# Responsibility for Character Traits

## –

## A Character-Centered Framework of Moral Responsibility

**Inaugural-Dissertation**

zur Erlangung des Doktorgrades der Philosophie

der Ludwig-Maximilians-Universität München

vorgelegt von

Susanne Schlee

aus

Bamberg

2024

Erstgutachter: Prof. Dr. Christof Rapp

Zweitgutachter: Prof. Dr. Erasmus Mayr

Tag der mündlichen Prüfung: 25.05.2023

# Acknowledgements

I would like to take the opportunity to thank my supervisors Christof Rapp and Erasmus Mayr for enriching this thesis with your advice and guidance throughout the last years. Your feedback has helped tremendously in shaping this thesis. Furthermore, I owe thanks to Laura Valentini for sharing your ideas and agreeing to be part of the examination committee. Thanks is also due to Philipp Brüllmann for taking on the role as mentor and providing me feedback and motivation. I would also like to express my gratitude to George Sher, who has been so kind as to discuss early parts of this project with me over Zoom.

Thank you also to Monika Betzler for allowing me to be part of the Munich School for Ethics in Practice. The group discussions and shared experience have helped tremendously in staying on track.

Lastly, I want to say thank you to my family and friends, who have given me the support and love one needs for such a project.

# Contents

# Introduction

## 1. The Importance of Character (Responsibility)

Character is utterly important to us and our daily lives. Consider the following (more or less formal) processes in the legal realm, in politics, at work and in social life: in many judiciary systems the length of legal sentences depends on the offender's character reflected in previous records. As such, character is much more relevant in determining a sentence than we usually assume. Our judiciary system relies heavily on character references as it is used to indicate the probability of repetition of the crime (cf. Fileva 2017a, p. xxi). Offenders with a previous record get higher sentences or are more likely to be convicted as it reveals something about their character and the likelihood that they might repeat the crime or may not regret the previous one (cf. Rhode 2019, pp. 41–95). Formerly convicted thieves for example, accused of yet another theft, will struggle to convince the judge of their innocence, as past offences are considered predictive of future behavior. And despite character evidence being banned in many criminal justice systems, its importance cannot be overstated. Judges will still assess the offender's character and base their ruling on it. The German legal system, for example, includes preventive custody, a measure which is not part of the sentence but may be imposed on those offenders considered dangerous to society. The central factor decisive for the imposition of this legal force is essentially the offenders' character, i.e., whether or not she is likely to commit the crime again and the gravity of the offence (cf. Laubenthal 2004).

The importance of character also extends to working in certain professions. Many job opportunities require references of good character. These are essential

statements, which have great bearing on whether the applicant will be successful. Especially lawyers in the United States, trying to pass the bar, must face a jury in an interview and convince them of their character fitness. Only those, who demonstrate good character in this thorough interview process, will be admitted to working as a lawyer. Yet, jobs in the judiciary sphere are not the only ones that involve an inquiry into character. Basically, every type of job application and interview process illustrates how fundamental character is in shaping the view we have of each other. When we ask superiors or colleagues for references, we signal to our potential future employer that we will behave adequately (cf. Powell 1959, p. 502). These references may refer to past behavior but in fact are much more than the summary of previous actions: these statements aim to capture who we are essentially as a person - what we have done, and what we are likely to do (cf. Miller and Knobel 2015, pp. 32–33). Furthermore, companies now tend to employ those with a "personal fit." The interview process oftentimes centers not only on the capabilities and training needed for the specific job but also on the character features of the applicant. Questions such as "Please describe your strengths and weaknesses." reflect the implicit inquiry into the character of those interviewing for the job. Thus, given its centrality in the legal system and the work environment, character has a wide-reaching influence on our life opportunities in general:

"Employers and occupational licensing authorities assume that the moral character required for a particular position is a stable attribute that they can predict with reasonably (sic!) accuracy based on past conduct. (...) Prior misconduct, particularly when it results in a criminal record, is seen as predictive of future dishonesty, theft, violence, breach of fiduciary obligations, and/or lack of work ethic" (Rhode 2019, p. 42).

Further, character takes on a special role in politics as politicians are eager to demonstrate good character to gain their voters' trust. And many voters consider moral character important in political leaders. But display of good character in politicians is not only an instrument to secure the vote of constituents - the list of politicians who were forced to resign because their character was questioned adds constantly - it also has symbolic and practical value. Symbolically, moral character in

political leaders is important as it defines the degree of confidence the public has in politicians. Practically, good character may ensure steady and level-headed leadership in the face of great challenges (cf. ibid., pp. 96-102).

The importance of character in life is also reflected in the way that schools implement programs to cultivate good character in children to prevent "(…) the social ills - sexually transmitted disease, teenage pregnancy, drug abuse, crime, and so on (…)" (Doris 2008, p. 121). Many educational institutions in the U.S. have therefore integrated *character education* programs into their curriculum (cf. Chen 2013, p. 346). These programs are based on the firm belief that society is better off with members "(...) who are intrinsically and actively pro-social" (Althof and Berkowitz 2006, p. 496).[1] Depending on agenda and program, students are taught certain values such as trustworthiness, respect, responsibility, fairness, caring, and citizenship (cf. Character Counts! 2022).[2]

Moreover, character is not only crucial for institutionalized areas of life. We place a lot of importance on character in our social interactions as well. Accepting a friend's promise involves the belief that we can trust that she will in fact honor that promise in the future (cf. Cohon and D'Cruz 2017, pp. 215–227). When I believe Hannah's promise to pick me up for dinner at p.m., I have faith in her that she can now make a predictable statement about the future, which I in turn can rely on. Even more so, character not only helps us know whether we can trust one another, but also determines what kind of person our friend is. Character reveals what Hannah

---

[1] There have been different approaches to character development in the past, e.g., values clarification, cognitive development theory, and care ethics. Values clarification, a non-normative approach, which was prominent in the 60s/70s aims to clarify the student's values by asking a series of questions without imposing the teacher's values on the student. Cognitive development theory is a concept based on Kohlberg's moral development stages which engages students in moral reasoning by confronting them with moral dilemmas. Care ethics focuses on educating students in attachment, love, friendship, listening and the like (cf. Sanderse 2013, pp. 27–71). Today, character education is the leading approach, which aims to teach predefined values (cf. Althof and Berkowitz 2006, pp. 499–500).

[2] These values do not necessarily reflect (moral) character traits in the philosophical sense. Chapter 1 introduces a classification which clarifies the philosophical fundamentals of character traits.

cherishes and holds dear, it is part of her identity (cf. Goodwin et al. 2015, pp. 110–112):

"Without using concepts of specific character traits we would be unable to continue to talk as we do presently about our colleagues, students, relatives, friends, lovers, and pets. We would also be unable to understand most films, novels, short stories, plays, and biographies. We constantly give common sense explanations of behaviour in terms of character traits" (Butler 1988, 215).

Viewing my friend Francine as just, means that I believe her to have the character trait of justice, explaining her past behavior and predicting that she will make just decisions in the future. Further, I assess her normatively, thinking that this trait makes her a good person. Thus, character has explanatory and predictive power, and functions as a normative assessment (cf. Miller and Knobel 2015, pp. 32–33).

Because character has these tremendous implications for our life, we care whether we are responsible for our traits. However, despite these grave consequences we have diverging intuitions to what extent we really are responsible. Some tend to think that we are responsible for our characters because it is up to us who we end up being. This view reflects our reaction to news of cases in which an offender has murdered several people after having tortured the victims for time on end. Horrific stories such as these raise serious questions about the crime but also the character of the offender: "How can someone be such a bad person?," we find ourselves thinking. Hearing of agents who are capable of such despicable crimes often evoke questions of fair punishment. Delinquents deserve severe punishment for these crimes, the reasoning goes, because they control who they are. But then again, we might come to think that we are not responsible after all. As soon as we learn of the shocking circumstances of the murderer's upbringing - the extreme physical and psychological abuse - we tend to shift some of the blame on the environment, for example the parents or the deprived community: "Our questions become: To what extent is [*s*]*he* responsible for becoming the sort of person [s]he now is? Was it *all* a question of bad parenting, societal neglect, social conditioning, and the like, or did [s]he have any role

to play in it?" (Kane 2007, p. 7). And the more we think about it, the question becomes inevitable of whether we have any control over our characters at all. For it seems that we all are a product of our environment and the circumstances that we happen to grow up in.

The conflicting intuitions we have about our character illustrate that there is a social interest to reflect on this question. That is because the practices we engage in - the legal consequences, job opportunities, political success and downfall, as well as social standing - better be well-grounded if they are to entail such grave consequences. If agents are not responsible for their characters, it stands to question what justifies these practices (cf. Coates and Tognazzini 2012, pp. 197–198).

## 2. Theoretical Developments

The question whether we are responsible for our character is certainly not new. On the contrary, research on character responsibility is extensive. Various authors have addressed the question, ranging from ancient philosopher Aristotle to contemporary authors who specialize on a variety of issues such as virtue ethics, moral psychology, and metaphysical questions of moral responsibility. Despite the difference in philosophical backgrounds, proponents and opponents usually base their assessment of whether agents are responsible on the question of whether they have control. Those who argue in favor of character responsibility mainly rest their argument on the claim that actions, which in turn habituate character, are voluntary, thus giving character a notion of control (e.g., cf. Aristotle 2002; Jacobs 2001). Responsibility-skeptics, on the other hand, question whether we do in fact have the required amount of control for building our character. The main support for this view lies in an infinite regress-argument, concluding that we are not *ultimately* morally responsible for our character, as we do not constitute ourselves (cf. Strawson 1994, p. 7). Taken seriously, the infinite regress takes place when considering that our choices, which habituate character, are

always based on the current state that we are in (cf. Cyr 2019, p. 206).[3] Hence, responsibility is restricted as our environment is a matter of chance and we lack self-constitution. Advocates of character responsibility do not take these worries to be damning, though. They employ several types of arguments to restore the argument that we do in fact have the relevant kind of control over our character. Some authors argue that we develop our character through prior choice, while others suggest that we can at least endorse the character we have. Others again maintain that agents are responsible by means of critically reflecting themselves. Even if we might not have been able to choose who we want to be in the past, future actions shape our character, which we can control (cf. Russell 2009, p. 379; 388).

But why revisit the question when there is such extensive research on the topic already? From a general viewpoint one motive to engage in the debate once more is that there is obviously not much agreement, and the question remains whether agents are responsible for their character. Further, despite (or, maybe because of) the quantity of authors who have discussed responsibility for character, some authors rather imply or simply assume responsibility for character. They seem to defend such an implicit responsibility-positive account in stating that "(…) character is not destiny" (Altshuler 2016, p. 5).

But there is another, rather more pressing reason to reconsider the arguments exchanged. A closer look reveals that existing analyses are one-sided and leave out relevant developments in research. This gap unfolds in three aspects: first, many of the analyses mainly rely on an understanding of character that stands in the virtue ethical tradition (cf. Jacobs 2001; Russell 2009)[4] or, more often, is undefined but resonates with a folk psychological conception of character. Both variations operate under the term *Global Traits* because, despite some differences, they consider character traits broad

---

[3] Some have also questioned the extent to which agents can be responsible when the prerequisites for character may be partly genetic (cf. Rhode 2019, p. 9) or "endowed by nature" (Chen 2013, p. 352). See chapter 5 for a brief consideration of the matter.

[4] There is also considerable contribution from Aristotelian scholars who provide interpretations of his approach (for example cf. Bondeson 1974; Meyer Sauvé 2011).

features of agents, which are stable over time and cross-situationally consistent (cf. Miller 2003, p. 368; Doris 2008, pp. 22–23). However, two contiguous developments have since complemented and complicated the debate on character. On the one hand, there are considerable (empirical) contributions from psychology, which shape the debates carried out in both fields. But as psychology has long been conflicted about character and been devoted to personality research (cf. Fleeson et al. 2015, p. 41; Miller et al. 2015, pp. 2–5), many contemporary discussions lack clarity with respect to their research object. Further, based on the influence of empirical psychological findings (e.g., Milgram 1963; Isen and Levin 1972; Darley and Batson 1973) and the resulting fundamental critique, called *situationism*, two other approaches, Local Traits (Doris 2008) and Mixed Traits (Miller 2014), have added to the Global Traits view. These conceptions of character depict traits as narrower or more fine-grained than the Global Traits account. This in turn limits what agents can control and be aware of. Yet, most analyses neglect the diversity in the character theory landscape and how they affect character responsibility.[5]

Second, not only do these analyses rely on a specific theoretical concept of character, by focusing on whether agents have sufficient control, the literature is narrowed down to a specific conception of moral responsibility, volitionism. Volitionism is a position that holds that responsibility requires that the agent has control and awareness (for example cf. Fischer and Ravizza 2000; McKenna 2008; Nelkin and Rickless 2017). However, this view has been challenged and a variety of theoretical approaches add to volitionism now. These other positions can roughly be grouped as attributionism (for example cf. Smith 2005; Talbert 2009; Arpaly 2015) and capacitarianism (for example cf. Sher 2009; Rudy-Hiller 2017; Clarke 2017b). Both negate that control and awareness are necessary conditions for responsibility, yet they differ with regards to how they explain the agent's responsibility in absence of these

---

[5] There is some fair amount of research on the implications of situationism on responsibility (e.g., Brink 2013; Doris 2008; McKenna and Warmke 2017; Miller 2019; Nelkin 2005; Vargas 2013b). However, these publications are primarily concerned with how responsibility for *actions* is affected by the possibility of situations playing a bigger part in explaining behavior than character.

conditions. This is particularly interesting for analyses of character responsibility as not only these alternative character conceptions potentially limit what agents can control and be aware of, character traits are not necessarily developed consciously. Hence, analyses of character responsibility must consider these alternatives and how they alter the arguments exchanged.

Third, past efforts to discuss character responsibility have done so with an implicit focus on agents' responsibility for the actions which habituate character (cf. Jacobs 2001, p. 2; Chen 2013, p. 353; Fileva 2017b, p. 194). The arguments in favor or opposition of character responsibility center on the question whether agents meet the conditions of responsibility necessary for *actions which in turn habituate character*. However, this neglects the ambiguity of the research question. To ask whether agents are responsible for their character can not only refer to this latter interpretation, it can also ask whether agents are responsible for having a certain kind of character (cf. Sher 2001, pp. 148–149).[6] Discussions of this latter sense of the question are rather rare exceptions (e.g. Sher 2006a). Since many analyses of character responsibility focus on only one of these questions, there is strong reason to also examine the other one. Additionally, there is further reason as to why discuss both variations of the question: the general question of moral responsibility often pops up with regards to actions, i.e., whether agents are responsible for what they *do*.[7] And even though it is not always explicitly mentioned, responsibility for actions seems to presuppose responsibility for character traits (cf. Brink 2013, p. 121). That might be because actions are often considered to derive from character.[8] In fact, influential philosophers such as David Hume have argued that responsibility for action requires that it stems from the agent's character (cf. Sher 2006a, p. 17; Kauppinen 2017, pp. 46–51). Hence, if agents are not responsible for having the character traits they have, it seems that we would need a

---

[6] Actually, it can also mean whether agents have an obligation to develop character. I will address this shortly and argue that this refers to another kind of responsibility not at issue here.

[7] Chapter 3 provides an outline of the most prominent theoretical strands encountered in debates on moral responsibility.

[8] See also Fileva 2017b, pp. 182–202, who discusses the connection between traits and reasons for actions.

separate explanation for the responsibility agents have for their actions. Hence, more generally, these three points reveal that there remains much to be said about the topic.

# 3. Research Question and Scope

This thesis aims to address these three aspects of the research gap identified. The one-sided focus on a globalist conception of character combined with a volitionist point of view leaves open room to explore what alternative models can tell us about character responsibility. On these grounds, debates about responsibility for character ought to consider analyzing how these new or alternative models affect responsibility, while also taking into account the two notions of the research question.

Based on these observations, we are now in a position to specify the research question as follows: how can we understand character responsibility considering what volitionist, attributionist, and capacitarian theories tell us about Global Traits, Local Traits, and Mixed Traits? Thus, I plan to bridge the gap and engage in a novel discussion of character responsibility considering the two interpretations of the research question and a diverse range of theories at hand. With a more detailed understanding of character and responsibility, I believe to enhance the question of character responsibility.

Nonetheless, in broadening the scope and thus exploring further options to discuss character responsibility, some preliminary remarks are in order: first, the type of responsibility at issue is moral responsibility. Moral responsibility differs from other forms of responsibility (cf. Talbert 2023). The news anchor may speak of the thunderstorm being responsible for the destruction of several housing units, a cat is responsible for knocking over an antique vase, or the sunburn on my back is responsible for the ache I feel. On this reading, responsibility is a purely causal concept. Roughly said, it *explains* what has caused the event in question (cf.

Zimmerman 2015, pp. 45–46).[9] Now, imagine that the building complex did not collapse due to environmental causes but that the civil engineer of the building complex had made a horrible arithmetic mistake. Or that it was not a cat but your friend who were to smash the vase against the wall in a heated argument that got out of control. Or that it is not the sunburn that causes your backache, but you were hit by a car which requires you to receive medical treatment. In these altered scenarios the civil engineer, your friend, or the reckless driver of the car respectively are responsible for what happened. This time, however, the type of responsibility in question is not only causal but also moral as we take fundamentally different attitudes towards moral agents than we take towards inanimate things and non-humans (cf. Wolf 1993a, p. 7). For example, we would certainly feel resentment or anger towards the civil engineer, our friend, and the car driver as opposed to the thunderstorm, the cat, or the sun. Indeed, moral responsibility is the essence of our differentiation between treating agents as persons instead of objects (cf. Strawson 2008). Consider again the cases of the thunderstorm, which has caused damage to family homes, the cat that has knocked over the vase, or simply the sun that has burnt your back. While we might feel anger or frustration at first, it would truly be odd to *morally* blame the thunderstorm, the cat or the sun respectively in the "elaborate" sense we would blame the civil engineer, your friend or the car driver (cf. Wolf 1993a, p. 7; Fischer and Ravizza 2000, p. 1). Even more so, we would not be able to *forgive* the thunderstorm, the cat or the sun as they simply bear no *moral* responsibility (cf. Smith 2007, p. 474). We can now see that while causal responsibility serves as an explanatory function, moral responsibility is a power held only by a moral agent. While causal responsibility describes a mere causal connection, moral responsibility includes a moral agent's capacity. Natural phenomena, animals, inanimate objects, adults with mental disabilities do not have the same capacities as morally responsible agents. Even though they may be causally

---

[9] I will refrain from any further elaboration on the complicated topic of causation. Here, I only take it to mean that something is the cause of some consequent event in the most basic sense.

responsible, they may not have moral responsibility (cf. Wolf 1993a, pp. 41–42). The type of responsibility I will consider in this thesis is moral responsibility.

Second, this thesis considers responsibility in the backward-looking sense. Another familiar way of speaking of responsibility is to say that someone *has a responsibility* to do something. It may be Iris' responsibility to take out the trash or Rachel's responsibility to watch the dog while her parents are out. Both Iris and Rachel have responsibilities in the sense that we expect them to attend to certain matters. However, their responsibilities correspond to sorts of obligations or duties. There would be no change in meaning if we swapped the terms responsibility and obligation in this context. However, responsibility as an obligation does not correspond to the responsibility, we ascribe to someone as a mental capacity. Iris' obligation to take out the trash differs from the responsibility that we ascribe to the civil engineer who is responsible for the collapse of the building. While the former describes the obligation that is associated with certain expectations to be met in the future (namely that Iris will take out the trash), the latter describes a capacity held in the past or currently that grounds social practices such as blame (namely that the civil engineer may be held accountable for miscalculating the building) (cf. Zimmerman 2015, p. 46). Both meanings of responsibility are easily conflated not only because they share linguistic similarities, but also because someone might have a responsibility for which we consider her responsible once she has (not) carried out said responsibility. The civil engineer's responsibility (obligation) may have been to present a well thought out and calculated concept to build a safe housing unit. The failure to do so is now grounds for the responsibility we ascribe to her. The difference between the cases (even though there is some overlap) is that Iris' obligation refers to a forward-looking type of responsibility, whereas the civil engineer's responsibility may be viewed as backward-looking. There are two reasons to focus on backward-looking responsibility in this thesis. For one, most of the work discusses the sense of responsibility as backward-looking. More importantly, however, the forward-looking notion of responsibility would alter the research question fundamentally. It would be to ask whether we have

an *obligation to develop* certain character traits. However, as mentioned before, this thesis discusses the question of whether agents are responsible for the character traits that they *have (developed)*. These are two distinct questions and thus need to be kept separate. Therefore, I will not pause to analyze *responsibility as an obligation* in more detail. From now on, I will take responsibility to mean responsibility in the backward-looking sense.

Third, traditionally, the debate on moral responsibility focuses on the kind of responsibility agents have for their *actions*.[10] This necessitates two clarifications: one, despite the previous focus on actions, recent discussions highlight that agents can also be responsible for omissions. In fact, the theoretical development in the debate on responsibility is driven by a shift to those cases in which agents unwittingly omit something (cf. Clarke 2017a; Sher 2017). The reason for this is the attempt to carve out those cases which are much more common than intentional actions but are not covered (or ruled out) by standard theories of responsibility. For purposes of simplicity, however, I will continue to speak of actions but imply the case of omissions if not indicated otherwise. Moreover, the focus on actions in responsibility theory naturally invites an objection to the aim of this thesis: it might be objected that since the theories of moral responsibility generally give accounts of agents being responsible for an action, the intended analysis cannot tell us anything fruitful about the responsibility had for *character*. Even worse, the conclusions drawn from this examination might be wrong. Nevertheless, there are two reasons why it is still sensible to investigate what these theories have to say about character. To begin with, none of the theories excludes character traits. On the contrary, even though only few passages indicate that their view also applies to character, there is a paragraph in which prominent contributors to the responsibility debate John Martin Fischer and Mark Ravizza explicitly say so:

---

[10] See this exemplary writing by R. Jay Wallace: "People who are responsible may be made to answer for their *actions*, in the sense that their *actions* render them liable to certain kinds of distinctively moral responses [italics S. Sch.]" (1998, p. 54).

"(…) someone is responsible *for a particular bit of behavior* (or perhaps a *trait of character* [italics S. Sch.]) to the extent that [s]he is an appropriate candidate for at least some of the reactive attitudes on the basis of that behavior (or *trait of character* [italics S. Sch.])" (2000, p. 6).

While the majority of cases that Fischer and Ravizza offer for consideration certainly concern actions, there are some passages in their writing in which they clarify that it includes character. And even though another contemporary, who will be of importance to this thesis, George Sher, neither provides a framework solely for character, he analyzes character responsibility (for bad character traits) in *In Praise of Blame* (2006a). More importantly, though, it is necessary to stress that to date there is no theory specific to responsibility for character traits. This is reflected in the fact that all existing analyses are based on action-centered theories. Hence, by broadening the scope of the theories considered, we engage in a novel discussion. So *even if* none of the theories turn out to be fully suitable to analyze character, it makes sense to examine them from a research perspective, nonetheless. Based on this preliminary analysis, we may be able to determine whether the current landscape of theories suffices or whether character as a variable requires further clarification. For these reasons I take the theories to provide a sufficient basis for analysis.

## 4. Chapter Overview and Theses

To give an answer to the research question and thereby address the research gap identified, this thesis consists of two parts. The first part, comprised of chapters 1-3, sets the stage. Here, I introduce essential concepts. Thereby, these chapters help illustrate the main debates and different views on character and responsibility. Part two, which is comprised of chapters 4 and 5, progresses to analyze what the theories outlined in part one can tell us about character responsibility and establishes a character-centered framework.

# Introduction

In pursuing the question of character responsibility, I argue for the following main theses: 1. Character is relatively more complex than actions and thus requires independent consideration. Therefore, I develop a character-centered framework which addresses this complexity. 2. With the exception to those who are manipulated, agents can be responsible for their character even when they lack control over their traits and the formation thereof and only have minimal awareness. 3. Building on this, I conclude that agents are generally responsible for *having* the character they have because their traits are attributable to their quality of will as they have a general idea of what they are like. Further they can be responsible for *developing* their traits because they can know that their actions develop their traits. This, I maintain, is independent of the character conception. Lastly, even though agents are generally responsible for their traits, some factors such as cultural membership and deprived childhood may diminish an agent's character responsibility. Further, since all agents vary to the degree that they satisfy the relevant conditions, I claim that 4. character responsibility comes in degrees and some agents might be less responsible for their character traits than others. In fact, some might be barely above the minimal threshold.

I develop these and other ancillary theses throughout the chapters, which are structured as follows: the first part starts with a classificatory approach to character. **Chapter 1** is a direct response to the first aspect of the research gap mentioned, namely that analyses on responsibility for character traits lack a clear notion of character. The resurgence of interest in character from various academic disciplines, especially psychology, has led to many contributions but also to much confusion about the terms personality and character. To approach the question of character responsibility meaningfully, it is necessary to explicate what is meant by character. Chapter 1 does so by arguing that previous attempts to differentiate character from personality fail because they either cannot explain why character is commonly associated with morality (e.g., the identity view) or the differentiation proposed lacks a comprehensible view of the relation between character and personality (e.g., the dichotomy view). A plausible classification of character, I maintain, should first

14

explicate the structural relation between both concepts. Second, it should elaborate on the conditions which distinguish character from personality. Third, it should unravel the association of character and morality. Incorporating these criteria, I essentially follow Christian Miller (2014) and propose that we ought to understand character traits as kind of personality traits, namely those that agents are normatively assessed by. Further, I suggest that there are moral and non-moral character traits (cf. Kupperman 1995; Miller 2014). Moral character traits are those that agents are morally assessed by (cf. Miller 2014).

Additionally, I provide an argument of why we ought to understand character traits as dispositions. A discussion of the majority view, *dispositionalism*, I maintain - following Maria Alvarez (2017) - reveals that character traits share some of the features of *paradigmatic* dispositions but differ with respect to manifestation and thus are not paradigmatic dispositions. Some paradigmatic dispositions lose their power as soon as they manifest, character traits do not. Not only that, I argue for the even stronger claim that character traits *must* manifest (cf. Alvarez 2017, p. 79). Therefore, I explore an alternative view to the dispositions view, which states that character traits are summaries of trait manifestations (cf. Brandt 1970; Buss and Craik 1983). Nonetheless, I reject this view because it runs contrary to the conceptual assumption of character traits, which is that agents have them even in absence of presently or having recently manifested them. Ultimately, I suggest that since character traits must have manifested in the past but are also real properties of the agent, making character traits a kind of disposition, namely manifesting dispositions.

**Chapter 2** follows on from the previous one and is motivated by the advances in character theory in recent years and thus also directly addresses the first aspect of the identified research gap. It provides an overview of three philosophical theories on character (Global Traits, Local Traits, Mixed Traits) which simultaneously reflects the historical development of each. The illustration in chronical order establishes the similarities and differences between the three views explored. Global Traits are depicted as a family of views of character that conceive of character as broad (stable

and cross-situationally consistent) dispositional features of the agent. Local Traits have emerged from a profound critique known as situationism. On this view proposed by John Doris, character traits are also stable dispositional features yet are not cross-situationally consistent but highly localized and situation-dependent (cf. 2008). Mixed Traits, as suggested by Christian Miller, are conceptualized as interrelated fine-grained dispositions whose manifestation is dependent on triggers, yet are stable and cross-situationally consistent, nonetheless (cf. 2014). In terms of content, all theories share the view that character traits are stable dispositions which are habituated by the agent. However, they differ with respect to their structural set-up. Character traits on the Global Traits view are coarse-grained; they are narrower in the Local Traits model; and the Mixed Traits account tells us that character traits are fine-grained.

**Chapter 3** shifts the focus from theories of character to theories of responsibility and thus addresses the second aspect of the research gap in the literature. As most analyses of responsibility for character utilize a volitional account, this chapter broadens the view of responsibility. The chapter identifies three strands of theories in responsibility debates: volitionism, attributionism, and capacitarianism. Volitionists maintain that agents need control and awareness, attributionists argue that an action must be attributable to the agent and is so when it is reflective of the agent's quality of will. More recently, contemporaries, who embrace capacitarianism, have investigated in which way agents can be responsible without any awareness. Volitionists usually take a *tracing* position which holds that agents must have some awareness or in the case of not knowing currently, the bad act must be traceable to a benighting fully knowing act. Anti-tracers, i.e., attributionists and capacitarians deny this (cf. Rudy-Hiller 2018; Talbert 2023). From each strand I select a theory that is representative of the respective family of theories. The joint work of John Martin Fischer and Mark Ravizza (2000) present one of the most prominent views in the debate. Their volitionist account establishes a condition for responsibility they call *guidance control*. Agents satisfy this condition when they take responsibility and the action in question issues from their moderately reasons-responsive mechanism.

Matthew Talbert (2009, 2013) represents attributionist thought. He maintains that for an agent to be responsible, the action needs to reflect her evaluative judgment. For this to be true the agent only needs circumstantial, yet not de dicto awareness. George Sher's (2006a, 2009) account is chosen as representative of capacitarianism. He stresses the connection between the agent's self and the action. Agents are responsible even for omissions when they could or should have been reasonably expected to be aware and the action's wrong-making features are suitably connected to the agent or when agents act unwittingly rightly despite having made enough cognitive contact with the action's right-making features.

**Chapter 4** explores what the theories of responsibility by Fischer and Ravizza, Talbert, and Sher can tell us about the character conceptions (Global Traits, Local Traits, Mixed Traits) introduced. But to do justice to the third aspect of the research gap identified - the lack of acknowledgment of the ambiguity of the research question -, I first provide a distinction between direct and indirect responsibility for character traits, which each relate to one interpretation of the research question. Direct responsibility for character traits concerns the agent's responsibility for having the character she has. Indirect responsibility for character traits inquires whether agents are responsible for their character as a result of being responsible for the actions that habituate their character. Further, I observe that character adds complexity in two further aspects. On the one hand, in contrast to action, character is much more multifaceted. All theories of character affirm that it is made up of many traits (even though each varies with regards to the level of *structural complexity*). On the other hand, the question of character responsibility requires investigating not only the responsibility had for the action but *dimensionally* extends to character, meaning that an analysis must consider whether agents satisfy the relevant conditions of responsibility for the actions that develop character.

After having clarified these issues, I continue with an analysis of each responsibility theory by applying the differentiation between direct and indirect character responsibility. A thorough discussion of each theory yields different results.

## Introduction

Fischer and Ravizza's volitionist account contrasts the question behind direct character responsibility and thus yields no results. However, their view accommodates indirect character responsibility for coarse-grained accounts (Global Traits). Nonetheless, their volitionist view entails a relatively strong epistemic condition and thus raises the question whether they can also do so for Local Traits and Mixed Traits. The joint control and epistemic condition is also the very reason they excuse agents when they are not aware or in control, yet character traits are not always developed consciously or are under an agent's control. While Talbert's attributionist structurally fits direct character responsibility, his view takes manipulated agents to be responsible, a controversial claim, which I ultimately argue takes too many agents to be responsible. Sher's capacitarian account negates that agents are directly responsible for their character but is generally more open for indirect responsibility for coarse-grained views. Yet the complexity of more fine-grained views also prevents his account to show how agents are indirectly responsible for their Local or Mixed Traits, even though it is constructed to accommodate those cases in which agents lack control and awareness.

The analysis, I maintain, exemplifies that character is relatively more complex than actions. Most importantly, I articulate the thesis that due to the complexity, character responsibility requires that character is addressed as a distinct phenomenon, whilst paying heed to the three challenges that emerge from the discussion: first, it must differentiate adequately between the two forms of character responsibility and specify the relationship between them. Building on this, since neither theory analyzed gives a positive, unrestricted, account of *direct* character responsibility, an analysis of character responsibility must illuminate how we can conceive of a viable version of this variant of the question. Second, especially volitionists and attributionists take opposing views on the importance of the agent's acquisition of her values for the assessment of her current psychological setup, entailing a difference in opinion on cases of manipulation. Therefore, it must delineate in what way the agent's history is essential to her responsibility attribution. Third, the discussion illustrates that all

theories of responsibility render agents not responsible for their narrow or fine-grained traits as these are much harder to be aware of, yet traits may not always be developed consciously. Based on this, they differ with respect to the class of agents they consider responsible. Thus, acknowledging the fact that character introduces more complexity to the question of responsibility, one must provide insight into how epistemic requirements translate to character considering these specifications.

**Chapter 5** is devoted to addressing the main thesis and challenges raised in chapter 4 by presenting a character-centered framework for responsibility. I begin by setting up the two conditions of character responsibility. The view that I develop aims to show that agents can be responsible for their character even when they lack control and only have minimal awareness. Regarding the control condition, I suggest that it is best conceived of as a defeasible voluntariness condition. This condition ensures that agents are not manipulated and thus have the right kind of history without resorting to the concept of control. Regarding the epistemic condition, I argue that agents only need minimal awareness of their character and its development. Generally then, agents can be responsible for their traits and the development thereof even when they lack control and only have minimal, more general awareness, when it reflects their quality of will. In what follows, I contend that agents are directly and indirectly responsible for their character traits. This may also apply to agents with certain cultural memberships or a deprived childhood. These agents are not necessarily exonerated as they must be differentiated from manipulated agents. Hence, they can be responsible for their character. Notwithstanding this assessment, I acknowledge that these agents do not exactly resemble "normal" agents with no such history. Building on this, I introduce the degrees-view, which holds that agents vary with the degree to which they satisfy the relevant conditions and thus the degree to which they are indirectly responsible. Cases of cultural membership and deprived childhood then reflect the way that despite fulfilling the basic conditions, some agents only barely meet the minimal threshold. Thus, on this proposal, the indirect character responsibility of agents with such a history is not *generally* called into question, yet it

can seriously diminish it. Further, this view explains cases of undeveloped agency and the fact that all agents vary with regards to what they know and control. Hence, I conclude that agents are responsible for their character traits, yet to a varying degree.

# I The Concept of Character

## Classification, Delineation and Fundamentals

---

## 1. Introduction

Character is both elusive and informative: we seem hard pressed to grasp it, but it tells us what kind of a person someone is, what her good and her bad features are. We use the word *character* so commonly in our everyday language that we usually do not think twice what it really means. It has taken the place of a blurry concept that has various connotations depending on context. For example, we might say that someone is "quite a character" (Brandt 1970, p. 24), inferring that she is a special type of person, bordering on being eccentric. Or, we might stress that someone "has character" when she has demonstrated especially integer behavior, we deem virtuous (cf. Brandt 1970, p. 24; Rhode 2019, p. 4). On the contrary, however, saying that someone "has *no* character" does not mean that they are bad but that they are simply unreliable and easily swayed by temptation. In yet another sense, we also use character to distinguish between individual differences (cf. Kupperman 1995, pp. 6–7; Cohen and Morse 2014, p. 45; Rhode 2019, p. 3). Having character indicates the possession of certain qualities that are distinct for that individual. Especially this latter form of usage seems to resemble the origin of the English word *character*, which stems from the Greek *charaktêr* meant to describe the impression upon a coin (cf. Frow 2016, p. 7; Homiak 2019). For example, while Hannah seems like the honest kind, Judith has proven to be a rather compassionate person. Judith might have habits that demonstrate her compassion such as volunteering at an animal shelter every Wednesday. When we think of Hannah as honest, it might be due to her strong identification with the trait of honesty, which she upholds not only by telling the truth regularly but also by advising her friends

truthfully when facing a predicament. Thus, character on this view - character as the distinctive feature of an individual[11] - is identical to personality (cf. Rhode 2019, p. 3). And this is not too far from how we usually use both terms interchangeably. Nonetheless, character may have moral meanings that personality does not have. When we speak of Francine as having a good *personality*, we usually mean that she is charming. Saying that Hannah has good *character* because she is very honest, however, indicates that she is a morally good person (cf. Kupperman 1995, pp. 3–5). Thus, on closer reading, character seems to refer to the realm of morality while personality does not.

The academic focus on character is traditionally located in philosophy but other disciplines such as psychology are also attentive to the subject matter. However, psychological research has been somewhat conflicted about character. Originally, there was profound interest, but the beginning of the 20th century marked a shift as several experiments fueled doubt about whether people actually had character traits. Further concerns regarding the subjective notion of character led some to refrain from dedicating more extensive research on the topic. Hence, it is only of recent that psychologists have begun to redirect their attention to character again (cf. Miller 2014, p. 9; Fleeson et al. 2015, p. 41; Goodwin et al. 2015, pp. 101–102; Miller et al. 2015, pp. 2–5).[12] The varying interest is mirrored in the fact that most psychological research still operates under the term *personality theory* and some use the term personality almost interchangeably with character (see for example Cohen et al. 2013, pp. 817–818). More lately, however, the latter practice has changed and psychologists are starting to quite consistently identify character with notions of morality (cf. Cervone and Tripathi 2009,

---

[11] While common speech indicates that we use personality and character to illustrate the differences among people (cf. Funder 1991, p. 33; Johnson 1997, p. 74), two people could theoretically have the same trait structure (cf. Miller 2014, p. 6). In addition to differentiating *between* persons, character may help to differentiate *within* a person to describe the variability of her trait related behavior (cf. Fleeson 2007, pp. 828–829; Cervone and Tripathi 2009, pp. 32–33).

[12] For example, psychological research contributes to general trait research (non-character specific) in various facets such as using empirical experiments to test for traits (e.g. Fleeson 2001; Cohen et al. 2013; Kalimeri et al. 2013).

p. 30; Fleeson et al. 2015, p. 41). Still, parallel developments in philosophy and psychology in the last century have led to imprecisions in the research area. Even though today there is a growing number of joint interdisciplinary research, "(…) there is nothing approaching consensus about how to use the terms 'character traits' and 'personality traits'" (Miller 2014, p. 9). And even though philosophers generally adhere to the term character and research dates back to as long as Aristotle, there, too, has been little effort in past years to *structurally* specify how character traits differ from other types of traits.[13]

In addition to the issue of terminological imprecision, there is considerable disagreement in the research area concerning the nature of character traits. The dispute derives from the question of how to tell if an agent *has* a certain trait. Studies show that we typically speak of others using conditional statements (cf. Wright and Mischel 1988, pp. 454-456; 465). So, if we think of Hannah as honest or Judith as compassionate, we explain and predict their behavior using if-then constructs. For example, *if* Hannah is asked by her friend whether she likes their new pants *then* Hannah will tell the truth. Similarly, *if* Judith sees a homeless person, *then* she will give them some money. Thus, we see character traits as tendencies (cf. Newman and Ulman 1989, p. 168) that explain what a person *would* do (cf. Wright and Mischel 1987, p. 1161). The academic research, however, has yet to come to an agreement to what character traits are. Philosophers often pose that character traits are a property that is causally effective in giving rise to thoughts and behavior (cf. Miller 2014, p. 23). This is contrasted with the (primarily psychological) view that character traits are essential summaries of trends in behavior (cf. Buss and Craik 1983, p. 105).

The shared interest in character by philosophers and psychologists has brought advances to the research field. Yet, structured approaches to character remain rare and there is much confusion about how to classify character (and thus differentiate it from personality). Many of the contemporary debates between

---

[13] One of the exceptions to this is Christian Miller (2014), who provides an attempt at a classification that differs from previous views. His classification is the starting point for this chapter.

psychologists and philosophers on these issues remain unresolved because no common conceptual basis has been created.[14] The lack of clear distinction has led to much confusion and, as many authors diverge in their understanding and usage of the terms, past research was stuck in debates that could have been avoided or at least solved by a better conceptualization of them. Instead, it seems, terms such as traits, dispositions, character, and personality are at the center of these discussions without a shared meaningful interpretation.[15] It is thus not hard to see why it is necessary to first clarify these concepts and how they relate to one another. Due to the dual interest, precise analyses of character must take into account psychological approaches and findings. Therefore, the chapter is structured as follows: first, I delineate a classification of character traits which digresses from other views held such as the identity view or the dichotomy view, which are the leading - albeit somewhat implicit - approaches. The former, I argue, does not substantiate why there is a specific term for character traits and thus what the conditions are to differentiate character traits from other types of traits. Further, it fails to make sense of the fact that character is often associated with morality. The latter does not sufficiently explain the relation between character and personality. I hold that a sensible classification should accommodate for three requirements. First, it should name the condition that constitutes character. In doing so, it should, second, elaborate on why character traits are commonly associated with morality. Lastly, it should explain the structural relation between character and personality. Incorporating these three criteria, I argue in line with Christian Miller (2014) that character traits are a specific kind of personality trait, namely those that agents are normatively assessed by (cf. p. 15). Further, I maintain that character traits

---

[14] Additionally to the questions of the moral valence of character and the dispute on trait attribution, other unresolved questions pertain to how character change happens and what kind of personality traits are needed as a precondition for a character trait (cf. Clement and Bollinger 2016, pp. 174–181). Refer to chapter 5 for some comments on the latter issue.

[15] This is particularly evident in a hotly contested debate on the validity of a fundamental criticism called *situationism*, which will be briefly introduced in the subsequent chapter. The problem arises partly because the arguments do not rest on the same conceptions of character. For an analysis of the different concepts at use see Elliott 2017, especially chapter 4, which illustrates the "(…) deep differences between virtues and psychological traits" (ibid., p. 11).

can be moral and non-moral (cf. ibid., pp. 32-34). The kind of character traits of interest for this thesis are moral character traits.

Second, I outline why it is sensible to think of character traits as dispositions. One reason that character traits are standardly characterized as such is because they seem similar to what is referred to in the debate as *paradigmatic* dispositions. A closer comparison of the two, however, reveals that character traits differ with respect to the feature of manifestation (cf. Alvarez 2017). The alternative summary view, favored by psychologists, affirms the claim that character traits must manifest. Nonetheless, I argue that we ought to reject this view because it fails to account for the fact that agents possess their traits also when they do not manifest them currently or openly, meaning that they are properties of agents. Since the features of character traits converge in large part with dispositions, yet are dependent on *past* manifestation, I endorse a version of dispositionalism, which holds that character traits are *manifesting dispositions*.

## 2. Character Trait Classification

### 2.1 The Identity and the Dichotomy View

One approach to character and personality is to state that they overlap significantly. This view, which we may call the *identity* account, conceives of both concepts as substantially similar. For example, prominent psychological concepts such as the Big 5[16] encompass both personality traits and character traits. Thus, the identity view refrains from differentiating between both terms. On the contrary, character and personality, on this view, are somewhat interchangeable. This is not to say that this view excludes the possibility that there are differences between the terms. But *if* there

---

[16] The concept is best known as Big 5 but is now called HEXACO as a further, sixth, trait adds to the previous five. The traits included in this model are honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, openness to experience (cf. Cohen and Morse 2014, p. 50).

are differences, then these are nuances that are irrelevant to describing the concept. And it is not hard to see that the identity view resonates with the fact that personality and character traits both are *mental* traits as they are: "(…) concerned with the mental life of a creature, i.e., the mental states and processes that constitute thinking" (Miller 2014, p. 4). As mental traits they form a certain class of general traits. There are many different forms of general traits that we know of, e.g., someone's height or their weight. Francine can have the trait of being tall, while Christine is a rather heavy person. These types of traits are not genuinely human as a car might possess the trait of being fast or an animal might have an oval shape (cf. ibid., pp. 3-5). Neither are these general traits necessarily mental. Hence, one reason to think that character and personality are identical is that they are both united by belonging to the mental sphere. But there is further reason that supports the identity view. For see how both are dispositions[17] to have specific thoughts, beliefs and desires, and act respectively (cf. Cohen and Morse 2014, p. 45; Miller 2014, p. 7)[18].[19] Hence, on the identity view, it could be argued, the relation between character and personality is that of sameness. The conditions that ground character would simply be the same as those that mark personality traits from other types of traits, i.e., that they are mental traits. The main advantage of the view is that it accounts for the intuition that character and personality are closely related. It also reflects how both terms are often used interchangeably.

However, because the identity view holds that the conditions and structural relation between character and personality are identical, it gives no sufficient explanation why there are two terms for the same concept. Additionally, it does not reflect the way in which character is commonly associated with morality. If character and personality traits are of the same class and share the same conditions, there is no

---

[17] See section 2 of this chapter for a discussion of why character traits are dispositions.

[18] Note that Miller refers to personality traits while Cohen and Morse speak of moral character. Notwithstanding the slight differences, the descriptions are very similar.

[19] While some traits involve beliefs, desires, and respective actions, not all need to. For example, being logical requires no bodily actions, and being generally anxious may not involve any actual belief about the appropriateness of being anxious (cf. Miller 2014, pp. 3–8).

telling why character seems to differ with respect to morality. Hence, the view cannot be used as a classification for this thesis because it neither provides a condition by which to tell both terms apart nor does it elucidate the connection between character and morality. However, it seems that a plausible classification that can be employed here should account for exactly these things: first, it should elaborate the specific conditions of character that differentiate it from personality. Second, it should explicate why character traits are often associated with morality.

Another approach to classifying character, which reflects these criteria, is the *dichotomy* view. The dichotomy view stresses that character traits are indeed different from personality traits as they are morally valenced (cf. Kupperman 1995, p. 5; Bleidorn and Denissen 2015, p. 700). This approach differentiates between character and personality traits in accordance to their moral valence (for a similar differentiation see for example Doris 2008, p. 19; Chen 2013, pp. 353–354; Fleeson et al. 2015, p. 42; Rhode 2019, p. 3). Most plainly, this view holds that character traits are mental traits like personality traits but differ with respect to their moral component. And there is good reason to endorse this view. In contrast to the identity view, it provides us with a condition of how to differentiate between various mental traits. At the same time, it explains why character traits are commonly associated with morality: on the dichotomy view, character traits are *moral* mental traits. Thus, it can simultaneously show how both are separate kinds of mental traits and account for the specification of morality. This view comes in handy for those, who take character to be the prime subject of philosophers because it deals with questions of morality.

However, the dichotomy view, despite it being a very common assumption, might not be the best fit for a classification of character traits for this thesis either. The view conceives of character as morally valenced, while personality is not. More specifically, the view suggests that character and personality are not identical and that there is in fact a difference. The difference, it claims, is the moral content and thus character traits form a different class. But, while it does give a condition for identifying character traits and simultaneously explains why character traits are so commonly

associated with morality, it is unclear just how to understand the structural relation between personality and character traits. The view does not explicate how these two classes interact. Do they share a generic term other than both being mental traits? And if so, how can we conceive of them as two separate entities? This latter argument of course is no reason to reject the dichotomy view. We could imagine a specification of the dichotomy view, which clarifies the structural relationship between character and personality, yet maintains its condition. Nonetheless, this view leaves the question of how to structurally conceive of both concepts unresolved. It then seems that a plausible classification that can be reasonably employed here should account for a third criterion: it should exemplify the structural relation between the two terms.

## 2.2 An Alternative Classification

### 2.2.1 Character Traits as a Kind of Personality Trait

The consideration of the two views introduced highlights that a classification that can be utilized for this thesis should account for three criteria. To account for these criteria, I follow a more recent alternative to the identity and dichotomy view proposed by Christian Miller (2014). I adopt his view with regards to the structural classification. But I digress slightly with regards to the conditions that support this classification.

Regarding the structural relation, Miller suggests that character is a *kind* of personality trait (cf. ibid., p. 9). Character and personality do belong to the same family of mental capacities; however, they are not identical, nor are they in dichotomous opposition. Rather, character traits are a *specific kind* of personality trait. Following this approach, not every personality trait is a character trait. This view helps explain why character traits and personality traits are both mental traits but gives a more satisfactory answer to how they are structurally related. Like on the identity view, personality and character are intimately linked. But like on the dichotomy view, they are separate. That being said, I do not claim that this view ultimately trumps the others,

but it does avoid the ambiguities contained in the two concepts presented before. With this structural classification we can visualize the linkage between the two concepts. Both are mental traits but are not identical.

But how can we tell whether a personality trait such as being fair, honest, kind, greedy, or vile[20] (cf. ibid., pp. 4; 8; 21) is also a character trait, i.e., what is the condition that differentiates character from personality? Miller, considers three strategies to differentiate between character traits and other non-character personality traits: the first possibility to assess whether a personality trait is a character trait is that the latter involves some kind of normative judgment on *behalf of the agent possessing* the respective trait. For example, Hannah, who possesses the trait of honesty, then is thought to make judgements concerning the moral appropriateness of telling the truth in various situations. However, Miller quickly rejects this approach as he believes that it might be too intellectual. The reason being that we can think of Hannah being honest without making certain types of normative judgements about telling the truth being appropriate. In fact, we often encounter people, who intuitively tell the truth without having made said judgement in the first place. It could simply be that they are honest because they were asked a question (cf. ibid., pp. 10-11). Yet, if the assessment is dependent on the agent's judgment, this means that the agent determines whether a trait is a character trait. Even more so, I think that building on this argument by Miller, there is further reason to believe that this strategy does not help advance a profound categorization. To elaborate more on this worry: a judgment on behalf of the agent possessing the trait is not only intellectual, it is also arbitrary as agents might have normative judgments about all of their mental traits, not only their character traits. For example, Julia might believe that her talkativeness is as beneficial as her courage. The same might be true for any belief that Julia holds (e.g., if Julia thinks that 2+2 is 5, she might think that this is correct). Thus, the perspective of the agent provides us with

---

[20] Further examples of personality traits are being forgiving, just, compassionate, loving, nefarious, understanding, courageous, talkative, expansive, artistic, dry, jovial, formal, clever, optimistic, tense, calm, nervous, extraverted, shy, sociable, imaginative, logical, witty, modest, brave, adventurous, and humble (cf. Brandt 1970, p. 23; Miller 2014, 4;8;21).

little information about the traits other than her personal assessment and therefore does not serve as a suitable condition. Hence, I agree with Miller that this strategy does not help us in classifying character traits. But while Miller is uneasy that this strategy ultimately depends on the agent, I additionally worry that the strategy simply gives us no condition at all. That is because taking this strategy at face value, we resort to the identity view as there is no telling a difference when agents assess both their character and non-character personality traits. In conclusion, no matter whether Miller's worry about this strategy being too intellectual or my additional point about it being arbitrary is striking, both worries illustrate that this strategy does not help differentiate between personality and character.

Miller offers two further strategies for consideration to differentiate between character and personality. First, he proposes normative responsibility[21] as a distinct feature of character. On this view, character traits are the specific traits, which agents are responsible for. And, agents are responsible if they are an appropriate candidate for the reactive attitudes: "Honesty, for example, can be a character trait (…) if the honest person in question is someone who is responsible for being honest and is (thereby) a fit candidate for praise and respect" (ibid., p. 12). Genetically acquired personality traits[22], Miller maintains, cannot count as character traits as they lack the element of control.[23] We can easily see how this reflects our intuition about toddlers and children (cf. ibid., pp. 11-12). Children may have certain personality traits, which are, however, not yet character traits. Think of three-year-old Francine, who stands out in her creativity when playing with other children in kindergarten. If normative responsibility - as envisioned by Miller - is the decisive element to tell whether this artistic trait is a character trait, it seems that it is not.

---

[21] Miller maintains only normative kinds of responsibility (including legal and epistemic responsibility) qualify, not causal or role responsibility for instance (cf. 2014, p. 11).

[22] Miller is not quite specific about this. However, it is suggested that some personality traits are genetically acquired while others, e.g., character traits, are not (cf. 2014, p. 12).

[23] Miller notes that he favors a volitional account of responsibility, requiring control (cf. ibid., 12 footnote). Chapter 3 gives an overview of different positions regarding the conditions of moral responsibility.

However, this condition, like the first one, is of little use considering the goal of this thesis. As elaborated in the introduction, most analyses of character responsibility rely on an understanding of responsibility that conceive of control as the main condition to determine an agent's responsibility. This shortcoming to consider other theoretical offers, however, is one main motivation to evaluate what different theories of responsibility can tell us about responsibility for character traits that do not necessarily utilize control as their conditional feature.

In foresight of the argument that one can have a character trait even without control, Miller takes a cautious stance on the applicability of his proposal but argues that this strategy can still accommodate the complication:

"For those who do have the intuition that the person could still have the trait of honesty in this case [even if the agent has no control; S. Sch.], technically speaking the proposal (…) need not challenge this. For there can be 'honesty the character trait' and 'honesty the non-character trait.' Whether in a particular case a person's trait of honesty gets to count as a character trait or not will depend on facts about normative responsibility" (ibid., p. 14).

In line with this argument, he suggests that honesty, compassion, humility, temperance, and others can be both, character and non-character traits (cf. ibid.). Whether these traits are in fact character traits depends on the agent's responsibility for them.

Nonetheless, for purposes of this thesis, I refrain from adopting this approach as it still introduces a second, more grave problem that is specific to the research question stated in this work: if character traits are those personality traits that agents are responsible for, there simply is no question as to whether we are responsible for our character (and, as this thesis and a vast bulk of literature dedicated to the topic illustrate, this is a highly disputed question). That is because following this strategy, character traits are per definition those that agents are responsible for. Hence, it anticipates the answer to the research question prior to the actual assessment. Adopting this strategy would entail losing all the analytical leeway to whether agents are in fact responsible for their character traits or not. Even though Miller maintains

31

that this strategy - which he does in fact ultimately endorse in combination with the upcoming third strategy - is a "rough test" (ibid., p. 17), his proposal seems to carry philosophical baggage that would jeopardize the discussion to come ahead. In fact, it would introduce circularity. Thus, I do not think that either of the two strategies encountered so far is suitable as a condition to differentiate between character and personality for this thesis.

The third strategy, which Miller introduces, is the idea of character traits pertaining to certain normative standards: "A character trait is a personality trait for which a person who has it is, in that respect, an appropriate object of normative assessment by the relevant norms" (ibid., p. 15). According to this classification, character traits are evaluative in nature. An advantage of this view is that one escapes the potential problem that it is too intellectual or arbitrary as criticized when considering the first option. When the assessment does not depend on the agent's judgment but some exterior normative assessment, the classification gains explicability through accessibility. Moreover, this view explicates traits not only in terms of descriptive properties but also normatively (e.g., being admirable for having the trait). Character traits on this view do not only describe agents but give a normative account of what they are like (cf. ibid, p. 15). This has special appeal because it reflects a social practice whilst also providing a sensible condition that supports the structural classification advanced so far. It explains not only why more typical examples such as courage, honesty, and kindness are often considered character traits but also why self-control, being logical and imaginative add to the list (cf. Brandt 1970, p. 23; Miller 2014, p. 10). The latter are also traits by which agents are regularly normatively assessed, meaning that they function not only as descriptions but also normative assessments. For example, to say that Christine is logical yields two sorts of explanations. It describes her as someone who has this trait and differs from those who are not logical. But it is also a normative assessment of her mental skills. That is, she is admirable for being logical (cf. Miller 2014, p. 15).

As far as Miller goes, he ultimately argues that the best version to conceive of a condition for character traits is simply to combine "responsibility" and "appropriate normative assessment" (cf. ibid, p. 17). An obvious reason for this is that it seems plausible that an agent is an appropriate object of normative assessment *because* she is responsible (cf. ibid, p. 17 footnote).[24] Due to the issues that arise in the context of strategy two for this thesis, however, I suggest that the only suitable condition of the three presented is the last. That is because the third option avoids the problems identified on the first two strategies. By opting for the first or second alternative, we lose our ability to differentiate between character and personality because we might end up with no classification at all. The third option on the contrary provides a standard by which to tell whether a personality trait is also a character trait.

### 2.2.2 Moral and Non-Moral Character Traits

Again, the view I advance conceives of character traits as *kinds* of personality traits. An advantage of the view, which is inspired by Christian Miller, is that it captures the essence of the intimate yet separate relation between character and personality. It seems that exactly this more complex relation has led to some difficulties to account for it previously. This view provides a convincing structural relationship between personality and character. But it also gives a plausible condition of how to identify character traits. Character traits not only describe but also assess agents.

It might be questioned that I have not yet said anything about the criterion, which requires an explanation of why character traits are commonly associated with morality. Worse, critics could raise the point that the view as presented so far is no

---

[24] Miller stresses that the term "appropriate" is of some significance. He does so because some agents such as very young children, those mentally handicapped or with severe mental illnesses may not (yet) have the capacities required as to be appropriately assessed by the relevant norm (cf. Miller 2014, p. 15). I agree with the cases described, yet it does not necessarily follow that the appropriateness of the assessment hinges on the notion of responsibility. Hence, refuting the responsibility strategy does not preclude us from endorsing the strategy relying on appropriate normative assessment.

improvement to the dichotomy view after all but in fact potentially retracts to a version of it, if changed accordingly. But how so? The view I propose accounts for two criteria. One, it suggests that the structural relation is such that character traits are a kind of personality. Two, it provides a condition, namely that character traits are those personality traits by which agents are normatively assessed by. Now consider how the skeptic might apply the third criterion of morality. The condition given could be changed to *moral* normative assessment. Character could be a personality trait which agents are morally normatively assessed by. This alteration would enable us to remain loyal to the structural classification while still being able to differentiate between both terms. Further, it incorporates the third criterion. But, again, this, the skeptic could argue, would simply amount to a version of the dichotomy view, though one that avoids the charge of being unclear about the structural relation. So why provide an alternative if the dichotomy view can be altered to work as well as the critic might suggest? We are back to square one, it seems.

However, this option is the less favorable one as it neglects that there are character traits that agents are normatively assessed by, however, which are not necessarily moral. For example, what can we say about traits such as grit, self-control, and resilience? Certainly, there are normative standards governing these traits, but they need not necessarily be moral. Hence, instead of re-creating a version of the dichotomy view, I propose that the classification remain as outlined. One simple solution to incorporate this last criterion without falling back to a version of the dichotomy view is to make a further differentiation between moral and non-moral character traits (cf. Miller 2014, pp. 32–34). While moral character traits - as the name suggests - pertain to morally relevant aspects, non-moral do not necessarily do so. The introduction of a further class provides the room to account for those traits which do not seem to fit neatly into the logic of the dichotomy view. Not all character traits seem to belong to the moral realm. Instead of classifying all character traits as moral, this further individuation allows for some traits to be non-moral. In fact, the additional differentiation between moral and non-moral character traits reflects a view that has

been suggested by others who maintain that there can be moral and non-moral character traits. For example, Kupperman offers the following examples for moral character traits: being truthful, refusing to steal, torture or commit murder. In contrast non-moral character traits are rebounding from failure, weakness and being "a depressive oaf" (cf. 1995, p. 8). And Rhode, too, who differentiates between moral and performance character, observes that performance character traits, such as self-control, grit, and perseverance, may be used to do good or bad things. Whereas moral traits, such as honesty and compassion, are unconditionally good or bad. She notes an important relation between the two as without moral character, grit and such can be used for the unethical. And without performance character, we cannot follow through on our ethical beliefs (cf. 2019, p. 4). A similar concept can be found in Piazza et al. (2014), who differentiate between core-goodness traits and value-commitment traits. While the former are unconditionally good, the latter are conditionally good. On their view, these include determination, courage, diligence, and dedication:

"These traits are 'default positive,' in that they enhance the goodness of a good or neutral agent. However, these traits are only conditionally good in that they may also amplify the *badness* of bad agents, in part, by enhancing the perception that such agents are committed to particular immoral values" (p. 529).

Even though I plan to employ this further differentiation between moral and non-moral character traits, it should be noted that the suggestions by Rhode and Piazza et al., which imply that there are conditionally and unconditionally good or bad traits, seem to not stand against the simple test when considering a trait such as honesty. Whilst honesty is certainly generally considered a good trait to have, it definitely can be used for the unethical. Consider for example an agent like Maxine who strongly dislikes her coworker Julia whom she knows to be very conscientious in completing her tasks. One day she notices that Julia has made a mistake. Recognizing that Julia was up for a promotion, Maxine reports her to their supervisor, costing Julia her hard-fought jump in her career. Thus, it seems that even traits such as honesty, which are often thought to be good, can be used for unethical behavior.

But how then can we differentiate between moral and non-moral character traits? How can we tell whether a character trait such as conscientiousness, courage, honesty, self-control (cf. Brandt 1970, p. 23) and perseverance is also a moral character trait? Recall that I have advocated the view that character traits are those that agents are normatively assessed by. To differentiate between moral and non-moral character traits, I now suggest that we follow a much simpler strategy, again adopting from Miller, namely sticking with the condition of normative assessment, but specifying it. A promising strategy to do so would be to test moral character traits by the varying *moral* assessment (cf. Miller 2014, pp. 8-18; 32-35). For example, Judith possesses the moral character trait of compassion because we believe that her helping her fellow female coworkers is morally relevant. Non-moral traits, on the contrary, can pertain to many different domains such as the epistemic (intellectual humility and intellectual courage), prudential (diligence and cleverness), religious (religious faith and religious devotion), or athletic (competitiveness and discipline) (cf. Miller and Knobel 2015, p. 34).

The view I endorse[25] is an alternative to the two prevailing approaches. I propose this option because it provides a clear condition of character traits, exemplifies the intricate relation to personality and accounts for a specific class of character traits which are non-moral. However, not everyone shares the enthusiasm to break loose from the dichotomy view: "Some (…) think there is no natural distinction in our language between morally loaded and morally neutral names for character traits; at best we invent the distinction" (Butler 1988, p. 216). Even though a more elaborate classification may be perceived as theoretical overload, I do not share this skepticism. On the contrary, I remind readers that much confusion in the research area stems from the identity and dichotomy view being used without further clarification. In addition, it captures our everyday practices more precisely than other views that have been

---

[25] Again, I follow Miller in the general structural classification ((he argues that character traits can be divided into moral and non-moral character traits) (cf. Miller 2014, pp. 32–34)), but my approach differs with respects to the condition, i.e., normative assessment.

suggested. Last but not least, as mentioned before, this research thesis particularly addresses the question of character responsibility and is therefore in need of precise specification. With this classification in mind, this thesis will be concerned with moral character traits. We can now move on to highlight further structural aspects of character traits.

## 3. Character Traits as Manifesting Dispositions

### 3.1 The Dispositional View and the Problem of Manifestation

In the contemporary philosophical debate on character, there is overwhelming agreement that character traits are dispositions (cf. Miller and Knobel 2015, pp. 20–25). On this view, character traits are real properties had by agents (cf. Miller 2014, p. 19). They are not simply third-party ascriptions but possessed by the agent. But why are character traits standardly characterized as dispositions? The reason may primarily lie in the similarities in features that character traits share with *paradigmatic* dispositions of physical objects such as a fragile glass. Despite debates about specifics, there is agreement that these paradigmatic dispositions usually have four features: directedness, a stimulus condition, a categorical/causal basis, and independence (cf. Alvarez 2017, pp. 72–74). Comparing character trait dispositions with paradigmatic dispositions reveals that the former share some of the features of the latter.

First, dispositions are *directed* in the sense that they are defined by their potential manifestation, i.e., their outcome (cf. ibid, p. 72). The glass, for example, manifests its disposition of being fragile when it is struck. And character traits, too, are directed. Compassionate Judith will assist her friend, who needs help with carrying her groceries. Hence, Judith's compassion has a specific outcome. In that case, Judith's disposition to be compassionate makes her help her friend by going shopping with her. Second, paradigmatic dispositions have a *stimulus condition* (or *trigger*), in the case

of the glass it is "being struck." Character traits can also have stimulus conditions (cf. Miller 2014, p. 21).[26] For example, Judith may help her friend because she notices that her friend struggles to lift the bags. In fact, there are many possible stimulus conditions. Therefore, character traits might be considered a prime example of what is called *multi-track* dispositions in the debate because they have more than one conditional (cf. Mumford 2003, p. 8).[27] The general idea is that multi-track dispositions - in contrast to single-track dispositions - have multiple conditionals as they cannot be described adequately by a single conditional (cf. Vetter 2013, p. 334). For example, there are multiple stimuli that might trigger Judith to help others. Judith will act compassionately when seeing her friend having a hard time with her groceries; but also, when given the chance to donate to a charity, or when asked for help by an elderly person to accompany her across the street. Thus, her compassionate behavior may be triggered by different conditional inputs, which vary in content. Similarly, having a character trait can manifest in various actions (cf. Powell 1959, pp. 497–498). For example, Judith manifests her compassion by comforting her friend, by being the bigger person in an argument, or by taking care of her infant sibling for the day. Further, dispositions have a *categorical/causal basis*. The glass' disposition itself is based on the crystals that make up the glass. And this, some maintain, also goes for character traits. For example, Miller argues that each trait disposition has in turn various underlying mental state dispositions relevant to the respective character trait (cf. Miller 2014, pp. 24–25). On this view, compassionate Judith might have a disposition to notice despair and believe that she is good at fixing things. Lastly, paradigmatic dispositions are *independent of manifestation* (cf. Alvarez 2017, pp. 72–74). The glass' disposition to be fragile exists independent of the whether it has been struck. Proponents of dispositionalism argue that character traits too need not manifest because they are properties had by the agent (cf. Driver 1996, p. 125). On this view, agents can have a

---

[26] However, it should be noted that character traits can potentially manifest without triggers (cf. Alvarez 2017, p. 85).

[27] In fact each behavioral output might also be caused by the interrelation of multiple traits (cf. Funder 1991, p. 33).

trait without exhibiting them. Thus, dispositionalists could well assume that character traits resemble paradigmatic dispositions.

Yet, contrary to these claims made by dispositionalists, there is some indication that paradigmatic dispositions actually differ with regards to the feature of manifestation. That is because paradigmatic dispositions not only never have to manifest (cf. Hampshire 1953, pp. 7–8), but *some* even lose their disposition as soon as they do. Consider again the glass which possesses the disposition to be fragile. As soon as it breaks, it actually loses its fragility (cf. Mumford 2003, pp. 42-45; 50-51). However, this is not the case for character traits. When agents manifest their character traits, they habituate that trait, they do not lose it. On the contrary, they probably strengthen that trait. Hence, while the glass loses its disposition once it has manifested it, character traits do not function this way; character traits allow for manifestation. Though I acknowledge that not all paradigmatic dispositions lose their disposition as soon as they manifest (e.g., the electric wire retains its disposition to conduct electricity even when it has done so prior), the observation that character traits are strengthened by their manifestation gives us a first indication that they differ with respect to the feature of manifestation.

However, I think there is a stronger, more interesting assertion to be made following Maria Alvarez (2017): character traits not only *allow* for manifestation, they require it, hence are *dependent* on it. In other words, manifestation is constitutive of having a character trait and independence is not a feature after all. For see how having a character trait requires trait-relevant behavior. This may include specific actions but also thoughts, desires, reasoning and more. Thus, manifestation need not be overt. For example, Hannah has the trait of honesty because she regularly tells the truth and Judith has the trait of compassion because she thinks of ways to help others and engages in community service (cf. Johnson 1997, p. 74; Alvarez 2017, p. 77).

At this point, a brief digression is in order: it is important to note that there are different interpretations to which the independence claim may relate. First, independence from manifestation may simply translate to the view that an agent can

have a character trait without it currently manifesting. For example, compassionate Judith retains her trait even in moments in which she does not specifically manifest the trait in action or thought. A second option to read the independence claim would be to state that an agent can have a particular trait, yet not manifest it even when respective external factors are present that would normally entail her manifesting that trait. If Hannah is considered honest, she is usually thought to tell the truth whenever external features prevail, such as someone asking for her opinion or her recollection of a chain of relevant events. Yet, the way I want to digress from the independence claim is not by rejecting either the first or the second understanding but rather by stating that character traits must have manifested in the past. Hence, I do not deny that an agent may have a character trait without currently displaying it (which seems common sense as agents cannot manifest all of their traits constantly and simultaneously) nor that agents can have a character trait yet not manifest it even though external triggers or factors would support the manifestation thereof. Instead, what I want to challenge is whether they must have manifested in the past.

In order to restore their independence-claim, proponents of dispositionalism often counter this with two arguments. First, they hold that an agent can have a character trait despite not having manifested it in the past as she simply may have *not been presented with the opportunity* to manifest the trait. Failing to manifest the trait then is not evidence for the lack thereof. Instead, Hannah may have simply not have had the chance to display her honesty (cf. Miller 2014, pp. 19–20). Even more so, agents can have a disposition not only when they do not manifest it but also when they act contrary to them, as certain external factors might have prevented the agent from manifesting the corresponding disposition (cf. Wright 1990-1991, p. 49). The latter proposition reflects a type of argument, which can be found in the discussion on paradigmatic dispositions. Glass might have the disposition to be fragile, i.e., to break once it falls down on the floor. However, if wrapped into paper it might not break. Some authors have argued that these are certain types of *mimickers* and *maskers*, which prevent the glass from having this disposition (cf. Manley and Wasserman 2008, p. 63).

Analogously, character dispositions might not manifest because they are inhibited. In the case of Hannah, she may have been pressured into lying in order not to hurt somebody's feelings.

But once we realize what that argument entails, it does not hold up. Let me be specific about the claim that dispositionalists actually make here: they maintain that agents can (have) *never act(ed) in character* yet still possess the respective character trait. But to see that this claim does not withstand, simply consider an agent, who has never smoked once in her life. We would not think of her as a smoker unless she has in fact smoked before (cf. Ryle 1949, p. 113). Similarly, if Hannah has never told the truth before, it begs the question of how and why we would speak of her as honest (cf. Alvarez 2017, p. 81). This is not to repudiate the valid point that agents can be prevented from manifesting their traits on occasion because inhibitors interfere (cf. Brandt 1970, pp. 35–36). Again, I do not subscribe to the claim that an agent must manifest a certain trait as soon as relevant external factors are present. However, there is a fundamental difference between occasionally not having had an opportunity to manifest a trait, even acting contrary to that trait, or never having manifested it at all. The point I am trying to make here (again, following Alvarez 2017) is that it is unclear what the trait would consist of. How can we conceive of Hannah having the trait of honesty without her ever even thinking an honest thought? And this is not a matter of whether there is a certain way that we can be sure to know that Hannah has this trait. I am not saying that we cannot attribute the trait of honesty to her because we have no tangible evidence as third-party observers. Rather, what I am claiming is that she does not possess this trait because having manifested it - be it internal or external - is constitutive of having that trait. Never having manifested a trait supports the lack thereof (cf. Alvarez 2017, pp. 79–80).

A second and slightly different argument proponents of dispositionalism give in favor of their view is that agents could *discover* that they have that trait. So, on this view Judith may not have been compassionate in the past. But one day she may find herself in a situation where she takes in a homeless person into her home to offer her

shelter and food because it is freezing outside. Yet, there are two reasons to reject this argument as well. First, one action does not make a character trait. Not every action is necessarily an instantiation of a trait. For example, Judith may sign up for extra community service because she wants to enhance her CV to apply for certain universities. This does not necessarily make her compassionate. Additionally, Judith's behavior may be out of character[28] even and not a manifestation thereof (cf. ibid., pp. 82-83). Second, I do not see how this aligns with the view that character traits are habituated.[29] How can we explain the way in which that character trait was acquired in the first place? Neither the argument from opportunity nor the argument from discovery sufficiently counter the claim that manifestation is not simply a possibility, it is a key feature of character traits.

The passionate dispositionalist may now reverse the argument and question that if manifestation is constitutive of character traits, then what kind of manifestation suffices? Does Hannah count as honest when she tells the truth twice? Or does this characterization only apply to those who are honest at least every week? How frequently must a trait manifest to count as such? What is more, what can we say about the required intensity? Is a handshake sufficient to count as manifestation of Judith's compassion or would only a hug combined with kind words count? While these questions certainly are legitimate and it might be favorable to have a definitive answer to this, this is no grounds to doubt the claim I am making alongside Alvarez (2017), namely that never manifesting a trait implies that the agent does not have this character trait.[30] That is because both are separate issues that are certainly linked; yet

---

[28] There is some considerable agreement that agents can act "out of character", (for examples of this view see Hampshire 1953, p. 7; Sher 2006a, p. 66; Grover 2012, p. 33; Elliott 2017, pp. 39–56; Fileva 2017b, p. 183). These agents have a certain character trait but act contrary to the behavioral output affiliated with that trait.

[29] This point certainly relates to the common claim agreed upon by all major character theories that character is primarily a matter of habituation. See chapter 2, which gives an outline of these theories and the claims entailed, including habituation.

[30] Alvarez (2017) provides a great discussion of whether character traits are dispositions. Here, she also argues that the unresolved question about the quantity or quality of manifestation does not challenge her main point, which I endorse here, that manifesting a trait is constitutive thereof.

the claim about dependence does not necessitate a definitive answer to the degree of manifestation (cf. ibid., p. 84). Since we do not need to elaborate on the issue to retain the claim that character traits must manifest, we can continue with the more fundamental question of what character traits are.

## 3.2 The Summary View

So far I have subscribed to the claim that manifestation is constitutive of character traits, i.e., character traits are characterized not by independence but dependence. Thus, since character traits differ from paradigmatic dispositions with respect to this feature, the worry arises that character traits might not be dispositions after all. An alternative is to explore a rival view to dispositionalism, which reflects the view that comes with the dependence claim, namely that character traits must manifest: the summary view (or act-frequency-approach). Mostly proposed by psychologists (cf. Wright and Mischel 1987, p. 1160)[31], the summary view states that character traits are not dispositional properties had by agents but merely summaries of tendencies to behave.[32] For example, consider how proponents of the summary view characterize an agent as talkative: on the view imagined the more an agent talks the more she is in that state[33], meaning that "density distributions" of displayed behavior help to characterize agents (cf. Fleeson 2007, p. 826). To say that Judith has the character trait of compassion

---

[31] See for example Hampshire 1953, pp. 5–7; Brandt 1970, pp. 25–26; Buss and Craik 1983; Newman and Ulman 1989, p. 163.

[32] For example, Hayes et. al. (cf. 2017, pp. 268–282) claim that character is nothing more than trait attribution made by an observer, which may serve different functions. Character then is no mental property.

[33] Note that Fleeson talks about states instead of traits here. In philosophy, character traits are considered dispositions, while this is not so clear in psychology (cf. Miller 2014, p. 18). For example, in psychology two contrary camps have evolved on traits and states. While some argue for trait theories (e.g. Johnson 1997, pp. 73–75; Cohen and Morse 2014, p. 45; Fleeson et al. 2015, p. 44), others propose that the mind includes mental states, which are momentary glimpses and need not be cemented in traits (e.g. Mischel 1968, p. 8; Kalimeri et al. 2013, p. 28). The different arguments are also reflected in prominent competing accounts such as the HEXACO (formerly Big 5), cognitive-affective personality system (CAPS), and Whole Trait Theory (e.g. Fleeson and Jayawickreme 2015).

then simply means that she has displayed acts of compassion in the past. A trait then is merely a summary of a *trend* in (past) behavior. This does not entail perfect behavior; an agent might still act out of character. Nonetheless, the view centers on the agent's past behavior, for example when predicting future behavior. Yet, this is detached from any property of the agent (cf. Buss and Craik 1983, pp. 106–107). Thus, the summary view accounts for the claim that character traits must manifest.

However, the summary view - despite its initial appeal - does not grasp the essence of character traits either. One problem of the view is that it fails to explain our practices of making trait attributions based on single instances of behavior (cf. Brandt 1970, p. 26). That is because contrary to these practices summary view theorists claim that behavior must be observed over a course of actions in order to make valid trait ascriptions (cf. Buss and Craik 1983, p. 107). However, this alone would not be problematic because our practices of trait attribution might simply be wrong (and again, one instance of manifestation does not necessarily constitute a trait). But this view entails a more fundamental issue: one of the view's main claims pertains to character traits being summaries of trends, meaning they are not considered properties of the agent. But since on this view character traits are merely attributions based on observed behavior it cannot account for the fact that agents possess these traits even in those instances when they do not (openly) display them. Hence, due to the focus on actual behavior and the view from the third-party observer, this seems to exclude cases in which agents are inhibited from acting currently and/or openly. For example, consider Julia, who has a record of being courageous in her late teens but has not had the opportunity to manifest that trait recently. On the summary view whether she can now count as courageous depends on behavior from the distant past. But since she has been inhibited from being courageous lately it is not at all clear whether the summary view can explain why she would have the trait now. Hence, it begs the question of why we speak of agents having a specific trait when they do not currently or have not recently manifest(ed) them. Similarly, stressing trait attributions instead of traits as properties neglects that some manifestations are invisible to the observer (cf. Miller

2014, pp. 19–20). Thus, the claim that character traits are not real properties of the agent, results in the inability to account for the cases in which agents do not currently and/or openly exhibit their traits but are still considered to possess them.[34]

## 3.3 Manifesting Dispositions

The claim I have advanced so far is that character traits share many features with paradigmatic dispositions but differ with regards to the feature of manifestation. The alternative summary view contests that character traits are dispositional properties. Instead, character traits are considered trends in past behavior. But since on the summary view character traits are not properties of the agent, it struggles to explain why agents *have* a trait when they have not recently or openly manifested it. Character traits then, it seems, are neither purely dispositional nor merely summaries of past behavior. Hence, a viable option must manage the balancing act. It must reflect that character traits are not exactly paradigmatic dispositions because they require manifestation whilst affirming that they share many features, most importantly that they are possessed by the agent.

One view which resonates with these claims is the *mixed summary view*, a version of the summary view, which states that character traits are dispositions but those that must manifest (cf. Brandt 1970, p. 25; Fleeson 2007, p. 826; Funder 1991, p. 32). This view seems more satisfying than the *pure* version of the summary view because it avoids the criticism the latter faces. However, despite its assertions being plausible, I refrain from adopting this terminology. (To be clear, I do not intend to challenge the content of the mixed summary view - on the contrary, I subscribe to its most basic claims). I do this for primarily two reasons. First, the mixed summary view

---

[34] Miller (2014) provides a further argument against the summary view: he argues that it fails to incorporate two essential features character traits are typically thought to have. Usually, they explain (causal) and predict behavior. Yet, if traits are not possessed, Miller maintains, the summary view does not "capture the familiar idea from ordinary thinking about character traits" (2014, 20 footnote).

does not sufficiently demarcate from the "original" summary view. The terminology still suggests that character traits are first and foremost summaries of past behavior. Second, building on the first point, it fails to positively point out that character traits are properties of agents.

A more attractive and obvious solution to the problem is to reformulate the conditions of dispositions. For example, we could stipulate that independence of manifestation is in fact not a feature of *all* dispositions (independence could be a sole feature of *paradigmatic* dispositions yet not *character* dispositions), therefore eliminating the hurdle to categorize character traits as such (cf. Alvarez 2017, pp. 85–86). Changing the conditions by discarding the feature of independence, we could allow for character traits to be a certain kind of disposition. This is why I suggest a terminology which better captures the significant overlap in features between dispositions and character: character traits are *manifesting dispositions*. Ultimately, this might be a matter of labels, but this term better captures the way in which character traits are dispositional properties of the agent but those that must manifest. This terminological specification focuses on the dispositional aspect while acknowledging that character traits must manifest.[35] It elucidates why character traits share many similarities with paradigmatic dispositions such as a water glass, which can be fragile sitting at the counter of a desk. Analogously, Francine is just even when she does not currently manifest that trait. Nonetheless, viewing character traits as a certain kind of disposition avoids the criticism that character traits differ from paradigmatic dispositions as they need to manifest. Aside from these obvious advantages, can this combination of the summary view and the view that character traits are dispositional features possessed by the agent withstand the criticism the pure summary view faces? I believe it can. That is because the manifesting dispositions view (again, a terminological choice which is built upon the mixed summary view) also gives a more satisfying answer to the problems that the summary view seems to struggle with.

---

[35] Again, this is not to imply that they can be inhibited from manifesting in some cases due to mimickers/maskers. However, they have to manifest at some point.

Recall that the pure summary view fails to account for one important issue: since the summary view does not consider character traits to be genuine properties had by agents, it cannot explain why we consider someone courageous who has only manifested that trait in the distant past but not recently or has/does not openly manifest it (cf. Miller 2014, pp. 19–20). The manifesting dispositions view, on the other hand, considers character traits to be dispositional features of an agent, hence the agent can possess the trait even though she has only manifested it a long time ago (or only internally). Francine, who has manifested her trait of justice on multiple occasions, yet not as of recent, she can still be considered to have that trait on this view. In conclusion, conceiving of character traits through the lens of the manifesting dispositions view makes sense of this case by illustrating that in order to count as a trait, the agent must have manifested it in the past but since it is now a property had by her, it need not currently manifest.

# 4. Conclusion

This chapter served to lay the groundwork for a general understanding of character. The term is often subject of confusion because philosophers and psychologists alike are interested in character. In order to avoid a similar problem in the later analysis, this chapter gave an account of how to classify character. On the view proposed, character traits are a kind of personality trait, namely those that agents are normatively assessed by. This classification digresses from popular (folk) psychological and philosophical views that conceive of character traits as synonymous to personality (identity view) or take all character traits to be moral personality traits (dichotomy view). Further, I maintained that not all character traits are moral, and it is plausible to differentiate between moral and non-moral character traits. Moral character traits are those that agents are morally assessed for. The character traits of interest for the subsequent analysis are moral character traits.

The second part of the chapter affirmed that character traits are dispositions. However, they do not exactly resemble paradigmatic dispositions because they are not independent of manifestation. The alternative summary view, which stresses that character traits must manifest, I argued, is not much of help either. This view faces problems because it runs contrary to the view that agents possess their traits also when they do not manifest them currently or openly, meaning that they are properties of agents. To illustrate that character traits resemble paradigmatic dispositions for the most part (they are directed, have a causal basis, and have stimulus conditions), yet must have manifested at some point, I suggested that character traits are best understood as a kind of disposition, namely *manifesting dispositions*.

# II Theories of Character

## Global Traits, Local Traits and Mixed Traits

---

## 1. Introduction

The last chapter introduced a general classification that demarcates character from personality. Further, the chapter introduced character traits are dispositions, joining the majority of philosophers. However, I specified that they are a kind disposition, namely manifesting dispositions because, contrary to paradigmatic dispositions, they must manifest.

Despite these preliminary clarifications, there is still much to be said about the essential features and content of character traits. This is why the aim of this chapter is to build on the previous one by taking a closer look at some of the philosophical accounts of character found in the literature. Much of the philosophical interest in character is embedded in virtue ethical thought and dates back as far as Aristotle (cf. Kamtekar 2004, p. 477). Today still, much of the research on character is related to virtue ethics. And many analyses of character responsibility are derived or explicitly based on a conception of character that stands in the virtue ethical tradition. But despite this focus, it is important to deal with the subject matter in more detail by looking at alternative models and adjacent research in related areas as character has quite recently gained more attention in contemporary (moral) psychology. Due to the aforementioned dual evolution of contemporary character research in psychology and philosophy, it is nearly impossible to paint a picture of philosophical character theory without also considering debates and experiments in psychology. For example, the virtue ethical strand of character theory has received some criticism, known as *situationism* born from psychological experimental research. This has led researchers

to either supplement the virtue ethical approach or reject it altogether. Thus, this chapter broadens the scope by outlining three leading philosophical accounts of character. The debates and theories on character are extensive, and therefore, it is necessary to contextualize them and establish a more detailed understanding of character. The structure of this chapter reflects the historical development of these character theories and illustrates their relation to psychological research. Yet, employing this procedure not only mirrors the chronological evolution of these theories. It also provides a better grasp on the reasons and origins of these accounts as well as the resonant critique. Finally, a closer look at each of these theories will illustrate their commonalities and differences and further illuminate the array of general questions on character, which is found both in philosophical thinking and psychological research.

The chapter starts off with an outline of Global Traits. Global Traits stand in the tradition and overlap with - yet are not synonymous to - virtue ethical conceptions of character. Global Traits as conceived in this thesis are broad, i.e., cross-consistent and stable features of an agent. The chapter then progresses to an overview of Local Traits. Local Traits are the result of the fundamental situationist critique of virtue ethical conceptions of character provided by John Doris (cf. 1989, 2008). For the most part, Local Traits deviate from the claim that character traits are broad in the sense that they are cross-situationally consistent. Instead, character traits are structurally narrower than Global Traits as they are assumed to be localized and applicable to specific situations only. However, they are still considered to be temporally stable. Lastly, the chapter provides a summary of Christian Miller's (cf. 2013, 2014) model of Mixed Traits. On this view, character traits are best explained as broad but neither in terms of virtues nor vices. Instead, character traits are made up of interrelated trait dispositions that pertain to specific moral domains. What can be observed is that all the theories affirm that character traits are stable dispositions that are habituated. However, they differ with respect to the degree of individuation. While the Global Traits account conceives of character traits as coarse-grained, broad features of an

agent, the Local Traits approach is narrow, and the Mixed Traits model specifies character traits as fine-grained.

# 2. Global Traits

## 2.1 Globalism and (Folk) Psychological Conceptions

The globalist conception is probably the most common approach of how to conceive of character. Not only is the view reflective of common language, it is also by far the most employed concept in philosophical research on character. Notwithstanding this extensive usage, it is far from easy to pin down the specifics of the approach. This is because often times it is referenced implicitly. Nonetheless, one feature of the concept is evident: as already indicated by the name, globalist conceptions conceive of character traits as global. This entails that character traits are broad features of the agent that are stable and consistent (cf. Walker 1989, p. 354; Doris 2008, pp. 22–23; Miller 2009, p. 249). To say that character traits are stable means that they temporally endure in an agent. For example, someone with the trait of kindness will have this trait over a longer period. Hannah has this character trait not only for a day or a week but significantly longer because it is a feature of her. But character traits are not only stable, they are also consistent. Someone with the trait of kindness is expected to be kind across different situations. Hannah will support her co-worker by listening to her problems, be understanding when her student was not able to turn in her homework in time because she was sick or give way to strangers in front of the checkout at the supermarket if they only have one item to scan. Furthermore, character traits have explanatory and predictive power. To say that Hannah is a kind person explains why

she is kind to others regularly.[36] It also means that we can expect her to be kind in a wide range of situations in the future:

"(…) ascriptions of character traits to individual agents are supposed to play two central roles on a globalist framework—they are meant to explain consistent and stable manifestations of trait-relevant behavior, and they are supposed to accurately ground predictions of such behavior in the future" (Miller 2009, p. 249).

These assumptions converge with some (folk) psychological thinking about traits. For example, in folk psychology, character traits are also considered broad, i.e., stable and cross-situationally consistent. This is supported by evidence that many believe in character stability. In addition, psychological research also suggests that people not only think of traits as stable, they also think that traits entail predictability (cf. Newman and Ulman 1989, 165–166; Funder 1994, p. 125; Fleeson 2007, p. 828)[37]: "Why else do prospective employers ask for accounts of people's character if not to gain some idea of what to expect from them?" (Powell 1959, p. 502).

## 2.2 Virtue Ethics and Global Traits

In philosophy, it is almost nearly impossible to speak of character without mentioning virtue ethics. That is because not only are virtue ethics the philosophical account

---

[36] Some question whether dispositions have explanatory power as it reveals a tautology (cf. Newman and Ulman 1989, p. 167; Tellegen 1991, p. 14; Mumford 2003, p. 136). Since character traits are a kind of disposition, the argument may apply. The statement that Hannah is kind because she has behaved kindly is a circular argument. However, even if the general statement does not offer an explanation, it seems that character traits can offer insights on the reasons why an agent has acted a certain way (cf. Fileva 2017b, p. 195).

[37] It should be noted that there is no explicit reference to character in these particular studies. Traits may also refer to personality as shown in the previous chapter.

primarily concerned with character[38], the globalist view traditionally stems from virtue

ethical conceptions of character and thus converges in large part. This is mirrored in

the fact that some philosophers use the term *globalism* to refer to the virtue ethical

conception of character (for example cf. Doris 2008, pp. 22–23). Also, the vocabulary

between global conceptions of character and virtue ethical approaches is significantly

similar (e.g., traits are "honesty," "kindness" etc.). Further, on both views it is

generally thought that agents acquire character through habituation (cf. Dougherty

2007; Annas et al. 2016; Swanton 2016).[39] Lastly, they are dispositional features of the

agent (cf. Annas 2011, pp. 8–11).

Despite this overlap, it would be a mistake to equate Global Traits and

(contemporary) virtue ethical accounts. For one, the latter are tied to normative

thinking, i.e., virtue ethical accounts conceive of character not only in broad terms but

virtues and vices (cf. Elliott 2017, 66–68). Virtues are exceptionally good character

traits. With this normative view, virtue ethicists maintain that virtuous character is an

ideal to strive for. One of the assumptions on this view is that all agents aim to become

better (cf. Annas 2006, pp. 523–525). This also comes with a special view of the

virtuous: virtuous attain a state in which they sufficiently combine practical wisdom

and emotion and are free of inner turmoil. Practical wisdom guides the agent through

life to pursue the ends she has set for herself. This know-how lets an agent make the

appropriate choices based on reflecting what she believes it means to live well, which

---

[38] Other major strands of theories, such as deontology and consequentialism have incorporated character concepts. These competing character concepts are, however, restricted to theories within their own ethics. Thus, they are limited to providing a notion of character and its role within deontology or consequentialism, respectively (cf. Hursthouse and Pettigrove 2018 2018). For example, consequentialists can stress the role of character in raising the overall good (for a discussion see Bradley 2017, p. 83), while those subscribing to Kantian ethics might argue that good character can increase the likelihood of adhering to duties (cf. Sanderse 2013, p. 74; Swanton 2013, pp. 315–316; Miller 2014, pp. 188–189).

[39] More specifically, some virtue ethicists draw an analogy between acquisition of a virtue and the acquisition of a skill (cf. Russell and Miller 2015, p. 105). Both require critical reflection and the ability to give reasons. Just as a skilled piano player should be able to state why she plays Mozart a specific way, a virtuous person can explain why there is reason to act a certain way. Hence, exercising virtue is different from completing simple tasks, which do not require this sort of reflection (cf. Russell 2009, p. 375; Annas 2011, pp. 22–25).

is an end in itself (cf. Foot 1997, pp. 112–115; Hacker-Wright 2015, p. 986; van Zyl 2015, 188;190; Russell 2009, p. 375). With this "special grasp" (Hills 2017, p. 147), the agent can apply her virtues to the situations she faces (cf. Sanderse 2013, p. 87). Ultimately, then, agents, who possess the virtues are characterized by integrating the cognitive and emotive parts as to leading her in the same direction (cf. Elliott 2017, p. 34): "A virtuous person is a good person, one who acts well, where 'acting well' does not merely involve performing certain kinds of actions (...) but acting with *wisdom and sound emotion* [italics S. Sch.]" (van Zyl 2015, p. 187). This emotional and cognitive integration gives virtue "(…) a special kind of psychological 'depth'" (Russell 2009, p. 374) because virtuous agents conceive of certain ends that they pursue. Hence, while there is some congruence with the globalist assumption that character entails reliability, virtue ethicists maintain that: "Virtue (…) is a disposition not just to act reliably in certain ways but to act reliably for certain reasons" (Annas 2011, p. 27). Yet, these assumptions are not necessarily true for Global Traits as employed here.

And there is further reason to keep Global Traits separate from virtue ethical accounts. Even though contemporary accounts have produced novel interpretations of character and its role in virtue ethics, it stands out that most refer to some version of the Aristotelian model of character (cf. Elliott 2017, pp. 12; 17; Kamtekar 2004, p. 477). Still today, Aristotle's impact on contemporary virtue ethics is remarkable[40]: "(…) any definition of virtue ethics having the implication of excluding Aristotle altogether from the genre of virtue ethics is fatally flawed" (Swanton 2013, pp. 319–320). But the Aristotelian version of character may not necessarily converge with the broadness of traits as on the globalist framework. To illustrate the point, it is worth

---

[40] Of course, Plato and other Greek philosophers have also greatly influenced virtue ethics. However, in many cases, contemporary virtue ethics is explicitly based on Aristotle's work. The current revival in popularity of Aristotelian ethics could be due to the fact that "(...) there are characteristics of Aristotle's ethics that make it more attractive for modern ethical theory; his account of virtue does not seem to carry the burden of Socratic intellectualism; it seems to be less dependent on metaphysical background theories than, e.g., Plato's ethics; it does not display the same hostile attitude towards emotions as the Stoic account of virtue; and it seems to be closer to some important common-sense convictions (...)" (Rapp 2006, p. 99).

noting that some contemporary scholars contend the broadness of character traits in Aristotle's thought. In fact, Aristotle's virtues could be narrower than on the globalist conception and refer to more explicit contexts, i.e., donating money instead of generosity simpliciter (cf. Kamtekar 2004, pp. 479–480).

In order not to fall prey to the intricate assumptions on virtue ethical accounts, I advance a view that is separate from virtue ethical conceptions. Global Traits as employed here are neither normative nor intended to provide an Aristotelian exegesis. Instead, they serve as an umbrella term for accounts that are united by an understanding of character traits as broad and robust, meaning that they are features of the agent, which are stable and cross-situationally consistent. Therefore, it is important to note for our analytical purposes the coarse-grained structure of Global Traits.

## 3. Local Traits

### 3.1 Situationism: The Doubt about Cross-Situational Consistency

The globalist conception of character has been the topic of much debate. Virtue ethical conceptions of character had been the predominant approach until the introduction of a fundamental critique called situationism. Situationism is a position that had originally first arisen in personality psychology as part of the "person-situation debate" and draws from multiple psychological experiments. Among them is one of the first experiments on moral character, Hartshorne and May's "Cheating and Lying" - study in 1928, which tested honesty and cheating behavior among schoolchildren. In this case, children were asked in a guessing game to answer in self-reports on which side a coin had landed on. The unsettling results were that most children cheated when given the opportunity. Another experiment, which is particularly infamous, is the *Milgram Experiment* conducted by Stanley Milgram (1963) to test for obedience. The

experiment's setup included a subject, who was to administer electric shocks to a "student," whenever said student gave wrong answers to a question (in reality there were no electric shocks). Each time the student failed to give the right answer, subjects were asked by the experimenter (dressed in a technician's coat) to increase the voltage of the electric shocks. In turn, the students would ask the subject to stop administering those shocks (sometimes even cry out for help). Even though the voltage would have been lethal at some point and subjects were visibly under stress and duress, they kept going just because an authoritative person gave them simple instructions (cf. ibid., p. 374). Further, a study called the *Good Samaritan experiment* (Darley and Batson 1973) had seminary students prepare a speech on the parable of the Good Samaritan (the control group talked about possible professions for seminary students). The students were then asked to go to another building to give the talk. Those students, who had to speak on the parable of the Good Samaritan, were told that they were late already and were thus asked to hurry (while the control group had no such cue). On their way to the other building, all students found a person in need of help sitting on the floor. After the incident, students progressed to give their talk. The study found that those in the group, who were asked to hurry, were much less likely to help. The content of their talk had little influence on their helping behavior. Even more so, a subsequent questionnaire taken by the seminary students revealed that in fact all subjects had noticed the person in need of help (cf. ibid., pp. 100-108). Lastly, another experiment, known as *dime-in-a-phone booth* (Isen and Levin 1972), found that subjects were much more likely to help someone in need, when they had previously found as little as a dime in a phone booth compared to those who had not.[41]

These experiments are among the most cited. And, although they seem to be only a few, proponents of situationism claim that these experiments "(…) are not aberrational, but *representative* (…)" (Merritt et al. 2010, p. 357) with many more studies to ground the skepticism in. The experiments listed generated uncertainty about

---

[41] For an extended outline and discussion of the most common experiments see for example Miller 2014 or Elliott 2017 (especially chapter 3).

character in numerous ways: the Milgram experiment seemed to show that people are generally not virtuous as many were willing to administer potentially lethal shocks to someone who had simply failed on a test (cf. Miller 2018, p. 93). The Good Samaritan experiment introduced the worry - which was later confirmed by the Dime in a phone booth experiment - that morally good behavior seemed to depend on insignificant situational factors such as hurry or finding money (cf. Darley and Batson 1973, pp. 100–108). Most importantly though, in the situationists view, these experiments reveal one problem: character traits are not cross-situational consistent (cf. Doris 1989, pp. 506–507). Neither had Hartshorne/May's schoolchildren demonstrated honest behavior across situations (cf. Cervone and Tripathi 2009, p. 34), nor had the Milgram subjects shown compassion towards the "students," nor had seminary students helped those in need, and nor had people consistently helped without situational cues (cf. Elliott 2017, pp. 61–91). The common denominator of these experiments then is serious doubt over character (cf. Elliott 2017, p. 6). Mainly because it seems that seemingly insignificant situational factors highly influence our behavior. More generally, though, these experiments have revealed a number of insights that have shaped our understanding of character. First and foremost, they challenge the dominance of character on behavioral outcome.[42] Rather, it seems, that situational factors drive behavior, and not traits (cf. Doris 1989, p. 508).[43]

---

[42] Despite this general finding, there is no one single position that *is* situationism. Both, in psychology and philosophy there are different understandings of what the experiments ultimately reveal and hence what the situationist challenge entails (cf. Elliott 2017, pp. 111–138; Miller 2019, p. 407). For example, some think that situations simply add to explaining behavior; others claim that character is less important while situational features are more important in explaining behavior. Further, it could just mean that there are no stable character traits (cf. Miller 2003, p. 370; Brink 2013, p. 126; Rhode 2019, p. 8). See Elliot (2017), chapter 5, for an outline of five different formulations of the situationist challenge. See also Miller (2019) for a short overview of both psychological and philosophical variations of situationism.

[43] One view that reflects these findings, known as *interactionism,* stresses the relation of traits, situations and behavior (cf. Cohen and Morse 2014, p. 54; Fleeson 2007, p. 856; Funder 1991, p. 36; Rhode 2019, p. 9).

## II Theories of Character

Some years later, in the 90s/early 2000s, the situationist position was then adopted by philosophers John Doris (1989, 2008) and Gilbert Harman (2000).[44] Still today, despite recent re-interpretations of the experiments and efforts to give the position new ammunition[45], in philosophy, the versions by Doris and Harman stand out. Doris and Harman took the experiments and the theoretical basis of situationism founded therein to introduce a fundamental critique of virtue ethics, more particularly, the predominant notion of character it had advanced. Against the backdrop of surprising findings in experimental studies, they have concluded that these experiments challenge the concept of global character traits which particularly virtue ethicists had advanced up to that point (cf. Miller 2003, p. 370; Lucas and Donnellan 2009, p. 147; Miller 2019, p. 410).[46]

Notwithstanding the shared general criticism, Doris and Harman have drawn different conclusions. While Harman has voiced his concerns regarding the existence of any form of character (cf. 1999, p. 316)[47], Doris believes that the findings call into

---

[44] Doris and Harman have made situationism popular in philosophy but their critique is actually preceded by a first phase of situationists such as Owen Flanagan (1991), who notes the importance of situational factors (cf. Upton 2009, pp. 107–108).

[45] Further formulations of the problem, specifically for virtue ethics, have been given by Vranas (2005) and a joint re-formulation by Merritt, Doris & Harman, who have more recently argued that there is an integration challenge due to moral dissociation (cf. Merritt et al. 2010).

[46] Today, the subsequent heated debate around (philosophical) situationism has "(…) gained so much attention over the last decade, that there are hardly a self-respecting virtue ethicist or moral philosopher left who has not added something to the literature on this topic. Doris's and Harman's publications have even yielded so many (mainly critical) reactions that the literature has reached a saturation point" (Sanderse 2013, p. 93). And, in psychology too, so many arguments have been exchanged in the person-situation debate that it has been declared over (cf. Fleeson and Noftle 2008). Nonetheless, there are people in both fields who believe that the debate still has left open some issues (see for example Lucas and Donnellan 2009; Miller 2017b). Regardless of the verdict on this, today there seems to be some consensus regarding the general significance of situationism. For example, in philosophy, many authors now take the discussion to reveal that it is neither sensible to just buy into the situationist scare, nor maintain that character theory is completely unproblematic (cf. Besser-Jones 2008). Further, there is broad acceptance in psychology and philosophy of the fact that situational factors at least partly influence our behavior (for some examples see Mischel 1968; Miller 2010; Sreenivasan 2013; Russell and Miller 2015).

[47] I refrain from expounding more on Harman's view, which is connected to the Fundamental Attribution Error - a position that holds that our ordinary trait attributions are heavily misled as people falsely attribute character traits without any evidence: "(…) it does illustrate the tendency of observers to infer wrongly that actions are due to distinctive robust character traits rather than to aspects of the situation" (Harman 2000, p. 223).

question whether character traits are as broad as on globalist conceptions. The experiments, Doris believes, reveal one fundamental misconception about character, namely that agents are generally consistent with regards to acting across a range of different situations. The empirical evidence collected in the experiments, he argues, proves that people generally lack consistency (cf. Doris 1989, pp. 506–507). The only consistency that was shown, he goes on, is within similar situations. Dissimilar situations, however, show a weak link to consistency (cf. Doris 2008, p. 64). Any behavioral regularity is to be explained primarily in terms of situational regularity, meaning that people act stable over time just because they typically look for similar situations (cf. ibid., p. 26). Hence, it is situations rather than character, he claims, which produce stability in outcome (cf. Doris 1989, p. 508). Thus, his main concern is that character seems to lack cross-situational consistency. However, this directly contrasts the globalist view that character traits are broad and robust, i.e., cross-consistent. The lack of behavioral consistency as demonstrated in the experiments, Doris concludes, deems virtue ethical conceptions in the Aristotelian tradition empirically inadequate (cf. 2008, p. 23).[48] This is especially important to Doris because he "(…) advocate[s] 'psychological realism' in ethics - roughly, the idea that ethical reflection should be predicated on a moral psychology bearing a recognizable resemblance to actual human psychologies" (ibid., p. 112). But even though Doris fiercely argues against robust, global conceptions of character, he remains positive about the general possibility of character.

---

[48] The fundamental critique of virtue ethical conceptions of character was in turn met by a storm of defenses. Generally, virtue ethical defense strategies against situationist forms of critique can be grouped into three types. Their first approach to avoid the criticism is by arguing that virtue is rare. Thus, the experiments do not show that people generally do not have the virtues because not many possess it to begin with (cf. Kamtekar 2004, p. 485; Sreenivasan 2013, pp. 296–298). Secondly, virtue ethicists question the validity of the situationist arguments and try to show that the experiments are flawed (cf. Miller 2003, p. 369; Sanderse 2013, pp. 81–82; Sreenivasan 2013, pp. 300–303; Fleeson and Furr 2017, pp. 231–240; Lamiell 2018, p. 251). The third strand of arguments to rebut situationism is that it employs a wrong conception of character (for examples cf. Sreenivasan 2002; Webber 2013, p. 1084; Miller 2014, pp. 191–198; Marmadoro 2001, p. 3; Kupperman 2001, pp. 239–245). At this point, I will not pause to consider the proposed defense strategies nor their prospect of success in more detail.

## 3.2 Locally Specified Character Traits

Based on this fundamental critique, Doris proposes an alternative to the Global Traits model of character, which, in his view, is better suited to explain the findings of the empirical experiments. According to him, instead of *Global Traits*, character consists of *local* traits, which are temporally stable, yet not cross-situationally consistent (cf. ibid., p. 25).[49] Local Traits then pertain only to "(…) trait-specific behavior in a narrowly construed set of circumstances" (Miller 2003, p. 368). The possession of certain character traits on this account may vary according to the situational setup. Hence, Local Traits are more specific than Global Traits (cf. Grover 2012, p. 27). Doris summarizes the origin and basic claim of his view as follows:

„(…) behavior will exhibit considerable temporal stability over iterated trials of highly similar situations. Where such temporal stability obtains, we are justified in attributing highly contextualized dispositions or 'local' traits. (…) It should now be obvious that a central challenge for any theory of personality is accounting for the remarkable situational variability of behavior. This variability is not easily explained by globalist theory; if human personalities were typically structured as evaluatively integrated associations of robust traits, it should be possible to observe very substantial consistency in behavior. I therefore contend personality should be conceived of as *fragmented*: an evaluatively disintegrated association of situation-specific local traits" (Doris 2008, p. 64).

Local Traits, as Doris puts it, are applicable to a certain set of situations only. For example, Judith demonstrates compassion when she meets homeless people on the street. She offers them money and even takes them to a nearby restaurant to have dinner with them. However, when Judith encounters Amnesty International volunteers in the city, who ask for donations to advance human rights' issues, she passes by. She is simply not swayed by the thought of helping others without seeing immediate and tangible results. As she possesses the Local Traits of compassion-when-

---

[49] Analogously, psychological approaches have also shifted to narrower trait concepts in the 90s (cf. Funder 1991, p. 31). Today, there is a slight backshift to global conceptions (see for example *Whole Trait Theory* by Fleeson (cf. 2015; 2017)).

facing-those-who-seek-help-personally, she is happy to assist the homeless person but refuses to donate money to a charity. Similarly, Hannah tells the truth when her friend asks her about her newest sweater but will tell a white lie when her mom asks her about her whereabouts the night before. Hannah's Local Traits pertaining to honesty is limited to honest-when-talking-to-her-friend. Hence, the conception presented by Doris is much more localized than Global Traits. While Global Traits are broad features of the agent, the structure of Local Traits is not as coarse-grained but narrow. Thus, on the Local Traits account, instead of robust and Global Traits, character is best described in terms of narrow, locally specified, traits.

What should be noted about Doris' account is that his remarks are slightly rudimentary. The above description of Local Traits is not an extremely boiled down version of Doris' remarks but reflects the extent of his elaboration on the topic. His main focus is not to detail the specifics of a positive alternative conception of character but lies in rejecting the globalist thesis concerning consistency (cf. Brink 2013, p. 127). Primarily, he argues that character traits exist and are stable features of persons but rejects cross-situational consistency. Since Doris' priority clearly concerns the validity of situationism, he provides little details about the claims he positively endorses. Even though the remarks suggest that Doris affirms most of the globalist claims such as dispositionalism (cf. Doris 2008, pp. 15–16), for example, he makes few comments with respect to habituation. The fact that Doris elaborates little about the positive claims of his account suggest that he adopts much of the globalist view and only differs in terms of cross-situational consistency. However, there is not enough information to confirm this with certainty. One reason for him to refrain from advancing his view more forcefully might be that his goal is not to provide a stand-alone account but rather to make a case for his main thesis, namely, to repudiate Global Traits' claim about cross-situational consistency. But despite the scant information available for a positive outline of Local Traits, it can be summarized as follows: character on this view includes a variety of locally narrowed traits that are stable, yet not cross-situationally consistent. With this narrower conception, Local Traits structurally differ from Global Traits as

agents may display similar behavior in similar situations but not usually when situations are different. Nonetheless, it is also likely that the account shares some of the features with Global Traits such as stability, dispositionalism, habituation and explanatory/predictive power. While Global Traits mirror how we often speak of people, Local Traits can explain the variety in behavior observed in the experiments. More importantly, it accounts for the difficulty that participants of the experiments have in displaying consistent behavior across a variety of situations. Thus, the Local Traits approach provides a potential explanation for the findings in the empirical experiments, accounting for why agents behave deplorably in many of the experiments.

## 4. Mixed Traits

### 4.1 A Review of Previous Concepts

So far, I have introduced two different accounts of character. The outline of the final account will call attention to a rather recent approach named Mixed Traits, which has been proposed by Christian Miller (2013, 2014). Miller is one of the latest authors to profoundly reconsider the theoretical concepts and empirical data available on character. As part of a large-scale research project called "The Character Project"[50] involving philosophers, psychologists, and theologians, he has redirected focus on character in hope of being able to provide a scientifically nuanced and accepted account, "(...) which [he] claim[s] is empirically more plausible as an account of the moral character traits had by most people (…)" (Miller 2014, p. 61). The result is a "bottom-up theory of character" (Bates 2015, p. 422), which takes both insights from virtue ethical globalism and situationism to inform it. The starting point of Miller's

---

[50] For an overview of the project and main contributors, see http://www.thecharacterproject.com/

account of character is the empirical data.[51] In fact, he argues that the empirical data does not support the conclusion that agents generally possess the virtues and vices. More specifically, there is reason to believe that only few moral exemplars are virtuous. That is because the empirical evidence does not match the expectations we have of virtuous or vicious agents as the virtues (and vices) require a minimal threshold (cf. Miller 2017a, p. 254). For example, a compassionate person would be expected to help in minor situations such as someone dropping papers. But the evidence shows that most people refrain from helping even in far more serious situations (cf. Miller 2018, pp. 74–75). The same also goes for the vices. The Milgram experiments give no reason to believe that most have the vice of cruelty despite the despicable behavior witnessed in most of the participants. Certainly, the general outcome of the experiments is shocking. But looking at the 18 different versions of the experiment that Milgram conducted, the evidence shows that participants were much less likely to induce lethal shocks when the authority figure abstained from making encouraging remarks to continue. A viciously cruel person, however, Miller maintains, would use the opportunity to do cruel things, regardless of the cues of an authority figure. Further, most participants in the Milgram experiment had severe internal conflicts. This contradicts the general notion of vices: a cruel person would be comfortable doing cruel things. Yet, most people reported that they felt distress throughout those experiments (cf. ibid., pp. 95-99). Thus, Miller holds that most are neither virtuous nor vicious (cf. 2014, pp. 92–94).

But simultaneously, there is also little support for the situationist claim that behavior is not cross-situationally consistent. On the contrary, empirical experiments show that people help in many situations. And, even though the experiments suggest high behavioral fragmentation at first glance, a closer look reveals behavioral patterns (cf. Miller 2009, pp. 262–263). Hence, the situationist claim that virtue ethical conceptions of character are proven wrong by the empirical data, simply is not true.

---

[51] In *Character & Moral Psychology* (2014), Miller provides a vast analysis of the empirical data and offers a fresh and intelligible re-interpretation.

Analyzing people's behavior across multiple situations provides reason to believe that behavior is holistic rather than fragmented. It thus indicates that there is evidence for stability and consistency in character (cf. ibid., pp. 247-275). Still, Miller acknowledges the widely held view that environmental features play a significant role in influencing behavior (cf. 2018, p. 148). However, this does not confirm the situationist claim about the lack of cross-situational consistency. The fragmentation in behavior observed in several studies might be explained by the discrepancy between the nominal standpoint, from which behavior might look inconsistent and the first-person perspective. Differences, he suggests, can be accounted for by the way in which agents perceive others (cf. Miller 2014, pp. 54–61). In consequence, Miller concludes that agents do possess Global Traits but these are "(…) a far cry from the traditional virtues like compassion" (2009, p. 248).

## 4.2 Fine-Grained Dispositions

Miller challenges whether (virtue ethical) Global Traits and situationism-derived Local Traits offer the best explanation of what most people are like on the grounds of the available empirical data. Re-evaluating both, he concludes that neither virtue ethical globalism nor the situationist critique - and consequently Local Traits - succeed in providing a theory which accurately describes what is evidently found in experiments. Miller argues that both initially raise correct claims but fail to bring them together in a sufficient manner. Based on the re-interpretation of experiments and subsequent re-assessment of the claims made by Global Traits theorists and situationists, Miller builds his own framework, Mixed Traits. The studies show that many will behave admirably in many situations but also deplorably in many others. Even more so, most will switch from one to the other in different situations (cf. Miller 2018, pp. 144–146). This strongly indicates that agents are generally neither virtuous nor vicious. Yet, the studies are compatible with the conclusion that agents have global, i.e., stable and consistent, character traits (cf. Miller 2014, pp. 52–57). Thus, he maintains that we

ought to rid ourselves of the terms as well. For instance, instead of the vice of cruelty, what most people possess is the Mixed Aggression Trait.[52] Similarly, most adults have the Mixed Helping Trait[53] relating to helping behavior (cf. Miller 2013, p. 154). Miller stresses that the Mixed Traits model does not introduce virtues and vices through the backdoor:

"It is important to be clear about the sense in which this trait is "mixed". The claim is not that the Mixed Aggression Trait is a virtue in some situations or contexts, and a vice in others. Rather the claim is that this trait is not a virtue in any situations or contexts. Nor is it a vice in any situations or contexts" (2017a, p. 256).

Notwithstanding the global feature of Mixed Traits, Miller maintains that character traits are made up of interrelated dispositions. On this view, character traits consist of fine-grained interrelated dispositions pertaining to specific moral domains which are best described as mixed. Thus, in contrast to Global Traits, which are considered coarse-grained, and Local Traits, which are narrower, the structure of Mixed Traits is actually fine-grained. In fact, trait dispositions themselves have underlying mental state dispositions (cf. Miller 2014, pp. 24–25; Miller and Knobel 2015, pp. 25–26). They are related to trait dispositions as they enable that trait: "This underlying *causal base* of the disposition, in other words, includes dispositions to form certain trait-specific beliefs and/or desires in the person's mind" (Miller 2014, p. 26). For example, Francine might be disposed to believe or desire something that is not currently occurrent to her, such as the day having 24 hours or that her favorite singer will star at a festival (cf. ibid., pp. 23-24). Hence, Miller concludes, „(…) our characters are piecemeal and fragmented (…)" (2018, p. 159).

However, it is important to note that these trait dispositions are not always active but are rather triggered as environmental features play a significant role in their manifestation (cf. Miller 2014, p. 88). Triggers are situational cues perceived by the

---

[52] For an outline of the Mixed Aggression Trait see chapter 2 in *Character and Moral Psychology* (Miller 2014).

[53] For an outline of the Mixed Helping Trait see chapter 7 in *Moral Character* (Miller 2013).

agent, which can enhance or inhibit some traits. The high sensitivity to situations may cause different behavior depending on each situation (cf. ibid., p. 100). For example, Judith may be angered by her friend, making it far more likely for her to behave aggressively. However, Miller is at pains to stress that the triggers he stipulates for Mixed Traits are different from situational factors. While situational factors induce behavior at that particular moment, triggers can lead to behavior in that situation but also other situations (cross-situationally) as long as the trigger remains active (cf. Miller 2009, p. 251). Instead of either being honest proper like on Global Traits or honest-when-talking-to-her-mom like on the Local Traits model, Hannah simply possesses the Mixed Trait that pertains to 'truth-telling.' She might be more inclined to speak the truth when she is positively triggered (enhanced) by an inspirational book she might have just read about the power of honesty, or just because she received a good grade at school and feels happy. Then, on other occasions, she is inhibited from being honest to her friend, as she is angry about an earlier comment by that particular person (cf. Miller 2018, pp. 142–165).

Despite the fine-grained makeup of Mixed Traits, Miller believes that agents, similarly to other character models, develop Mixed Traits through habituation. There are more and less promising strategies to cultivate good character. Among the more promising strategies are using moral role models, selecting situations, and "getting the word out." For example, role models can inspire us to do better in school, be more empathetic to those in need, or have a more positive outlook in life. And just as much as role models may enhance some character strengths, selecting situations can be helpful as well. Instead of putting oneself in situations that have the potential of temptation, we can make choices to avoid these situations. However, this strategy has its limitations as many situational influences are unconscious. Lastly, "getting the word out" can help improve our character by reminding ourselves of our desires and therefore overcoming bad patterns. Miller envisions this strategy as a combination of self-education and pause of sorts. Before passing by the person who drops her papers, we can ask ourselves what we already know of these types of situations and re-

evaluate what we should do. In contrast to selecting situations, this strategy is not about avoiding certain situations but being more aware once we are in them (cf. ibid., pp. 169-216).

Irrespective of the individual success of these strategies, Miller's account of character stands out "(…) because of its serious empirical credentials" (Bates 2015, p. 422). The view provides a plausible explanation of why behavior can differ in situations, but character may still be consistent. Miller explains this with the structure of Mixed Traits that are essentially made up of trait dispositions, which themselves have underlying mental state dispositions, giving rise to trait-relevant desires and believes. But if we can explain someone's mental makeup entirely by dispositions, the critic might be tempted to ask what traits add? It is not at all obvious why there is any need for Mixed Traits when much of the "work" is done by mental state dispositions (cf. Miller 2014, pp. 24–25; Bates 2015, pp. 422–423). Assuming our characters are as fine-grained as Miller claims, Mixed Traits do not seem to add any value. Still, Miller is not concerned. Even if Mixed Traits add nothing of substance to trait dispositions, they still serve as a general framework to cluster them. By conceiving of these interrelated dispositions as traits pertaining to specific spheres, we gain additional predictive and explanatory power (cf. Miller 2014, pp. 25; 27). With this easier grasp we have the proper vocabulary and "(…) a deeper and more psychologically satisfying explanation using familiar mental categories that are already found to be illuminating in explaining behavior" (ibid., pp. 27-28).

# 5. Conclusion

This chapter briefly outlined three accounts of character found in the philosophical literature. The reason being that many analyses of character responsibility focus on one particular conception, (virtue ethical) globalist models. However, mainly two competing models have been proposed ever since that add to the standard view. The

depiction of the three theories reflects the historical development of the debate on character starting with the globalist account to encompassing Local Traits and lastly a re-interpretation of the research that has been done over the years, mounting in Mixed Traits. Each of the accounts is a result of dissatisfaction with the previous one. Nonetheless, all three accounts share some assumptions and build on the insights gained in the previous chapter. For example, on all three accounts character is seen as temporally stable. Further, character traits are conceived of as dispositions that have explanatory and predictive power. And, lastly, character is a matter of habituation (though this is only implied on the Local Traits account).

The views outlined share some essential features, but it is important to note the differences. The accounts provide views of character that differ with regards to the degree of individuation. Global Traits are coarse-grained features of the agent, which share similarities with common sense conceptions of character. Global Traits are often described as broad because they are applicable across a wide range of situations. That is because they are temporally stable and cross-situationally consistent. This coarse-grained conception of traits differs from the Local Traits account. Situationism challenges the globalist claim about cross-situational consistency. The alternative Local Traits view is similar to Global Traits inasmuch as Local Traits are temporally stable features of a person. But in contrast to Global Traits, they are more fragmented, pertaining to individual situations only. On the Local Traits account, character traits are not cross-situationally consistent. Consequently, they are more localized and narrower. Lastly, Mixed Traits are made up of many interrelated dispositions pertaining to particular moral domains. Thus, Mixed Traits are much more fine-grained than Global and Local Traits.

# III Theories of Moral Responsibility

## Guidance Control, Evaluative Judgment and Origination

---

## 1. Introduction

The aim of the prior two chapters was to outline basic features of character and the general theoretical landscape, including alternative conceptions. The reason for this is to address one aspect of the research gap, namely that most of the research on character responsibility *predominantly* focuses on globalist, coarse-grained conceptions of character and does not explicitly refer to recent advances in character theory. In order to acknowledge these advances, part one of this thesis established a classification, clarified the dispositional, yet manifesting, nature of character, and introduced three prevailing accounts found in the literature. With these different theories of character at hand, this chapter progresses to the second dimension of the research question, the responsibility aspect. However, a similar challenge emerges: the majority of analyses of character responsibility primarily addresses the issue by means of discussing whether agents have *control* over their character, thereby adhering to a specific interpretation of responsibility. Depending on the assessment thereof, they argue in favor or against character responsibility. Nevertheless, it is important to stress that the responsibility landscape is just as much as, if not more, diverse than the character debate. Thus, this chapter, analogously to the previous one, provides an overview of various theories of responsibility.

For some time, there had been wide agreement in the philosophical debate that there are usually two distinctly necessary and combined sufficient conditions of responsibility: control and awareness (also known as the freedom-relevant and epistemic/ knowledge/ cognitive/ mental condition) (cf. Barnes 2016, pp. 2311–2312).

One strand in the debate, which focusses on just that, is *volitionism*. Volitionist accounts take control and awareness to be necessary conditions of responsibility. Another strand, however, *attributionism*, questions whether this is the case. They argue that for an agent to be responsible the action must be rightly attributable to her. Mainly, they maintain that agents often hold beliefs or have certain attitudes, which they do not control, but these still reflect the agent's quality of will (cf. Smith 2008, p. 371).

Until fairly recently, the focus in debates was the control condition, with the epistemic condition receiving much less attention than the former (cf. Sher 2009, pp. 3–4). But lately, the latter has attracted more attention, leading to the development of a third position, called *capacitarianism*. Capacitarians believe that neither volitionists nor attributionists can account for a specific type of unwitting wrongdoing[54]: agents can be responsible for *forgetting*, even though they generally do not control what they forget, nor does forgetting necessarily reflect their quality of will. To show that agents can be responsible for forgetting and the consequences thereof, capacitarian approaches suggest that the reason agents are responsible for these unwitting omissions is that they have capacities which make it reasonable to expect them to realize that they could and should have not forgotten (cf. Rudy-Hiller 2018). Thus, to account for the wide array of theoretical approaches to responsibility, this chapter serves to introduce its main strands. To illustrate these positions, the chapter is structured accordingly: each section gives an outline of the respective view and its main claims regarding the control and awareness condition. However, since some approaches negate that control and/or awareness are a necessary condition, the sections follow the argumentative flow of each. Hence, I will discuss the conditions separately where applicable, but refrain from doing so rigorously.

After providing an overview of the family of theories, I select a specific theory that is exemplary of the respective view and best illustrates its main claims. Selecting a specific theory helps advance this thesis because the subsequent chapter will explore

---

[54] Cases of unwitting wrongdoing are best described as those in which: „(…) the agents aren't aware on these occasions that they aren't doing the things in question" (Clarke 2017a, p. 63).

what the representative accounts can say about the responsibility agents have for their character traits. By looking at individual accounts rather than families of theories we can gain a clearer picture of what character responsibility looks like in detail without discarding the possibility to make inferences about the family of views respectively. The view most representative of volitionism is that of John Martin Fischer and Mark Ravizza. Fischer and Ravizza (2000) argue that for agents to be responsible, they need guidance control. Matthew Talbert's (2009, 2013, 2017a) view is exemplary of attributionist thinking. He maintains that for an agent to be blameworthy, the action must reflect the agent's evaluative judgment. Lastly, George Sher (Sher 2006a, 2009) proposes a capacitarian account which stresses the suitable connection between the action and the agent herself.

## 2. Volitionism

### 2.1 Historicism and Tracing

One natural intuition is that responsibility presupposes control. One side of the academic debate has focused on just that: volitionism is a position that holds that for agents to be responsible, they need control. The element of control is primary to any rightful assessment of responsibility on this type of view. Depending on context, volitionists stress the agent's prior, present or future choice, i.e., either past chosen actions make something attributable to that person, their current voluntary endorsement or that they may be able to voluntarily choose in the future (cf. Smith 2005, pp. 238–240). Recall the example given in this thesis' introduction of your friend Hannah who smashes a vase against the wall in a heated argument. Now imagine an alteration of the scenario: Hannah has no intention of breaking anything but simply wants to take a close look at the vase. However, in this variation of the case she lets

loose as she is shaken from a sudden earthquake. It now seems that Hannah is no longer responsible for smashing the vase because she seems to lack control.

The claim that control is a necessary condition of responsibility is essential for volitionists. This view not only reflects this natural intuition many have about responsibility, proponents of volitionism take this claim to explain a problem frequently raised in the debate: agents who have been subject to manipulation do not seem to exert the right kind of control. This also goes for agents who are "(…) raised in an environment that made it difficult to avoid acquiring the bad values that inform her judgments" (Talbert 2017b, p. 19). On the volitionist view, it simply seems that agents, that suffer from manipulation, brainwashing, or hostile environments in their upbringing, are not responsible (cf. Fischer and Ravizza 2000, p. 187). Think of Hannah again, who throws your vase against the wall in a heated argument in the initial case. Now assume that Hannah was typically of mellow nature but an evil scientist was able to plant a device in her brain overnight which would render her especially aggressive; it now seems that she is not responsible anymore (cf. Mele 2020, p. 3153). This is why volitionists typically introduce a strategy that revolves around the historical acquisition of the agent's values (historicism). To illustrate what is meant, volitionists often invoke the example of the drunk driver, who might not be directly responsible for crashing into another car, but who is responsible for getting drunk in the first place (cf. Vargas 2013a, pp. 267–301). In the case of the drunk driver, it is very intuitive to see why she is responsible for the accident: she had control over whether she would intoxicate herself knowing that driving under the influence would leverage the risk of harming others. Hence, volitionists maintain that agents must control the process of acquiring their current makeup. By introducing the historical condition, volitionists are able to maintain that the agent can be responsible when she had the relevant control at a certain point in time, whilst asserting that manipulated and brainwashed agents are not responsible as they lack historical control (cf. Mele 2020, p. 3153).

Regarding the control condition of responsibility, volitionists maintain that an agent needs some kind of (historical) control to be responsible. And similarly,

regarding the epistemic condition, volitionists hold that agents require some kind of awareness. Agents, who lack awareness, are not responsible, unless their ignorance was culpable. Recall the civil engineer, whom we have also encountered in the introduction, who miscalculates relevant portions of a building, which collapses subsequently. She might, for example, have thought about the very real possibility of miscalculation but decided not to have her work double-checked. Volitionists contend that only if the ignorance of action is the result of another act from ignorance, the agent is exculpated. Thus, the civil engineer is responsible for her miscalculation from ignorance only if that ignorance stems from "(…) some prior failure to discharge one of [her] procedural epistemic obligations" (Rosen 2004, p. 303). Most volitionists thus argue that the present ignorance must be due to some culpable, "benighting"[55], act (cf. Smith 1983, p. 547; Zimmerman 1997, p. 412; Levy 2009, p. 730). This benighting act is akratic as the agent must know that her act is bad, yet decide to still do it (cf. Ginet 2000, p. 273; Rosen 2004, p. 307). Hence the agent

„(…) not only *recognizes* the very real risk of later memory or attention failure, but also *embraces* the risk by deciding to do something that will increase, or by deciding not to do something that will decrease, the probability of such failure" (Nelkin and Rickless 2017, p. 113).

Thus, volitionists conclude that the agent needs occurrent (cf. Zimmerman 1997, pp. 421–422) and de dicto beliefs about the action's overall wrongness (cf. Rudy-Hiller 2018). Agents that lack this kind of awareness are exculpated. There are differences in how proponents of this view spell this out exactly - for example some argue that to be responsible, an agent must be conscious of the facts that give an action its moral significance (cf. Levy 2014, p. 37) while others propose that instead of conscious agency, an agent needs to have had a past opportunity to avoid unwitting omissions (cf. Nelkin and Rickless 2017, p. 116) - but what unites these views is the requirement of being able to trace back the current bad action to a prior non-ignorant failure in time. As an agent can only rightfully be responsible on this view if we can

---

[55] The term was originally coined by Holly Smith (1983).

trace back her bad act to a prior benighting act, which she was aware of, these accounts are often referred to as *tracing accounts.* However, it is important to note that tracing is not the default explanation for volitionists, as it is not needed in ordinary cases. Rather, tracing is employed to account for those cases, in which the agent is apparently not aware currently or lacks sufficient control (cf. King 2017, p. 266).

## 2.2 Fischer and Ravizza: Guidance Control

The appeal of volitionist theories derives from them accounting for the potent intuition that only those agents are responsible, who have some sort of control over and awareness of their actions. This is also, why for the longest time the majority of debates on responsibility was contextually situated in questions of free will (cf. Talbert 2023). Since the intuition that control is necessary is so intuitively appealing, the obvious adjacent question is, whether we have free will at all and thus control our actions. Fischer and Ravizza (2000) are prominent advocates of volitionism and have dedicated much of their work to the the idea that responsibility presupposes control. Their work is exemplary of volitionism and neatly joins the discussion against the backdrop of the debate on free will. One potential problem - and probably the most threatening - to free will is (causal) determinism. Causal determinism holds that every event in question is caused by a prior event. Knowing of these events in combination with the laws of nature would make every future event predictable (cf. Fischer 2006, p. 322). In fact, it would entail that there is only one possible future (Kane 2007, p. 5). But how can we square the truth of causal determinism with free will? In providing an answer to this question, essentially two camps have emerged: incompatibilists - as the name

suggests - maintain that freedom of will is incompatible with causal determinism.[56] On the contrary, compatibilists believe that the truth of causal determinism does not matter much to our responsibility practices. They defend the view that even if causal determinism is true, it is still compatible with agents having free will. This in turn has led to a somewhat deadlock discussion between incompatibilists and compatibilists. Both share the view that control is a necessary condition of responsibility and that the absence of control would seriously challenge our status as responsible agents. However, while incompatibilists are eager to show that causal determinism is a real threat to responsibility, compatibilists make their case for their view that causal determinism does not have these implications for responsibility (cf. McKenna and Coates 2021).

As compatibilists, Fischer and Ravizza want to avoid the threat posed by causal determinism and thus aim to find a viable solution that can make way for this particularity. Therefore, their account tries to settle the question of what kind of control exactly an agent must exercise to be responsible. Their strategy is to maintain that control is indeed a necessary condition of responsibility but that the control in question does not require alternatives for the agent to choose from (cf. Fischer and Ravizza 2000, pp. 29–30). Thus, they target the *Principle of Alternative Possibilities*. The principle holds that an agent must have *alternative* possibilities in order to exercise control over her actions (cf. Fischer 2006, p. 334; McKenna and Coates 2021; Robb 2023). See how this resonates with the intuition about cases such as the friend shaken from the earthquake described at the beginning of this section. It seems that our friend lacks responsibility for destroying the vase precisely because she had no alternative:

---

[56] Incompatibilists come roughly in two camps: hard determinists, such as Pereboom (2014) (though he refrains from calling himself a hard determinist in earlier work and states that he remains agnostic to the cause (see 2007)) and libertarians, such as Robert Kane. Hard determinists maintain that causal determinism is true and that we can thus not have free will. Libertarians, too, hold that free will and causal determinism are incompatible. However, they believe in the existence of free will and that causal determinism is false (cf. Kane 2007).

"Typically, we think of ourselves as responsible precisely in virtue of exercising a distinctive kind of freedom or control; this freedom is traditionally thought to involve exactly the sort of 'selection' from among genuinely available alternative possibilities (…)" (Fischer 2007, p. 46).

Yet, the problem of causal determinism illustrates that this might not be an actual possibility. Hence, Fischer and Ravizza argue that there are two types of control: r*egulative* and *guidance* control. Agents who have alternative possibilities and control their actions have regulative control. To see how regulative control plays out, imagine an agent who drives in her car and encounters a crossroads. She wants to get to work and needs to make a right turn. Thus, she turns her steering wheel to the right. Assume that she has formed an intention to turn right but could go to the left just as well. She is in control over the car and its movements (cf. Fischer 2013, p. 6). Insofar as the agent has the power to guide the car in any direction she likes, Fischer and Ravizza maintain, she has regulative control. But regulative control is not the kind of control *necessary* for responsibility. Instead, Fischer and Ravizza, argue that one needs to look closer at the actual sequence of actions (cf. 2000, p. 37).[57] The following analogy to guidance control outlines how the actual sequence of actions may come into play: imagine yet again an agent, who plays with racecars. She has a controller with which she can make the cars accelerate and a joystick with which she can guide the car along the play tracks. The racetracks are lined out, the cars can only take the route of the tracks. This agent cannot choose from alternate paths for her racecar to go. If the racetracks make a left turn, her

---

[57] Fischer and Ravizza base this claim on a groundbreaking thought experiment introduced by Harry Frankfurt (1988), which challenges the idea that agents need alternative possibilities to be responsible. Many compatibilists take Frankfurt's argument to show that an agent may exercise control without having alternative paths available to her. Frankfurt's example goes as follows: Jones wants to murder the city's mayor. He tells his friend Jack about his plans, who is just as enthusiastic about Jones' plan. However, Jack is not convinced that Jones will go through with his plan. Thus, unbeknownst to Jones, he employs a secret device in Jones' brain, which monitors all of his brain activities. If Jones were to hesitate just for a fraction of a second, the device would ensure that Jones carries out his original plan. At the next town meeting, Jones does not show any signs of hesitation and kills the mayor. As it turns out, Jones does not change his mind. But even if he had, the device would have ensured him carrying out the plan, leaving him with no alternative. Thus, Jones does not satisfy the requirement of alternative possibilities but still is responsible. Frankfurt uses this example to trigger the intuition that Jones is responsible for killing the mayor even though he had no alternative course of action. Hence, Frankfurt concludes, that alternative possibilities are actually not necessary for responsibility (cf. 1993, pp. 286–287).

racecar, too, can only take a left turn. Thus, she does not have regulative control, as she does not have the power to guide the car in any direction she likes. Instead she has what Fischer and Ravizza dub *guidance control* (cf. Fischer 2013, p. 6). In the crossroads-example above, the agent can turn any direction she chooses to get to work. She can either go along her standard path or instead turn another direction. In the racecar-example, she can only drive along the tracks. Yet, in both scenarios, the agent can freely perform her actions. But while regulative control amounts to free choice, guidance control refers to control over an action independent of alternate possibilities (cf. Fischer and Ravizza 2000, pp. 31–32). Since Fischer and Ravizza have ruled out that regulative control as necessary for responsibility (and generally follow a compatibilist approach), they claim that the freedom-relevant kind of control required for responsibility is guidance control (cf. Fischer 2013, p. 7).[58]

But how are we to consider guidance control as a real form of control when the agent possibly has no other alternative? The answer lies in the view that

"(…) the *agent chooses freely, and acts in accordance with [her] choice* [emphasis S. Sch.], in just the way [s]he would have, had there been no 'counterfactual intervener' at all. The Frankfurt-type examples [see footnote 63; S. Sch.] highlight the fact that, as long as no responsibility-undermining factor actually operates, an agent may be responsible, even though such a factor would have played a role in the alternative scenario (and thus the agent lacks alternative possibilities)" (Fischer and Ravizza 2000, p. 38).

Nonetheless, Fischer and Ravizza stumble upon a potential problem: it is easy to see why agents that are subject to manipulation, hypnosis or brainwashing are intuitively not responsible. Consider Francine who gives money to charity. We may genuinely praise her for her action. However, it would seem odd to praise Francine if her decision were to stem from brainwashing or hypnosis. That is because agents under these types of duress are "like marionettes" (2000, p. 36) and thus not responsible. More specifically, these cases are instances of exemptions or excuses

---

[58] Fischer and Ravizza note that regulative and guidance control often go together. However, there are instances, as illustrated by the Frankfurt-type examples, in which an agent only has guidance control and retains responsibility (cf. Fischer and Ravizza 2000, p. 32).

respectively.[59] But why is this so? Have Fischer and Ravizza not committed to the claim that agents do not need alternative options to be responsible by ruling out regulative control (cf. 2000, pp. 28–36)? At first glance, it seems that contrasting cases of hypnosis and the like with those "(…) in which there is the 'normal,' unimpaired operation of the human deliberative mechanism[60]" (ibid., p. 36) shows that only those agents are responsible, who have regulative control. Thus, to restore their claim that responsibility only requires guidance control (which entails that the action in question may be brought about by external features), Fischer and Ravizza need to square it with cases in which agents, who are subject to instances of external duress such as hypnosis and the like are not responsible. So, how can Fischer and Ravizza accomplish this task? Their strategy is twofold: first, they argue that cases such as hypnosis, brain washing or irresistible urges undermine guidance control, as the agent is *not responsive to reason*. Second, as volitionists, they appeal to a condition which takes the history of the agent to be central to her responsibility. Agents under the types of duress just mentioned fail to exhibit the right kind of history (cf. Fischer and Ravizza 2000, pp. 230–231). Let us take each into consideration:

Fischer and Ravizza argue that an agent has guidance control when her mechanism is reasons-responsive. There are multiple ways to interpret reasons-responsiveness. According to Fischer and Ravizza, an agent may be strongly reasons-responsive or, to the contrary, weakly reasons-responsive. For a mechanism to be strongly reasons-responsive, the agent would - in light of sufficient reason to do otherwise - "(…) *recognize* the sufficient reason to do otherwise and thus *choose* to do

---

[59] Exemptions pertain to those agents who are hypnotized, under great strain or not a member of the moral community (e.g., small children or mentally ill). Excuses on the other hand pose no general threat to the responsibility of an agent. Rather, these refer to single instances due to accidents, ignorance or force. So while exemptions pertain to the agent's capacity as a responsible agent in general, excuses are limited to specific events (cf. Fischer and Ravizza 1993a, p. 20).

[60] Take note that Fischer and Ravizza shift their wording from "agent" to "the agent's mechanism." This is due to their argument following the Frankfurt-type examples. Recall that the agent in the actual sequence is reasons-responsive, while in the alternative scenario - which does not take place - she would not be, as she would be manipulated into carrying out the action. This is grounds for Fischer and Ravizza to focus on the mechanism, which leads to the actual action rather than the agent. Only agents with reasons-responsive mechanism are responsible (cf. Fischer and Ravizza 2000, p. 38).

otherwise and *do* otherwise" (ibid., p. 41). This entails that there are three options in which an agent's mechanism can fail to be strongly reasons-responsive. Most fundamentally, the agent can fail to recognize something as sufficient reason. Fischer and Ravizza take those suffering from delusional psychosis to be candidates of this sort. A second possibility, in which an agent's mechanism is unresponsive to reason, is a failure to choose to act upon a reason, even though she may recognize it as sufficient reason. This particularly goes for compulsive or phobic neurotics but also those with weakness of will. Though Fischer and Ravizza admit that there are cases of weak-willed actions for which the agent can be held responsible. Third, an agent's mechanism is not strongly reasons-responsive if she recognizes something as sufficient reason, chooses to act on that reason but fails to put that choice into action. Agents, who pertain to this sort, include those with physical incapacities and, again, those with weakness of will. However, Fischer and Ravizza argue that responsibility does not require strong reasons-responsiveness, mainly because it would render agents not responsible who we typically take to be responsible. For example, a thief may know (*recognize)* that stealing is wrong. She has sufficient reason not to steal the book but does so anyhow. Now having sufficient reason to refrain from stealing the book, the thief would fail to choose an action in alignment with that recognition and hence not be responsible. This clearly contradicts the intuition most people have about these cases (cf. ibid., pp. 41-45). Hence, Fischer and Ravizza test how weak reasons-responsiveness fares. In the case of weak reasons-responsiveness, again, an agent has sufficient reason to do otherwise, the agent chooses to do otherwise and does otherwise. However, the authors now look at *some possible scenario* in which the agent has this reason:

"Under the requirement of strong reasons-responsiveness, we hold fixed the actual kind of mechanism and ask what would happen, if there were a sufficient reason to do otherwise. In contrast, under weak reasons-responsiveness, we (again) hold fixed the actual kind of mechanism, and we then simply require that there exist *some* possible scenario (or possible

world) in which there is a sufficient reason to do otherwise, the agent recognizes this reason, and the agent does otherwise" (ibid., p. 44).

If we replace the requirement of strong reasons-responsiveness with weak reasons-responsiveness, we can readily see that the thief may be responsive to reason and is thus responsible. For example, if she knew that her friend will give her the book which she was about to steal as a birthday gift a week later, she had plenty of reason to refrain from stealing. Therefore, Fischer and Ravizza claim that weak reasons-responsiveness is necessary for responsibility. However, weak reasons-responsiveness does not capture the range of all cases as "(…) the envisaged fit [between reasons and action], is, unfortunately, too loose" (cf. ibid., p. 65) and is thus not sufficient. Why this is so, we need only think of the thief again, who knows that stealing is wrong and does so, nonetheless. If she steals the book at any price but decides to buy it instead if it costs exactly $4000 (but would steal again if it cost $4001 and so forth), we have difficulty comprehending that particular reason but may also be puzzled by her general pattern of responding to reasons. Hence, weak reasons-responsiveness does not suffice because it takes agents to be responsible who fail to display an understandable pattern of reasons (cf. ibid., pp. 66-68).

Therefore, Fischer and Ravizza maintain that *moderate* reasons-responsiveness is both necessary and sufficient for responsibility. In order to exemplify this view, they stress the difference between recognizing reasons (receptivity) and reacting to reasons (reactivity).[61] Moderate reasons-responsiveness requires that the agent is strongly reasons-*receptive* and weakly reasons-*reactive*. More specifically, strong reasons-receptivity requires that agents have a general, understandable pattern of reasons (cf. ibid., pp. 68-71). Agents must "(…) recognize certain reasons as moral reasons" (Fischer 2013, p. 187). But it does not mean that the agent must actually act on those reasons in the actual scenario. Instead, it suffices for the agent if she would react to the

---

[61] For purposes of simplicity, the authors take reasons-reactivity to include choosing a reason and acting on a reason (hence converging two steps which were discussed separately prior) (cf. Fischer and Ravizza 2000, p. 69).

reasons in a possible scenario. Still, if strong reasons-*receptivity* is required, why does weak reasons-*reactivity* suffice? The reason for the asymmetry lies in the view that if an agent would have reacted differently to a sufficient reason to do otherwise in some possible world, it shows that the actual mechanism in place has the capacity to react differently to the actual reason. This general capacity, Fischer and Ravizza argue, grounds responsibility (cf. 2000, pp. 73–74).

Coming back to our initial challenge, one reason for Fischer and Ravizza to maintain that manipulated agents are not responsible is to state that their mechanism is not reasons-responsive. That is because the agent would act as she does independent of the reasons she had. Hypnotized agents, for example, do not respond to reasons even if they were confronted with very strong reasons (cf. ibid, pp. 36-37). However, there is further reason to believe that manipulated agents are not responsible. That is because in accordance to Fischer and Ravizza, guidance control also requires ownership of the reasons-responsive mechanism (cf. Fischer 2013, p. 187). Hence, an agent has guidance control when the action is a result of her *own* reasons-responsive mechanism. To ensure this connection, Fischer and Ravizza argue that responsibility is an essentially historical notion (cf. ibid., p. 186). That is, that the agent's makeup is a function of the agent's history: "(…) someone's being responsible requires that the past be a certain way" (Fischer and Ravizza 2000, p. 207). An agent's past is this certain way if she has *taken responsibility* when she was a child. She does so by understanding that she is the source of actions and that these actions have consequences. She must also come to acknowledge that based on these actions, she may be the proper target for the reactive attitudes. Based on experiencing that she is the source of actions, that they have consequences and that she may be held responsible, the agent takes responsibility (cf. ibid., pp. 210-213). It is a process that originates in the agent by which they come to realize that their actions shape who they are. Taking responsibility is not a single verbal commitment or action rather it is an accumulation of beliefs. Nonetheless, it is important to note that it is not something an agent decides or does consciously (cf. ibid., pp. 217-219).

With this two-step solution, Fischer and Ravizza are certain that they can account for the problem of manipulation, which raised the question of how the authors can keep up the claim that agents only need guidance control, not regulative control. Now we can see that an agent who has been manipulated is not responsible because she would act as she does independent of what reasons she had. Further, she has not *taken* responsibility and thus lacks historical control (cf. ibid., p. 232).

So far, the spotlight has been on the control condition of responsibility on Fischer and Ravizza's account. And this reflects much of the work on responsibility by these authors. Generally, Fischer and Ravizza do not elaborate a great deal on their stance on the epistemic condition, though they do acknowledge its relevance. Nonetheless, their focus lies on the control condition (cf. Fischer 1999, p. 99). However, some passages reveal that they favor a volitionist *tracing approach*, which is connected to the reasons-responsiveness dimension of guidance control: "(…) we must be able to *trace back* [italics S. Sch.] along the history of the action to a point (…) where there was indeed an appropriately reasons-responsive mechanism" (Fischer and Ravizza 2000, pp. 50–51). How this tracing approach comes into play, is again best illustrated by the example of a drunk driver mentioned above, who kills an innocent child. The drunk driver might not be reasons-responsive at the time of the accident; however, she is reasons-responsive prior to drinking and getting into the car after (cf. ibid., p. 50). Admittedly, this type of tracing puts the focus on a period in time where an agent had guidance control (cf. Fischer and Tognazzini 2009, p. 533)[62], yet a closer look reveals that guidance control includes epistemic elements as the agent must be in a state that allows her to put her unclouded reasons-responsive mechanism to use. This dual setup illustrates how reasons-responsiveness has both a control condition and an epistemic condition built into it:

"An agent must first be able to recognize what good reasons there are and assess them for whether they are sufficient for pursuing a course of action. This is the cognitive dimension

---

[62] In this article, Fischer is much more straightforward about the necessity of a tracing component in a sufficient theory of responsibility (cf. Fischer and Tognazzini 2009).

whereby reasons are 'received.' But second, she must also be able to react appropriately so as to guide her own actions in light of her recognition of the good reasons that there are. This is the volitional dimension whereby an agent reacts to the reasons after she receives them" (McKenna and Warmke 2017, p. 708).

The first part of reasons-responsiveness, reasons-receptivity, then actually reflects the epistemic features of guidance control. Thus, while the focal point of Fischer and Ravizza's account is the control condition, this includes the satisfaction of the epistemic condition (cf. Mele 2010, p. 107). Hence, a closer look at morally responsibility actually reveals that it contains both volitional and epistemic components.

# 3. Attributionism

## 3.1 Attributability, Structuralism, and Anti-Tracing

Fairly recently, contemporaries have questioned the implications of volitionist accounts. Generally, they disagree that only agents, who are in control, are responsible. They support their view by pointing out that volitionist accounts face a problem: it is not hard to see that most agents do not act in clear-eyed akrasia or have consciousness of the action's moral significance. Hence, on the tracing account very few are responsible (cf. Mason 2015, p. 3039).[63] Some are happy to accept this - somewhat revisionist[64] - conclusion (cf. Levy 2008, p. 217, 2017, p. 252), while others think it „(…) is a bitter pill to swallow, and numerous theorists have attempted to achieve better consilience with common sense" (Nelkin and Rickless 2017, p. 107). One such attempt is that of attributionists, who argue that volitionists cannot account for paradigmatic

---

[63] Rosen is more optimistic regarding the possibility of akratic action: "And to the extent that it is actual, it is available to terminate the regress we have identified" (2004, p. 308). However, he concedes that it will usually be close to impossible to pin down a specific act of akrasia (cf. 2004, p. 309).

[64] Put to the extreme, volitionist tracing accounts are associated with skepticism about moral responsibility (cf. Rudy-Hiller 2017, p. 403).

cases of responsibility, such as agents who have certain attitudes or beliefs. Consider for example *forgetting* a friend's birthday (cf. Smith 2005) or holding the *belief* that women are inferior to men (cf. Hieronymi 2008)[65]. In both cases, most will agree that we are in fact open for moral assessment and responsible for being a bad friend or having sexist beliefs respectively. My friend may well be angry with me for forgetting her birthday and we may demand of the sexist that he reconsider the arguments and facts he takes to ground his beliefs in. However, we do not consciously choose to forget our friend's birthday or have a sexist bias (at least in most cases). Nonetheless, we appear just as responsible for passivity as for voluntary choices (cf. Smith 2005). Hence, volitionists will have a hard time to show that an agent had control over implicit attitudes or beliefs (cf. Hieronymi 2008, pp. 363–371), yet attributionists take these to be paradigm cases of exercising responsibility (cf. Smith 2008, pp. 369–370).[66] Attributionists do not deny that control plays a role in responsibility assessments, but they challenge the necessary function volitionists make it out to have. Thus, while the volitionist account implies that agents are far less often responsible than we usually tend to think, attributionist accounts hold that agents are responsible far more often than the volitionist view entails.

The appeal of the attributionist view becomes particularly evident when we consider the many cases in which we take others to be responsible when in fact they had little to no control over their attitudes or beliefs. To account for the intuition that these agents are responsible nonetheless, a number of views have invoked judgment-sensitive attitudes (cf. Scanlon 2000), evaluative judgment (cf. Smith 2005, 2008), or giving reasons (cf. Hieronymi 2008), which are rightfully attributable to the agent. These views can be considered an advancement of what Susan Wolf has coined *real self-views* (1993a) and has also been referred to as *self-disclosure* views (cf. Watson 1996). More recently, these views have been called *attributionist*. The specifics of the

---

[65] Hieronymi does not offer one specific paradigmatic case. Rather, she elaborates on any kind of belief one might hold.

[66] See for example Brownstein (2016), who argues that agents are responsible for their implicit biases because it is attributable on the grounds of being reflective of how one is as a moral agent.

attributionist views diverge; what unites these views, though, is that they all show how the action is attributable to the agent. Angela Smith, for example, argues that forgetting a friend's birthday does so because

"(…) if one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one's tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person. If this is so, then the fact that a person fails to take note of such factors in certain circumstances is at least some indication that she does not accept this evaluative judgment" (2005, p. 244).

Hence, one fundamental claim attributionists make pertains to the connection between the agent and her actions. Attributionists argue that agents are responsible for many cases in which they do not actually have control. And similarly, they maintain that agents neither need a clear-eyed akratic act in order to be responsible. Instead, they argue that agents can be responsible for actions because they are attributable to them. They can be responsible for beliefs and attitudes because it is a display of their quality of will. More specifically, they hold that beliefs and attitudes, which are not necessarily voluntary[67], reflect the agent's quality of will and thus ground the agent's responsibility. Thus, many attributionists employ a Strawsonian line of thought that for agents to be blameworthy, actions have to express their quality of will (cf. Strawson 2008). In the case of forgetting your friend's birthday, you simply did not care enough. It is then about the display of lack of concern (cf. Talbert 2017b, pp. 22–23). Agents such as the civil engineer, who believes that double-checking her work is unnecessary, may simply lack concern for the gravity of her actions and their consequences. Hence, there is no need for a benighting act as culpable ignorance can be direct.[68] In contrast to tracing views then, attributionists do not require tracing back to a point in time in which the civil engineer decided not to re-check her calculations,

---

[67] Note the point of contention about the specific conception of voluntarism at stake. Some argue that the appeal of volitionism derives from its apparent association with bodily actions (cf. Smith 2005, pp. 264–265). However, proponents of volitionism not only refer to direct control, choice, and deliberation, but rather all intentional actions (cf. McKenna 2008, p. 36).

[68] See chapter 4 for a differentiation between direct and indirect responsibility with regards to character.

despite the probability of potential mistakes and grave consequences. Her ignorance to do so does not excuse her on this view. Thus, attributionists are typically *anti-tracers*. However, most attributionists agree that agents need at least some kind of awareness. Recall that volitionist tracers maintain that the agent must have de dicto and factual awareness of the action's moral significance (cf. Rudy-Hiller 2018). Applied to the case of the civil engineer, this amounts to her believing that her calculations are correct. But she also needs to be aware that a miscalculation can injure, if not kill, many. Attributionists usually negate the former (cf. Mason 2015) and affirm the latter (cf. Talbert 2013, p. 226; Arpaly 2015, p. 145; Harman 2017).[69] Blameworthiness, they maintain, does not require awareness of the action being wrong (cf. Arpaly 2015, pp. 146–147). This is because in order for an action to reflect the agent's quality of will, the agent only has to be aware of the factual consequences. Lacking moral de dicto awareness does not exculpate (cf. Rudy-Hiller 2018). If the civil engineer knew that a mistake in her calculations might have fatal consequences but does not belief that double-checking is worthwhile, she would be considered responsible.[70]

While the volitionists' emphasis of the control condition is intuitive at first sight, attributionists account for everyday ascriptions of responsibility. But attributionists are hard-pressed on the matter of manipulation. That is because they negate that control is a necessary condition for responsibility. And the case of manipulation seems to be exactly a case in which an agent has no control over her makeup. Recall that volitionists argue for a historicist position, which requires that the agent has acquired her values in a specific way. This allows them to argue that agents,

---

[69] An exception is Angela Smith (2005, 2008) who argues that the lack of factual awareness (e.g. forgetting a friend's birthday) also reveals a lack of concern and therefore an agent may be responsible given that she is answerable. Not everyone shares this view. Talbert, for example argues that even though there are many cases in which forgetting does not excuse, there are others in which it does. There are instances in which "(…) people forget things about which they care deeply and as much morality requires. In such a case, the forgetful agent ought to be excused" (2017a, p. 58).

[70] For example, slaveholders were aware of all the facts but were morally ignorant of their doing. Despite this potential moral ignorance at the time, they are responsible for holding slaves (cf. Harman 2017). For a discussion of cultural membership, which may diminish the agent's awareness of moral facts, see chapter 5.

who are subject to manipulation, are not responsible. Since attributionists claim that responsibility does not require control, they cannot employ the same strategy. On the contrary, they argue for an *ahistorical* position, also known as *structuralism*. The view is derived from Harry Frankfurt's mesh theory:

"(…) an agent is responsible for an action as long as there is an appropriate mesh between how the agent wants to be moved and the desires that actually produce her actions. Thus, mesh theories are ahistorical: responsibility is simply a matter of the contemporary structure of an agent's will and it does not matter how this structure came about" (Talbert 2009, p. 4).

While volitionists argue that the agent's history is highly relevant to the assessment of her current responsibility, attributionist structuralists, in contrast, deny that this is the case (cf. Mason 2015, p. 3052). Therefore, structuralists must show how a manipulated agent can still possess ill will (cf. Talbert 2009, pp. 2–3). They typically do so by arguing that the manipulated agent's values still reflect herself, specifically *her* quality of will. Hence, on the structuralist view agents are responsible as long as their will is their own (cf. Zimmerman 2003, p. 639).

Despite this explanation, some critics are discontent nonetheless with attributionism as it leaves proponents of the view unable to differentiate between bad and blameworthy agents (cf. Levy 2005, pp. 5–6; McKenna 2008, p. 35). Attributionists typically take bad agents[71] to be blameworthy agents. That is because a moral assessment of an adult being is to make the demand for justification, which "(…) *by its very nature* implies responsibility, for it is directed at [her] judgmental activity, activity for which we must regard [her] as responsible if we are to regard [her] as a moral agent in any sense" (Smith 2008, p. 388). Volitionists on the other hand stress the difference between the two, since on their view responsibility is a matter of having (had) a choice. Take the example of agents who undergo sudden character change (e.g., cases of mental illness and manipulation): these cases mirror the way in many agents lack control over their constitution. Volitionists can argue that these agents are bad but they

---

[71] Note that this term actually reflects an imprecision as it may refer two meanings, namely to a "'bad person who acts' and 'person who acts badly'" (Sher 2006a, p. 9).

are certainly not blameworthy ones (cf. Levy 2005, p. 8). However, attributionists argue that even though they generally do not differentiate between bad and blameworthy agents, their view allows for some specific cases in which there is a slight difference between a negative assessment of an agent and them being blameworthy, for example animals and small children. In fact, they point out that this case counts in favor of their view. That is because animals and children can have bad dispositions but are not responsible and thus blameworthy. Yet, the same cannot be said for "regular" agents because saying that someone's judgment is bad whilst holding that they are not responsible, would be to regard her as someone not to be reasoned with (cf. Smith 2008, p. 388). Hence, attributionists maintain that their view captures this essential difference.

The disagreement between volitionists and attributionists has led the former to express a further, more fundamental, critique, namely the worry that the attributionist view can only account for superficial types of responsibility. Critics argue that real self-views do not offer an account of *deep* responsibility (cf. Wolf 1993a, 1993c). This in turn has led Gary Watson in a landmark essay to differentiate between *Two Faces of Moral Responsibility* (1996): accountability and attributability.[72] In his essay, he re-affirms the criticism, especially the need to clarify that real self-/ self-disclosure-/ attributability-views are not sufficient for practices of holding others accountable as accountability requires control. While, in his view, accountability addresses questions of fairness, attributability does not. Thus, the self-disclosure-view, he maintains, is concerned with attributability instead of accountability (cf. ibid., p. 235).[73] Watson's essay has sparked a debate on the depth of attributability accounts. Taking this

---

[72] David Shoemaker argues that we ought to consider a third *face*, namely answerability. Some authors, Shoemaker criticizes, have used attributability in terms of answerability and thus failed to differentiate between both concepts. Attributability, he suggests, pertains to the sense in which an agent deserves aretaic appraisal. Answerability on the contrary reflects the notion of responsibility that is connected to the agent's evaluative judgments, i.e., the reasons she may cite for her course of actions. However, both ought to be kept distinct. Most importantly, Shoemaker concludes that there may not be any unifying theory of responsibility but rather a pluralistic approach (cf. Shoemaker 2011).

[73] See Brink/Nelkin (2013, p. 286) for the same argument.

criticism by face value, attributability accounts would certainly be no alternative to volitionism (cf. Smith 2008, p. 377). However, contemporary attributionists reject this claim as they take their view to imply accountability (cf. Talbert 2017b, p. 18). Nonetheless, Watson's differentiation is interesting because it illuminates the grave differences between those who stress the necessity of control and those who do not.

## 3.2 Talbert: Evaluative Judgment

Matthew Talbert's account is representative of attributionism. He picks up on the claim that there are many paradigmatic cases of responsibility, for which volitionists simply cannot account and which leave them unable to mark many wrongdoers as blameworthy. He argues that it is unintelligible that an agent needs to be aware of the action's wrongness for her to be responsible. But, while an agent need not believe that her action is wrong, Talbert maintains, she still needs to be aware of the possible consequences, i.e., have circumstantial awareness (cf. 2013, p. 242). The reason Talbert takes circumstantial awareness to be important for responsibility is that the lack thereof does not reflect the agent's *evaluative judgment* of the significance of the action. This is why he disagrees with other attributionists (e.g., Smith) who argue that forgetful and inattentive agents are responsible. These agents, Talbert believes, are *not* responsible because their actions do not necessarily reflect their evaluative judgment as they do not know about the obligations thus lacking circumstantial awareness (cf. 2017a, p. 56).[74] In other words, Talbert disagrees with Smith's assessment that forgetting a friend's birthday shows that the agent simply does not care enough. Instead, he argues that if the lack of concern plays no role in the explanation of the agent's forgetfulness, then she is no proper target of blame (cf. ibid., pp. 58-59).

---

[74] As mentioned above, attributionists employ a Strawsonian line of thought which takes the agent's quality of will to be essential. Talbert takes this view to inspire his own (cf. Talbert 2009, p. 2). While Talbert tends to prefer the term *evaluative judgment*, he also at times refers to an agent's quality of will (e.g., Talbert 2012).

However, Talbert does not believe that the same holds true for agents who lack awareness about the action's wrongness itself. Therefore, agents, who lack moral awareness, can be blameworthy. That is because it displays a disagreement between the agent and those who object her behavior. Hence, while an agent needs circumstantial awareness to be responsible, moral ignorance does not exculpate (cf. Talbert 2013, p. 236). For an agent to be responsible, the action needs to reflect her evaluative judgment. And for an action to reflect her evaluative judgment, the agent needs circumstantial, yet not moral awareness:

"(…) awareness that one's actions are wrong is not necessary for the presence or expression of such judgments. Much more relevant in this context is knowledge of the effects that one's actions will have on others. If a person knows that her action will injure someone who objects to this result, then it is reasonable to attribute to her the judgment that the other's injuries and objections are not a decisive reason to refrain from the action. Judgments like this help to ground blaming responses because of the way they call into question the standing of the other to raise objections to certain forms of treatment and to cite her interests as normative considerations" (ibid., p. 242).

And, as the lack of circumstantial awareness cuts the connection between the agent's evaluative judgment and the action, it does not reflect the agent's quality of will making the agent not blameworthy (cf. Talbert 2017a, p. 56).

Up to now, we can see that this view contrasts the volitionist view regarding awareness. Yet, Talbert's account also contrasts volitionist theories with regards to the agent's history. Recall how volitionist accounts - and Fischer and Ravizza in particular - introduce a historical condition of responsibility in order to account for the intuition that agents, who have no control over how they came to be the way they are (e.g., those who are subject to manipulation), are not responsible. Contrary to Fischer and Ravizza, Talbert proposes a structuralist view of responsibility. He does so for two primary reasons. First, historicists face an issue: typically, they argue that manipulated agents are not responsible because they have no historical control over their makeup. But on this view, there is no telling what would have to happen for that agent to become

responsible again. And certainly, these agents would not stay "unblameworthy" forever (cf. Talbert 2009, p. 14). Hence, historicists cannot show how agents that suffer from some sort of manipulation become responsible (again). And second, Talbert believes that manipulated agents are blameworthy for their actions because they can still possess ill will: "(…) what matters in these cases is not how an agent came to have her desires but whether, when she acts to satisfy a desire, she acts because of considerations that she counts in favor of so acting" (ibid., p. 9). Talbert concedes that agents are not blameworthy in manipulation cases if the agent forgoes her capacities for reflective self-government. However, Talbert insists that this does not support a historicist conclusion as the majority of cases does not entail a loss of these capacities (cf. ibid., pp. 9-10). The basic thought for this structuralist argument stems from Harry Frankfurt, who argues that manipulated agents are responsible as long as their higher-order desires significantly overlap with their first-order desires. Introducing desires at this point helps Talbert to show what agents take as motivation for action. Depending on which of the desires an agent acts on reflects what she takes to be a reason "we care about the contexts in which a person takes his desires to be reasons in part because we care how agents value other things in comparison with the satisfaction of their desires" (ibid., p. 5). Thus, he argues, it matters whether an agent's desires reflect her reasons for action. Most importantly, it reflects the agent's quality of will (cf. ibid., p. 6).

Still, historicists do not contend that the manipulated has values; however, they doubt whether in the case of manipulation, the values are actually the agent's own. The previous section on volitionism highlighted how introducing the concept of *taking responsibility* is Fischer and Ravizza' strategy to ensure that actions issue from the agent's own reasons-responsive mechanism. So how can Talbert maintain that the agent's actions express not only *any* quality of will, but are actually *the agent's own* if he denies that the agent's history matters? They are her own, Talbert argues, simply because just like in ordinary non-manipulation cases, the values at stake play the same explanatory role. Even agents, who do not play a role in accumulating the values that they have, can still act according to these values (cf. ibid., p. 12). Manipulated agents

91

can be responsible because they are still open to the reactive attitudes, i.e., moral blame. They are open to such attitudes because even after manipulation "(…) the origin of her values and dispositions does not entail that her actions fail to display the core features of blameworthy behavior" (ibid., p. 2). Hence, he defends a structuralist view on which an agent is responsible as long as the action is reflective of her evaluative judgment and the agent has circumstantial awareness.

# 4. Capacitarianism

## 4.1 Capacities and the Reasonable Standard

Up to now the chapter considered two positions regarding the control and epistemic condition of responsibility: volitionists maintain that some type of control and awareness are necessary for an agent to be responsible, which must be traceable to a point in the agent's history in which she was aware that her act would be benighting. Attributionists take up an opposite view arguing that tracing is wrong because it does not capture many common-sense cases. They claim that agents can be responsible when an action is rightfully attributable to the agent. More specifically, attributionists argue that their view better accounts for cases in which an agent is typically considered responsible (e.g., for having certain beliefs or attitudes) but has no control or awareness of them. As anti-tracers, attributionists believe that agents can be responsible without having to trace back to an earlier point in time in which the agent had control or was aware.

In light of the debate between volitionist tracers and attributionist anti-tracers, a third strand has emerged, which can be best described as capacitarianism (cf. Nelkin and Rickless 2017, pp. 107–108). Capacitarians are dissatisfied with either approach. They hold that agents can be responsible for cases where there is neither exercise of control nor display of ill will. Along with attributionists, they argue that tracing is

wrong because the cases described by tracers only capture one type of rare wrongdoer, namely the clear-eyed akratic wrongdoer (cf. Rudy-Hiller 2017, pp. 398–399). But it is important to note that ignorant wrongdoing is much more common: "After all, there are an indefinite number of things each of us is (unconsciously) not doing at this very moment, but most of us would hesitate to describe all of them as 'omissions' on our part" (Smith 2017, p. 36). For only think about how many things we omit daily without thinking about its possible moral significance and its consequences: forgetting to call a friend, who relies on us; not checking the mail and thus missing to pay some due invoice; the civil engineer failing to attend a meeting in which significant challenges with the substance matter of the building are discussed and which in turn alter her calculations, ending in the death of others. Tracing fails to account for many ordinary cases such as these. Hence, going down the volitionists' road and trying to define a moment in which the agent was aware, "(…) we would (…) fly in the face of common experience" (Sher 2009, p. 25).

However, capacitarians are also dissatisfied with the attributionist explanation proffered.[75] Recall that attributionists have different opinions regarding forgetting cases: Talbert holds that agents are *not* responsible in forgetting cases because they do not necessarily reflect the agent's quality of will (cf. 2017a, p. 56), while Smith takes these to be paradigmatic cases of responsibility maintaining that they reflect the agent's evaluative judgment, hence the extent to which they care (cf. 2005, p. 244). Capacitarians agree with Smith's position that these agents are indeed responsible, but they disagree that the reason is a lack of care. That is because even if the agent cared enough, she could be distracted. Thus, the lack of concern cannot really "be the whole story" (Sher 2017, pp. 7–9). Hence, capacitarians argue that forgetting cases show that agents can be responsible not only despite any previous failure to discharge some obligation (cf. Rudy-Hiller 2017, pp. 402–403) but also despite any lack of ill will. If the

---

[75] Even though anti-tracers are united by the view that tracing is incorrect, they are no homogenous group but come in many different shades, which differ significantly with regards to their actual view on the epistemic condition (cf. Clarke 2017b, p. 246).

civil engineer had just gotten an important phone call about the passing of a close relative, which distracted her from double-checking her calculations, at what point did she "fail to discharge [herself] from the epistemic obligations" (Rosen 2004, p. 303) in question? Certainly, her forgetting is neither traceable to some point in time in which she willingly forgot (anti-tracers are at pains to stress that the case shows just how hard it would be to define the particular moment). But, at the same time, her forgetting is no display of ill will. Nonetheless, the civil engineer is responsible for the subsequent collapse of the building.

Capacitarians hold that these cases of "unwitting wrongful conduct" (Clarke 2017b, p. 233) illustrate that agents do not need to be aware nor display ill will in order to be responsible. But what exactly grounds the unwitting wrongdoer's responsibility if it is not some type of awareness or ill will on part of the agent? Capacitarians, as the name suggests, appeal to the agent's capacities to locate responsibility. These agents are responsible because they have certain capacities or abilities which make it reasonable that they could have been expected to be aware of the action's significance (cf. Clarke 2017a, p. 66):

„(…) we have what I think are good grounds for finding agents in many cases of unwitting wrongful conduct to be blameworthy. Given their possession of these capacities and abilities, it was reasonable to expect them to have realized that their conduct was wrong, and they were able to avoid it. They then satisfy conditions that plausibly suffice for direct blameworthiness for wrongful conduct despite lacking awareness of its wrongness. And (…) they can be blameworthy for that conduct even if their blameworthiness doesn't trace back to blameworthiness for their ignorance" (Clarke 2017b, p. 240).

Hence, an agent can be responsible even for fully unwitting omissions where there is no consciousness as long as she has certain capacities that make it reasonable to expect that she should or could have been aware. Stressing the agent's capacities and abilities, capacitarians establish a standard to which the agent is held. One central challenge for the capacitarian view is to define what support the claim that an agent should or could have known better (cf. Rudy-Hiller 2018). In the case of unwitting

wrongdoing failing to exercise these "responsibility-relevant capacities" (Rudy-Hiller 2017, p. 399) grounds responsibility. These capacities are manifold and include active and passive ones. Turning one's attention to an event or deciding on something are examples of active capacities. Remembering something or noticing something about a situation, on the other hand, are examples of passive capacities (cf. Clarke 2017b, pp. 242–245). Unwitting wrongdoers retain these capacities even if they forget to do something. Thus, consciousness is not necessary. On the capacitarian view, the epistemic condition is satisfied if the agent could or should have been aware of factual and moral considerations. The civil engineer may not have remembered to double-check her calculations, but she could and should have been reasonably been expected to, because it is her job to make sure that buildings built upon her instructions are safe. Since the civil engineer could have been expected to know that her failure to double-check is wrong and she was not incapacitated, she has no excuse on the capacitarian view (cf. ibid., pp. 244-245).

Capacitarians are standardly grouped with other anti-tracers. The reason for this is that most of them share the view that responsibility for unwitting omissions is direct (cf. Nelkin and Rickless 2017, pp. 107–108). However, it is important to note that there are differing views on this. While some such as Randolph Clarke argue that agents can be directly blameworthy for their ignorance, Fernando Rudy-Hiller denies this (cf. 2017, p. 415). Yet, in contrast to volitionist tracers, who look for a prior point in time in which the agent was in control and aware, capacitarians trace the agent's responsibility to a point in time in which she was *capable* of exercising control "(…) and, had [s]he done so, [s]he would have avoided being ignorant of the relevant considerations" (ibid., p. 412). Nonetheless, capacitarians can avoid the regress volitionists face because tracing in this case does not have to go back to a knowing, benighting, act (cf. ibid., pp. 411-412). However, this creates a natural doubt regarding the conception of responsibility advanced thus far. For agents generally do not control what they are aware of. This seems to weaken the capacitarian conception because it seems to render it to chance. However, the type of control they advance is *capacitarian*

*control.* For example, Rudy-Hiller argues that agents can be in control of their ignorance and thus have moral obligations to remember or be aware of some features of situations which is not limited to intentional actions/omissions (cf. ibid., p. 418). And Clarke too argues that agents generally have the ability to perform or omit and thus have the required kind of control without necessarily exercising control (cf. 2017a, p. 70).

## 4.2 Sher: Origination

George Sher has been one of the authors to advance an influential capacitarian view. As such his account shares the criticism regarding volitionist and attributionist accounts. Since standard volitionism clearly fails to account for many cases of responsibility, Sher concludes that the epistemic condition does not require consciousness (cf. 2009, p. 95). Thus, he shares the criticism attributionists typically have about volitionist accounts. Nonetheless, Sher's view is far from a generic attributionist view.[76] Instead he believes the capacitarian view to be appealing, which holds that an agent is responsible when she can reasonably be expected to have been aware of some fact. On the one hand, this approach avoids the volitionist regress problem, while on the other hand simultaneously managing to justify the intuition that agents are blameworthy in forgetting cases. However, critics argue that capacitarians then need to show how these norms disclose anything of importance about the agent. Digressing from a standard one has been reasonably expected to abide by cannot show how unwitting omissions relate to the agent (cf. Smith 2017, p. 38). That is because failing to meet a standard conveys nothing positive about the agent herself. But that is - according to these critics - exactly what blame implies (cf. Smith 2008, p. 374). Further,

---

[76] Sher's account has been called or been compared to attributionist views (for example cf. McKenna 2008, p. 30), however, it seems that this description stems from a misinterpretation of his arguments. Some of the confusion might stem from his premise that the action must be rightfully *attributable* to the agent. The probably most fitting description of Sher's view is that it is capacitarian, borrowing from volitionism and attributionism alike (cf. Levy 2008, p. 214).

they raise doubts about how these standards explain an agent's *moral* responsibility. If it is by the fact that an agent failed to meet some obligation which she could have been reasonably expected to meet this seems not to be enough to ground blameworthiness (cf. Talbert 2017a, p. 57).

Picking up on this criticism capacitarians are faced with, Sher argues that the capacitarian notion is a step towards the right direction, but since the standard view is a "mere nonevent" (Sher 2009, p. 85), which does not relate to the agent herself, it is not to Sher's satisfaction. To account for the problem, Sher introduces the origination requirement: the action's wrong-making features must be suitably connected to the agent. In Sher's view, the connection is correctly established when the agent either knowingly performs an action or is unaware, but it is caused by her traits. He maintains that agents are responsible as long as the action is suitably related to the *agent's self* (cf. Sher 2009, pp. 85–88). What Sher's account requires then is not simply that an agent should or could have been reasonably expected to perform an action but also that the omission of it directly reflects the agent's self:

"(…) someone who is unaware that he is acting wrongly or foolishly is only responsible for doing so if his failure to recognize his act as wrong or foolish is both defective in relation to some applicable standard ("he should have realized") and *due to some combination of his own constitutive attitudes, dispositions, and traits* [italics S. Sch.]" (ibid., p. 87).[77]

Sher adds another layer, namely attitudes, dispositions and mental states, which in turn give rise to the evaluative judgments or choices that issue in actions or omissions. These in turn are situated between the agent's dispositions and mental states and the actual resulting acts or omissions. Even though in many instances the

---

[77] By referencing the agent's connection to the action, Sher consequently also provides a variant of the reasonable standard, which has been critizied fails to relate to the agent in the right way. After rejecting some options, among them a subjective standard applying to the agent solely, a statistical standard comparing an agent against others, or what one has done in the past (cf. Sher 2009, pp. 97–110), he settles on the view that the standard is rooted both, in the agent's capacities and what she is obligated to be aware of "(…) if [s]he is to discharge [her] moral or prudential duties" (ibid., p. 111). The combination of the two elements is made visible through understanding that the agent's cognitive capacities are the precondition for reasons-responsiveness; simultaneously, moral and prudential "(…) demands address us precisely in our capacity as reason-responders" (ibid., p. 115).

agent's evaluative judgments and choices explain her behavior, Sher maintains that it can be bypassed by the same dispositions or attitudes, which often inform her evaluative judgment or choices. Hence, Sher's position neither necessitates that the agent has control nor must the action be reflective of her quality of will (cf. 2017, pp. 9–10). In forgetting cases, Sher proposes that these agents are blameworthy not because of some choice or ill will but because the failure is attributable to their selves. Hence, Sher counters the criticism that generic capacitarian views say little about how the standards by which the agent is assessed relate to her by introducing the origination requirement. On his account, an agent can be blameworthy for an unwitting omission only when it was reasonable to expect from her that she should and could have known better but also when it is causally explained by some constitutive features of herself. As this relation is purely causal, not all acts necessarily manifest a trait. Some states are constitutive and not necessarily morally bad but may give rise to morally bad actions (cf. ibid, p. 10).

But what makes a feature constitutive of the agent? On Sher's view features are constitutive when they are "(…) part of the larger constellation of states and traits whose members interact in ways that causally support the agent's broader rational capacities" (ibid., p. 11). Since agents generally have these capacities, they are not exculpated in case that they are *locally* incapacitated. That is because they retain their general rational capacity. In contrast, agents who lack these capacities altogether - i.e., they are not locally but *globally* incapacitated - are not blameworthy. That is because these agents cannot recognize and appropriately respond to reasons. In this case, none of the features is constitutive and the agent does not stand in a suitable relation to the blameworthy action (cf. ibid., pp. 11-12; 16).[78] For example, since we can assume that the civil engineer is generally capable of responding to reasons, she retains the relevant capacities to ground her responsibility. She may be locally incapacitated as her

---

[78] Sher foresees the possible criticism that on his account, the same features, which explain the success of recognizing and responding to reasons are also the reason for failure to responding to them. He replies that this is true for all kinds of system failures (cf. 2017, pp. 11–12).

judgment is clouded by the grief, she experiences from the passing of a family member. This does not altogether excuse her though. And this is also the reason why on Sher's account agents who have these general capacities are responsible and have no excuse when they cannot recognize or respond to reasons in any given specific case. Having a general rational capacity to recognize and respond to reasons suffices to show that the agent has the required psychological setup to be responsible:

"(…) and that will be enough to guarantee that some subset of [her] features is constitutive in the specified sense. Because we will thus be able to make sense of the idea that some of [her] features are constitutive, we will also be able to make sense of the idea that [her] current failure to recognize or respond to a given reason is due to the anomalous interaction of some members of the constitutive set" (ibid., p. 16).

Sher's motivation derives from forgetting cases, which he maintains that neither volitionists nor attributionists can sufficiently account for. Agents, he argues, are responsible either, when they are aware of their bad acts or are not aware but fall below a standard which made it reasonable to expect them to be aware the failure originates from the agents' self. However, Sher notes that this setup still does not cover all cases because there is a second class of unwitting agents: unwitting rightdoers. Consider how the cases presented so far focus primarily on bad acts. The knowledge condition Sher constructs captures these cases but in the positive equivalent, it does not. That is because, prudent acts must be done for the right reasons. However, since unwitting rightdoers are not aware of the features that make their actions right, their actions cannot stem from their constitutive psychology and hence are not performed for the right reasons (cf. Sher 2009, pp. 137–143). To account for this variation, Sher adds a third clause to his epistemic condition, which then reads as follows:

"(…) When someone performs a wrong or foolish act in a way that satisfies the voluntariness condition, and when [s]he also satisfies any other conditions for responsibility that are independent of the epistemic condition, [s]he is responsible for [her] act's wrongness or foolishness if, but only if, [s]he either

(1) is aware that the act is wrong or foolish when [s]he performs it, or else

(2) is unaware that the act is wrong or foolish despite having evidence for its wrongness or foolishness his failure to recognize which

(a) falls below some applicable standard, and

(b) is caused by the interaction of some combination of [her] constitutive attitudes, dispositions, and traits; or else

(3) is unaware that the act is right or prudent despite having made enough cognitive contact with the evidence for its rightness or prudence to enable [her] to perform the act on that basis" (ibid., p. 143).

With this quite lengthy outline of the epistemic condition, which even Sher calls "complicated and unlovely" (ibid., p. 144), he best captures "the full range of our intuitions" (ibid.).

As noted above, some capacitarians argue for a capacitarian notion of control which does not require consciousness. In contrast to his fellow capacitarians, Sher rejects the control condition.[79] This further step derives from the aim to uphold a principle that has been noted more frequently lately, namely that control requires knowledge (cf. Sher 2006a, p. 53, 2006b, p. 286; Clarke 2017a, pp. 65–66). That is because it is hard to imagine how an agent can exercise control without having any knowledge. On the contrary, agents lack control when they do not have knowledge of their actions (cf. Levy 2005, p. 5). Hence, since Sher argues that agents can be responsible for fully unwitting omissions, he must show that agents can be responsible without any control, else he would violate the principle.[80] But the view that agents are not responsible for anything that is beyond their control is persistent. Thus, he calls into question the control requirement by means of arguing that its appeal stems from underlying premises which can be rebuffed. To do so, he reminds us of his argument that any suitable relation between the agent and her action requires origination (cf. Sher 2009, pp. 145–147). This requirement can be transformed into the control

---

[79] In an earlier paper (Sher 2006b), Sher is skeptical of the control condition, yet does not fully reject it. Instead he argues for a "watered-down substitute"(2009, p. 146). He later withdraws from this strategy and instead opts to reject the control condition fully.

[80] This is not to say that other capacitarians cannot account for the principle. Yet, I do not assess the success of their theoretical constructions here.

condition „(…) as soon as we add that in order to stand in the origination relation to a given feature of a given act, an agent must choose to perform that act while fully aware that it will have that feature" (Sher 2009, p. 148). However, it is the origination requirement not the interpretation of it which is "forced upon us by a deep structural fact about responsibility" (ibid.). Further, any explanation why origination should require conscious choice retracts to volitionist arguments. Hence, he concludes that we need not accept the control requirement. Nonetheless, he admits that it would be absurd to maintain that an agent is responsible, whose actions are not voluntary (cf. ibid., pp. 146-149). This is why Sher continues to argue for a stripped-down version of voluntariness which severs its ties with the control requirement in the sense of choosing. Even though Sher does not specify which alternative conception of voluntariness he endorses, he maintains that there are various interesting contenders such as those that believe that voluntariness is satisfied when the agent is free from coercion, compulsion, or insanity, or when it reflects the agent' character, or when the agent is responsive to reasons (cf. ibid., pp. 150-151).[81]

However, rejecting the control condition (while sticking to the voluntariness condition), what can Sher tell us about cases of manipulation? Recall that Fischer and Ravizza employ the control condition to illustrate that manipulated agents are not responsible, while Talbert argues that these agents are responsible nonetheless because their values play an explanatory role. Sher acknowledges the worry that responsibility seems to be undermined when the agent is a product of manipulation, by stating that if this worry is correct, then his view can accommodate the requirement that the agent not have been produced in these ways (cf. Sher 2006b, 298 footnote). Moreover, he mentions that agents with a certain cultural membership such as "slave or caste societies" are not responsible, instead they are (at least to some degree) excused as they are given no reason as to why their actions are objectionable. Since their environment

---

[81] As one of the examples, Sher cites Fischer/Ravizza's aspect of the mechanism from which the action results itself must be reason-responsive as a possible version of the voluntariness requirement (cf. 2009, p. 150).

(including their main care takers and other members of this culture) is their main source of information with respect to the acceptability of their actions, these agents have no reason to question that her actions might be wrongful (cf. Sher 2009, 116 footnote). And this verdict also goes for those brainwashed as "(…) they have been rendered incapable of responding to certain sorts of reasons (…)" (Sher 2009, p. 136). Hence, he seems to share the view with Fischer and Ravizza that manipulated agents are not responsible.

## 5. Conclusion

This chapter served to give a brief overview of the vast field of responsibility because just as in the case of character, most of the analyses of responsibility for character rely on a particular understanding of responsibility, which depends on the control condition. But the literature on this field is broad and dense at once. And while there is no way in doing justice to any of the intricate discussions in this thesis, this overview helps establish a common ground for the main disputes in the debate. The main dissent lies on what the exact conditions are that ground responsibility. In order to explore what different theories of responsibility can tell us about character in the upcoming chapter, I chose three accounts that reflect the main claims of these theories respectively, those by Fischer and Ravizza, Talbert, and Sher. More specifically, volitionists typically endorse control as a necessary condition for responsibility (Fischer and Ravizza (2000) suggest that agents need guidance control), while attributionists deny that this needs to be the case (Talbert (2013, 2017a) stresses the agent's evaluative judgment as essential to her assessment). The difference between these two views becomes most visible when addressing the problem of manipulation. While attributionists typically emphasize that they can account for cases in which agents seem to be responsible but have no control, volitionists challenge this view by arguing that manipulated agents are not responsible. Hence, their plead for a historical

notion of responsibility seems to counter the structuralist view advanced by many attributionists (cf. Talbert 2023). Analogously to the control condition, volitionists and attributionists have defended opposing views on the epistemic condition. The former argue for a tracing account, whereas the latter (joined by capacitarians in the matter) reject the idea of tracing. While tracing views satisfy the intuition that agents need some kind of awareness, anti-tracers are at pains to stress that this view fails to account for many ordinary cases. Despite the unity regarding tracing, attributionists and capacitarians dissent on the issue of the unwitting wrongdoer. While attributionists typically retain a certain amount of awareness (e.g., circumstantial awareness), capacitarians suggest that agents can be responsible also in cases of fully unwitting omissions such as forgetting (Sher (2009) argues that the agent could or should have been aware and the action must originate from her). This case is particularly interesting because agents do not generally choose to forget something, neither does forgetting necessarily reflect their quality of will (cf. Rudy-Hiller 2018).

# IV Putting the Theories to Test

## An Analytical Examination of the Main Theories

---

## 1. Introduction

Part I of this thesis - comprised of the previous three chapters - served to establish an overview of the theoretical landscape in debates on character and moral responsibility. Moving on, this current chapter seeks to explore what we can learn from the theories introduced about the responsibility agents have for their character. In doing so, it not only combines the insights gained from the previous chapters, it primarily confronts the third and final aspect of the research gap: standard analyses of character responsibility implicitly focus on one particular interpretation of the research question, namely whether agents are responsible for developing their traits. However, as mentioned in the introduction of this thesis, the question whether agents are responsible for their character is ambiguous. It can be interpreted as inquiring whether agents are responsible for developing their traits or also for *having* them. Therefore, I begin by briefly outlining a view of how to conceive of this distinction in terms of character by means of connecting it to two forms of responsibility, direct and indirect responsibility standardly differentiated when considering whether agents are responsible for their actions. For reasons of clarity, I will discuss both concepts distinctly in the subsequent analysis. Moreover, character, I argue, introduces complexity with regards to structure (i.e., the composition of traits) and dimensions (i.e., the analytical focus), which must be addressed respectively. Considering each theory of responsibility selected, I then continue to explore what they can tell us about of how we can conceive of agents being directly and/or indirectly responsible for their character traits.

I contend that Fischer and Ravizza's theory does not "fit" direct character responsibility due to the specific historical setup of their view. Regarding indirect responsibility, their view can generally show how agents are responsible, yet this is confined to more coarse-grained traits because their volitionist position entails a stronger epistemic condition. Additionally, due to their tracing approach, many agents are excused when they lack awareness, yet character traits are not necessarily developed consciously. Due to his ahistorical position, broadly conceived, Talbert yields positive results with regards to agents being directly and indirectly responsible for their traits (yet, also more fittingly for Global Traits rather than Local or Mixed Traits). However, employing his structuralist approach also means taking agents who suffer from manipulation to be responsible, a claim refuted by many. Lastly, Sher's own discussion concludes that agents are not directly responsible for their bad traits. I suggest that this also extends to good traits. I then discuss the application of his account to whether agents are indirectly responsible for their character and argue that agents are not for their Mixed Traits but that the argument why they are not has less clout for coarse-grained conceptions of character.

The chapter closes with a short framing of the lessons learned throughout the analysis and clusters them into three challenges: first, an analytical examination of responsibility for character traits must distinguish between the two forms of responsibility and spell out its conditions. More importantly, it should show how agents can not only be indirectly but also directly responsible for their character traits. Second, it must specify in what way the agents history is essential to her responsibility assessment. Lastly, character adds more structural and dimensional complexity and, depending on the particular conception, agents may know less about their character and the formation thereof. This is reflected in the fact that the theories that take agents to be responsible, even when they lack control and awareness, also fare better with Global Traits. An analysis that focusses on character must therefore present a plausible epistemic condition which clarifies the added dimensions of awareness needed and reflect that character traits are not always developed consciously. These three

challenges, I argue, lead us to endorse the thesis that in order to account for character responsibility, analyses must address character as a distinct phenomenon.

## 2. The Two Forms of Responsibility and the Complexity of Character

There is overarching consensus that there are mainly two ways in which agents can be responsible: agents can be *directly* and *indirectly* responsible (for example see discussions in cf. FitzPatrick 2008, p. 591; Clarke 2014, p. 107; Nelkin and Rickless 2017, p. 107; Mele 2021, p. 562). To illustrate the difference, consider again the case of the civil engineer introduced earlier. The civil engineer miscalculates fundamental parts of the housing unit, which consequently collapses, injuring many. If we find out that the civil engineer has taken drugs before working on her sketches of the supporting pillars, we cease to view her as fully capable of doing her job. Her incapacitation alters her responsibility; she is thus no longer *directly* responsible for the miscalculation and the tragic collapse of the building. While she is directly responsible for taking the drugs, her subsequent failure to do her job correctly is a product of the former action. She is thus *indirectly* responsible for the miscalculation.

Generally, the two types of responsibility can be distinguished by means of inquiring whether the action is the result of some prior action. Direct responsibility refers to the agent's responsibility for some action that is *not* the result of some other action by the agent. In other words, direct responsibility for an action means that it *is not* inherited (cf. Mele 2021, p. 562). On the contrary, indirect responsibility means that it *is* in fact inherited because it is a consequence of the responsibility for another action, which renders agents responsible for the subsequent action respectively. While authors differentiate regarding the wording (e.g., direct responsibility is also referred to as "basic" or "original," indirect responsibility is also termed "derivative" (for examples of these different terms used see FitzPatrick 2008, p. 591; Nelkin and Rickless 2017, p. 107), there is overall agreement that direct responsibility does not depend on

a previous action or omission, while indirect responsibility does. Again, the civil engineer is indirectly responsible for the collapse of the building because she is directly responsible for intoxicating herself before working on the calculations.[82]

Regarding actions this general differentiation is pretty straightforward and usually necessitates no further clarification.[83] However, this does not seem to be the case for character as there are mainly two ways in which confusion might arise: first, there is an asymmetry between what is the standard object of investigation and what agents are directly responsible for. In the case of actions, direct responsibility concerns the way agents are responsible for an action. In the case of character, too, most analyses inquire whether agents are actually responsible for the *actions* which habituate their character (see for example Jacobs 2001, p. 2; Chen 2013, p. 353; Fileva 2017b, p. 194). That is, they provide arguments for or against the view that agents are responsible for developing their character. However, analyzing whether agents have direct responsibility for the actions, which lead to character acquisition, in fact does not focus on direct responsibility for character at all but only provides an answer in the indirect sense. That is because the responsibility in question is dependent on a prior responsibility had for the actions that develop character. Therefore, the sense in which direct responsibility applies is different for action and character responsibility.

Second, confusion might arise because there are two ways in which responsibility can be indirect. One version is what I have referred to as the standard interpretation, which asks whether agents are responsible for developing their

---

[82] Another prime example of indirect responsibility is the case of the drunk driver, who is responsible for the subsequent accident because she is responsible for drinking before driving (cf. Mele 2021, p. 563).

[83] While the standard case of responsibility for actions certainly is direct responsibility and the classification of the respective action as such is mainly uncontested, recent discussions on the epistemic condition have shown that there is disagreement whether agents can be directly responsible for unwitting omissions or whether these are always instances of indirect responsibility. Attributionists and capacitarians share the view that agents can be directly responsible for unwitting omissions (cf. Clarke 2017a) (with the exception of Rudy-Hiller, who argues that agents can only be indirectly responsible in these cases (cf. 2017, p. 415)). Tracing accounts, on the contrary, require an akratic benighting act and thus maintain that agents can only be indirectly responsible for unwitting omissions (cf. Rosen 2004, 300; Nelkin and Rickless 2017, p. 117). Despite this disagreement, there is no difference in opinion about the actual distinction between the two concepts, i.e., that direct responsibility requires no responsibility derived from a prior action or omission while indirect responsibility does.

character. However, a further variant of indirect responsibility emerges, adding to the standard interpretation. This becomes evident when we remind ourselves that indirect action responsibility is the responsibility agents have for the consequences of their actions. Taking this to apply for character would be to inquire the responsibility agents have for the consequences of their traits, not the development of them. This type of responsibility might be considered "derivative character responsibility." However, this latter question is different from asking whether agents are responsible for acquiring the traits they have. Yet, we need to realize that "derivative character responsibility" is not a new type of responsibility which has simply been disregarded in the discussion so far. On the contrary, "derivative character responsibility" reflects the way in which agents are responsible for their actions and therefore resembles direct responsibility for actions. And recall the reason cited in the introduction of this thesis of why direct character responsibility seems important: loosely put, since actions are generally thought to derive from character, it seems to imply that responsibility for actions also stems from agents being responsible for their traits (cf. Sher 2006a, p. 17; Kauppinen 2017, pp. 46–51). Hence, if agents are not directly responsible for their character, it begs the question how their responsibility for their actions is generated. Despite this issue being a relevant and surely adjacent question, due to the focus on character responsibility in this thesis, the upcoming analysis will not be concerned with this form of responsibility in more detail.

The ambiguity of the research question entails two possible interpretations which resonate with direct and indirect responsibility. By connecting the two sorts of responsibility with the two notions of the question and specifying which type of responsibility is at issue, we gain analytical clarity. However, since the distinction between direct and indirect responsibility for actions does not *exactly* translate to character (again, direct action responsibility would be indirect character responsibility), but is actually potential cause for confusion, it is necessary at this point to outline how I plan to employ the two forms of responsibility with regards to character. In the case of actions, direct responsibility is the responsibility agents have

for the actions themselves. This reflects the way in which direct action responsibility does not depend on a prior action. Utilizing a conceptual application, in the case of character, agents would be directly responsible for *having* a character trait. Hence, I will use direct character responsibility as the responsibility agents have simply for having that specific character. In the case of indirect responsibility, there are two ways in which responsibility can be indirect. Yet, since only one inquires about character responsibility in the specific sense intended for this thesis, namely the standard interpretation, I use indirect character responsibility as the responsibility agents have for the actions which ultimately develop their character.

Moving on, the differentiation between the two forms of responsibility also highlights a more general fact: adding the variable of character to the equation increases the level of complexity for this analysis. But it does so not only in terms of direct and indirect responsibility, i.e., the two types of responsibility just differentiated. Complexity also enters with two further regards, on a structural and dimensional level: first, structural complexity enters because in contrast to actions, character is much more multifaceted. We can see this when we recall that all character theories encountered in chapter 2 stress that character is made up of many traits. Therefore, an investigation of the question whether agents are responsible for their character is already different from inquiring whether agents are responsible for their actions. In addition, and more importantly for this analysis, each theory differs with regards to the individuation of these traits. Recall that Global Traits are thought to be broad, i.e., temporally stable and manifest in cross-situational consistent behavior (cf. Doris 2008, pp. 22–23; Miller 2009, p. 249; Rhode 2019, p. 10), while Local Traits generally lack cross-situational consistency and apply to narrow scenarios only (cf. Doris 1989, p. 508, 2008, p. 64), and Mixed Traits are highly individualized fine-grained dispositions, which can be triggered and stay active even if certain situations have

passed (cf. Miller 2009, p. 251, 2017a, p. 255, 2018, p. 159).[84] The complexity in structure thus prompts us to clarify in what way these different theories will be discussed in the upcoming section. Since theories of responsibility mainly pose two conditionals (or debate whether these are really necessary), there is reason to do so by means of what can be controlled or known about them. Hence, in order to make meaningful assertions about the way agents can be responsible considering the varying theories of responsibility, I will analyze them based on their most salient differences that relate to the conditionals of moral responsibility, i.e., control and awareness. Even though this methodology amounts to a somewhat simplification of the theories of character encountered, it allows us to pursue our analytical purposes more evenly and across all theories to arrive at conclusions that help us establish common assessment.

Second, the dimensional complexity enters even more so in the case of indirect character responsibility because now the question is not limited to the action itself but extends to character. In other words, the case of indirect character responsibility seems to ask about the agent's responsibility for the action *and its consequences* (i.e., resulting in a particular trait) simultaneously. While it is common for theories of action responsibility to specify to which extent agents also need awareness of the consequences of their actions (for example Fischer and Ravizza maintain that an agent needs guidance control of them (cf. Fischer and Ravizza 2000, pp. 92–122)[85]; and Talbert states that, even though moral ignorance does not exculpate, agents do need

---

[84] It should be mentioned that even though Miller suggests that character traits (more specifically, the model he proposes) are essentially fragmented (cf. 2018, p. 159), he also assents to the view that character traits are broad (cf. 2009, p. 250, 2014, pp. 49–61). However, the upcoming analysis of character responsibility provided highlights the former feature, depicting Mixed Traits as fine-grained. This is in line with the criticism mentioned, namely the question what trait dispositions add to the fine-grained mental dispositions (cf. Bates 2015, pp. 422–423). And since Miller himself notes that he endorses a property monist view (in contrast to a dualist view), which holds that trait dispositions are identical to the mental state dispositions (2014, pp. 27–32), and therefore much more individualized, I take it that Mixed Traits are best conceived of as fine-grained.

[85] Fischer and Ravizza specifically attend to the matter of responsibility for consequences (cf. 1993b, 2000, pp. 92–122). Mainly, they hold that it too requires guidance control of the agent such that both, the agent's „(...) 'inner mechanism' leading to the agent's bodily movement, and the 'outer path' leading from the bodily movement to the relevant event in the external world" (Fischer and Ravizza 2000, p. 107) are responsive.

circumstantial awareness (cf. Talbert 2013, p. 226)), in the case of character responsibility standard analyses typically do not *explicitly* call attention to this point. At times it is implied, other times it seems neglected. However, as the concepts of character introduced vary strongly with respect to their degree of individuation, there is not only the question of what agents can be aware of in terms of structural but also dimensional complexity. Thus, in order to discuss character responsibility distinctively (and therefore not simply responsibility for actions), the analysis that lies ahead, pays special attention to this aspect of the dimensional complexity of character responsibility. Further, it should be apparent by the way I have presented these two types of complexities that both interact, i.e., discussing whether an agent can be indirectly responsible for a specific trait requires consideration of whether the agent has sufficient control and awareness of the trait structure simplicter and how an action would develop that trait. Considering that character introduces structural complexity, it seems that agents would have to have knowledge that an action develops a specific Global, Local or Mixed Trait, which seems progressively more difficult, *and* what kind of character they have. Therefore, the upcoming analysis will reflect these additional complexities (structural and dimensional) by considering structurally diverse theories of character and by means of broadening the scope to include how actions develop traits and the specific trait structure. With these terminological and conceptual specifications in mind, we are equipped to explore what the theories chosen can disclose about responsibility for character traits.

## 3. Fischer and Ravizza

The previous chapter introduced Fischer and Ravizza's account as representative of volitionist thinking. Recall the volitionist requirement that for an agent to be responsible she needs to have some sort of control. Fischer and Ravizza particularize that an agent needs *guidance control*, which is comprised of two features (cf. Fischer

2013, p. 187). First, an agent has guidance control when her mechanism is her own. Agents make their mechanism their own by taking responsibility. This includes understanding that one is the source of actions and the actions' consequences as well as being the praised or reprehended for one's behavior (cf. Fischer and Ravizza 2000, pp. 210–213). Hence, guidance control is ultimately historical. Second, an agent has guidance control when her mechanism is moderately reasons-responsiveness. This in turn, too, includes two elements. Agents are moderately reasons-responsive when they display an understandable pattern of reasons, i.e., they are strongly reasons-receptive. However, agents only need to be weakly reasons-reactive (cf. ibid., p. 69). I also highlighted that despite Fischer and Ravizza's outspoken focus on the control condition, they do refer to the epistemic condition. For not only do they propose a tracing approach, reasons-receptivity entails a cognitive component.

## 3.1 Historical Control and the Current Makeup

I start my analysis by taking a look at what the volitionist account by Fischer and Ravizza tells us about direct character responsibility. I suggested that for an agent to be directly responsible for their character, the responsibility in question cannot derive from another action for which she is responsible. Because direct character responsibility asks about the agent's responsibility for her character that is independent of the prior acquisition thereof, the agent is only directly responsible for her traits if it is not a consequence of being responsible for something else. It would have to be *original* responsibility so to speak. However, this clashes with the setup of Fischer and Ravizza's account for two reasons: first, guidance control involves taking responsibility, which ensures that the agent has the right kind of history (e.g., that she is free from manipulation). But the question behind direct responsibility for character traits is not concerned with historical control in the sense suggested by Fischer and Ravizza. Instead, it asks of the agent's responsibility simpliciter, her current makeup, and does not depend on some prior *controlled* acquisition of the agent's values. Hence,

it seems that there is no room for Fischer and Ravizza's account to accommodate for agents being directly responsible for their character. Second, even if this were not the case, it is doubtful how the remaining features of the account would produce other results. That is because the other aspect of guidance control is moderate reasons-responsiveness. Yet, agents do not *have* (as opposed to develop) traits for a reason. On the one hand, agents do not demonstrate a reasonable pattern of reasons-receptivity simply by having a trait. On the other hand, trying to square the concept of direct character responsibility with Fischer and Ravizza's approach neither seems conceivable when paired with reasons-reactivity. For how would an agent (weakly) react to reasons, which are nonexistent, for only *having* a certain kind of character? And these questions can be raised for all kinds of character conceptions. Independent of the structure, whether it be coarse- or fine-grained, the same problems would arise. However, it is important to note that Fischer and Ravizza's account of guidance control does not yield negative results for direct character responsibility. Rather, they produce *no results at all*. That is because the concept simply does not match the question that direct character responsibility entails. This is not to imply that there is something wrong with Fischer and Ravizza's account. As I said, the account is designed to account for action responsibility first and foremost. Yet, it illustrates that the view is insufficient to accommodate direct character responsibility. The problem seems to stem from an overall structural misfit between direct character responsibility and Fischer and Ravizza's volitionist (and essentially historical) account.

## 3.2 Questionable Awareness

Despite the apparent misfit between direct character responsibility and the volitionist account proposed by Fischer and Ravizza, this does not necessarily mean that the same holds for indirect character responsibility. In a short digression, Fischer and Ravizza even address this variation of the question. Consider their take on the matter in this shortened paragraph:

"It is clear how an agent can exercise guidance control in *forming* a trait. (…) Further, (…) [s]he may control whether [s]he *retains* that trait. Perhaps a child raised in an orthodox religious environment has acquired certain pious tendencies; as an adolescent or a young adult, however, [s]he may be able critically to reflect on these tendencies" (Fischer and Ravizza 2000, p. 88).

Even though their analytical focus is on the actions produced by traits, Fischer and Ravizza provide some ideas on how to conceive of agents as indirectly responsible for their character.[86] Agents, they argue, can exercise guidance control over the formation and retention of traits. They can do so by means of critical reflection, engaging their moderately reasons-responsive mechanism. Even though they do not explicitly state what kind of conception of character this view is based on, their argument resonates with standard thinking about those who endorse responsibility for character traits. Since they employ globalist language by speaking of "pious tendencies" (recall that language around Global Traits often stresses broad traits such as honesty, compassion, justice or the like), this also seems to support the conclusion that Fischer and Ravizza have a somewhat coarse-grained understanding of character. And the features relevant for this discussion identified are that Global Traits are broad - meaning stable and cross-consistent - features of an agent (cf. Walker 1989, p. 354). And if this is so, then I agree that their account can produce positive results for Global Traits. To confirm this, let us reconsider the two components of their theory:

First of all, it is plausible that agents take responsibility. Agents take responsibility, when they are aware that they are the source of their actions and recognize that their actions have consequences, which in turn make them an appropriate target of the reactive attitudes. Lastly, they must be aware of evidence of that (cf. Fischer and Ravizza 2000, pp. 210–213). Applied to character, we can assume

---

[86] Note that Fischer and Ravizza introduce this in terms of responsibility for non-reflective behavior that issues from traits. As mentioned in section 2 of this chapter, responsibility for actions issued from traits - which I have called "derivative" - is different from responsibility for actions that result in the formation of traits. Indirect character responsibility, as employed here, is concerned with the latter. Nonetheless, the digression the two authors offer also provides a short discussion of indirect character responsibility as of interest here.

that agents have experienced (more or less consciously) being the source of actions and that they have consequences, i.e., that they also habituate character. Even though one could interject that taking responsibility as outlined involves epistemic steps, which would potentially interfere with more fine-grained accounts, Fischer and Ravizza are quite clear that taking responsibility is not a matter of decision (cf. ibid., pp. 217-219). Thus, we can assume that agents not only take responsibility for their Global Traits. Due to the nature of the condition, agents also take responsibility for their traits even in the face of the complexity of Local Traits and Mixed Traits.

Second, regarding moderate reasons-responsiveness, I too agree that Fischer and Ravizza's view can yield positive results for Global Traits. But first, recall that when considering indirect responsibility the dimensional complexity of character requires an analysis of the satisfaction of the conditions for the action, yet also how they translate to character. Hence, agents would need to be moderately reasons-responsive with regards to their traits as well. Considering the broad structure of Global Traits, it is plausible enough how an agent would know of her traits. Further, it would be easier to know how an action translates to character, i.e., how to habituate them as the structure of Global Traits should enable agents to have enough cognitive contact with them. For example, Hannah will probably have some kind of understanding of her character as she regularly speaks the truth and knows that being upfront contributes to strengthening her honesty. Thus, agents can display an understandable pattern of reasons-receptivity and therefore be moderately reasons-responsive for their Global Traits because the coarse-grained structure would allow them to be aware of these traits and how to habituate them.

However, the same fit seems to be called into question considering narrower and fine-grained character models. For one, when we bring back to mind how the Local Traits approach acknowledges that situational features are the main explanation for behavioral manifestations (cf. Doris 1989, p. 508), this may be worrisome. If situational factors causally influence agents' behavior, to what extent can agents be moderately reasons-responsive to them (cf. Nelkin 2005; Brink 2013; Ciurria 2013;

McKenna and Warmke 2017)[87]? The problem may be put like this: situational factors may impair the agent's[88] ability to recognize and react to reasons (cf. Nelkin 2005, p. 199):

"And how deep might the rabbit hole go? If situational features of one sort or another are constantly bombarding our reasons-responsive capabilities and either seriously impairing them or taking them temporarily offline, for how much of our behavior are we really responsible" (McKenna and Warmke 2017, p. 714)?

Because the Mixed Traits model similarly stresses the relation between triggers and the agent's behavior (cf. Miller 2018, p. 148), the same argument may apply. In fact, the argument may prevail even stronger as triggers may enhance or inhibit behavior even after the passing of a particular situation (cf. Miller 2009, p. 251). Notwithstanding the initial complication, one might argue that the problem lies in the difference between "competence and performance errors" (McKenna and Warmke 2017). In the former case, agents lack the relevant capacity to reason and reflect and thus are exempted or excused respectively. In the latter case, agents retain their capacity, yet fail to perform accordingly. Agents that respond to situational factors or even triggers may only be a demonstration of performance errors: "(…) situationism addresses situational patterns in *performance,* not issues of *competence"* (Brink 2013, p. 140).[89] And this generally resonates with the claim that moderate reasons-responsiveness requires strong reasons-receptivity but only weak reasons-reactivity.

---

[87] Nelkin, Brink and McKenna/Warmke specifically review reasons-responsiveness theories with an overall positive conclusion. Ciurria takes a more skeptical stance regarding deep-self-views and expresses sympathy for a reasonable-person view.

[88] For Fischer and Ravizza, to be specific, this actually concerns the agent's *mechanism (*cf. Nelkin 2005, 201 footnote).

[89] McKenna/Warmke, who endorse this problem particularly with a focus on the situationist literature, argue that even if taken seriously, agents do not lack the type of reasons-responsiveness required for responsibility. However, they are not blind to consequence of triggers diminishing responsibility because agents'"(…) reasons-responsive capabilities may be a lot more poorly put together than we had expected or hoped. Even if most of us usually meet the threshold requirements for having free will and moral responsibility, we still might fall much closer to the minimum threshold than we'd like" (2017, p. 727).

Nonetheless, while situational factors and triggers may not completely cloud the agent's abilities to respond to reasons generally, there is reason to question whether the application of the responsibility concept as proposed by Fischer and Ravizza translates as easily to narrower or more fine-grained traits. That is, even if we were to subtract situational factors and triggers (which, again, may not ultimately prevent the agent from having the capacity required, yet they do pose additional challenges), it stands to question whether agents can really be responsive to reasons considering the structural and dimensional complexity more generally. Let me expound more on the reason to cite this worry:

As mentioned before, one way in which complexity in analyses of character responsibility enters has to do with the structure of character. Now character conceived of as Mixed Traits increases the complexity even more as the model stresses the fine-grained nature of trait dispositions and their underlying mental dispositions (cf. Miller 2014, p. 26, 2017a, p. 255). Hence, it raises the worry whether agents can display a pattern of strong reasons-receptivity when they might not even be aware of the nature of their character due to the intricacy of the structural setup, i.e., when it is composed of a multitude of interrelated fine-grained dispositions. To that, one might object that agents never have this kind of detailed awareness, even for actions and any resulting consequences. And certainly, this does not repudiate Fischer and Ravizza's view on action responsibility and neither does their control condition entail a revisionist view about responsibility generally. The problem in this case specifically, however, is generated through the subject matter, meaning that comparing the different conceptions of character illustrates the different levels of awareness an agent most probably has. Hence, relatively speaking the conditions of the volitionist account combined with the structural complexity of Mixed Traits specifically would be at tension as agents it would be much harder for agents to have sufficient awareness to display an understandable pattern of reasons-receptivity when character traits are made up of a variety of fine-grained interrelated mental state dispositions (and not broad traits simpliciter).

Second, building on this difficulty, there is further reason to doubt whether agents are responsible for more fine-grained traits when considering volitionist accounts such as Fischer and Ravizza's. That is because in the case of indirect responsibility, I have claimed that an accurate analysis must consider the structural *and* dimensional complexity of character. Hence, we need to take an additional look at the action and the character trait which is developed thereby. The added dimensional complexity might prevent the agent from being aware of how to habituate these less coarse-grained traits. Again, taking the condition of moderate reasons-responsiveness at face value would require that an agent can display an understandable pattern of how certain actions habituate specific traits. Paired with Mixed Traits, this would at least seem very hard to achieve. In fact, consider Hannah with her Mixed Trait pertaining to truth-telling, which is made up of many fine-grained mental dispositions such as beliefs and desires about the telling the truth in some situations rather than others and which may include many instances of telling the truth, yet also some of *not* telling the truth or blatantly lying. It would require that Hannah knows not only of the structure of her complex Mixed Trait, but also how any one action might contribute to changing the specific structure of this trait (and this is without even considering the further complication of triggers). Hence, looking at the problem from this point of view, it might not seem as straightforward that the agent's mechanism can be moderately reasons-responsive for her Mixed Traits.

In suggesting these complications, I would like to stress that this is not to assert that this view *cannot* accommodate Local or Mixed Traits in any way. However, a lot less agents will be moderately reasons-responsiveness for them. Therefore, I maintain that while their volitionist account can generally be well applied to Global Traits, Fischer an Ravizza have a harder time showing how agents are responsible for their less coarse-grained traits, especially Mixed Traits. Despite some reservations and questions regarding the epistemic condition on narrow and fine-grained conceptions, their view is open to indirect responsibility for coarse-grained accounts such as Global Traits. Nonetheless, the epistemic requirement leaves the question whether they do so

for all kinds of character traits independent of the broadness/narrowness. The difference then lies in the specific understanding of character.

Before ending this section, I would like to offer a last, brief, observation, which builds on the last point and casts some general doubt on this approach with regards to its suitability for character. As volitionists, Fischer and Ravizza generally hold that agents need control and awareness to be responsible, which they specify as guidance control. Hence, on this view, agents can be excused when they are not aware of their actions or cannot trace back to a benighting act. As noted before, this usually leaves volitionists unable to account for many cases (cf. Nelkin and Rickless 2017, p. 114). This worry might also apply to indirect character responsibility as character traits are not always necessarily developed consciously. On the contrary, many agents acquire their traits without being aware or being able to trace back to a moment in which they were. And this not only goes for Local or Mixed Traits but all character conceptions. Thus, instead of claiming that Fischer and Ravizza cannot show that agents are indirectly responsible for their traits, I would like to point out that employing their approach means acknowledging that not many agents may in fact be (indirectly) responsible for their character.

## 4. Talbert

Talbert's account centers on the question of how an action is attributable to the agent. The general idea behind this is that agents are responsible for those actions that reflect their quality of will, or their evaluative judgment.[90] In consequence, he defends the view that - contrary to volitionist claims - responsibility does not require control or moral knowledge. Instead it suffices that an action is attributable to the agent, given that she has circumstantial awareness (cf. Talbert 2013, p. 236). Further, employing a structuralist argument, he maintains that manipulated agents can be responsible if the

---

[90] Henceforth, I will use the two terms interchangeably. Talbert uses both himself.

agent's values play an explanatory role and thus her actions still display her quality of will (cf. Talbert 2009, p. 13).

## 4.1 Character and Evaluative Judgment

As before, I begin with an analysis of direct character responsibility: the main reason why Talbert's account seems to fit well with this version of the question right from the get-go is the structuralist aspect of his account. Recall that one of the main reasons Fischer and Ravizza have difficulty with the concept of direct character responsibility is because their account requires that the agent has a specific kind of history that centers on the notion of control, which does not align with the question behind direct character responsibility. That is because direct character responsibility asks of the agent's responsibility for having the character she has. Simultaneously, this is why the very setup of Talbert's account seems promising for direct character responsibility: taken by itself, direct character responsibility precludes the agent's history. For the agent to be directly responsible for their character, we only need to ask whether she is responsible for her traits simpliciter, shielding the way she came to be who she is. This aligns with Talbert's own conviction that responsibility assessments are independent of the agent's history (cf. 2009, p. 4).

Considering whether next to this preliminary structural fit, there is also content-related reason to think that Talbert can accommodate direct character responsibility, we have to assess whether the agent's character is reflected in her evaluative judgment. To do this, we need to first establish the kind of relationship between character and evaluative judgment.[91] I hold that most importantly, the basic nature of the relation between character and evaluative judgment is that the former forms the latter. By this I mean that one's evaluative judgment is at least partly a result

---

[91] Please note that this is by no means a general standard I aim to establish. Rather, I want to expatiate on my own understandn of the link between both concepts for the present purpose of providing an analytical extension of Talbert's view to direct character responsibility.

of one's character. For example, considering that Judith is a compassionate person, she probably cares about others and judges helping them to be important. This basic relation, however, contrasts the question we are concerned with here. It does not ask whether we can draw conclusions from character to evaluative judgment but whether evaluative judgment is reflected in character. Nonetheless, one could argue that while the direction might primarily originate in character, which leads to the formation of a particular evaluative judgment in a given situation, the question is not about the direction but whether knowing of one's evaluative judgment informs us about her character: hence, even though an agent's evaluative judgment might derive from her character, knowledge thereof allows an inference of her character. And, therefore, broadly conceived, I maintain that evaluative judgment does reflect character.

Further, Talbert claims that for actions to reflect the agent's evaluative judgment, they also need circumstantial awareness. Talbert contends that the agent's quality of will is only mirrored in her actions when she knows of the consequences (cf. 2017a, p. 56). Applied to character, this might mean that Hannah knows that her Global Trait of honesty makes her speak the truth often or that Judith knows that her compassion regularly results in her helping others. While I believe - in line with my previous argument when considering Fischer and Ravizza that it seems much more plausible for agents to know of their Global Traits than their Mixed Traits - that it is more likely for agents to have circumstantial awareness of their Global Traits than their Mixed Traits, I do want to point out that the requirement does not translate exactly from action to character. That is because in the case of actions, the consequence is a direct result of the action. But in the case of having character, the consequence would be a *potential* manifestation. Thus, in the latter case, the awareness agents need would be much more hypothetical. Nonetheless, a favorable and more forgiving interpretation may yield positive results. Moreover, it is not even clear whether lack of circumstantial awareness would really limit the agent's responsibility as, unfortunately, Talbert does not further specify the exact status of the condition. While at times, Talbert remarks that circumstantial awareness plays a significant role in the

assessment of agents' blameworthiness (especially for cases of inattentive or forgetful agents), he also notes that it is not necessary (cf. ibid., p. 47). Hence, even if we cannot determine whether agents have circumstantial awareness of their traits simply for having them, it seems that Talbert's attributionist account can accommodate direct character responsibility at least for Global Traits. Consequently, Talbert's structuralist account seems better-suited to account for direct character responsibility than Fischer and Ravizza.

## 4.2 Structuralism and Manipulation

The critical discussion of Talbert's structuralist view concludes that a broadly conceived interpretation of the relation between character and evaluative judgment would allow a positive, albeit tentative, conclusion to direct character responsibility. Moving on, to analyze what his account can tell us of how to conceive of agents as indirectly responsible for their traits, there is an argument to be made that agents' actions do reflect their evaluative judgment. For example, Talbert tells us that when an agent acts "(…) on [her] deepest and most authentic values (…)" (2017a, p. 53) her action reflects her evaluative judgment and thereby makes her responsible. Thus, we could consider Judith responsible for her trait of compassion, because acting compassionately reflects that she considers being compassionate to be important. However, recall that in the case of indirect character responsibility, I have argued that the complexity of character requires that we not only consider the action itself but also how the action develops character (dimensional complexity) and whether the agent can be aware of her character structure (structural complexity). Thus, it seems that in the case of indirect character responsibility, we have to consider circumstantial awareness (at least for the purpose of analytical consideration of Talbert's account) in order to determine whether agents are responsible for their traits. As mentioned various times before, it would be harder to imagine how agents have awareness of the structure and formation of their traits, when depicted as Local or Mixed Traits. For an

agent would have to be aware that specific actions develop particular traits, which seems harder to attain given narrower or more fine-grained traits. More specifically, Talbert's account would require that these actions reflect the agent's evaluative judgment. Thus, Judith would not simply have to judge being compassionate to be important but rather that it would result in her acquiring a specific Local Trait of being compassionate-when-engaging-with-her co-worker-on-sunny-days, or her Mixed Trait pertaining to the moral domain of compassion which consists of many interrelated dispositions to be compassionate at times, yet not at others or indifferent many times in between. But, again, taking circumstantial awareness to be relevant for indirect character responsibility, it would mean that on Talbert's account an action cannot quite specifically reflect the agent's evaluative judgment when the agent is not completely aware of the consequences (cf. Talbert 2017a, 47; 56). Nonetheless, in principle, Talbert can accommodate indirect responsibility for Global Traits. And since his account does not require control, it excuses less agents than Fischer and Ravizza.

However, the initial advantage that Talbert's account has entails a strong commitment to an - at least - controversial claim, which pertains to direct and indirect responsibility for character traits alike. That is because Talbert argues for a structuralist account, which disregards the agent's history. This leads him to subscribe to the claim that even agents who are manipulated are responsible (cf. Talbert 2009). Yet, this contrasts a common intuition had by many. For example, consider Hannah who we have learned to be generally honest and who fully endorses this trait. If overnight she were manipulated to resemble the psychology of Nancy, a dishonest and deceitful person, some will have trouble viewing her as responsible for her new trait(s). Of course, cases of manipulation differ in gravity and many of the arguments exchanged stem from varying understandings of what the manipulation in question is actually about. In some of the discussed cases, agents are only manipulated with regards to a

specific feature, while others employ situations of "full-blown" manipulation.[92] Then again, some of the differences lie in the question what changes if a manipulated agent endorses these new values or whether the manipulation in question resembles an irresistible urge (cf. ibid, pp. 9-10). Hence, intuitions and arguments are likely to diverge. Thus, let me be more specific about the kind of manipulation relevant for our purposes: here, cases of manipulation mean that an agent is completely alien to her character because it is nearly implanted in them. At one point Talbert similarly describes manipulation cases as such where "(…) a person's desires or values are directly implanted in her instead of being acquired as part of a process in which the agent participates" (ibid, p. 9). This external force consequently rids the agent of a history with her character. "Hannah" is such an example of manipulation. She has no historical relation with her new character. In fact, she has no relation to this trait at all other than it being implanted in her. Initially, she is as alien to her character as she is to any other "value" she were to have by external force, such as being held at gunpoint. Yet, structuralists such as Talbert try to establish that newly dishonest "Hannah" is responsible, nonetheless. Most importantly, the argument tries to make a case for the view that the manipulation she suffers from does not preclude her from possessing ill will (cf. ibid, pp. 2-3). Further, Talbert argues that the correct values are retained in a manipulated agent. In doing so, he refers to the explanatory role of these values. "Hannah" is responsible because her actions are still governed by her internal values. And they are in fact her values, despite the manipulation, Talbert goes on, because they inform her judgment and guide her actions (cf. ibid, pp. 12-14). The reason being that

"(…) post-manipulation [Hannah's] values have their explanatory power in virtue of informing her judgments about how to behave. These judgments are internal to (…) [her] psychology, so the values in question explain (…) [her] actions *from the inside.* This is very

---

[92] See the discussion between Talbert (2009), Cyr (2020) and Mele (2020) who analyze variations of a case originally introduced by the latter (cf. 1995, pp. 145–146). Chapter 5 will come back to the issue and discuss cases of cultural membership and deprived childhood, which, I believe to differ from the stylized case of manipulation like that of "Hannah."

different from a case in which certain values explain (…) [her] actions only because they are the values of external manipulators who are directly causing (…) [her] to act in certain ways" (ibid., p. 12).[93]

I believe that this explanation proffered why "Hannah" should be considered responsible to be implausible. For first, in the outlined case, even if the manipulation pertains to only one specific aspect of her psychology, "Hannah's" new trait is not indicative of what she is normatively committed to with regards to that manipulated feature (cf. ibid., p. 7). Instead, the manipulation in question completely "(…) bypasses the victim's capacities (…) for assessing and modifying [her] values and principles (…)" (Mele 2020, p. 3148). Further, "Hannah" certainly has no authorship over her newly acquired trait. That is because the values "Hannah" now has are *not her own* simply by virtue of being action-guiding. Her dishonesty is a product of her manipulation. Yet, in adopting the structuralist position as presented by Talbert we would have to consider her responsible for her trait. Thus, employing Talbert's approach would mean that many agents are responsible for their character, even those manipulated.

Notwithstanding this assessment, it should be noted that Talbert concedes that some agents are not responsible when they are manipulated. This includes agents that are specifically manipulated to reproduce certain "verbalizations and bodily movements" (Talbert 2009, p. 9). Talbert compares these agents to those who have an unavoidable urge. In these instances, agents' rational capacities are forgone and thus the agent's evaluative judgment plays no significant role in the explanation of her actions. Thus, he argues that these types of agents in fact are not responsible. However, their responsibility is not undermined because historicism is correct but because in

---

[93] In giving this explanation, Talbert shifts from his general agreement with Frankfurt, who stresses the mesh between higher-order dispositions and the desires that move the agent. This latter view, however, Talbert maintains, is a potential target for historicists, who may call into question where the higher-order dispositions get their authority from. More specifically, the way the agent obtained these dispositions may inhibit her responsibility. Therefore, Talbert argues that the crucial point is rather the explanatory value the desires play for the agent in taking them to be a reason for action (cf. Talbert 2009, pp. 4–5).

these cases the agents' desires are not manipulated and consequently are irrelevant to the explanation for actions the result thereof (cf. ibid, pp. 9-10). While Talbert only takes those specific kinds of manipulation to render agents not responsible, it seems that there are many cases of manipulation that are not exactly as those described, yet are cases in which the values at stake are manipulated and thus seem to at least call into question the agent's responsibility. This is to say that in other manipulation cases imaginable, agents are less prone to being subjected to repeating certain phrases but rather to embracing new values and therefore acting upon them. In cases like these, it is hard to see how deceitful "Hannah's" values are not affected and how it would still be *her* values. Nonetheless, Talbert takes these manipulated agents to be responsible.

In summary, there are two observations we can make in this discussion. First, Talbert's view is generally suited to account not only for indirect but also direct responsibility for character traits. However, second, endorsing this view entails maintaining that manipulated agents are responsible, even if they have acquired their traits by means of implantation or other kinds of manipulation such as brainwashing. Thus, his account stands in sharp contrast to Fischer and Ravizza's conviction that these agents are not responsible because they lack the right kind of history and thus do not own their character.

## 5. Sher

In the previous chapter I introduced Sher's account as representative of a family of views known as capacitarianism. Capacitarians, and Sher in specific, are discontent about volitionist and attributionist handlings of a specific class of cases, namely "forgetting cases." In Sher's opinion, neither volitionists nor attributionists can sufficiently account for the intuition that agents are blameworthy even if the omission on their part is *fully* unwitting. By ways of extending these thoughts, he arrives at a version of capacitarianism, which essentially takes the action originating in the agent

to be relevant to connecting it to a reasonable standard and therefore any responsibility assessment (cf. Sher 2009, p. 117). Agents can be responsible in forgetting cases when they should or could have known and the failure is due to some combination of dispositional features that are constitutive of their self. To account for those cases in which agents act unwittingly rightly, they are responsible when they have made enough cognitive contact with the actions right-making features that allow her to act on that basis (cf. ibid., p. 143).

## 5.1 Lack of Causal Influence and Reasons

Even though Sher provides an account of responsibility which mainly focusses on actions, in his book *In Praise of Blame* (2006a), he also discusses whether agents are responsible for their character, yet with a focus on bad traits. His analysis includes a short depiction of character that seems to reflect fine-grained models such as Mixed Traits. This is interesting because, as I have argued, most analyses and discussions of responsibility for character are based on a coarse-grained understanding of character that resonates with the view I have introduced as Global Traits. Yet, according to Sher, character "(…) encompasses [a person's] (…) whole characteristic set of cognitive, affective, and behavioral dispositions - the whole collection of interrelated tendencies that together make [her] the person [s]he is" (ibid., p. 20). Similar to the Mixed Traits model, he understands character as an accumulation of various mental features, which function in interrelation. Further, he states that constitutive features of the agent, which may result in morally good or bad behavior need not be morally good or bad themselves (cf. ibid., p. 47). This mirrors the claim that Mixed Traits are neither good nor bad but rather a collection of many interrelated mental dispositions. With that, it seems, Sher - without specifically mentioning it - very closely captures the understanding of character I have introduced as Mixed Traits.

Based on this conception, he provides us with reasons of why agents are not responsible for their bad traits. Interestingly, his central question pertains to whether

agents are responsible for *having* bad character traits. This translates to what I have referred to as direct character responsibility so far (again, not only do standard analyses focus on coarse-grained character but also on whether agents are responsible for developing their traits, i.e., indirect character responsibility). In posing the question, he is quick to give the preliminary answer. For an agent to be responsible, she must causally influence that outcome (cf. ibid., pp. 67-69):

"(…) the easiest way to show that responsibility for traits is incoherent is simply to point out that where a person's merely possessing (as opposed to causing, failing to prevent, or manifesting) a trait is concerned, the basic necessary condition for responsibility is never met" (ibid., p. 67).

Yet, he shares an even more fundamental worry regarding agents being directly responsible for their traits: even if an agent had causal influence over having her character, it seems implausible to think that she had specific reasons in the outcome of her current character. However,

"(…) to be responsible for something is precisely to be subject to a request for one's reasons for causing or not preventing it. However, merely having (as opposed to cultivating) a particular character trait is obviously not something one does for a reason; so on these grounds, too, it seems unintelligible to suppose that someone might be responsible merely for having a bad character" (ibid., pp. 67-68).

Sher is swift to brush off direct responsibility for bad traits for two reasons. First, agents lack causal influence and second, they generally cannot cite reasons for having that particular character. I share Sher's worries: for first, his account requires that for an agent to be responsible, the action must originate in the agent (cf. Sher 2009, p. 117). The requirement of origination introduces causality. Agents have no causal influence over their bad traits simpliciter. This is enough to show that even if the agent should or could have been aware or did know of her character, her *having* a character is not suitably connected to her. But Sher also stresses that agents have to be able to respond to reasons. He does so because those agents, who do not retain their rational capacities, i.e., those who are globally incapacitated, are exculpated as they cannot recognize and appropriately respond to reasons. Thus, the features are not constitutive

of the agent and there is no suitable connection between the agent and the action (cf. Sher 2017, pp. 11–12). In consequence, I agree that Sher is correct in denying direct responsibility for bad traits as his view requires a causal connection that is not established simply by having a trait.

However, focusing on bad traits, Sher does not specifically attend to the matter whether this also pertains to character generally. In order to account for those instances in which an agent unwittingly acts rightly, Sher introduces the last and third clause of his epistemic condition, which holds that these agents are responsible, when they have made enough cognitive contact with the action's right-making features that allow her to act on that basis (cf. Sher 2009, p. 143). Sher adds this extension because agents must perform those actions for the right reasons. Yet, it is not hard to see that the arguments against direct responsibility for bad traits apply to good traits equally. Agents neither causally influence their good traits, nor have them for a reason simply for having them. And this applies to all kinds of character conceptions. That is because both arguments are independent of the specific structure of character. Rather, they refute the applicability of the concept. Therefore, even though for different reasons, Sher's view, like that of Fischer and Ravizza, overall negates that agents are directly responsible for their character.

## 5.2 Reasonable Expectation and Cognitive Contact

Even though Sher provides us with an argument of why agents are not directly responsible for their bad traits, he does not *explicitly* address whether they are indirectly responsible. Nonetheless, he discusses whether agents have any effective control over their character traits, which he ultimately denies. He argues that in order to have effective control over the development of these traits, one would have to believe that some actions would prevent them from developing, and, even more so, put these actions into practice. Neither can be expected from a child or adolescent; yet agents primarily develop their traits in their childhood (cf. Sher 2006a, pp. 53–55).

However, this argument need not necessarily mean that Sher's account forecloses indirect character responsibility. For recall that it is motivated by cases in which agents are intuitively responsible yet either forget, lack any exercise of ill will, or do not recognize the action's wrongness or underestimate the action's moral weight. In fact, his own account allows for agents without control or awareness to be responsible for their actions as long as it is suitably related to the agent and she could and should have known better or the agent has made enough cognitive contact with the action's right-making features (cf. Sher 2009, p. 143). We might apply this to indirect character responsibility as follows: considering the case of the development of bad traits, we first need to analyze whether there is a suitable connection between the agent's actions, which develop her traits, and herself. Afterwards we then need to check whether the agent could have known that these actions would habituate her bad character.

Beginning with the former, it is readily apparent that the argument regarding the lack of causal influence for direct character responsibility does not apply in the case of indirect character responsibility. On the contrary, agents are generally suitably related to their actions that develop their traits because their constitutive features can be thought to be causally explanatory of them. Nonetheless, Sher shares a concern regarding the structural and dimensional complexity of agents' character traits and the formation thereof that still inhibits them from being indirectly responsible for them: "Our characters develop slowly and by accretion, and their development is influenced not only by the decisions we make and the situations into which we enter, but also by our innate tendencies and the innumerable unchosen exigencies with which life presents us. Given the complexity of each factor, and given the exponentially greater complexity of the ways in which the different factors can interact, we rarely make decisions with the clear understanding that they will cause us to acquire traits or habits that are markedly worse than normal. (...) We generally have little idea of which traits we will develop if we do or do not marry a certain person, undertake a certain career, or put down roots in a certain part of the country" (2009, p. 38).

Taking this paragraph by heart initially paints a pessimistic picture for Sher's account. Even if the actions simpliciter are suitably related to the agent, due to the

circumstances and dimensional complexity of character development in general, combined with the structural complexity of fine-grained character, agents could not have been reasonably expected to know that certain actions habituate specific traits. And in fact, I actually believe that the argument not only pertains to bad traits but also those that are not markedly worse than normal. To see why, recall that the third clause of his epistemic condition is introduced to account for those who act unwittingly rightly. This does not require awareness or that the agent could have been reasonably expected to be aware but rather solely that the agent has made enough cognitive contact. Those, who act rightly but are unaware, „(…) accurately but unconsciously process […] the information to which they have access" (ibid., p. 143). Applied to character, this would mean that the agent must have made enough cognitive contact to be able to perform the actions that habituate their traits on that basis. Yet, even though it suffices that these agents have general evidence of these traits and the actions which habituate these traits, we can readily see that it would prove significantly harder to make enough cognitive contact which would enable one to perform the actions that habituate the agent's traits. Thus, considering this complication, it would render agents not indirectly responsible for their character traits (neither good nor bad).

However, the argument rests on a conception of character which I believe resembles Mixed Traits. Thus, the problem might not strike as deeply for structurally more coarse-grained conceptions of character. For consider that agents potentially have more awareness of these traits, because they have more cognitive contact with them as they manifest in a diverse array of situations. This is an argument I have mentioned before when considering Fischer and Ravizza's and Talbert's accounts; Global Traits are structurally more coarse-grained and manifest not only in temporally stable but also cross-situational behavior. This makes them less complex and increases the chance of agents being aware of these traits and how to habituate them. So, if, similar to the concept of Global Traits, the structural complexity is lowered, it seems that Sher could potentially account for this variation. This is not to refute the argument from complexity, but instead to point out that it seems relatively less powerful for

Global Traits than Mixed Traits as chances are higher that agents have more awareness or make enough cognitive contact with them. If these coarse-grained traits manifest in cross-situational consistent behavior and are discernible as such, it seems that agents can be reasonably expected to be aware of them and how they develop. Hence, I believe that the argument does not carry through with the same force to Global Traits. Hence, it seems, Sher's view could account for indirect responsibility at least for coarse-grained character traits if interpreted accordingly.

## 6. Lessons Learned: Framing the Challenges

In the general introduction, I identified a research gap pertaining to the topic of responsibility for character traits. Many analyses focus on globalist and volitionist theories, neglecting recent theoretical developments that have added to these views. Furthermore, most of the analyses consider a particular interpretation of the research question, which I have referred to as indirect character responsibility. Hence, the desideratum of this thesis, which directly builds on the research gap, is to analyze what alternative theoretical approaches can tell us about responsibility for character traits, both direct and indirect. I defended this approach against the potential objection that it may be futile as theories of responsibility generally focus on actions and therefore may not be suitable for an analysis with such a focus. I maintained that exploring how these theories respond to different conceptions of character is fruitful nevertheless, because not only is there no character-specific theory available and existing analyses, too, employ action-centered theories, the envisioned approach can actually enrich the discussion and point to new directions. And, as a matter of fact, the analysis shows that character generally *and* specifically conceived of as Global Traits, Local Traits, and Mixed Traits poses additional challenges to these theories as each yields different results, not only regarding the different concepts of character but also the two different forms of responsibility introduced at the beginning of this chapter. Therefore, I now

want to take a moment to frame the challenges that have emerged so far more generally, i.e., what can we learn from the discussion in this chapter? A clear understanding of the challenges will provide us with a sound basis of understanding what issues we need to address when considering character responsibility.

The first challenge I want to draw attention to pertains to the two forms of responsibility. Looking at both questions separately reveals that Fischer and Ravizza are optimistic regarding indirect character responsibility, yet their account seems troubled regarding direct character responsibility. Then again Talbert provides positive answers for both interpretations of the question. And finally Sher is pessimistic about either version of character responsibility (though this may not carry through for indirect responsibility regarding more coarse-grained traits). Hence, the assessments differ with regards to what each has to say about responsibility standardly conceived of as indirect but also in the way responsibility for character can be direct, requiring no prior responsibility had for actions that develop character. Moreover, not only do the theories vary with regards to whether they take agents to be responsible directly and indirectly, they also vary with regards to why they answer the question of direct character responsibility positively or negatively. Indeed, it is noticeable that neither Fischer and Ravizza nor Sher accommodate direct character responsibility, though each for separate reasons. Fischer and Ravizza preclude that agents are directly responsible for their character as they conceive of responsibility as essentially linked to historical control (therefore not denying direct character responsibility but rather yielding no results at all), and Sher rules out that agents are directly responsible for their traits on the grounds of the lack of reasons and causal involvement of the agent. On the contrary, Talbert's attributionist account fares better in terms of a general fit. However, I pointed out that his reasoning for a structuralist approach leads him to endorse the view that manipulated agents are responsible - a claim rejected by many. Hence, while Talbert may be the only one to account for direct character responsibility, his view may ultimately take too many agents to be responsible. These observations point us to our first, twofold, challenge. On the one

hand, it tells us that an analysis of character responsibility must clearly differentiate between the two types of responsibility and demonstrate how both concepts relate to one another. On the other hand, since analyses of whether agents are directly responsible for their traits are rare (one of those exception is Sher (cf. 2006a, p. 68), yet only with a focus on bad traits) and the responses of the action-centered theories are rather pessimistic[94], a character-centered analysis must fathom whether there is a viable version of how to conceive of this variant of the question. As mentioned in the introduction of this thesis, the latter interpretation of the research question is not only less often subject of consideration, there is also genuine philosophical interest to inquire agents' direct responsibility for their character traits.

The second challenge concerns the control condition. This chapter mainly juxtaposes two positions: while volitionists argue for historical accounts, attributionists often propose structuralism, denying that the agent's history is relevant to her current assessment. This reflects the way in which the latter structurally fits well with direct character responsibility and the former with indirect character responsibility. History-focused views such as those favored by Fischer and Ravizza align with the concept of indirect character responsibility because the agent's history, and thus the acquisition of her values, is naturally embedded in their account. In contrast, attributionist accounts like those of Talbert fare well with agents being directly responsible for their traits because they "only" require that the agent's character must reflect her quality of will. At first sight, adopting Fischer and Ravizza's historical approach also reflects an intuition had by many. Namely, that since an agent's environment and thus her history strongly correlate with how she will turn out - yet is not controlled by her - it should matter to her assessment of her responsibility. However, if we opt for a historical account with focus on control, we may excuse many agents and not be able to show how they can be directly responsible for their traits. If, on the other hand, we opt for a structuralist account, we may be able

---

[94] Talbert provides a positive conclusion, yet may include too many agents considering his take on the case of manipulation.

to account for direct responsibility, yet end up taking manipulated agents to be responsible, a hard pill to swallow. Hence, the discussion of the different accounts stresses the need to clarify in which way the agent's history is relevant to the responsibility she has for her traits. Therefore, we need to investigate whether there is a version of character responsibility that acknowledges the importance of the agent's history whilst showing how they can be directly responsible, not excusing too many.

The third challenge, which became eminent in this chapter's discussion, concerns the epistemic condition. Volitionists such as Fischer and Ravizza maintain that agents need awareness to be responsible, while attributionists like Talbert argue that this only pertains to circumstantial awareness. And capacitarians such as Sher believe that agents can be responsible for fully unwitting omissions. The three theories of responsibility introduced vary with regards to what they require of the agent. Despite the diversity of the epistemic condition embroidered in these theories, there is surprising overlap in the analytical conclusion regarding the different conceptions of character. All of them fare better with coarse-grained structured character, even those theories that usually take agents to be responsible even when they lack awareness. A likely explanation for this is the structural and dimensional complexity of character which raises the question in what way the epistemic condition translates to character. I argued that it requires us to broaden the scope in two ways. First, a thorough analysis must pay close attention to the structure of character. Second, and this pertains to indirect character responsibility specifically, we need to investigate not only whether agents satisfy the necessary conditions for the actions, which habituate their character traits but also whether they do so for their character simpliciter as well, i.e., expand the dimension of examination. Overall, the analysis illustrates that the structural and dimensional complexity of character may limit what the agent can actually know about her traits. This is especially the case for narrower and fine-grained concepts: situationism - and the Local Traits concept based on it - highlights the way situational factors bear on agents (similarly, Mixed Traits rely on triggers). Yet, irrespective of the conception used, it opens up the question of what agents need to know about their

characters to be considered responsible for them. Furthermore, character traits might not necessarily always develop consciously. Hence, an adequate analysis must exemplify the way in which agents can be responsible for their traits without consciousness if it is not to excuse a significant number of agents. Therefore, an analysis of responsibility for character must be attentive to these complications and specify the epistemic condition to account for the intricacies of character. Overall, the variable of character may require a tailored epistemic condition.

Because all theories face these difficulties in light of the complexity the variable character introduces, I propose the following claim: analyses must discuss character responsibility as an independent phenomenon which requires special attention to these challenges respectively. Therefore, we are tasked with no less than formulating a view of responsibility of character traits which shows that agents can not only be indirectly but also directly responsible for their traits while explaining why manipulated agents are not responsible and their history is in fact essential without excusing too many and whilst configuring the epistemic condition as such that agents can also be responsible for their fine-grained traits. Consequently, this means constructing a framework that explains character responsibility considering a whole array of restrictions arising from complexity.

## 7. Conclusion

This chapter explored in what way three representative theories of responsibility react to the concepts of character found in the literature. Yet, in order to gain maximal clarity, I first made note of the complexity of character and introduced a differentiation between two types of responsibility, direct and indirect responsibility for character traits. Furthermore, this complexity is mirrored not only in the ambiguity of the research question but also the composition of traits (structure) and what needs to be considered on each (dimensions). With these clarifications I went on to discuss what

insights we can gain from different theories of responsibility about agents' direct and indirect responsibility for their traits. Based on the critical discussion of each of the accounts, I continued to carve out and group the challenges that we are faced with. As my main thesis, I concluded that character adds significant complexity (structurally and dimensionally) and thus must be viewed as an independent phenomenon considering these challenges. First, an analysis of character responsibility must not only explicate in what way agents can be responsible for their character indirectly but also whether they can be directly. Second, it must present a plausible control condition. More specifically, it must demonstrate the way in which the agent's history is essential to her responsibility. Lastly, it must explicate in what way agents need awareness and delineate an epistemic condition that incorporates the complexity of character. If responsibility for character requires the satisfaction of an epistemic condition, an analysis focusing on character must elucidate what agents need to know about their character, specifically whether this also requires awareness of the nature of character and the formation thereof, while acknowledging that traits are not always developed consciously. Hence, the next chapter establishes a character-centered framework that aims to incorporate the challenges laid out.

# V Towards a Character-Centered Framework

## Incorporating the Three Challenges

---

## 1. Introduction

In the previous chapter I reviewed what theories of responsibility representative of the debate can tell us about agents' responsibility for their traits considering various conceptions of character that mainly differ with regards to their structural composition. The analysis, I maintained, exemplifies that character is relatively more complex than actions, suggesting that analyses must consider character as a distinct phenomenon.

This chapter takes an attempt at addressing this first thesis by outlining a character-centered framework, taking into consideration the three challenges outlined above. Here, I make my case for my second thesis, namely that responsibility for character traits does not require control and only minimal awareness. To do so, I start by discussing the applicability of the two conditions of responsibility. Regarding the control condition, I argue for a defeasible structuralist version which focuses on the voluntary nature of character acquisition and ensures that manipulated agents are not considered responsible. Regarding the epistemic condition, I maintain that agents need only minimal, coarse-grained, knowledge of their character and the formation thereof. Overall, and this is my third thesis, agents are responsible for their character when its development has been voluntary and it reflects the agent's quality of will. Thus, I propose that an attributionist approach best suits our needs in order to account for both types of responsibility, yet needs some adjustment to account for the importance of the agent's history and the limitations of awareness generally had. Further, I examine cases of cultural membership and deprived childhood which are

often thought to excuse agents due to the epistemic restrictions they endure. I argue that these agents *can* be responsible for their traits nonetheless but acknowledge that these agents do not exactly resemble "normal" agents with no such limitations. Based on the discussion I advance my final thesis, namely that character responsibility comes in degrees. Biographical specifics about an agent's past can mitigate the degree of her responsibility, yet seldomly exonerate her. Returning to "normal" agents, this degrees-view additionally explains cases of undeveloped agency and the fact that all agents vary with regards to what they know and control. Thus, while some agents may be closer to the minimal threshold others satisfy the conditions more easily. Nonetheless, agents are responsible for their traits, yet to a varying degree.

## 2. The Conditions of Character Responsibility

### 2.1 Character and the Control Condition

#### *2.1.1 Ultimate Moral Responsibility and Self-Constitution*

Tasked to find a viable control condition suited for character specific needs, we can draw from the theoretical repertoire to see what will be helpful for constructing the framework. The last chapter outlined the challenge regarding the control condition as follows: due to the contrasting views of historicist and structuralist accounts entailing different conclusion to what class of agents can be considered responsible, we need to outline the way in which the agent's history is important to her current assessment. Historicists stress the relevance of the agent's history for her responsibility, yet tend to emphasize a control condition that may potentially excuse many agents. Structuralists, on the other hand, have a hard time handling cases of manipulation, deeming even those responsible who have obtained their traits via external force such as implementation. Therefore, in order to avoid this latter conclusion, it seems that at first glance a character-centered framework is best advised to employ a historical account

even though we may be lead to excuse many agents and take them not to be responsible. Yet, this strategy comes with some difficulties that go beyond potentially excusing agents that are in fact responsible. That is, opponents of historicism sometimes state a powerful objection: they stress that the historicist narrative is faulty because on the historicist view envisioned we need to find a time in the agent's history in which she has control over her makeup. But this, critics argue, only shows that agents are not responsible because they lack *ultimate* moral responsibility. Ultimate responsibility pertains to the problem that no one can ever be *truly* responsible because no one is the cause of herself (cf. Strawson 1994, pp. 5–6). The basic argument goes back to Galen Strawson, who maintains that

"it is undeniable that one is the way one is, initially, as result of heredity and early experience, and it is undeniable that these are things for which one cannot be held to be in any responsible [sic!] (morally or otherwise)" (ibid, p. 7).

It opens the pressing question in what way agents' histories really matter and if they can become responsible when they can only start from what has already been given. The problem stresses a truism, namely that agents are not *ultimately* responsible because they have no control over the environment they are born into or the personality they end up having. Yet, both factors largely contribute to the development of agents' character traits (cf. Russell 2009, pp. 380–383). Hence, agents' future choices are always colored by their current character which they had no control over (cf. Katsafanas 2017, p. 138). The problem of ultimate responsibility can also be framed in terms of constitutive luck.[95] Agents are simply lucky regarding their general constitution, which "(…) concerns the aspects of agents that make them who they are, such as their traits and dispositions" (Cyr 2019, p. 197), and their environment, which influences the education they are given (cf. Sher 2006a, p. 54). Hence, the lack of self-

---

[95] The term refers to Thomas Nagel's general concept of *moral luck*, which describes the seemingly contradictory observation that agents are often assessed for things beyond their control. This is best illustrated in terms of two reckless car-drivers, of whom only one runs over a child. While both do the same, only the driver who consequently kills an innocent child will face severe blame and punishment (cf. Nagel 2012, pp. 25–26). Constitutive luck specifically comprises the sense in which all agents are lucky in terms of their initial constitution (cf. Nelkin 2023).

construction "(…) jeopardizes the notion that one's character could be 'one's' own, in any richer sense than that one's character is located in one rather than in others"(Russell 2009, p. 377). Yet, the peril of the challenge is looming even more so as other authors (e.g., Taylor Cyr) have recently introduced a further complication of the issue: agents' first actions stem entirely from character which depends on constitutive luck. Thus, if agents are ever to be responsible, they have to be responsible for actions, which stem from constitutively lucky character. This, however, demonstrates that every agent's history resembles that of a manipulated agent. Neither controls the process of acquisition of their character. The manipulated agent's character is installed via some type of external configuration; the "normal" agent is constitutively lucky with regards to her initial constitution and her environment. Hence, looking at the issue from this perspective only goes to show that there is no relevant difference between manipulated agents and those who are constitutively lucky (cf. Cyr 2020, pp. 2386–2387). Thus, even if the "normal" agent becomes responsible, she is still not different from the manipulated one with respect to her initial constitution; meaning that we cannot simply trace back in time because any form of the agent endorsing her current constitution will be determined by her prior constitution, we would face a regress (cf. Cyr 2019, p. 206). In consequence, since all agents are constitutively lucky, it stands to question how we can square the truth of constitutive luck with agents having control over their character. Hence, when considering the control condition for character, we are faced with the following challenge: on the one hand, we can see that the standard structuralist approach cannot account for the intuition regarding manipulation. Yet, this difficulty illustrates that the agent's history is extremely relevant to the assessment of her responsibility. On the other hand, it stands to question how the lack of ultimate responsibility is compatible with agents having control at all. If anything, the complication seems to suggest that if (historical) control is a condition of character responsibility then agents do not satisfy it.

*2.1.2 Voluntariness of Character Acquisition*

Even though the problem of ultimate responsibility primarily concerns questions of free will, it does relate to our present inquiry. For contrasting the facts of ultimate responsibility and constitutive luck with the intuition many have about manipulation cases begs the question of how "normal" agents, who are constitutively lucky, differ from those who are manipulated (cf. Barnes 2016, p. 2315; Cyr 2020, p. 2387).[96] The reason for this is that both lack control over their initial constitution. But if there is no relevant difference we are confronted with a choice: either we retract our initial assessment that manipulated agents are not responsible or we find a suitable explanation of how "normal" agents can become responsible despite being limited by their initial constitution (cf. Talbert 2009, p. 14).

One frequently employed strategy to counter the worry about ultimate responsibility from those who aim to uphold historicism is to argue that self-construction is not necessary to own one's character and ensure that the agent has the right history:

"(…) faced with the ineliminable influence of constitutive luck on character, we find the option of abandoning ascriptions of responsibility for character altogether unattractive. To do so is to abandon a significant portion of the attempt to understand one's life as one's own. We can acknowledge the role of influences, background, circumstances, and the like without abandoning the conception of people as voluntary agents" (Jacobs 2001, p. 11).

The argument accepts the challenge of ultimate responsibility and constitutive luck but questions the implications (cf. Moody-Adams 1990, pp. 111–112; Jacobs 2001, pp. 20–21; Russell 2009, p. 388; Fischer 2013, p. 21). Those endorsing this strategy argue that instead of control over their constitution, agents need to have a suitable degree of control over the process by which their character is formed (cf. Barnes 2016, p. 2312). At best, it is a process of critically distancing (cf. Russell 2009, p. 392) and "(…)

---

[96] Others point out that the facts about manipulations cases do not differ from those that "obtain in an ordinary deterministic world" (Tognazzini 2014, p. 358). Hence, the task is to carve out which external interferences compromise responsibility and which do not (cf. Fischer 2004, p. 145).

reflectively embrac[ing] and identify[ing] with the values at issue (…)" (Mele 2020, p. 3149). It is not to refute the valid point that our environment heavily influences our social reality, possibilities, and hence the situations we find ourselves in (cf. Annas 2011, pp. 22–25). But

„(…) there are infinitely many factors over which I had no control, which are such that, if they had occurred, I would not be as I am. (…) I am thus not 'ultimately responsible.' And yet this does not in itself seem to expunge or etiolate my agency and my moral responsibility" (Fischer 2013, p. 171).

While I agree with the general direction of the argument, its aim is to align the fact that agents have no control over their initial constitution with the claim that this does not prevent them from exercising control over the process of character acquisition. However, even though this strategy acknowledges that agents do not in fact control every aspect of their character formation, it does not address an important worry, namely that it still focuses on the agent's control over some part of her self-creation (whether it be at the time of her initial constitution or in her later development). Yet, as mentioned before, character traits, in contrast to single actions, are much more complex (especially considering the way Local and Mixed Traits are thought to function). Under these conditions, only a very limited number of agents may exercise the control required (even if it does not extend to all aspects of life) to be responsible for these traits. Hence, while this strategy may curtail the issue of ultimate responsibility - and of constitutional luck respectively - it still carries the burden of potentially excusing many agents because it still centers on the idea of control, a conditional that may not be satisfied by many agents, if at all by any. Therefore, the extenuated version of historicism, which merges questions of ultimate moral responsibility and constitutive luck with process control over the formation of character, may not be applicable for the purposes of a character-centered model as envisioned here.

At this point, one might reconsider whether a structuralist approach might be the solution to the issue after all and question whether proponents of this family of

views really have the problem I have been trying to outline here as essential to our inquiry. Since they require that an agent's action must reflect her quality of will in order for her to be responsible for it, the counter-argument against my criticism on their take on the problem of manipulation might be as follows: the agent could easily maintain that due to her manipulation, the action no longer reflects her quality of will. Hence, taking this explanation at face value, it seems that structuralism does not have an issue with manipulation. However, the problem is generated because attributionists argue that these agents can be responsible simply by endorsing their psychological structure after the manipulation occurred[97] (recall a slightly amended, yet similar view presented by Talbert (2009), who argues that the agent is responsible when her values play an explanatory role in her actions). Historicists take issue with this explanation proffered because the agent might endorse her new self even though she might have been reluctant or even completely against being manipulated prior to the externally ensured configuration of her traits (cf. Zimmerman 2003, p. 650). To be more specific, recall Hannah, whom we have encountered as a nice, pleasant person who has been the unfortunate victim of overnight manipulation in order to resemble a deceitful version of herself - "Hannah." If Hannah were to be asked whether she would want to transform from a lovely, honest person to a devilish counterpart, she would probably resist this course of action. However, "Hannah" may fully endorse her evil side. Yet, speaking in terms of Talbert's explanation given, these new, despicable, values will play an explanatory function in her actions. Thus, she will satisfy the structuralist conditions of responsibility. However, something seems off about this case (cf. Zimmerman 2003, p. 650). And even though structuralists typically acknowledge that manipulated agents are "cheated of something valuable, namely the opportunity for naturalistically realized self-creation," (Zimmerman 2003, p. 649) they contend that these agents do not lose their reasoning capacities and thus still can be responsible (cf. Zimmerman 2003, p. 649). However, again, this view does not sufficiently explicate

---

[97] For example Frankfurt suggests the idea of endorsement by saying that agents must identify with their desires (cf. Fischer and Ravizza 2000, p. 184; Zimmerman 2003, p. 639).

how the agent comes to own her values when she is manipulated as it seems evident that her reasoning capacities would be affected by the manipulation in question to approve of these values. Therefore, there seems to be something right about the historicist handling of (character) responsibility: it simply seems that how an agent chooses, decides or carries herself cannot be viewed without how she came to have these capacities for reasoning (cf. Zimmerman 2003, pp. 640–641).[98]

One proposal that stresses the relevance of the historical process, yet does not resort to control, can be found in John Christman's writings on the adjacent topic of autonomy (2009).[99] Christman takes an agent's characteristics to be autonomous iff she "Were (…) to engage in sustained critical reflection on C [the respective characteristic; S. Sch.] over a variety of conditions in light of the historical process (adequately described) that gave rise to C; and (…) She would not be alienated from C in the sense of feeling and judging that C cannot be sustained as part of an acceptable autobiographical narrative organized by her diachronic practical identity; and (…) The reflection being imagined is not constrained by reflection-distorting factors" (2009, p. 155).

Borrowing from this thought, in the context of character responsibility, this may translate to an agent being responsible iff she were not to feel alienated from or reject her character traits were she to reflect the way she has come to have them.[100] However, if the agent were to reflect upon the historical process in which she has come to have the traits and upon learning of the heteronomous nature thereof, she rejects the process of acquisition, she is not responsible. For example, Hannah can be responsible for her traits because upon reflecting on her history, she does not feel alienated. "Hannah" on the contrary, may feel alienated when finding out that her traits, as much as she might endorse her relentlessly evil characteristics now, are not in fact *hers* as they are the product of overnight manipulation. This alternative view offered by Christman captures the way in which an agent's history is relevant:

---

[98] Structuralists do not deny that responsibility cannot be a matter of process but they maintain that the agent's internal features are sufficient (cf. Zimmerman 2003, p. 642).

[99] Thank you Laura Valentini for suggesting to consider this account.

[100] Note that the agent need not in fact reflect on these traits as Christman takes this to be a hypothetical (cf. 2009, p. 145).

considering the process, the way an agent comes to be who she is, does matter to her assessment. Those manipulated may feel discomfort, or - in Christman's term "alineation" - upon learning that their traits are external to them in a fundamental way. However, even though this proposal is a step in the right direction, it still does not sufficiently differ from other structuralist approaches considered before, which stress the agent's present psychological makeup as sufficient to ground responsibility. That is because on this view, the agent may possibly *not* feel alienated even upon learning that she is the product of manipulation, inviting the historicist criticism mentioned before: the post-manipulated agent will endorse her new traits by default; and thus maybe even the acquisition thereof. Hence, this view, while addressing the correct issue, is unfortunately ambiguous regarding manipulation.

As the importance of the process of acquisition is apparent, other authors, who endorse structuralism, have considered ways in which to incorporate the agent's history into her responsibility assessment. One of these alternatives that might be of interest here is the suggestion by Taylor Cyr (2020). He argues that the fact about constitutive luck makes a strong case for structuralism, yet that it is evident that the agent's history is relevant to her assessment. He arrives at a version of structuralism, which he dubs "history-sensitive." This view entails that structuralism suffices to ground the agent's responsibility, but the agent's history may vary the *degree* to which the agent is responsible. Hence, he does agree with other structuralists that manipulated agents are responsible. Instead, where he disagrees concerns the claim that manipulated agents are *as responsible as* "normal" agents not suffering from manipulation (pp. 2381-2394). One point in favor of this view is that it combines features of both, historicism and structuralism, and, most importantly, acknowledges that the agent's history does seem to influence the way she is responsible. Hence, taking this view to apply to the case of "Hannah" would mean that she is responsible even though she is manipulated into being deceitful. However, since her history is far from being that of Hannah, she may not be *as* responsible. Nonetheless, while this approach is enticing, and certainly comes closer to including the relevant factors of

character responsibility outlined, it still leaves us unable to account for the intuition that manipulated agents, and specifically "Hannah" in this case, are not responsible (or that their responsibility is at least called into question). Yet, most basically, "global induction of values is so disturbing […] because it involves the obliteration of a person's entire evaluative identity and its replacement by a radically different one" (Zimmerman 2003, p. 650). Hence, the problem of manipulation supports the view that structuralism simpliciter and the history-sensitive version presented by Cyr do not quite capture what is needed to explain our intuition regarding manipulation, whilst accounting for the complication that agents may not always control their traits, especially considering the character models that stress the more fine-grained nature of trait dispositions. Most importantly, it illustrates the need to incorporate the process by which the agent came to have her character more fundamentally.

The whole problem is now before us: character traits are inseparable from how one has come to acquire them, which suggests a strong focus on the historical process. However, standard volitionist accounts and variations thereof tend to potentially excuse many agents, an issue that is intensified considering the different models of character, which suggest to conceive of trait dispositions as narrower or more fine-grained. Structuralists have a more inclusive grasp on the matter, yet all variations considered fail to adequately rule out those who are manipulated. Hence, so far, neither an adjusted version of historicism nor structuralism has succeeded in explaining the full range of cases. It seems we are at a deadlock that forces us to bite either one bullet outlined before.

Despite this bad outlook, I believe that there may be an alternative, which can combine both our intuition that manipulated agents are in fact not responsible and which relies on the agent's history as relevant to her assessment, yet that does not require control. The alternative I have in mind takes insight from the claim advanced by Sher, namely that control is not necessary for responsibility but that voluntariness is: "(…) to say that agents can be responsible for acts that they have not in *any* sense performed voluntarily would indeed be to distort our concept of responsibility beyond

recognition" (cf. Sher 2009, p. 149).[101] Similarly, the view I hope to convince readers of is based on the notion of voluntariness. One way in which we may introduce voluntariness but refrain from reverting to control may be realized by appealing to the encountered structuralist view that an action must reflect the agent's quality of will in order to be attributable to her (cf. Talbert 2023) with the additional feature of voluntariness. By opting for such an attributionist account, we are be able to argue that agents do not need control, neither present nor historical, yet are in the position to consider the agent's history.

In considering my proposal, one might be quickly inclined to question the difference of this modification to the standard explanation given by structuralists which renders manipulated agents responsible as long as they endorse their post-manipulation make-up or their values play an explanatory role for their actions (e.g., recall Talbert's view on this in chapter 4): would the proposal advanced not ultimately resort to the same kinds of issues that were the reason for rejecting these alternative options of structuralism in the first place? That is because introducing voluntariness does not necessarily alter the outcome. For see that it could be argued that structuralists already take voluntariness to be a given conditional as endorsing one's character after the manipulation would also amount to some type of voluntariness. Or alternatively that the agent satisfies the voluntariness requirement when her present values play an explanatory role even though the acquisition thereof prior has been involuntary. Yet, from the case of "Hannah" we have learned that endorsing her values *after* the manipulation does not suffice to escape the historicist criticism, namely that pre-manipulated Hannah would not want to assent to the manipulation in question,

---

[101] Since Sher argues that agents can be responsible without consciousness, he is adamant about this entailing that agents neither need control. That is because if agents need control, they also need awareness. Thus, agents lack control if they have no awareness (cf. Sher 2001, p. 149; Clarke 2017a, pp. 65–66). In consequence, he argues for a voluntariness condition that is stripped off the control requirement. He does so by showing that the control requirement is not in fact what is needed. Instead, what is needed, he argues, is that the action originates from the agent voluntarily. By doing so he is in a position to maintain that agents can be responsible for fully unwitting omissions, i.e., have no consciousness (cf. ibid., pp. 146-149). For the complete argument see chapter 9 in *Who Knew* (cf. Sher 2009, pp. 138–154, specfically pp. 145-151).

but that post-manipulation "Hannah" does in fact endorse her newly acquired traits (cf. Zimmerman 2003, p. 650). Therefore, it seems that no progress is made with the suggestion given. Thus, simply appealing to voluntariness does not fully help to discern between those who have the right and wrong kind of history. Yet, this is why the proposed modification takes precisely this complication as a starting point. By referencing the relevance of the voluntary nature of *the process* of character acquisition, this variation simultaneously construes the structuralist condition as a defeasible one: in regular cases, conditions of structuralism suffice to explain an agent's responsibility but if the agent's history involves forms of involuntary trait acquisition, she is not responsible. Hence, the proposal is the following: agents do not need to control their character trait (development) as long as the actions that develop their traits or their traits simpliciter are attributable to their quality of will, meaning the acquisition thereof must be voluntary. With such a view, we can rule out those as responsible that have been manipulated and support a notion of character responsibility that takes the agent's process of how she came to have her traits to be essential to her assessment, while holding that what makes an agent responsible is not control but rather whether the traits are a reflection of her quality of will.

The critic might now think that this modification is a promising strategy but press her worry that there is still no relevant difference between manipulated agents and those who are constitutively lucky. Neither acquires their initial constitution voluntarily. Hence, appealing to the voluntary process by which an agent's character must be acquired leaves us unable to differentiate between those, who are manipulated and those who are constitutively lucky. While I take this worry seriously, there is a sense in which both types of agents do in fact differ: one way in which manipulation seems so severe that it ultimately compromises responsibility is in cutting the ties between the agent and what she takes to be reasons. Appealing to this complication, Fischer and Ravizza note that what is so disturbing about manipulation cases is that the agent does not respond to any kind of reason because she would act as she does despite any other reason given (cf. 2000, p. 232). This problem is not limited

to volitionists. Attributionists such as Smith conceive of responsibility in terms of answerability (cf. 2008, pp. 380–381) and therefore maintain that: "States that are not even in principle answerable to a person's judgment, therefore, would not be attributable to a person in the relevant sense" (Smith 2005, p. 256). The same restriction does, however, not apply to agents who are constitutively lucky. Constitutively lucky agents can in principle engage their reasons-capacities. Further, the condition of voluntariness also connects the agent to her character and thus gives a sensible explanation of why the "normal" agent - contrary to the manipulated one - owns it, while refraining from falling back to a control condition.[102] This form of voluntariness, which, again, is not to be conflated with control, takes the agent's traits to be hers because they are in fact attributable to *her*, despite all agents being constitutively lucky by construction. In consequence, I maintain that this strategy reflects the way agents do not constitute themselves yet avoids the problems proprietary to structuralism simplicter that standardly endorses the view that manipulated agents can be responsible.[103]

## 2.2 Character and the Epistemic Condition

### *2.2.1 Complexity and the Limitations of Awareness*

Now that we have addressed the challenge which asks us to outline the way in which the agent's history is relevant to her character responsibility assessment, there are two challenges left to consider. Therefore, I now want to progress by addressing the challenge regarding the epistemic condition. In essence it demands a specification of

---

[102] The concept of taking responsibility as proposed by Fischer and Ravizza also addresses the way an agent must own her mechanism (employing the respective term). However, as volitionists, Fischer and Ravizza introduce this idea within the broader view that an agent needs historical control (cf. Fischer and Ravizza 2000, p. 207).

[103] I would like to particularly thank Erasmus Mayr for suggesting to construe the condition as a defeasible one and consider the applicability of attributionism.

the awareness required, especially with regards to the structural and dimensional complexity of character that may limit what can be known about these traits and the further complication that character traits may not always be developed consciously. Hence, adhering to the challenge means spelling out in what way the epistemic condition translates to character paying attention to the specific "needs" of character.

In considering this complication for Fischer and Ravizza, I argued that it limits what their account can accommodate. I maintained that their volitionist view can account for Global Traits because it is plausible enough how an agent would know of the traits and how to habituate them as the structure of Global Traits should enable agents to have enough cognitive contact with them to moderately reasons-responsive. Further, it would be easier to know how an action translates to character. However, this view puts serious restraints on alternative character conceptions. Given the structural complexity and the added dimensional awareness needed, it is easy to see that it would be much harder for this condition to be satisfied. Depending on the conception at hand, this would prove infinitely more difficult. How could agents ever be *that* aware? Further, the account has a serious mark against it, as it potentially excuses too many agents (analogously to considering their account regarding control) and generally does not reflect that many traits are developed unconsciously.

However, not only Fischer and Ravizza are faced with the challenge to explain how agents can be responsible for their narrower or more fine-grained character traits in combination with the dimensional complexity. Neither attributionists nor capacitarians yield different results. This may come more as a surprise since these theories are eager to stress that a strong epistemic condition in the case of actions entails that it might not be able to account for all cases and thus construct alternatives which require little to no awareness (cf. Mason 2015, p. 3039; Nelkin and Rickless 2017, p. 107; Rudy-Hiller 2017, p. 403). For example, attributionists maintain that agents can be responsible for beliefs and attitudes (cf. Smith 2005). Yet, the difficulty for Talbert is to show how agents have circumstantial awareness of their character, more specifically for their Local Traits or Mixed Traits. And capacitarians, motivated by

forgetting cases, argue that an adequate theory of responsibility must show why agents are responsible for unwitting omissions over which they had no awareness. Thus, at first sight it seems that this latter view can accommodate any intricacy posed by fine-grained character. Nonetheless, even though Sher advocates for a view that poses minimal restrictions on what agents need to consciously be aware of to be responsible, he stresses the complexity of character, which puts restraints on the agent. Therefore, agents seldom foresee how their actions turn out to affect their characters in the long run (cf. Sher 2009, p. 38). Ultimately, this may limit significantly what agents can be reasonably expected to know about their character.

Even though his account aims to include those agents who have no awareness, I contend that, a closer look at Sher's interpretation of the reasonable standard and the cognitive contact reveals that the condition requires too much in the case of character. To see this, recall that I have argued that his account entails that agents cannot be reasonably expected to know about habituating their Mixed Traits. Even though the reasonable standard only requires that the agent *could* have been aware, it still requires that agents could be aware what actions will result in their character which is depicted as fine-grained. Thus, they must be able to have a "clear understanding" (ibid.) of which traits they acquire. And, similarly, even though the cognitive contact-clause is also constructed to account for those agents who are unaware, yet those who act unwittingly rightly, when translating it to character, it requires that agents can have enough cognitive contact to be able to perform the actions that habituate their traits on that basis. Hence, since character is not only structurally complex but also dimensionally, having intricate awareness thereof cannot be reasonably expected from the agent nor do they make enough cognitive contact which enables them to perform the actions that habituate their traits on this basis. This prevents his account from giving a positive answer to the question of character responsibility.

Generally, I agree with the pessimism expressed in Sher's argument from complexity: not only do agents often not know which actions will result in which trait, literally *no one* knows nor could be reasonably be expected to know or even makes

enough contact to be able to perform on this basis, which actions have which *exact* effect on character (and this also goes for any character conception; even considering the coarse-grained nature of Global Traits, there certainly are limits what agents can be aware of). Some traits might depend on the quality of an action, while others crystalize through the sheer quantity of repetition. However, this reveals a more general truth: the issue applies to all kinds of facts in the world. Agents are generally very limited regarding the things they are aware of, also with respect to their actions. And that is simply because no agent has all the information available to her at any given time:

"Given the many limitations on what we can know, it is impossible for any agent to be aware of every morally and prudentially relevant fact about every act that he might perform. Thus, if being fully responsible requires being aware of all such facts, then no agent is ever fully responsible for what he does. Still, because agents vary widely in the sorts of things of which they are aware, there remains ample room for the view that how much responsibility any given agent has for what he has done is a direct function of the range of relevant facts of which he was aware" (ibid., p. 5).

In this short paragraph, Sher is eager to stress that no one can ever be aware of all facts. Yet, this does not call into question their general responsibility. And not only capacitarians such as Sher stress his fact; even volitionists have to admit that they consider agents responsible for actions even though their awareness is *always* confined to some extent. This, however, calls into question why the standard would be higher for character? Of course, one might argue that the comparison does not apply. This objection employs a claim I have already made multiple times throughout this thesis, namely that character is more complex than action and therefore actions and character may well have distinctly shaped conditions. The argument might point out that actions and their consequences are much more foreseeable and thus more manageable than character. Taking this argument by heart, a higher standard for character might be justified by the relative difficulty of having adequate awareness to ground responsibility, the reasoning might go. However, the argument misses its mark: that is

because it neglects the connection between character and actions. Even if one were to stress that actions simpliciter are simply less complex to substantiate the claim that the reasonable standard or the cognitive contact (or any epistemic requirement for that matter) ought to be higher for character than actions, it begs the question why agents can be reasonably expected to have this awareness for actions or can have this cognitive contact, which may have some arbitrary consequence but not the actions which habituate a specific character trait. Rather, it seems that since actions have consequences, of which one may be the contribution to the development of a specific trait, it is unclear why the epistemic standard for the latter would be higher than for the former. Hence, why postulate that agents need specific knowledge of the consequences of their actions with regards to specific character traits which may be fine-grained interrelated dispositions that manifest in a multitude of behavioral output? This requirement can obviously not be sustained. On the contrary, since it is an undeniable truth that agents in many instances are unable to foresee their specific characteristics, yet these traits are inherently linked to the agent, I do not think this compels us to consequently conclude that agents are not responsible for these narrower or more fine-grained traits. Instead, I think it provides us with reason to think that since no one can ever be aware of all facts, it does not matter whether this pertains to action with any consequence, the actions which habituate character, or character simpliciter. If anything, this insight supports the claim that the bar for epistemic condition must be lowered. Hence, we need an epistemic condition that acknowledges the general limitations of agents' awareness and that reflects that agents can be responsible for their character traits despite the complexity of character. However, I do not think it requires the construction of a completely new epistemic condition. Rather, if we want to find a positive account of character responsibility, we need to focus on those accounts that already require little to no knowledge of the agents to be responsible and fiddle with the epistemic condition suggested.

# V Towards a Character-Centered Framework

## *2.1.2 General Awareness of Character and its Formation*

So far, I have endorsed the view that an agent can be responsible for her traits when it is attributable to her quality of will, given the defeasible condition that she has acquired the traits in question voluntarily. Hence, the task now is to determine what kind of awareness agents need for their traits (or the development of them) to reflect their quality of will. Since, as mentioned before, Talbert's account is too specific for our purposes, as taking the requirement of circumstantial awareness at face value would potentially rule out alternative character models that depict character as Local or Mixed Traits, there is a natural drift to consider alternative attributionist views that require even less awareness. For example, Angela Smith argues that agents can be responsible even when the agent forgets. In essence, on her view, this shows a lack of care (cf. Smith 2005). This point is definitely worth considering as it guides us into the direction that seems most promising for constructing an epistemic condition that reflects agents' general limitations of awareness by lowering its requirements for agents to be responsible. In adopting such a view, we could maintain that what makes agents responsible for their character traits is that they (or the actions that develop them) are attributable to them because they reflect their quality of will even when agents are unaware of them. Now, on this view, we could take Judith's traits of compassion, which she has no idea of, to reflect her quality of will because she judges these parts of her character not to be important. And her compassionate actions would reflect her quality of will even though she is unaware that they develop her compassionate traits because she simply does not care whether she is compassionate or not. Hence, with such an epistemic condition, we are in a position to maintain that agents can be responsible even when they are unaware. And since this version requires no awareness of agents, it would demonstrate how they can be responsible for their narrower and more fine-grained dispositions as the composition of agents' traits were irrelevant to them caring about it or not.

Yet, even though this interpretation would help us explain a significant class of cases, there is some need to clarify it to some extent. That is because it is unclear whether agents' traits or the actions that develop them are really attributable to their quality of will when agents are *completely* unaware. For see that when Hannah fails to be aware in *any way* that she has a tendency to be dishonest in some situations, it is unclear how it would relate to her quality of will. While the forgetting cases outlined by Smith (2005)[104] implies that at some point in time the agent did in fact know of her friend's birthday (else she could not have forgotten (cf. 2005, p. 236)), it is important to stress that there is a minimal sense in which awareness is required. Those endorsing a view similar to Smith might interject that the very fact that she is unaware of her traits would be an instance of her not caring about her character and therefore in fact be attributable to her quality of will. I agree that this seems to be the ready explanation; yet, it seems that for an agent to judge whether something is important, she must have *some* idea of what it is and that that the very thing exists in the first place. If Hannah has no clue whatsoever that she has a habit of lying, how can she judge it to be important for her character? Nonetheless, my wish is not to discard Smith's view entirely. Instead, what I mean to say is that while agents are sometimes unaware of specific parts of their character, there is a difference between these agents not being aware, forgetting or even ignoring these tendencies and them not having any idea that these traits exist in the first place. Thus, in order for the traits to be reflected in the agent's quality of will, it seems that agents do need *some type of awareness*.

But recalling our insight that agents' awareness is generally limited, what then might this adjusted epistemic condition look like that fits with the view that the an agent's traits and the actions that develop them are attributable to their quality of will when they have some - minimal - type of awareness yet not the specific kind encountered on other views? Most basically, the adjusted, lowered, epistemic condition I have in mind entails that agents only need to have general, coarse-grained,

---

[104] Smith does not only focus on forgetting cases but also lists instances of spontaneous attitudes and involuntary responses as relevant to her account (cf. 2005,pp. 240-242).

awareness of their traits and how they are formed. This need not be specific. Instead, it represents a more general understanding of the agent's traits and the formation thereof. Acknowledging the two types of responsibility, let us briefly break down what this would imply for each variant of the question. Beginning with direct character responsibility, agents need to satisfy the epistemic condition regarding their character itself, i.e., what traits and/or dispositional features they have. Therefore, an agent's traits can be said to reflect her quality of will when she has some coarse-grained understanding of them. For example, this might entail knowing that one generally lies more often than not. Nancy, who we have previously learned to be a rather unpleasant and dishonest person, need not be aware that she has a particular trait structure of, say, dispositions to be honest when she is in a good mood but telling lies whenever she feels like it to reflect her quality of will. We can conceive of her as responsible because she is aware that she tends to be untruthful, irrespective of whether this pertains to very specific circumstances. We can take this to reflect her quality of will because she has no intention of changing anything about it, instead she judges these dispositions to be part of her. In colloquial terms: she is ok with being who she is. And similarly with indirect character responsibility, agents need only general awareness that their actions develop their traits. However, we need to be a little more specific to clearly define the awareness agents need with regards to this other form of character responsibility. That is because I have stressed that the dimensional complexity of character traits requires us to inquire not only what kind of epistemic requirements agents must meet with regards to their actions but also whether they must be aware of the resulting traits. Taking this dimensional complexity to apply, in order to do justice to it, agents would have to have some general knowledge of the formation of their traits *and* their character. Yet, what character responsibility seems to require is not specific knowledge of the traits an agent develops via some particular action but general knowledge of traits being acquired through habituation. For example, that lying helps reinforce dispositions to be dishonest. Hence, the way we can comprehend

Nancy to be indirectly responsible for these traits because she is aware that her lying habitually may result in her reinforcing dishonest dispositions.

Thus, in order for an agent to be responsible for her traits and the development thereof, I argue that she must satisfy conditions of awareness. However, again, these requirements are minimal in the sense that they are not to be confused with any specific knowledge of an agent's trait structure or how it is developed in detail. Rather, what this view proposes is that incorporating these elements of minimal general knowledge illustrates in which way the traits and actions that develop the agent's traits reflect her quality of will and thus are attributable to her.

## 3. The Two Forms of Character Responsibility

Now that we have clarified two of the three challenges, we can progress to the remaining one, which demands of us to clearly differentiate between the two types of responsibility for character traits. This challenge requires of us to show in what way agents can be directly and/or indirectly responsible for their traits and shed light on the relation of both concepts. It derives from the observation that most analyses do not address the two concepts distinctly of one another or only focus on indirect character responsibility without further clarification. However, differentiating between direct and indirect character responsibility is important as they pertain to two separate questions.

### 3.1 The Case for Direct Character Moral Responsibility

The previous chapter explored what representative theories of responsibility can tell us about responsibility for character traits. Despite the large differences between the individual theories, none of them provided a satisfying account of how to conceive of agents as directly responsible for their traits, even though each for a different reason.

While Fischer and Ravizza seemed to struggle with the concept on a structural basis, Sher denies that agents are directly responsible for their character because the minimal basic conditions of responsibility he proposes are not met. And even though, Talbert's account would provide a plausible interpretation of the connection between character and quality of will, I have been tentative about it because he only does so for Global Traits and his theoretical setup does not avoid the problem of manipulation. The general difficulty all of the theories experience with the concept compelled us to investigate how we can conceive of agents as directly responsible for their traits by setting up a framework more suitable for the task. And in fact, I believe that the conditions proposed in this thesis yield more positive results regarding the remaining challenge. For consider that the "control condition" spelled out proposes that a default-linked structuralist view explains agents' character responsibility, i.e., that in the regular case, in which agents are not manipulated, and they have acquired their traits by means of voluntariness, they can be responsible when their traits reflect their quality of will. It does so, when agent have a general kind of knowledge of their traits. To reiterate, this does not mean that agents must be aware of every disposition that they have or even how their characters look as a whole. Instead, it suffices that agents have a sense of what they are like. Thus, the question now becomes whether agents, in light of seemingly lacking specific knowledge, *can* have the minimal, coarse-grained, knowledge needed to reflect their quality of will.

Turning to psychological research on character for empirical evidence paints a rather pessimistic picture. Some psychologists interested in how to determine an agent's character (cf. for example: Newman and Ulman 1989; Tellegen 1991; Kalimeri et al. 2013; Helzer et al. 2014) stress that research on trait inference is hard pressed to find a balance between third-person attributions and self-reports. While the former raise questions of reliability (cf. Mischel 1968, pp. 41–72), the latter include biases (cf. Funder 1991, p. 35). However, disregarding the possibility of methodological errors in these studies, this potential bias does not preclude agents from having a general understanding of their own character, even if it is varnished. Agents usually have

ample evidence of their character as they witness its manifestation constantly. In fact, the sheer quantity of situations, in which they observe themselves acting in combination with them being the only ones having access to all of the desires and beliefs etc. makes them prime candidates for having at least coarse-grained awareness of their traits. This general idea of *how one is* certainly is not an exact measure but for sure allows agents to know their basic commitments, ideals, desires and so forth. Hence, agents generally have enough confirmation of tendencies in specific matters. Agents such as Hannah, who has the Global Traits of dishonesty, has this kind of awareness because she can notice these broad traits and recognize their cross-situational manifestation when she lies to her friend, cheats on her math test, or withholds important information from her mother. And similarly for Local Traits, there is good reason to believe that agents have this awareness. Although Local Traits, in contrast to Global Traits, do not manifest in cross-situational behavior, they are still stable and thus applicable to a specific set of situations. And despite the influence of situational factors, agents can make enough cognitive contact to be able to realize that in certain situations they tend to engage in some behavior more often than not. Judith, who has the Local Traits of compassionate-when-speaking-to-a-loved-one, need not be aware of every situational cue to be able to notice that when engaging with her mother she tends to be more empathetic than when speaking to her colleague. And even though I have described Mixed Traits as a multitude of fine-grained interrelated dispositions whose manifestation can be enhanced or inhibited by triggers, since only coarse-grained awareness is required, which, again, need not be specific but rather represents a general understanding, it seems plausible that agents could have that kind of awareness irrespective of the conception. That is, because the Mixed Traits model also acknowledges that the interrelated dispositions are still attributable to moral domains (cf. Miller 2017a, p. 255).[105] Since agents are so familiar with themselves, it

---

[105] Also recall that Miller proposes various strategies to develop good character (cf. 2018, pp. 169–216). While the sole fact that Miller makes these suggestions is not exact proof that these work, they would at least require some awareness of the agent with respect to her character and the habituation thereof. Hence, it seems to imply that Miller at least assumes that agents can have sufficient awareness.

would be truly odd to say that they do not have this type of minimal awareness. Consequently, agents can be directly responsible for their traits because they typically have the general knowledge required for their traits to reflect their quality of will (with exception to manipulated agents). Agents such as Hannah and Judith are responsible because being aware of their general constitution - irrespective of the trait structure in fact had - can be thought to signify that their characteristics reflect their quality of will.

Notwithstanding this awareness had, the skeptic might object based on two arguments that seem to tell against direct character responsibility. The first argument concerns the way in which my proposal can in fact give a sensible interpretation of direct character responsibility, while the latter one is aimed at the concept of direct character responsibility more generally. To begin with the first one, in considering my explanation in light of the discussion of other approaches, confusion might arise as I have rejected the applicability of Fischer and Ravizza's volitionist account to direct character responsibility in the previous chapter on the grounds of them providing a historicist view, which, I maintained, contrasts the question which asks of agent's responsibility for having the traits they have. How does my view then, which also takes the agent's history to play a fundamental role in her assessment, now evade the problem? This objection addresses a substantial point, namely that one of the challenges of this thesis is to provide a framework of character responsibility that in fact fathoms the balancing act between showing how agents are directly responsible for their traits without also admitting to the view that manipulated agents are responsible. So, if the account by Fischer and Ravizza fails to bridge the gap because of its historical notion, there seems legitimate worry that the modified control condition I have suggested cannot succeed either. However, even though the proposal as outlined here is history-focused and thus conceptually resembles the approach given by Fischer and Ravizza at first glance, it differs significantly from the historical view provided by these authors. To see why, consider the argument why Fischer and Ravizza's account seems ill-suited to fit direct character responsibility. The reason is not that it focuses on the agent's history simpliciter but rather that the approach

primarily promotes a view that requires a *positive* history of the agent (cf. Zimmerman 2003, p. 647). Recall their claim that "(...) someone's being morally responsible requires that the past *be a certain way* [italics S. Sch.]" (Fischer and Ravizza 2000, p. 207). On the contrary, the proposal, which I claim to be more successful, suggests that a *negative history* suffices to account for the problem of manipulation. As opposed to positive historicism, this does not rely on the notion that the "past be a certain way" but rather that the past *not* be a certain way.[106] This is to say that the view as presented Fischer and Ravizza puts too much focus on the actual positive process of character formation, which must be controlled by the agent, and thus steers away from what direct character responsibility is concerned with. Instead, by constructing a defeasible condition, which only asks that the agent lacks a history of manipulation but grants that otherwise structuralist conditions suffice, we are in a position to take the agent's history into consideration, yet leave sufficient room to not preclude direct character responsibility.

The second argument against direct character responsibility, in contrast, is not geared towards the approach as presented in this thesis but against the concept more generally. Recall that Sher opposes the view that agents can be directly responsible for their traits because they lack the right kind of causal involvement. That is because one of Sher's most fundamental claims pertains to the action and its connection to the agent, which he takes to be causal (cf. 2009, p. 143). Sher constructs this causal relation in terms of the origination requirement. However, he maintains that it is obvious that

---

[106] Zimmerman differentiates between positive and negative historicism as follows: "Negative historicism stresses the extent to which the acquisition or retention of responsibility-grounding psychological properties does or does not bypass the person's capacities for rational control *once the latter are in place*. Positive historicism, by contrast, emphasizes the manner in which the development of these very properties *meshes with the development of the child's capacities for practical rationality as they unfold*. It requires that the person herself play *an appropriately active role* in the development of her autonomous adult personality" (Zimmerman 2003, p. 648). The way I employ negative historicism here, differs slightly from the description by Zimmerman in as much as it takes the negative history to pertain to exactly the acquisition of the *responsibility-grounding psychological properties*. Yet, my slightly different interpretation coincides with negative historicism as conceived by Zimmerman inasmuch as it takes, "neurological fiddling and other invasive procedures" (Zimmerman 2003, p. 640) to excuse agents.

*having* character is not causal at all (cf. Sher 2006a, p. 68). And since on his view causality is further thought to involve being able to give reasons "to exhibit the right sort of causal involvement (…)" (ibid., p. 68), he himself holds that agents are not responsible for *having* (bad) traits. (And this is irrespective of any further epistemic specifications added). But even if it were not for this particular interpretation, Sher argues that any form of responsibility requires that the agent have a causal influence: „Under its standard interpretation, responsibility is a causal notion. To be responsible for an outcome, a person or thing must play some role in causing or failing to prevent that outcome. This is true whether the relevant notion of responsibility is merely causal, as it is when a design flaw is said to be responsible for the collapse of a bridge, or whether it is moral, as when an engineer's negligence is said to be responsible for the design flaw" (ibid., 67).

The argument by Sher coincides with the intuitive view held by many that causality is a basic tenet of responsibility because if the agent stands in no causal relation to the action/character in question, it simply is not true that *she* is responsible for it. Hence, the question becomes whether we need to go back and revise our conditionals set up or even reject the concept of direct responsibility for character altogether? I do not think so. That is because, first, it is not clear that agents do not really satisfy a causal condition and, second, that causal influence as proposed by Sher is even necessary. Granted, considering the question whether agents are responsible for having the traits that they have, they may not have *causally authored* them in this sense that Sher may refer to. However, this is not to say that there is no form of causality involved. For see how the argument neglects that agents can causally contribute to their traits' further development. For example, considering Francine and her tendencies to be unjust. When deliberating whether she can be responsible for these parts of her simply for having them, we can obviously not appeal to her bringing them about. However, what we might do is note that she may not necessarily endorse these dispositions and try to become more nuanced in her decision-making. In this sense, even though this might not amount to the kind of causal relationship that Sher envisions, character traits are attributable to her because not only is she intimately

linked to them (they are hers not somebody else's (cf. Jacobs 2001, p. 10), having character traits does reflect the agent's quality of will.[107][108]

## 3.2 The Case for Indirect Character Moral Responsibility

Having considered the way in which agents can be thought to be directly responsible for their character traits, we are now in a position to continue our investigation to inquire in which way agents can also be indirectly responsible for them. We can do so by looking into whether they meet the conditions proposed for the character framework. In order for agents to be indirectly responsible for their traits, the actions which develop them must reflect the agent's quality of will. They do so, I maintained, when the acquisition of their character traits is voluntary and agents have general knowledge that their traits form their character. So naturally, at this point again, the question becomes whether agents have the kind of awareness. And, quite simply, I do not see why they would not.[109] However, as it may not be apparent to everyone that agents in fact are aware of this relation, there seems reason to underpin the view with some arguments. For one, taking an analogy from other things that are formed through repetition, Jacobs (2001) for example notes that agents normally do not know of either one drink that turns them into an alcoholic (and there is not one that actually *does*). But agents can know that the regular consumption of alcoholic beverages contributes to the risk of making it a habit, if not addiction. Applied to the case of character, we can see that agents neither know that one particular lie causes them to become a liar (and there neither is one that actually *does*) (cf. p. 22). And similarly, displaying dishonest behavioral patterns may contribute to the development of dispositions to be dishonest

---

[107] Russell (2009) also suggests that even though agents may not bring about their traits in hindsight of ultimate moral responsibility, they can influence their future selves (pp. 377-378).

[108] I would like to thank Erasmus Mayr, Christof Rapp, and Laura Valentini for asking me elaborate on direct character responsibility more specifically bearing in mind one of the main motivations for this thesis in particular.

[109] A similar claim can already be found in Aristotle (cf. 2002, NE1114a9-11), who maintains that agents generally know that their dispositions are a result of what they do.

(whether it be a Global Trait of dishonesty or a Local Trait of lying to one's parents regarding test results or a Mixed Trait of fine-grained interrelated dispositions to truth-telling). This certainly does not constitute precise awareness of the development of the actual structure of one's character. But again, this is not what is required. What is required is that agents have a general idea that their actions will develop their traits. Consider the following example: it is reasonable to expect that Iris could be aware that telling a white lie about her incapacity to take out the trash is morally comprehensible and may result in an argument with her live-in partner. Simultaneously, this white lie may be one of many and reinforce her disposition to be dishonest. Perpetual lying makes one prone to forming certain beliefs about lying or even a liar. But why would it be less reasonable to expect that Iris could be aware that telling white lies may have certain consequences (of which there could be countless such as upsetting her partner to the point of them thinking that this is the final straw), yet not be aware that doing so regularly will corrupt her character? Surely, considering the structure of Mixed Traits or even Local Traits and the way in which their manifestation depends on situational factors and triggers, it seems unattainable that agents have this awareness down to the last disposition. Hence, an agent may not know of either one single action that will develop a disposition and neither how that would exactly contribute to her trait structure, „(…) but is capable of the sort of awareness that would enable her to judge whether it is desirable that that pattern should become second nature" (Jacobs 2001, p. 22). Certainly, agents are limited in what they know about their traits and may not always develop them consciously - after all, the epistemic condition is constructed to account for the fact that no one can know or be expected to know that certain actions will have *specific effects* on character - but they are not oblivious to them. And since agents usually have this awareness due to the many situations in which they manifest their traits, I contend that the actions which develop their traits reflect their quality of will as they demonstrate what the agent cares to be like. In conclusion, I assert that agents cannot only be directly responsible for their traits, but also indirectly

responsible for them because the traits that they develop and they have are attributable to them.

Yet, before ending this section, similarly to the discussion of direct character responsibility, there are again counter-arguments, I would like to pay some attention to: the first one questions the way in which the proposed framework can include all classes of relevant agents. More specifically, how does this view handle an aspect of character development, I have repeatedly mentioned throughout the last two chapters, which is that character traits are often developed unconsciously? Does the view I propose not exclude those, who are unaware? And would this not in fact exclude a significant class of agents, simply due to the fact that agents are often unaware? To answer the last question first, I do agree that such a view would take a serious amount of agents to be not responsible. However, to come back to the first question, I do not think that my account has these implications. That is because what this view requires, as has hopefully been made abundantly clear, is that agents have a general understanding of who they are and that their actions form their traits. This does not, however, entail that agents need conscious awareness whenever they act or with regards to any particular state that they are in. In fact, the way the epistemic condition has been set up is particularly open to including those that develop their traits unconsciously. How so? General knowledge of how one is or that one's actions develop one's traits is completely compatible with the fact that traits are often developed unconsciously. That is because one can be aware that telling the truth may contribute to forming truthful dispositions, yet not consciously think about this relation any time one acts. Such a view would seriously overburden agents and result barely anyone being responsible. Thus, agents usually have this type of awareness and can be thought to satisfy it even when they are not aware of every single disposition they acquire.

The second point that I would like to consider is that certain psychoanalytical factors may make it easier - or harder respectively - for some agents to develop specific

character traits (cf. Jacobs 2001, p. 19).[110] For example, someone with high anxiety levels will have difficulty meeting others and may thus be prevented from developing more compassion. Alternatively, someone shy may be overwhelmed by having to speak up for others in order to be honest. This, however, raises the question of whether these agents should not be considered responsible for having and developing the traits that they have (cf. Jacobs 2001, p. 25). The worry might stem from the view that if an agent cannot have or develop a trait, then she neither may not be held responsible for it. More precisely, as these may not even reflect the agent's quality of will, this seems to be about fairness and may be much harder to put to rest. Nonetheless, I would like to offer the following thoughts for consideration. To begin with, note that all agents not only differ with regards to the environment they are born into but also with respect to their initial constitution, their "(…) temperament is mainly endowed by nature (…), and consists of built-in factors or tendencies in us which are out of our control" (Chen 2013, p. 352). However, this very fact, namely that all of us differ in our initial constitution, which may bear great impact on what life chances one has and what kind of traits one can develop, is an indication that these agents can be responsible for their traits despite their limitations. For first, if we were to maintain that agents are not responsible due to certain psychoanalytical factors, how do we differentiate between those that excuse and those that do not? After all, we are all equipped with personality traits or dispositions that interact with the dispositions that we habituate as our character traits. Yet, even if one were to reply that agents may not be responsible for these single traits but can be for other ones (cf. Jacobs 2001, p. 19), there is some doubt to how this would apply. Not only is there the issue of having to tell apart those

---

[110] In following my argument in support character responsibility, one might be inclined to recall the general classification of character traits introduces in chapter 1. Here, I have made the case for the view that there are moral and non-moral character traits. The discussion so far has not considered this differentiation in more detail. So naturally, one might ask, are agents also responsible for their non-moral traits? I believe there is a brief answer to this. Character responsibility considers the way in which agents are morally responsible for their character traits. Hence, agents are not morally responsible for their non-moral traits. This is not to say that they have a kind of other normative responsibility I have not discussed. But agents are not morally assessed for their non-moral traits.

preconditional psychoanalytical traits that present a constraint to responsibility, there is also the question of how the interdependence of traits would allow to dissect those traits specifically that issue directly and without further influence from one specific personality trait. Moreover, and more importantly, even though these traits are beyond the agent's control and may determine the extent to which she can develop certain character traits, the character traits she forms are still attributable to her quality of will. That is because even if the traits they develop stem from actions which are expressions of these psychoanalytical factors, they can be responsible because it is a reflection of what agents take to be reasons and thus reflects their quality of will[111]:

"While it would be most implausible to hold people responsible for their temperaments, it is plausible to ascribe to them responsibility for the actions and characteristics through which temperament is expressed in choices, judgments, and reactions. We cannot choose our temperaments, though there is voluntariness at work in what temperament counts for in acts, intentions, and responses" (Jacobs 2001, pp. 25–26).

Notwithstanding these varying predispositions had by everyone, there are, however, some cases such as extreme phobias, in which agents are not responsible. The reason being that the expression of these does not reflect the agent's quality of will because she cannot exercise her reasons-responsive capacities. Hence, these instances may excuse the agent or seriously diminish her responsibility. However, these cases are very different from the general case of someone with a tendency to be more outgoing or less talkative as these agents do seem to lack any kind capacity to respond to reasons.

Generally, some may find the explanations unconvincing and resort to a skepticist view about responsibility. However, let me stress the importance that responsibility plays in our lives. Wolf for example notes that even if there were no metaphysical facts about agents being responsible, it seems that abandoning these practices outlined in the introduction seems strange (cf. 1993b). Others have also

---

[111] Chen argues that agents are not responsible for their "temperamental dispositions" but for being in control over their expression (cf. 2013, p. 353).

highlighted the way in which our responsibility practices are linked to conceiving of "one's life as one's own" (Jacobs 2001, p. 11) and our interpersonal relationships (cf. Fischer and Ravizza 2000, p. 16). And even though some, e.g., Sher, maintain that responsibility seems to only conceptually apply for actions (cf. Sher 2006a, p. 67), this position seems to neglect the strong correlation between the two. Instead, I believe that by construing a framework that takes agents to be directly and indirectly responsible account for the observation of how important character and the responsibility we have for it plays in our lives. On the contrary, skepticists must demonstrate why we place such importance on our traits if we are not responsible for them. Therefore, I believe the character-centered framework to provide an addition to an essential part that responsibility is by showing that agents are responsible for their traits.

## 4. Cultural Membership and Deprived Childhood

The conditions introduced allow for most "normal" agents to be responsible for their traits as it requires no control and minimal awareness of them. Nonetheless, my general line of argument for indirect character responsibility so far includes one exception, namely manipulated agents. I have stressed that manipulated agents are not responsible. There are several reasons to think so: not only do they lack control over their manipulated traits (which, on this view, is not even a necessary condition), they neither have any sort of authorship over their character and thus these character traits are not attributable to them as they are simply not *their* character traits or their values. Mainly, these agents have a wrong kind of history as their characters are completely alien to them. However, clear-cut cases of manipulation such as those in which an evil scientist implants a new trait overnight or an agent is constantly held at gunpoint may be less common than a variation of the manipulation problem: what about those brainwashed, with a certain cultural membership or deprived childhood (cf. Vogel 1993, p. 129)? Among the cases most cited are slave-owners in the 1850s,

German Nazis in the 1930s, "JoJo", the son of an evil dictator, or male chauvinists of the 1950s. Cases of this sort have been discussed by numerous authors, most famously Wolf, who suggests that these agents are insane and thus exonerated (cf. 1988, pp. 56–57). These are examples of agents whose moral competence, more specifically what they perceive as wrong, is impeded by their cultural membership. Thus, some conclude that „(...) cultural membership can be ethically disabling: (…) our moral education (...) can prevent us from grasping important moral facts" (Levy 2003, p. 160). Because these agents lack access to moral facts by virtue of their cultural membership, some advance "the inability thesis about cultural impediments" (Moody-Adams 1994, p. 293). Proponents of the inability thesis stress the difficulty had by these agents to acquire the moral knowledge necessary to be considered responsible. They argue that these agents could not have known better because most oppressive systems are characterized by "covering up evidence"(Benson 2001, p. 611). Also, agents in these societies "(…) have been brought up to internalize the central values and prevalent rationalizations of that institution" (ibid., p. 614). Most importantly, though, they argue that moral facts are not self-presenting (cf. Levy 2003, pp. 157–158). Hence, it is generally called into question whether it really can be expected of these agents to know of the moral facts.

Those authors who endorse the inability thesis argue that common practice supports the view that the agent's history is essential to our assessment of her. Further, they claim that we generally do not consider agents responsible (for their character) when we learn about their adverse history. If agents grow up brainwashed or are the product of manipulation, we usually refrain from thinking of them as responsible for their traits. And while I agree with regards to the question about clear-cut manipulation, I disagree with regards to *all* cases of cultural membership. While I do *not* mean to say that these agents are definitely responsible, I do mean to say that these cases require individual assessment, more specifically whether these agents own their values and thus their traits(i.e., whether they have obtained them voluntarily) and the actions that develop them reflect their quality of will in the relevant sense. The reason

to be more tentative about these variations of the manipulation case are manifold: for one, there is little empirical proof that supports the conclusion that agents with such a history are exonerated. On the contrary, it requires the construction of extremely contrived cases that include total isolation of the agent and near-to-manipulation propaganda in order to illicit the intuition that these agents are not responsible (cf. Faraci and Shoemaker 2010, p. 324).

In addition, and more importantly, consider how these cases actually differ from the clear-cut manipulation case. For first, while clear-cut manipulation cases usually involve a single instance, in which an evil scientist implants new values in the agent (the unrealistic version) or an agent is brainwashed (the more realistic version), cultural membership and deprived childhoods are characterized as environments in which the agent acquires her traits procedurally. But what, aside from the temporal focus, which allows agents to develop their traits over time, is the exact difference here? In cases of clear-cut manipulation, agents acquire their traits involuntarily and do not own them. Without diving into the topic of voluntariness into more detail, we can conceive of involuntary acquisition in this context to circumvent or completely bypass the agent's capacities (cf. Zimmerman 2003, p. 648; Mele 2020, p. 3148) to critically reflect on reasons. While this seems to apply to clear-cut manipulation cases, it does not automatically so in those where agents grow up with a deprived childhood or a certain cultural membership. These agents' actions can still reflect their quality of will because they may still be able respond to and recognize reasons. That is because there is a difference between complete deprivation and those kind of early childhood environments that may make life harder. The difference lies between the complete incapacitation from experience and obstacles that do not deprive or destruct those capacities (cf. Jacobs 2001, p. 18). Their values can thus thought to be really theirs because they have authored them with their reasons-capacities. Hence, the difference between manipulation that exonerates and cultural membership or a deprived childhood is whether she retains her capacities to recognize and respond to reasons and therefore in fact reflect *her* quality of will. For example, Wolf's case of JoJo, the son

of an evil dictator, is said to fully identify with the values of his father. Taking him as a role model, JoJo copies many of his father's gruesome orders, which include despicable methods such as torture and enforcing the death penalty. He is not coerced but instead expresses these values as his deepest desires (cf. Wolf 1988, pp. 53–54). And even an extreme example of a deprived childhood full of physical and mental abuse and neglect like that of Robert Harris, a brutal murderer (cf. Watson 1993), differ from the clear-cut manipulation case. Despite all the hardship endured by Harris, manipulated agents are completely alien to their character, while „(…) everyone was raised in and influenced by *some* environment" (Vogel 1993, p. 136). This latter point not only reiterates a truth mentioned before, namely that no agent has any control over their environment (cf. Talbert 2009, p. 15), some are eager to stress that while both types of environments employ conditioning (cf. Zimmerman 2003, pp. 656–657), there is a difference:

"Genuine education-for-autonomy is a process that awaits and deploys those developmental moments when children can begin to make the sorts of cognitive judgments that in turn enable them to *make up their own minds* about which values to embrace or reject. Mere indoctrination-for-heteronomy, by contrast, disregards, suppresses, even runs rough-shod over these cognitive-developmental transitions, by continuing to employ conditioning techniques long after the child is capable of making cognitive judgments which might otherwise affect the content of the values he or she eventually might embrace" (Zimmerman 2003, p. 657).

However, the conclusion, I believe that we should draw, is not that any environment that may resemble a certain cultural membership or deprived childhood exonerates but rather that it depends on whether the agent has acquired her reasons capacities voluntarily and can thus employ them to form *her* quality of will.

This is not to say that a childhood full of neglect and mental and/or physical abuse have no bearing on the agent, however, to put it in Smith's words: "we should not make the patronizing mistake of assuming that a person who has had a bad upbringing is not really responsible for her attitudes and for the judgment they reflect" (Smith 2005, p. 270). Even though I disagree with Smith's conclusion that these agents

are fully responsible but their circumstances may render them to less serious moral criticism (cf. Smith 2008, p. 390), I agree that these agents can be responsible because it seems untrue that these agents have no access to any moral considerations (cf. Vogel 1993, p. 133). Even though some cultural memberships may almost fully block access to moral considerations and thus what the agent can be reasonably expected to know, others may allow for more exchange of ideas: "(…) they [the slave-owner, German Nazi and male chauvinist; S. Sch.] are [not] hermetically sealed off from the intuitions that form the core of our morality" (ibid., p. 141). These cases illustrate that these agents may be more limited than we are today, yet they can be reasonably expected to have had access to these moral facts. And even if most of our evidence on moral facts comes from our environment (cf. Sher 2009, 116 footnote 1), this is not to say that agents in most cases never have any possibility of exchanging ideas and considering opinions alternative to the ones they are typically surrounded by. For example, male chauvinists of the 50s may well understand that their victims are human beings and thus deserve the say treatment as any other male counterpart (cf. Vogel 1993, p. 138). While, if we imagine an immensely contrived case like JoJo, there may be initial sympathy we feel when confronted with the epistemic restrictions these members of certain cultural communities or deprived childhoods face. However, this is different to these agents being unable or incapable of having this knowledge (cf. Vogel 1993, p. 142). What I am at pains to make clear here then is this: while the clear-cut case of manipulation, which resembles the implantation of characteristics, excuses agents from their responsibility, this is not automatically so in the case of cultural membership or bad childhood. Whether it does is an extremely individual case, yet mostly these agents' traits still reflect their quality of will.

Nonetheless, some are inclined to exempt or excuse these agents respectively as the same might happen to us: it is likely that future generations will similarly condemn some of the practices that are acceptable today, revealing our personal blind spots (cf. Wolf 1988, pp. 60–61). However, first, this actually gives us reason to believe that these agents are in fact responsible. If we base our assessment on the fact that

neither past nor present cultures have access to all moral facts, we would never be able to consider someone responsible. Ultimately, whether cultural membership and a deprived childhood in fact resemble manipulation and thus exonerate is an extremely individual matter. The prevalent question then becomes whether an agent's environment is of the former or the latter sort and whether they are able to form and apply her capacities to recognize and respond to reasons, which then reflect *her* quality of will and. Hence, the task is to define when such an environment leads to "culturally induced moral blindness" (Moody-Adams 1994, p. 294), which limits whether the agent is still able to reason and what can be expected from an agent. Irrespective of some potential rare cases, which would be extremely hard to imagine, I do not think that cultural membership or deprived childhood *automatically* exonerate the agent. These agents usually differ from clear-cut manipulation cases when their capacities are intact and thus can reflect *their* quality of will.

## 5. The Minimal Threshold and the Degrees-View

The approach I have advocated so far conceives of agents as responsible for their traits when they have acquired their traits voluntarily and thus their traits reflect their quality of will, given they have minimal awareness of them and the formation thereof. Defining a minimal threshold which must be satisfied is standard in responsibility debates; those who meet them are generally considered responsible. According to this minimal threshold, responsibility is something that an agent either has or has not depending on whether she satisfies the relevant conditions (cf. Coates and Swenson 2013, p. 630). And to some extent, this is true. Defining the conditions of responsibility subsequently leaves some agents responsible, while others are not. In fact, this is the whole idea behind setting up these conditions in the first place: To sort out which agents satisfy them, and which do not. The critic may now strongly oppose my view and protest that agents with certain cultural memberships and deprived childhoods

certainly cannot be as responsible for their character traits as "normal" agents. After all, the extraordinary circumstances of these agents reveal that even though they do not necessarily resemble manipulated agents, they neither resemble "normal" agents. How agents come to have the character they have *does* play a fundamental role. And I agree. Indeed, agents of the former sort spark different intuitions. Obviously, there is a difference between those environments outlined and the kinds of education that are the foundation of sound psychological development (cf. Zimmerman 2003, p. 647). However, I suggest that it usually does not excuse but only mitigate the agent's responsibility for her traits. The agent's history does not generally call into question their responsibility for their character traits, yet it can seriously diminish it. This realization may be part of a bigger claim, namely that character responsibility comes in degrees (cf. Coates and Swenson 2013, p. 630). Understanding character responsibility in degrees means digressing from the all-or-nothing-mentality we usually have regarding responsibility. This alternative may be more popular among those who are discontent with the dichotomous assessment of agents. Agents are generally not simply responsible or not responsible. Instead, with the degrees-view we are in a position to maintain that some agents are less responsible, maybe even only just above the minimal threshold, but are responsible for their character nonetheless. Hence, the view that character responsibility comes in degrees acknowledges that some factors may not exonerate but still mitigate the agent's responsibility (cf. Coates and Swenson 2013, p. 643). Thus, while agents with a deprived childhood or certain cultural memberships are usually responsible for their traits to a (significant) lesser degree, they are responsible for their traits nonetheless.

Some might refuse to fully buy into the degrees-view. Instead, they might argue that these agents are simply more or less blameworthy and that character responsibility is something an agent either has or has not (cf. Cyr 2020, 2391 footnote). For example, some suggest that agents with a deprived childhood or certain cultural memberships are responsible but might not be blameworthy (cf. Vogel 1993, p. 140; Smith 2008, p. 390). In fact, the idea that blameworthiness, not responsibility, comes in

degrees feels more natural to some authors.[112] Yet, this position neglects that conceiving of responsibility in degrees helps us also make sense of cases of undeveloped agency: children and adolescents are usually regarded responsible for some of their actions and/or character. However, we refrain from ascribing full responsibility to them (cf. Coates and Swenson 2013, p. 632). These agents meet the minimal threshold conditions, yet "(…) only to a slight degree" (Cyr 2019, p. 210). It is generally agreed upon that young adults may be responsible in some regards but less in others. Their degree of responsibility rises as adolescents mature. Hence, they are relatively less responsible compared to adults (cf. Coates and Swenson 2013, p. 632). Extending the view of a minimal threshold to conceiving of character responsibility in degrees accounts for these specific cases that portray kind of a grey area. Further, the degrees-view also accounts for the barely disputed fact that *all* agents differ with respects to how much they know and control (cf. ibid., p. 209). This latter aspect applies to character with two regards. First, depending on the conception of character, agents may be limited the agent can actually know. While the Global Traits model, due to the broad structure, may make it easier to have awareness of, environmental influences such as situations and triggers may lessen what an agent can actually know and also be expected to know about Local Traits or Mixed Traits. Thus, the potential of this view is to explain that agents are generally responsible for their Global Traits, Local Traits, and Mixed Traits but that depending on the conception they might satisfy the conditions more easily. Second, the view not only exemplifies the difference between the three theoretical approaches, it also makes sense of the way in which agents have different access to the various aspects of their character more generally: "That responsibility for character is to be ascribed to agents is not to say that they can be fully or equally responsible for all of their characteristics" (cf. Jacobs 2001, p. 19). Hence,

---

[112] By bringing this up, I would like to stress that the topic of blameworthiness for traits deserves much more consideration than I am able to give here. For a lengthy and insightful consideration of the desert of blame for character traits see Sher 2006a, most specifically chapter 4, who dedicates an entire book to the issue.

agents can be more responsible for some of their traits or specific trait interrelations than others.

The view I propose suggests that agents are generally responsible for their traits. However, comparing two agents of whom one has more awareness and control exemplifies how both can satisfy the conditions of character responsibility but one is responsible to a higher degree (cf. Coates and Swenson 2013, p. 630; Nelkin 2016, p. 357; Cyr 2019, p. 194).[113] For example, Judith is a self-reflected person who enjoys going on retreats where she meditates about herself in order to become more compassionate. She journals to observe her own behavior and how she may be perceived by others. She might notice that she tends to be more empathetic to her colleagues when she has had a nourishing breakfast. Hannah, on the contrary, does no such thing. She does not engage in these self-improving techniques. While she does care to be more compassionate, she does not take the time to reflect on past actions nor how she wants to behave in the future. Hannah and Judith are both responsible for their character traits, meaning that they both satisfy the minimal threshold conditions laid out. Yet, it is not hard to see that Judith has more awareness of her character than Hannah and thus is more responsible with regards to that trait. Therefore, the degree of character responsibility is a function of the satisfaction of the conditions. Hence, depending on the degree of satisfaction of the relevant conditions, some agents may be more or less responsible than others. In fact, some may only barely satisfy the minimal threshold.

---

[113] Note that even though my proposal considers agents responsible for their traits even if they do not control their character development. However, I challenge the necessary function some make it out to have. Similarly, Cyr, who defends a history-sensitive structuralist view, argues that the conditions for structuralism suffice as a minimal threshold. But the agent's history can affect the degree to which the agent is responsible (cf. 2020, p. 2392).

## 6. Conclusion

This chapter established a view of how to conceive of agents as responsible for their character even when they lack control and have minimal awareness. Directly building on the previous chapter, I took the emerging challenges to assemble a character-tailored framework. To do so, I began by exploring the construction of the control as well as the epistemic condition. First, I argued that a framework dedicated to character responsibility must exemplify in which way the agent's history is important to her assessment. Contrasting the problem of ultimate responsibility with structuralism revealed that neither simpliciter is favorable as each yields complications of its own. Since structuralism standalone is no remedy for the problem of manipulation, I maintained that we need to consider how the agent came to have the character she has. I instead opted for a historical condition that ensures the agent's manipulation-free history by retaining the voluntariness requirement. Consequently, I proposed a defeasible structuralist view. Building on this condition, I proceeded to inquire how to conceive of the epistemic condition. Since neither theory accounts discussed in the previous chapter shows that agents are responsible for their Local Traits or Mixed Traits and agents are generally epistemically limited, I suggested that agents can be responsible for their character traits even though they have only minimal general knowledge of their traits and the formation thereof. On the view proposed, the agent's traits and the formation thereof reflect her quality of will when she has minimal, coarse-grained, knowledge of them and the formation thereof.

After having addressed the first two challenges, I provided arguments for the view that agents are directly responsible for their character, with exception to manipulated agents, because the traits they have reflect their quality of will. Building on this, I went on to introduce a positive account of how to conceive of agents as indirectly responsible for their traits. Agents are indirectly responsible for their character - again with exception of manipulated agents, when their actions which develop their traits are attributable to them. They are when agents know of the general

relation between actions and character traits. For the direct and the indirect case, I held that agents usually can be considered to have the necessary awareness. This also applies to agents with certain cultural membership or a deprived childhood. Even though these agents are more limited in what can be reasonably expected from them, I claimed that they can be responsible for their character nonetheless because they differ from cases of manipulation in which the agent acquires her traits involuntarily and thus they are not attributable to *their* quality of will. Notwithstanding this assessment, I acknowledged that these agents do not exactly resemble "normal" agents. Building on this, I introduced the degrees-view, which holds that agents vary with the degree to which they satisfy the relevant conditions. Cases of cultural membership and deprived childhood then reflect the way that despite fulfilling the basic conditions, they do so to a (significantly) lesser degree. Further, this view explains cases of undeveloped agency and the fact that all agents vary with regards to what they know and control. Hence, I concluded that agents are responsible for their character, yet to a varying degree.

# General Conclusion

This thesis pursued the question of how to conceive of agents as responsible for their character. The main motivation to address this long existing topic once again was the one-sided approach to character responsibility, which neglects developments of character as well responsibility theories in recent years. In order to bridge the research gap, I introduced various theories from both debates, which had not been taken into account previously. The claim I have put forth is that the variable of character introduces several levels of complexity. Thus, I advocated for the thesis that character must be addressed as a distinct phenomenon. To do justice to the claim, I subsequently proposed a character-tailored framework. The framework shows that agents can be responsible for their character traits even when they lack control and only have minimal awareness of them and the formation thereof. Further, the framework clarifies that agents are responsible for their traits, not only for developing them (indirect character responsibility) but also by means of having them (direct character responsibility). Finally, I suggested that character responsibility comes in degrees, acknowledging not only the epistemic restrictions of environmental factors such as a deprived childhood or cultural membership but also the general differences in awareness and control had by agents.

Now, by means of concluding this thesis, instead of summarizing my findings once more in detail, I would rather like to point out what I see as its main contributions. Overall, this thesis joins the accounts that suggest that agents are responsible for their character. But it differs from other analyses of character responsibility for three main reasons: First, most obviously, this account takes agents not only to be indirectly responsible for their traits but also to be directly responsible for them. In doing so, it simultaneously provides a solid precondition for the view generally held that in order to be responsible for their actions, agents must be responsible for their character.

Second, it addresses character as a distinct singular issue by developing a character-centered framework. Most importantly, this framework accrued from the realization that the variable character is relatively more complex than actions. Thereby, this thesis provides a new angle from which to discuss the matter. Those engaging in the debate ought to be mindful of the structural and dimensional complications added to the question. This, I maintain, applies to all character theories encountered, but intensifies with conceptions that depict character as narrow or fine-grained. Third, in acknowledgment of the complexity introduced by the variable, this thesis provides an argument of how we can conceive of agents as responsible for their character without resorting to the view that they need control or exact awareness. By doing so, it pays tribute to the fact that character traits are not always habituated consciously, nor do agents necessarily control their development. The theories of character most prominent in the debate differ primarily with respect to their structural makeup and thus what can be known about them. But not only do the different conceptions of character have significant consequences on the awareness and control had by agents, agents neither control the environment they grow up in, yet it has strong impact on them. Therefore, this thesis makes a suggestion of how to conceive of character responsibility despite the real-life limitations had by all agents. Moreover, by introducing the degrees-view, the framework makes sense of the differences between agents and thus accounts for the often raised worry that some agents have been brought up in the wrong environment. On the view developed, these agents can be responsible for their traits, nonetheless.

Despite the arguments made and points raised, obviously, this thesis cannot once and for all settle the debate. The framework I have established certainly only is a starting point to address the issue from a different perspective. Clearly, it cannot be defended against all objections. For example, some might find the conditions set up implausible. They might insist that agents are not responsible even if they satisfy the conditions proposed and alternatively maintain that they are only responsible when they in fact do have control or specific awareness over the development of their traits.

The idea that character responsibility requires that agents have control and detailed awareness of their characters is persisting. However, such views must either explicate why so many agents are excused consequently or show that they in fact do satisfy these requirements even though our characters might be much more fine-grained than generally assumed. Both, I think, is a hard task to achieve.

Then yet again, others may agree that the conditions I have suggested are in fact correct, yet insist that agents do not satisfy them, nevertheless. While those in favor of such a view might not find this worrying, I would like to remind readers of the numerous social, professional, and legal practices that surround character as outlined in the introduction. As mentioned before, the implications of character on agents' lives are tremendous. Character has effects on all aspects of agents' lives ranging from the way they are perceived by others to the possibilities they have in their career and other institutionalized processes. If agents are not responsible for their characters, it stands to question why it plays such a crucial role in determining their life chances. Hence, such views must substantiate what justifies the practices.

Notwithstanding these objections, the aim of this thesis was not to show that this is the only framework conceivable or that the conditions proposed are, stand-alone or even in combination, the only correct ones. Rather, it was to provide sufficient arguments for the view that character must be considered as a distinct phenomenon and that agents can be responsible for their traits even if they do not satisfy the conditions typically assumed. Therefore, I am certain that future research benefits from adopting the view that approaching the question of character responsibility must include its complexity while granting that agents can be responsible for their traits even when they lack control and specific awareness.

# Publication Bibliography

**Althof, W.; Berkowitz, M. W.** (2006): Moral education and character education: their relationship and roles in citizenship education. In *Journal of Moral Education* 35 (4), pp. 495–518.

**Altshuler, R.** (2016): Character, Will, and Agency. In Masala, A.; Webber, J. (Eds.): From personality to virtue. Essays on the philosophy of character. Oxford: Oxford University Press, pp. 62–80.

**Alvarez, M.** (2017): Are Character Traits Dispositions? In *Royal Institute of Philosophy Supplement* 80, pp. 69–86. DOI: 10.1017/S1358246117000029.

**Annas, J.** (2006): Virtue Ethics. In Copp, D. (Ed.): The Oxford Handbook of Ethical Theory. New York: Oxford University Press, pp. 515–536.

**Annas, J.** (2011): Intelligent virtue. Oxford: Oxford University Press.

**Annas, J.; Narvaez, D.; Snow, N. E.** (Eds.) (2016): Developing the virtues. Integrating perspectives. New York: Oxford University Press.

**Aristotle** (2002): Nicomachean Ethics. transl. Christopher Rowe. New York: Oxford University Press.

**Arpaly, N.** (2015): Huckleberry Finn Revisited. In Clarke, R.; McKenna, M.; Smith, A. M. (Eds.): The Nature of Moral Responsibility. New Essays. New York: Oxford University Press, pp. 141–156.

**Barnes, E. Christian** (2016): Character control and historical moral responsibility. In *Philosophical Studies* 173 (9), pp. 2311–2331. DOI: 10.1007/s11098-015-0610-2.

**Bates, T.** (2015): Mixed Traits and Dispositions: Critical Discussion of Christian Miller, 'Moral Character: An Empirical Theory' and 'Character and Moral Psychology'. In *Ethical Theory and Moral Practice* 18 (2), pp. 421–424. DOI: 10.1007/s10677-014-9554-4.

**Benson, P.** (2001): Culture and Responsibility. In *Journal of Social Philosophy* 32 (4), pp. 610–620.

**Besser-Jones, L.** (2008): Social Psychology, Moral Character, and Moral Fallibility. In *Philosophy and Phenomenological Research* LXXVI (2), pp. 310–332.

**Bleidorn, W.; Denissen, J. J. A.** (2015): Virtues in action - the new look of character traits. In *British journal of psychology* 106 (4), pp. 700–723. DOI: 10.1111/bjop.12117.

**Bondeson, W.** (1974): Aristotle on Responsibility for One's Character and the Possibility of Character Change. In *Phronesis* 19 (1), pp. 59–65.

**Bradley, B.** (2017): Character and Consequences. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 78–88.

**Brandt, R.** (1970): Traits of Character: A Conceptual Analysis. In *American Philosophical Quarterly* 7 (1), pp. 23–37.

**Brink, D. O.** (2013): Situationism, Responsibility, and Fair Opportunity. In *Social Philosophy and Policy* 30 (1-2), pp. 121–149. DOI: 10.1017/S026505251300006X.

**Brink, D. O.; Nelkin, D. K.** (2013): Fairness and the Architecture of Responsibility. In Shoemaker, D. (Ed.): Oxford Studies in Agency and Responsibility. Oxford: Oxford University Press (1), pp. 284–313.

**Brownstein, M.** (2016): Attributionism and Moral Responsibility for Implicit Bias. In *Review of Philosophy and Psychology* 7 (4), pp. 765–786. DOI: 10.1007/s13164-015-0287-7.

**Buss, D. M.; Craik, K. H.** (1983): The Act Frequency Approach to Personality. In *Psychological Review* 90 (2), pp. 105–126.

**Butler, D.** (1988): Character Traits in Explanation. In *Philosophy and Phenomenological Research* 49 (2), pp. 215–238.

**Cervone, D.; Tripathi, R.** (2009): The Moral Functioning of the Person as a Whole: On Moral Psychology and Personality Science. In Lapsley, D. K.; Narvaez, D. (Eds.): Personality, Identity, and Character: Explorations in Moral Psychology. Cambridge: Cambridge University Press, pp. 30–51.

Character Counts! (2022): The Six Pillars of Character. Available online at https://charactercounts.org/character-counts-overview/six-pillars/, checked on 7/22/2022.

**Chen, Y.-L.** (2013): A Missing Piece of the Contemporary Character Education Puzzle: The Individualisation of Moral Character. In *Studies in Philosophy and Education* 32 (4), pp. 345–360. DOI: 10.1007/s11217-012-9331-6.

**Christman, J.** (2009): The Politics of Persons: Cambridge University Press.

**Ciurria, M.** (2013): Situationism, Moral Responsibility and Blame. In *Philosophia* 41 (1), pp. 179–193. DOI: 10.1007/s11406-012-9382-5.

**Clarke, R.** (2014): Omissions: Agency, Metaphysics, and Responsibility. New York: Oxford University Press.

**Clarke, R.** (2017a): Blameworthiness and Unwitting Omissions. In Nelkin, D. K.; Rickless, S. C. (Eds.): The Ethics and Law of Omissions. New York: Oxford University Press, pp. 63–83.

**Clarke, R.** (2017b): Ignorance, Revision, and Commonsense. In Robichaud, P.; Wieland, J. W. (Eds.): Responsibility: The Epistemic Condition. Oxford: Oxford University Press (1), pp. 233–251.

**Clement, S.; Bollinger, R.** (2016): Perspectives on Character Virtue Development. In *Research in Human Development* 13 (2), pp. 174–181. DOI: 10.1080/15427609.2016.1172445.

**Coates, D. J.; Swenson, P.** (2013): Reasons-responsiveness and degrees of responsibility. In *Philos Stud* 165 (2), pp. 629–645. DOI: 10.1007/s11098-012-9969-5.

**Coates, D. J.; Tognazzini, N. A.** (2012): The Nature and Ethics of Blame. In *Philosophy Compass* 7 (3), pp. 197–207.

**Cohen, T. R.; Morse, L.** (2014): Moral character: What it is and what it does. In *Research in Organizational Behavior* 34, pp. 43–61. DOI: 10.1016/j.riob.2014.08.003.

**Cohen, T. R.; Panter, A. T.; Turan, N.; Morse, L.; Kim, Y.** (2013): Agreement and similarity in self-other perceptions of moral character. In *Journal of Research in Personality* 47, pp. 816–830.

**Cohon, R.; D'Cruz, J.** (2017): Promises and Consistency. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 215–230.

**Cyr, T. W.** (2019): Moral Responsibility, Luck, and Compatibilism. In *Erkenntnis* 84 (1), pp. 193–214. DOI: 10.1007/s10670-017-9954-7.

**Cyr, T. W.** (2020): Manipulation and constitutive luck. In *Philosophical Studies* 177 (8), pp. 2381–2394. DOI: 10.1007/s11098-019-01315-y.

**Darley, J. M.; Batson, C. Daniel** (1973): From Jerusalem to Jericho. A Study of Situational and Dispositional Variables in Helping Behavior. In *Journal of Personality and Social Psychology* 27 (1), pp. 100–108.

**Doris, J. M.** (1989): Persons, Situations, and Virtue Ethics. In *Noûs* 32 (4), pp. 504–530.

**Doris, J. M.** (2008): Lack of character. Personality and Moral Behavior. Cambridge: Cambridge University Press.

**Dougherty, K. Poorman** (2007): Habituation and Character Change. In *Philosophy and Literature* 31 (2), pp. 294–310. DOI: 10.1353/phl.2007.0025.

**Driver, J.** (1996): The Virtues and Human Nature. In Crisp, R. (Ed.): How should one live? Essays on the virtues. Oxford: Clarendon Press, pp. 111–129.

**Elliott, J. R.** (2017): Character. London: Bloomsbury Academic.

**Faraci, D.; Shoemaker, D.** (2010): Insanity, Deep Selves, and Moral Responsibility: The Case of JoJo. In *Review of Philosophy and Psychology* 1 (3), pp. 319–332. DOI: 10.1007/s13164-010-0026-z.

**Fileva, I.** (Ed.) (2017a): Questions of character. New York, NY: Oxford University Press.

**Fileva, I.** (2017b): Two Senses of "Why". Traits and Reasons in the Explanation of Action. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 182–202.

**Fischer, J. M.** (1999): Recent Work on Moral Responsibility. In *Ethics* 110 (1), pp. 99–139.

**Fischer, J. M.** (2004): Responsibility and Manipulation. In *The Journal of Ethics* 8 (2), pp. 145–177.

**Fischer, J. M.** (2006): Free Will and Moral Responsibility. In Copp, D. (Ed.): The Oxford Handbook of Ethical Theory. New York: Oxford University Press, pp. 321–354.

**Fischer, J. M.** (2007): Compatibilism. In Sosa, E. (Ed.): Four views on free will. Malden, MA: Blackwell (Great debates in philosophy), pp. 44–84.

**Fischer, J. M.** (2013): Deep control. Essays on free will and value. Oxford: Oxford University Press.

**Fischer, J. M.; Ravizza, M.** (1993a): Introduction. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 1–44.

**Fischer, J. M.; Ravizza, M.** (1993b): Responsibility for Consequences. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 322–347.

**Fischer, J. M.; Ravizza, M.** (2000): Responsibility and control. A theory of moral responsibility. Cambridge: Cambridge University Press.

**Fischer, J. M.; Tognazzini, N. A.** (2009): The Truth about Tracing. In *Noûs* 43 (3), pp. 531–556.

**FitzPatrick, W. J.** (2008): Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge. In *Ethics* 118 (4), pp. 589–613.

**Flanagan, O.** (1991): Varieties of moral personality. Cambridge: Harvard University Press.

**Fleeson, W.** (2001): Toward a structure- and process-integrated view of personality: Traits as density distributions of states. In *Journal of Personality and Social Psychology* 80 (6), pp. 1011–1027. DOI: 10.1037//0022-3514.80.6.1011.

**Fleeson, W.** (2007): Situation-based contingencies underlying trait-content manifestation in behavior. In *Journal of personality* 75 (4), pp. 825–861. DOI: 10.1111/j.1467-6494.2007.00458.x.

**Fleeson, W.; Furr, M. R.** (2017): Do Broad Character Traits Exist? Repeated Assessments of Individuals, Not Group Summaries from Classic Experiments, Provide the Relevant Evidence. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 231–248.

**Fleeson, W.; Furr, M. R.; Jayawickreme, E.; Helzer, E. G.; Hartley, A. G.; Meindl, P.** (2015): Personality Science and the Foundations of Character. In Miller, C. B. et al. (Eds.): Character. New directions from philosophy, psychology, and theology. Oxford, New York, NY: Oxford University Press, pp. 41–71.

**Fleeson, W.; Jayawickreme, E.** (2015): Whole Trait Theory. In *Journal of Research in Personality* 56, pp. 82–92. DOI: 10.1016/j.jrp.2014.10.009.

**Fleeson, W.; Noftle, E.** (2008): The End of the Person-Situation Debate: An Emerging Synthesis in the Answer to the Consistency Question. In *Social and Personality Psychology Compass* 2 (4), pp. 1667–1684. DOI: 10.1111/j.1751-9004.2008.00122.x.

**Foot, P.** (1997): Tugenden und Laster. In Leist, A.; Vetter, H.; Wolf, U. (Eds.): Die Wirklichkeit des Guten. Moralphilosophische Aufsätze. Frankfurt am Main: Fischer Taschenbuch Verlag, pp. 108–127.

**Frankfurt, H. G.** (1988): Alternate possibilities and moral responsibility. In Frankfurt, H. G. (Ed.): The Importance of What We Care About: Philosophical Essays. Cambridge: Cambridge University Press, pp. 1–10.

**Frankfurt, H. G.** (1993): What We Are Morally Responsible For. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 286–295.

**Frow, J.** (2016): Character and person. Oxford: Oxford University Press.

**Funder, D. C.** (1991): Global Traits: A Neo-Allportian Approach to Personality. In *Psychological Science* 2 (1), pp. 31–39.

**Funder, D. C.** (1994): Explaining Traits. In *Psychological Inquiry* 5 (2), pp. 125–127. DOI: 10.1207/s15327965pli0502\_6.

**Ginet, C.** (2000): The Epistemic Requirements for Moral Responsibility. In *Nous* 34 (14), pp. 267–277. DOI: 10.1111/0029-4624.34.s14.14.

**Goodwin, G. P.; Piazza, J.; Rozin, P.** (2015): Understanding the Importance and Perceived Structure of Moral Character. In Miller, C. B. et al. (Eds.): Character. New directions from philosophy, psychology, and theology. Oxford, New York, NY: Oxford University Press, pp. 100–126.

**Grover, L.** (2012): The Evaluative Integration of Local Character Traits. In *The Journal of Value Inquiry* 46, pp. 25–37.

**Hacker-Wright, J.** (2015): Skill, Practical Wisdom, and Ethical Naturalism. In *Ethical Theory and Moral Practice* 18 (5), pp. 983–993.

**Hampshire, S.** (1953): Dispositions. In *Analysis* 14 (1), pp. 5–11.

**Harman, E.** (2017): When is Failure to Realize Something Exculpatory? In Robichaud, P.; Wieland, J. W. (Eds.): Responsibility: The Epistemic Condition. Oxford: Oxford University Press (1), pp. 117–126.

**Harman, G.** (1999): Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. In *Proceedings of the Aristotelian Society* 99, pp. 315–331.

**Harman, G.** (2000): The Nonexistence of Character Traits. In *Proceedings of the Aristotelian Society* 100, pp. 223–226.

**Hayes, T. L.; Hogan, R.; Emler, N.** (2017): The Psychology of Character, Reputation, and Gossip. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 268–282.

**Helzer, E. G.; Furr, R. Michael; Hawkins, A.; Barranti, M.; Blackie, L. E. R.; Fleeson, W.** (2014): Agreement on the perception of moral character. In *Personality & social psychology bulletin* 40 (12), pp. 1698–1710. DOI: 10.1177/0146167214554957.

**Hieronymi, P.** (2008): Responsibility for believing. In *Synthese* 161 (3), pp. 357–373.

**Hills, A.** (2017): Virtue and Cognition. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 147–157.

**Homiak, M.** (2019): Moral Character. Edited by Edward N. Zalta (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/sum2019/entries/moral-character/, checked on 10/24/2020.

**Hursthouse, R.; Pettigrove, G.** (2018): Virtue Ethics. Edited by Edward N. Zalta (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/, checked on 5/2/2019.

**Isen, A. M.; Levin, P. F.** (1972): Effect of Feeling Good on Helping: Cookies and Kindness. In *Journal of Personality and Social Psychology* 21 (3), pp. 384–388.

**Jacobs, J. A.** (2001): Choosing character. Responsibility for virtue and vice. Ithaca, N.Y: Cornell University Press.

**Johnson, J. A.** (1997): Units of Analysis for the Description and Explanation of Personality. In Hogan, R.; Johnson, J. A.; Briggs, S. R. (Eds.): Handbook of personality psychology. San Diego: Academic Press, pp. 73–93.

**Kalimeri, K.; Lepri, B.; Pianesi, F.** (2013): Going beyond traits. Multimodal classification of personality states in the wild. In Epps, J. et al. (Eds.): Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13. the 15th ACM. Sydney, Australia, 09.12.2013 - 13.12.2013. New York, New York, USA: ACM Press, pp. 27–34, checked on 5/28/2020.

**Kamtekar, R.** (2004): Situationism and Virtue Ethics on the Content of Our Character. In *Ethics* 114 (3), pp. 458–491. DOI: 10.1086/381696.

**Kane, R.** (2007): Libertarianism. In Sosa, E. (Ed.): Four views on free will. Malden, MA: Blackwell (Great debates in philosophy), pp. 5–43.

**Katsafanas, P.** (2017): Autonomy, Character, and Self-Understanding. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 132–146.

**Kauppinen, A.** (2017): Character and Blame in Hume and Beyond. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 46–62.

**King, M.** (2017): Tracing the Epistemic Condition. In Robichaud, P.; Wieland, J. W. (Eds.): Responsibility: The Epistemic Condition. Oxford: Oxford University Press (1), pp. 266–280.

**Kupperman, J. J.** (1995): Character. Oxford: Oxford University Press.

**Kupperman, J. J.** (2001): The Indispensability of Character. In *Philosophy* 76 (296), pp. 239–250.

**Lamiell, J. T.** (2018): On the concepts of character and personality: Correctly interpreting the statistical evidence putatively relevant to the disposition–situation debate. In *Theory & Psychology* 28 (2), pp. 249–254. DOI: 10.1177/0959354317748374.

**Laubenthal, K.** (2004): Die Renaissance der Sicherheitsverwahrung. In *Zeitschrift für die gesamte Strafrechtswissenschaft* 116 (3), pp. 703–750.

**Levy, N.** (2003): Cultural membership and moral responsibility. In *The Monist* 86, pp. 145–163.

**Levy, N.** (2005): The Good, The Bad and The Blameworthy. In *Journal of Ethics & Social Philosophy* 1 (2), pp. 1–16.

**Levy, N.** (2008): Restoring Control: Comments on George Sher. In *Philosophia* 36, pp. 213–221. DOI: 10.1007/s11406-007-9090-8.

**Levy, N.** (2009): Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick. In *Ethics* 119 (4), pp. 729–741.

**Levy, N.** (2014): Consciousness and Moral Responsibility: Oxford University Press.

**Levy, N.** (2017): Methodological Conservatism and the Epistemic Condition. In Robichaud, P.; Wieland, J. W. (Eds.): Responsibility: The Epistemic Condition. Oxford: Oxford University Press (1), pp. 252–265.

**Lucas, R. E.; Donnellan, M. Brent** (2009): If the person–situation debate is really over, why does it still generate so much negative affect? In *Journal of Research in Personality* 43 (2), pp. 146–149. DOI: 10.1016/j.jrp.2009.02.009.

**Manley, D.; Wasserman, R.** (2008): On Linking Dispositions and Conditionals. In *Mind* 117 (465), pp. 59–84.

**Marmadoro, A.** (2001): Moral Character *versus* Situations. An Aristotelian contribution to the debate. In *Journal of Ancient Philosophy* V (2), pp. 1–24.

**Mason, E.** (2015): Moral ignorance and blameworthiness. In *Philosophical Studies* 172 (11), pp. 3037–3057. DOI: 10.1007/s11098-015-0456-7.

**McKenna, M.** (2008): Putting the lie on the control condition for moral responsibility. In *Philosophical Studies* 139 (1), pp. 29–37. DOI: 10.1007/s11098-007-9100-5.

**McKenna, M.; Coates, D. J.** (2021): Compatibilism. Edited by Edward N. Zalta (Stanford Encyclopedia of Philosophy), checked on 10/26/2023.

**McKenna, M.; Warmke, B.** (2017): Does Situationism Threaten Free Will and Moral Responsibility? In *Journal of Moral Philosophy* 14 (6), pp. 698–733. DOI: 10.1163/17455243-46810068.

**Mele, A.** (1995): Autonomous Agents. From Self-Control to Autonomy. New York: Oxford University Press.

**Mele, A.** (2010): Moral Responsibility for Actions. In *Philosophical Explorations* 13 (2), pp. 101–111. DOI: 10.1080/13869790903494556.

**Mele, A.** (2020): Moral responsibility and manipulation: on a novel argument against historicism. In *Philosophical Studies* 177 (10), pp. 3143–3154. DOI: 10.1007/s11098-019-01363-4.

**Mele, A.** (2021): Direct Versus Indirect: Control, Moral Responsibility, and Free Action. In *Philosophy and Phenomenological Research* 102 (3), pp. 559–573. DOI: 10.1111/phpr.12680.

**Merritt, M. W.; Doris, J. M.; Harman, G.** (2010): Character. In Doris, J. M. (Ed.): The Moral Psychology Handbook: Oxford University Press, pp. 355–401.

**Meyer Sauvé, S.** (2011): Aristotle on Moral Responsibility. Character and Cause. New York: Oxford University Press.

**Milgram, S.** (1963): Behavioral Study of Obedience. In *Journal of Abnormal and Social Psychology* 67 (4), pp. 371–378.

**Miller, C. B.** (2003): Social psychology and virtue ethics. In *The Journal of Ethics* (7), pp. 365–392.

**Miller, C. B.** (2009): Empathy, social psychology, and global helping traits. In *Philosophical Studies* 142 (2), pp. 247–275. DOI: 10.1007/s11098-007-9185-x.

**Miller, C. B.** (2010): Character Traits, Social Psychology and Impediments to Helping Behavior. In *Journal of Ethics & Social Philosophy* 5 (1), pp. 1–36.

**Miller, C. B.** (2013): Moral character. An empirical theory. Oxford: Oxford University Press.

**Miller, C. B.** (2014): Character and moral psychology. Oxford: Oxford University Press.

**Miller, C. B.** (2017a): A New Approach to Character Traits in Light of Psychology. In Fileva, I. (Ed.): Questions of character. New York, NY: Oxford University Press, pp. 249–267.

**Miller, C. B.** (2017b): Character and Situationism: New Directions. In *Ethical Theory and Moral Practice* 20 (3), pp. 459–471. DOI: 10.1007/s10677-017-9791-4.

**Miller, C. B.** (2018): The character gap. How good are we? New York: Oxford University Press (Philosophy in action).

**Miller, C. B.** (2019): Situationism, Social Psychology, and Free Will. In Timpe, K.; Griffith, M.; Levy, N. (Eds.): The Routledge Companion to Free Will. London and New York: Routledge, pp. 407–422.

**Miller, C. B.; Furr, M. R.; Knobel, A.; Fleeson, W.** (2015): Introduction. In Miller, C. B. et al. (Eds.): Character. New directions from philosophy, psychology, and theology. Oxford, New York, NY: Oxford University Press, pp. 1–16.

**Miller, C. B.; Knobel, A.** (2015): Some Foundational Questions in Philosophy about Character. In Miller, C. B. et al. (Eds.): Character. New directions from

philosophy, psychology, and theology. Oxford, New York, NY: Oxford University Press, pp. 19–40.

**Mischel, W.** (1968): Personality and assessment. New York: Wiley.

**Moody-Adams, M.** (1990): On the Old Saw That Character is Destiny. In Flanagan, O.; Oksenberg Rotry Amélie (Eds.): Identity, Character, and Morality. Essays in Moral Psychology. Cambridge: MIT Press, pp. 111–132.

**Moody-Adams, M.** (1994): Culture, Responsibility, and Affected Ignorance. In *Ethics* 104 (2), pp. 291–309.

**Mumford, S.** (2003): Dispositions. Oxford: Oxford University Press.

**Nagel, T.** (2012): Mortal Questions. Cambridge: Cambridge University Press.

**Nelkin, D. K.** (2005): Freedom, Responsibility and the Challenge of Situationism. In *Midwest Studies of Philosophy* 29 (1), pp. 181–206.

**Nelkin, D. K.** (2023): Moral Luck. Edited by Edward N. Zalta, Uri Nodelman (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/spr2023/entries/moral-luck/, checked on 12/4/2023.

**Nelkin, D. K.; Rickless, S. C.** (2017): Moral Responsibility for Unwitting Omissions. In Nelkin, D. K.; Rickless, S. C. (Eds.): The Ethics and Law of Omissions. New York: Oxford University Press, pp. 106–130.

**Nelkin, D. Kay** (2016): Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness. In *Nous* 50 (2), pp. 356–378. DOI: 10.1111/nous.12079.

**Newman, L. S.; Ulman, J. S.** (1989): Spontaneous Trait Inference. In Uleman, J. S.; Bargh, J. A. (Eds.): Unintended thought. New York: Guilford Press, pp. 155–188.

**Pereboom, D.** (2007): Hard Incompatibilism. In Sosa, E. (Ed.): Four views on free will. Malden, MA: Blackwell (Great debates in philosophy), pp. 85–125.

**Pereboom, D.** (2014): Free will, agency, and meaning in life. Oxford: Oxford University Press.

**Piazza, J.; Goodwin, G. P.; Rozin, P.; Royzman, E. B.** (2014): When a virtue is not a virtue. Conditional virtues in moral evaluation. In *Social Cognition* 32 (6), pp. 528–558.

**Powell, B.** (1959): Uncharacteristic Actions. In *Mind* 68 (272), pp. 492–509.

**Rapp, C.** (2006): What use is Aristotle's doctrine of the mean? In Reis, B. (Ed.): The Virtuous Life In Greek Ethics. Cambridge: Cambridge University Press, pp. 99–126.

**Rhode, D. L.** (2019): Character. What It Means and Why It Matters. Oxford: Oxford University Press.

**Robb, D.** (2023): Moral Responsibility and the Principle of Alternative Possibilities. Edited by Edward N. Zalta, Uri Nodelman (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/win2023/entries/alternative-possibilities/, checked on 12/31/2023.

**Rosen, G.** (2004): Skepticism about Moral Responsibility. In *Philosophical Perspectives* 18, pp. 295–313.

**Rudy-Hiller, F.** (2017): A Capacitarian Account of Culpable Ignorance. In *Pacific Philosophical Quarterly* 98 (S1), pp. 398–426. DOI: 10.1111/papq.12190.

**Rudy-Hiller, F.** (2018): The Epistemic Condition for Moral Responsibility (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/fall2018/entries/moral-responsibility-epistemic/, checked on 9/21/2021.

**Russell, D. C.** (2009): Practical intelligence and the virtues. Oxford, New York: Oxford University Press.

**Russell, D. C.; Miller, C. B.** (2015): How Are Virtues Acquired? In Alfano, M. (Ed.): Current Controversies in Virtue Theory. New York and London: Routledge, pp. 91–122.

**Ryle, G.** (1949): The Concept of Mind. Reprint 1970. Harmondsworth: Penguin Books.

**Sanderse, W.** (2013): Character Education. A Neo-Aristotelian Approach to the Philosophy, Psychology and Education of Virtue. Delft: Eburon.

**Scanlon, T.** (2000): What we owe to each other. Reprint. Cambridge, Mass.: Belknap Press of Harvard University Press.

**Sher, G.** (2001): Blame for Traits. In *Noûs* 35 (1), pp. 146–161.

**Sher, G.** (2006a): In Praise of Blame. Oxford: Oxford University Press.

**Sher, G.** (2006b): Out of control. In *Ethics* 116 (2), pp. 285–301.

**Sher, G.** (2009): Who Knew? Responsibility without Awareness. New York: Oxford University Press.

**Sher, G.** (2017): Unintentional Omissions. In Nelkin, D. K.; Rickless, S. C. (Eds.): The Ethics and Law of Omissions. New York: Oxford University Press, pp. 3–16.

**Shoemaker, D.** (2011): Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. In *Ethics* 121 (3), pp. 602–632.

**Smith, A. M.** (2005): Responsibility for Attitudes: Activity and Passivity in Mental Life. In *Ethics* 115 (2), pp. 236–271.

**Smith, A. M.** (2007): On being responsible and holding responsible. In *The Journal of Ethics* 11 (4), pp. 465–484.

**Smith, A. M.** (2008): Control, responsibility, and moral assessment. In *Philosophical Studies* 138 (3), pp. 367–392.

**Smith, A. M.** (2017): Unconscious Omissions, Reasonable Expectations, and Responsibility. In Nelkin, D. K.; Rickless, S. C. (Eds.): The Ethics and Law of Omissions. New York: Oxford University Press, pp. 36–60.

**Smith, H.** (1983): Culpable Ignorance. In *The Philosophical Review* 92 (4), pp. 543–571.

**Sreenivasan, G.** (2002): Errors about Errors. Virtue Theory and Trait Attribution. In *Mind* 111 (441), pp. 47–68.

**Sreenivasan, G.** (2013): The situationist critique of virtue ethics. In Russell, D. C. (Ed.): The Cambridge Companion to Virtue Ethics, vol. 81. New York: Cambridge University Press, pp. 290–314.

**Strawson, G.** (1994): The Impossibility of Moral Responsibility. In *Philosophical Studies* (75), pp. 5–24.

**Strawson, P. Frederick** (2008): Freedom and resentment and other essays. London: Routledge.

**Swanton, C.** (2013): The definition of virtue ethics. In Russell, D. C. (Ed.): The Cambridge Companion to Virtue Ethics. New York: Cambridge University Press, pp. 315–338.

**Swanton, C.** (2016): Developmental Virtue Ethics. In Annas, J.; Narvaez, D.; Snow, N. E. (Eds.): Developing the virtues. Integrating perspectives. New York: Oxford University Press, pp. 116–134.

**Talbert, M.** (2009): Implanted desires, self-formation and blame. In *Journal of Ethics & Social Philosophy* 3 (2), pp. 1–18.

**Talbert, M.** (2012): Moral Competence, Moral Blame, and Protest. In *Journal of Ethics* 16 (1), pp. 89–109. DOI: 10.1007/s10892-011-9112-4.

**Talbert, M.** (2013): Unwitting Wrongdoers and the Role of Moral Disagreement in Blame. In Shoemaker, D. (Ed.): Oxford Studies in Agency and Responsibility. Volume 1. Oxford: Oxford University Press, pp. 225–245.

**Talbert, M.** (2017a): Akrasia, Awareness, and Blameworthiness. In Robichaud, P.; Wieland, J. W. (Eds.): Responsibility: The Epistemic Condition. Oxford: Oxford University Press (1), pp. 47–63.

**Talbert, M.** (2017b): Omission and Attribution Error. In Nelkin, D. K.; Rickless, S. C. (Eds.): The Ethics and Law of Omissions. New York: Oxford University Press, pp. 17–35.

**Talbert, M.** (2023): Moral Responsibility. Edited by Edward N. Zalta (The Stanford Encyclopedia of Philosophy). Available online at https://plato.stanford.edu/archives/fall2023/entries/moral-responsibility/, checked on 12/31/2023.

**Tellegen, A.** (1991): Personality Traits: Issues of Definition, Evidence, and Assessment. In Cicchetti, D.; Meehl, P. E.; Grove, W. M. (Eds.): Thinking Clearly about Psychology. Essays in honor of Paul E. Meehl. Minneapolis: University of Minnesota Press (2 Personality and Psychopathology), pp. 10–35.

**Tognazzini, N. A.** (2014): The Structure of a Manipulation Argument. In *Ethics* 124 (2), pp. 358–369.

**Upton, C. L.** (2009): Virtue Ethics and Moral Psychology: The Situationism Debate. In *J Ethics* 13 (2-3), pp. 103–115. DOI: 10.1007/s10892-009-9054-2.

**van Zyl, L.** (2015): Eudaimonistic virtue ethics. In Besser-Jones, L.; Slote, M. (Eds.): The Routledge companion to virtue ethics. New York: Routledge, pp. 183–195.

**Vargas, M.** (2013a): Building Better Beings: A Theory of Moral Responsibility. Oxford: Oxford University Press.

**Vargas, M.** (2013b): Situationism and Moral Responsibility: Free Will in Fragments. In Clark, A.; Kiverstein Julian; Vierkant Tillmann (Eds.): Decomposing the will. Oxford: Oxford University Press, pp. 325–349.

**Vetter, B.** (2013): Multi-track dispositions. In *The Philosophical Quarterly* 63 (251), pp. 330–352. DOI: 10.1111/1467-9213.12024.

**Vogel, L.** (1993): Understanding and Blaming: Problems in the Attribution of Moral Responsibility. In *Philosophy and Phenomenological Research* 53 (1), p. 129. DOI: 10.2307/2108057.

**Vranas, P. B. M.** (2005): The Indeterminacy Paradox: Character Evaluations and Human Psychology. In *Noûs* 39 (1), pp. 1–42.

**Walker, A. D. M.** (1989): Virtue and Character. In *Philosophy* (64), pp. 349–362.

**Wallace, R. Jay** (1998): Responsibility and the moral sentiments. 2. print. Cambridge, Mass.: Harvard University Press.

**Watson, G.** (1993): Responsibility and the Limits of the Evil: Variations on a Strawsonian Theme. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 119–148.

**Watson, G.** (1996): Two Faces of Responsibility. In *Philosophical Topics* 24 (2), pp. 227–248.

**Webber, J.** (2013): Character, Attitude and Disposition 21 (1), pp. 1082–1096. DOI: 10.1111/ejop.12028.

**Wolf, S.** (1988): Sanity and the Metaphysics of Responsibility. In Schoeman, F. (Ed.): Responsibility, Character, and the Emotions. New Essays in Moral Psychology. Cambridge: Cambridge University Press, pp. 46–62.

**Wolf, S.** (1993a): Freedom Within Reason. New York: Oxford University Press.

**Wolf, S.** (1993b): The Importance of Free Will. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 101–118.

**Wolf, S.** (1993c): The Real Self View. In Fischer, J. M.; Ravizza, M. (Eds.): Perspectives on moral responsibility. Ithaca, NY: Cornell University Press, pp. 151–169.

**Wright, A.** (1990-1991): Dispositions, Anti-Realism and Empiricism. In *Proceedings of the Aristotelian Society* 91, pp. 39–59.

**Wright, J. C.; Mischel, W.** (1987): A Conditional Approach to Dispositional Constructs: The Local Predictability of Social Behavior. In *Journal of Personality and Social Psychology* 53 (6), pp. 1159–1177.

**Wright, J. C.; Mischel, W.** (1988): Conditional Hedges and the Intuitive Psychology of Traits. In *Journal of Personality and Social Psychology* 55 (3), pp. 454–469.

**Zimmerman, D.** (2003): That was then, this is now: personal history vs. psychological structure in compatibilist theories of autonomous agency. In *Noûs* 37 (4), pp. 638–671.

**Zimmerman, M. J.** (1997): Moral Responsibility and Ignorance. In *Ethics* 107 (3), pp. 410–426.

**Zimmerman, M. J.** (2015): Varieties of Moral Responsibility. In Clarke, R.; McKenna, M.; Smith, A. M. (Eds.): The Nature of Moral Responsibility. New Essays. New York: Oxford University Press.

# Deutsche Zusammenfassung

Der Charakter von Menschen beschäftigt uns im Alltag beinahe täglich, wenn auch meist unbewusst. Mit Charaktereigenschaften beschreiben wir einander, erklären spezifische Verhaltensmuster und prognostizieren (mehr oder weniger zuverlässig), was jemand in einer gegebenen Situation tun wird. Dabei verlassen wir uns auf Charakter nicht nur in sozialen Interaktionen, auch in beruflichen Kontexten (bspw. im Bewerbungsprozess), in juristischen Angelegenheiten (bspw. der Entscheidung darüber, ob ein:e Täter:in in Sicherheitsverwahrung verbleibt) und auch in der Politik (bspw. der Vertrauenswürdigkeit von Politiker:innen) spielt Charakter eine tragende Rolle. Die Omnipräsenz von Charakter in unterschiedlichen Aspekten des Lebens spiegelt die Bedeutung wider, die wir ihm beimessen. Entsprechend wichtig ist uns auch, dass wir für unsere Charakterzüge verantwortlich sind. Dennoch finden sich zu dieser Frage sowohl in der akademischen Debatte als auch im gesellschaftlichen Diskurs gegensätzliche Meinungen wieder. Einerseits rechtfertigen wir viele unserer genannten sozialen Praktiken damit, dass wir kontrollieren können, wer wir sind. Andererseits sind wir stark durch unser soziales Umfeld geprägt. Die vorliegende Arbeit beschäftigt sich mit diesen unterschiedlichen Argumenten und besonders der Fragestellung, inwiefern wir uns für unseren Charakter verantwortlich verstehen können.

Ausgangspunkt ist die Beobachtung, dass gängigen philosophischen Analysen zu Charakterverantwortung ein spezifisches Verständnis von sowohl Charakter als auch Verantwortung zugrunde liegen. Unter Charakter wird oft die Gesamtheit grobzügiger Eigenschaften verstanden, das tugendethischen Überlegungen entspringt - jedoch nicht mit diesen gleichzusetzen ist - welche ich als Global Traits einführe. Zudem beschränken sich diese Analysen meist mit der Bestimmung, ob Akteure Kontrolle über ihren Charakter haben, was wiederum einem

besonderen Verständnis von moralischer Verantwortung, bekannt als Volitionismus, entspricht. Aus diesen beiden Interpretationen haben sich die genannten Positionen entwickelt: Die, die meinen, dass Akteure für ihren Charakter verantwortlich sind, argumentieren, dass diese Kontrolle haben; während die, die Akteuren diese Kontrolle absprechen, auch deren Verantwortung negieren. Allerdings haben jüngere Entwicklungen sowohl in der Diskussion zu Charakter als auch moralischer Verantwortung zu neuen Ansätzen und Modellen geführt. Hinsichtlich der Charakterdebatte ist besonders die situationistische Kritik der 90er Jahre hervorzuheben, die auf psychologischen Experimenten basiert und das Global Traits-Modell in Frage stellt. Im Zuge dieser Kritik ist die Entstehung alternativer Charaktermodelle vorangeschritten. Diese weichen von dem groben Verständnis von Global Traits ab und beschreiben Charakterzüge als strukturell enger und feingliedriger. Und in Bezug auf die Debatte um moralische Verantwortung ist eine ähnliche Entwicklung zu beobachten. Eine Vielzahl an Vertretern hat sich für theoretische Ansätze ausgesprochen, die Kontrolle nicht als notwendige Bedingung für Verantwortung sehen. Die Frage, ob wir moralisch verantwortlich für unseren Charakter sind, erhält deshalb wieder an Bedeutung. Diese Entwicklungen erfordern eine neue Bewertung der Argumente. Besonders kommt die Frage auf, inwiefern diese Entwicklungen Auswirkungen auf die Beantwortung der Fragestellung haben.

Diese Arbeit leistet einen Beitrag zu der Debatte, indem Modelle betrachtet und verglichen werden, die sich aus der relativ kürzlichen Kritik ergeben haben. Sie nimmt sich deshalb der Aufgabe an zu untersuchen, wie unterschiedliche Theorien moralischer Verantwortung mit den verschiedenen Charakterkonzeptionen umgehen können. Über die Arbeit hinweg werden deshalb der Charakterbegriff und das Konzept moralischer Verantwortung aufgefächert. Hierbei wird herausgearbeitet, dass Charakter sowohl strukturell als auch dimensional Komplexität mit sich bringt. In diesem Rahmen zeige ich, dass Charakter als eigenständiges Phänomen eines gesonderten Ansatzes bedarf. Die Arbeit versucht deshalb durch die Entwicklung eines solchen charakterspezifischen Ansatz zu zeigen, dass Akteure auch dann

verantwortlich für ihre Charakterzüge sein können, wenn sie keine Kontrolle und kein Wissen über deren Entwicklung haben.

Kapitel 1 beschäftigt sich hierzu mit der Einführung einer Klassifizierung von Charakter. Charakter wird dabei von dem Begriff/Konzept der Persönlichkeit abgegrenzt. Die Notwendigkeit dessen ist vor allem einer dualen Entwicklung in der Psychologie und Philosophie geschuldet. Beide Forschungsgebiete haben zu unterschiedlichen Zeitpunkten an dem Charakterbegriff gearbeitet. Gleichzeitig hat keine ausreichende Abgrenzung der Begrifflichkeiten Persönlichkeit und Charakter stattgefunden und deshalb zu uneinheitlicher Verwendung geführt. Zuletzt hat der Einfluss psychologischer Studien und der daraus resultierenden situationistischen Kritik zu einer Neu-Orientierung in der philosophischen Diskussion um Charakter geführt. Ich argumentiere, Christan Miller folgend, für ein Verständnis von Charakter, das als *eine Art* Persönlichkeitszug verstanden wird. Ich wehre mich jedoch gegen eine Auffassung, die Charakter als „moralische Variante" von Persönlichkeit sieht. Vielmehr lässt sich zwischen moralischen und nicht-moralischen Charakterzügen unterscheiden (Miller 2014). Anschließend schließe ich mich der vorherrschenden Ansicht an, dass Charakterzüge Dispositionen sind. Jedoch spezifiziere ich, dass Charakterzüge als solche Dispositionen zu verstehen sind, die sich manifestieren müssen (Alvarez 2017).

Basierend auf der in Kapitel 1 etablierten Definition, werden in Kapitel 2 drei Charaktertheorien eingeführt. Die drei Modelle stimmen darin überein, dass Charakterzüge Dispositionen sind, die von Akteuren habituiert werden. Jedoch unterscheiden sie sich hinsichtlich ihrer Struktur. Während *Global Traits* Charakterzüge als breite, also stabil und situationsübergreifend, begreifen, verstehen *Local Traits* (Doris 2008) Charakterzüge enger und lokalisiert, also nicht situationsübergreifend, sondern situationsabhängig. *Mixed Traits* (Miller 2013) versteht Charakter als Zusammenhang feinkörniger Dispositionen, die allerdings dennoch stabil sind und situationsübergreifend zum Tragen kommen.

# Deutsche Zusammenfassung

Mit diesem Strauß an unterschiedlichen Charakterauffassungen widmet sich Kapitel 3 der Einführung dreier Theorien moralischer Verantwortung, die die führenden Theoriestränge in der Debatte repräsentieren. Auffällig ist dabei, dass der Fokus der Debatte auf Handlungen/Unterlassungen liegt und keine charakterspezifischen Theorien vorhanden sind. Die Theorien sind sich insbesondere hinsichtlich der Notwendigkeit von Kontrolle und Wissen als Bedingungen moralischer Verantwortung uneins. Das führt dazu, dass Volitionisten, die Kontrolle und Wissen zur Voraussetzung der Erfüllung moralischer Verantwortung machen, zwar Fälle von Manipulation erklären können (Fischer and Ravizza 2000), allerdings bewerten sie Fälle anders, in denen Akteure intuitiv als verantwortlich angesehen werden ohne diese Bedingungen zu erfüllen (bspw. etwas glauben) (Zimmerman 1997; Rosen 2004; Smith 1983). Attributionisten hingegen können diese Fälle leichter beschreiben, weil sie Kontrolle und Wissen nicht als notwendige Bedingungen erachten. Sie argumentieren, dass der *quality of will* entscheidend für die Bewertung ist (Smith 2005; Hieronymi 2008). Jedoch führt das dazu, dass sie Akteure als verantwortlich ansehen, die manipuliert sind. Ein dritter Theoriestrang, Capacitarianism, ähnelt dem attributionistischen Ansatz. Sie teilen die Auffassung, dass Kontrolle und Wissen keine notwendigen Bedingungen für moralische Verantwortung sind. Jedoch sehen sie den Vorschlag der Attributionisten als nicht ausreichend. Denn Akteure können auch dann verantwortlich sein, wenn sie keinerlei Kenntnis einer Unterlassung ihrerseits haben, bspw. im Fall, dass sie etwas vergessen (Clarke 2017b; Sher 2009). Deshalb ist die Verantwortung auf die unausgeübten Fähigkeiten zurückzuführen, die diese Akteure besitzen.

Auf Basis dieser Darstellung und Einteilung werden drei repräsentative Theorien gewählt. Die Selektion der jeweiligen Theorie basiert auf deren Einschlägigkeit. John Martin Fischer/Mark Ravizza verfolgen einen volitionistischen Ansatz, der auf der Erfüllung von *guidance control* basiert. Matthew Talbert hingegen schlägt einen attributionistischen Ansatz vor. Akteure sind dann moralisch verantwortlich, wenn die Handlung das *evaluative judgment (quality of will)* des Akteurs

widerspiegelt. Zuletzt positioniert sich George Sher als Vertreter eines Fähigkeiten-Ansatzes (*Capacitarianism*). Danach sind Akteure moralisch verantwortlich, selbst wenn sie unwissentlich handeln, wenn die Handlung/Unterlassung angemessen mit ihnen verbunden ist und es von ihnen erwartbar gewesen wäre, dass sie von der Falschheit ihrer Handlung wissen oder sie unwissentlich richtig gehandelt haben und sie ausreichend kognitiven Kontakt gemacht haben.

Darauf aufbauend, adressiert Kapitel 4 die Forschungslücke in besonderem Maße. Auf Basis der in Kapitel 2 und 3 etablierten Theorien wird erörtert, was die ausgewählten Theorien moralischer Verantwortung über die eingeführten Charaktertheorien sagen können und wo sie sich unterscheiden. Um die Analyse fruchtbar im Sinne eines Ausgangspunktes für weitere Überlegungen zu machen, wird zunächst eine Standardunterscheidung zwischen direkter und indirekter Verantwortung auf den Fall von Charakter gemünzt, die zugleich die Ambiguität der Forschungsfrage und damit die Komplexität von Charakter deutlich macht. Betrachtet man den Fall von Charakter zeigt sich, dass Akteure dann *direkt* verantwortlich sind, wenn sie dafür verantwortlich sind einen Charakter zu haben, d.h. wenn sich die Verantwortung nicht aus einer anders gelagerten Verantwortung speist. *Indirekte* Verantwortung ergibt sich, wenn Akteure für die Handlungen verantwortlich sind, die Charakterzüge habituieren. Außerdem wird auf zwei weitere Aspekte der Komplexität von Charakter eingegangen. Einerseits unterscheiden sich die eingeführten Charakterkonzeptionen hinsichtlich ihrer Struktur, also ob sie eher von groben oder fein Charakterzügen ausgehen. Andererseits bedingt die Variable Charakter die Berücksichtigung zusätzlicher Dimensionen, nämlich nicht nur des Charakters selbst, sondern im Fall von indirekter Verantwortung auch der Handlungen, die Charakter habituieren. Aufbauend auf dieser Unterscheidung wird untersucht, wie sich die drei Theorien moralischer Verantwortung hinsichtlich der unterschiedlichen Theorien von Charakter verhalten. Die Analyse ergibt, dass keine der Theorien ohne Einwände dem Phänomen direkter Verantwortung vollends gerecht wird. Hinsichtlich indirekter Verantwortung können Volitionisten wie Fischer

und Ravizza besonders gut mit breiteren Theorien (bspw. Global Traits) umgehen, werfen bei feineren Theorien (bspw. Mixed Traits) die Frage auf, ob die epistemische Bedingung erfüllt werden kann. Zudem wird darauf hingewiesen, dass Volitionisten potentiell eine große Anzahl an Akteuren entschuldigen, da Charakterzüge nicht immer unbedingt wissentlich habituiert werden. Zusätzlich zeigt sich, dass Attributionisten wie Talbert zwar zeigen können, dass Akteure keine Kontrolle und Wissen benötigen, allerdings können sie Manipulationsfälle nicht ausreichend abdecken und schließen somit zu viele Akteure als verantwortlich ein. Zuletzt können Capacitarians wie Sher zwar darstellen, dass Akteure indirekt verantwortlich sind, allerdings scheitern auch sie an der Komplexität von engeren oder feingliedrigeren Charakterdarstellungen. Generell leite ich aus der Diskussion die These ab, dass Charakter aufgrund der erhöhten Komplexität eines eigenen Ansatzes bedarf. Auf Basis der Analyse werden drei Herausforderungen hervorgehoben, die durch die Variable Charakter deutlich werden. Zunächst muss ein charakterzentrierter Ansatz zeigen, wie Akteure zu den beiden Formen moralischer Verantwortung (direkter und indirekter) stehen. Zusätzlich muss erörtert werden, welche Relevanz die historische Akquirierung der Charakterzüge hat. Zuletzt muss dargestellt werden, welches Wissen Akteure über ihren Charakter haben müssen und wie sich die epistemische Bedingung auf die Komplexität von Charakter übersetzen lässt.

Kapitel 5 versucht sich an einem solchen charakterspezifischen Ansatz von moralischer Verantwortung. Hierfür erarbeite ich zunächst die Bedingungen, die erfüllt werden müssen. Hierbei zeige ich, dass Akteure auch ohne Kontrolle und mit nur minimalem Wissen verantwortlich sein können. Insbesondere argumentiere ich, dass Akteure dann verantwortlich für ihren Charakter sind, wenn sie diesen freiwillig erworben haben und sie minimales, allgemeines Wissen über ihren Charakter und die Habituierung dessen haben. Unter Berücksichtigung der herausgearbeiteten spezifischen Bedingungen, die sich im Kontext von Charakter ergeben, zeige ich, dass Akteure sowohl direkt als auch indirekt für ihren Charakter verantwortlich sind. Ein anschließender Abgleich ergibt, dass diese Bedingungen mit allen vorgestellten

Charakterkonzeptionen vereinbar sind. Zwar lösen besondere Fälle, wie etwa die Zugehörigkeit zu spezifischen Kulturen oder eine benachteiligte Kindheit oft die Intuition aus, dass diese Akteure nicht verantwortlich sind, weil ihnen der Zugang zu bestimmten moralischen Fakten besonders erschwert wird; allerdings argumentiere ich, dass diese Fälle von eindeutiger Manipulation unterschieden werden müssen, da Akteure mit solchem Hintergrund in einzelnen Fällen die Möglichkeit haben ihre Charakterzüge freiwillig zu erwerben und diese ihre evaluative Werturteil (evaluative judgment/quality of will) widerspiegel können und es trotz Schwierigkeiten erwartet werden kann, dass sie Wissen über bestimmte moralische Fakten haben. Um zu zeigen, dass diese Akteure dennoch nicht Akteuren ohne eine solche persönliche Biographie gleichen, schlage ich vor, Charakterverantwortung in Graden zu verstehen. Dieser Ansatz hat zusätzlich den Vorteil, Fälle von Kindern zu erklären, denen wir zwar Verantwortung zuschreiben, jedoch nur zu einem gewissen Grad. Zudem spiegelt er die Annahme, dass sich alle Akteure darin unterscheiden, inwieweit sie die notwendigen Bedingungen von Charakterverantwortung erfüllen. Deshalb hilft der charakterspezifische Ansatz dabei zu erklären, dass Akteure generell verantwortlich für ihren Charakter sind, jedoch biographische Besonderheiten und Faktoren den Grad an Verantwortung schmälern bzw. vergrößern können.