responsible for his behavior within the deep self picture, a non-manipulated wrongdoer would likewise be less responsible. Therefore, once again, these deep self views lack internal resources to account for a principled distinction between egregious and non-egregious wrongdoing from a historical perspective.

Alternatively, one might embrace part of the conclusion drawn from the structural requirements of deep self views and agree that there is no principled difference between an egregious evildoer and a wrongdoer who is less bad, yet arrive at the opposite verdict—that both are fully responsible for their behaviors. Consider McKenna and van Schoelandt's (2015) discussion, which could support this verdict. In evaluating Frankfurt's concept of a wanton—someone who does not adopt any second-order desires about their first-order desires regarding a practical matter and is thus excluded from our assessment of their responsibility—they argue that one can be a wanton in certain areas of activity while adopting second-order desires in others (50-52, 58). For example, an agent might evaluate and govern their motivations and sources of action when it comes to discipline and work productivity in the way that is freedom and responsibility-conferring in the deep self picture, thus indicating their ability for such self-governance, yet fail to do so when it comes to addictive vaping. In this case, they could still be considered free and responsible when

they vape and fail to exercise their ability to be more moderate. This is because a wanton may be able to adopt and act on higher-order preferences over their own motivations but fail to do so due to their own negligence, therefore satisfying the minimal requirements for responsible agency.

Tony can be understood in this light: with respect to caring about others, he is a compartmentalized wanton. His care for his family indicates his ability for having and acting on the concern, yet he fails to exercise that ability and extend the same care to his enemies and their families due to his own faults. Thus, he is both responsible for protecting his family and for causing harm to his enemies. Similarly, an ordinary wrongdoer who is compartmentalized might also be a compartmentalized wanton, and therefore, be considered fully responsible in the same way as Tony. However, appealing to this approach requires an account of abilities necessary for responsibility that standard deep self views lack, therefore, they do not have the internal resources to sympathize with such a diagnosis.

These compelling, alternative diagnoses can only be defended in the deep self framework if deep self theorists can explain why the structural requirement for responsible agency is sensitive to non-structural features in our assessment of responsibility, such as a history dimension of responsibility or certain abilities necessary for responsibility. To do so, they would need to radically modify core features of their views. Or they would have to bite the bullet and accept the counterintuitive claim that all compartmentalized wrongdoers are less responsible for their wrongdoing. So, arriving at the judgment that all compartmentalized wrongdoers, including those who are not as severe as Tony, are less responsible for their wrongdoing undermines the plausibility of deep self views.

In conclusion, I have elaborated a novel criticism of standard deep self views of moral responsibility. I argued that deep self views commit to a unified conception of responsible agency, which faces a challenge from the phenomenon of compartmentalized responsible agency. To address this challenge without compromising their requirement for unified responsible agency, deep

self theorists would need to judge that all compartmentalized wrongdoers are less responsible. I have shown that this undermines the plausibility of their views.

Statements and Declarations

The author has no relevant financial or non-financial interests to disclose.

Acknowledgment

I am grateful to Dong An, Cong Chen, Davide Fassio, Xiaoyu Ke, Neal A. Tognazzini, Jason Turner, Sean Whitton, Yili Zhou, the participants of the Work-in-Progress Workshop at Zhejiang University organized by Dong An, and the two anonymous reviewers for their valuable suggestions, comments, and discussions on this paper.

References

- Badhwar, Neera (1996). The limited unity of virtue. *Noûs* 30 (3):306-329.
- Bratman, Michael E. (1997). Responsibility and planning. *The Journal of Ethics* 1 (1):27-43.
- Bratman, Michael E. (2004). Three theories of self-governance. *Philosophical Topics* 32 (1/2):21-46.
- Bratman, M. (2005). Planning agency, autonomous agency. In *Personal Autonomy* (pp. 33–57). Cambridge University Press.
- Coates, D. J. (2017). A wholehearted defense of ambivalence. *The Journal of Ethics*, 21, 419-444.
- Cyr, T. W. (2023). Why history matters for moral responsibility: Evaluating history-sensitive structuralism. *Philosophical Issues*, 33(1), 58–69.
- Duckworth, A. L., & Tsukayama, E. (2015). Domain specificity in self-control. *Character: New Directions From Philosophy, Psychology, and Theology*, 393-411.

- Frankfurt, Harry G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1):5-20.
- Frankfurt, H. (1992, November). The faintest passion. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 66, No. 3, pp. 5-16). American Philosophical Association.
- Gorman, August (2019). The minimal approval view of attributability. In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility* 6. Oxford University Press.
- Gorman, August (2022). Demystifying the deep self View. *Journal of Moral Philosophy* 19 (4):390-414.
- Haji, I. (1998). *Moral appraisability: Puzzles, proposals, and perplexities*. Oxford University Press.
- Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: cookies and kindness. *Journal of Personality and Social Psychology*, 21(3), 384.
- Latané, B., & Darley, J. M. (1970). Social determinants of bystander intervention in emergencies. *Altruism and Helping Behavior*, 13-27.
- Levy, Neil (2007). The responsibility of the psychopath revisited. *Philosophy, Psychiatry, and Psychology* 14 (2):pp. 129-138.
- Matheson, B. (2019). toward a structural ownership condition on moral responsibility. *Canadian Journal of Philosophy*, 49(4), 458-480. doi:10.1080/00455091.2018.1480853
- McKenna, Michael (2012). Moral responsibility, manipulation arguments, and history: Assessing the resilience of nonhistorical compatibilism. *The Journal of Ethics* 16 (2):145-174.
- McKenna, M., & van Schoelandt, C. (2015). Crossing a mesh theory with a reasons-responsive theory: Unholy spawn of an impending apocalypse or love child of a new Dawn? In *Agency, Freedom, and Moral Responsibility* (pp. 44–64). Palgrave Macmillan.
- Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy*. Oxford University Press on Demand.

- Miller, Christian B. (2014). *Character and moral psychology*. Oxford, GB: Oxford University Press.
- Schauber, Nancy E. (2014). How bad can good people be? *Ethical Theory and Moral Practice* 17 (4):731-745.
- Shoemaker, D. (2015). Ecumenical attributability. In *The Nature of Moral Responsibility: New essays* (pp. 115–140). Oxford University Press.
- Sreenivasan, Gopal (2008). Character and consistency: Still more errors. *Mind* 117 (467):603-612.
- Sreenivasan, Gopal (2009). Disunity of virtue. *The Journal of Ethics* 13 (2-3):195-212.
- Sripada, Chandra (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies* 173 (5):1203-1232.
- Stichter, Matt (2021). Virtues as skills, and the virtues of self-regulation. *Journal of Value Inquiry* 55 (2):355-369.
- Watson, Gary (1975). Free agency. *Journal of Philosophy* 72 (April):205-20.
- Watson, Gary (1987). Free action and free will. *Mind* 96 (April):154-72.
- Watson, Gary (2004). *Agency and answerability: Selected essays* (pp. 219–259). New York: Oxford University Press.
- Wilson, Alan T. (2021). Unity of the intellectual virtues. *Synthese* 199 (3-4):9835-9854.
- Wolf, S. (1987). Sanity and the metaphysics of responsibility. In F. D. Schoeman (Ed.), *Responsibility, Character, and the Emotions: New essays in moral psychology* (pp. 46–62). Cambridge University Press.
- Zhang, K. (2024). Bringing the deep self back to the racecourse: Rethinking accountability and the deep self. *Analytic Philosophy*, 00, 1–17. <https://doi.org/10.1111/phib.12354>