WHAT MAKES CHARACTER CHANGE POSSIBLE?

By
Shiman Luo

Submitted to Central European University - Private University
Department of Philosophy

In partial fulfilment of the requirements for the degree of Master of Arts

Supervisor: Professor Katalin Farkas

Vienna, Austria
2024

TABLE OF CONTENTS

INTRODUCTION

In André Gide's novel *La Porte Étroite* (1924[1909]), Jérôme, the narrator, tells the readers how annoyingly agitated he and his cousins find Aunt Félicie to be: "She was in a continual state of breathless bustle; her gestures were ungentle and her voice unmusical; she harried us with caresses and at odd moments of the day, when the need for effusion seized her, she would suddenly overwhelm us with the floods of her affection." (18) One evening, when the kids contrasted Félicie as such with more contemplative souls which they think are superior, M. Bucolin heard the conversation and smiled sadly:

> "My children," said he, "God will recognize His image even though broken. Let us beware of judging people from a single moment of their lives. Everything that displeases you in my poor sister is the result of certain events, with which I am too well acquainted to be able to criticize her as severely as you do. There is not a single pleasing quality of youth which may not deteriorate in old age. What you call 'agitation' in Félicie, was at first nothing but charming high spirits, spontaneity, impulsiveness, and grace. We were not very different, I assure you, from what you are today. I was rather like you, Jérôme—more so, perhaps, than I imagine. Félicie greatly resembled Juliette as she now is—yes, even physically—and I catch a likeness to her by starts," he added, turning to his daughter, "in certain sounds of your voice: she had your smile—and that trick, which she soon lost, of sitting sometimes, like you, without doing anything, her elbows in front of her and her forehead pressed against the locked fingers of her hands." (20)

M. Bucolin reminds the kids that Félicie used to be as graceful and capable of contemplation as they currently are. In his view, Félicie has undergone a character change over age. To defend his sister from the kids' criticisms, M. Bucolin provides a causal explanation of

her change by appealing to certain previous events in her life. However, the causal account raises two further questions. First, external influences sometimes do not change a person's character. Presumably, Félicie's character could have remained unchanged after all those events. She could have remained as highly spirited, spontaneous, impulsive and graceful as when she was younger, rather than turning into agitation and ungentleness. In other words, Félicie's character change is a contingent event. What interests us are the mechanisms behind this contingent change, i.e., the way in which it (and other similar ones) unfolds through time, and how such cases are different from cases in which no change occurs. Second, the explanation does not successfully defend Félicie. Because, even if we grant that circumstantial factors have played an important causal role, it may still be the case that Félicie herself has contributed to the change, and so is not wholly irresponsible for the result.

In this thesis, we will address the first question about what makes diachronic character change possible. The second question about one's responsibility for character will not be our primary focus. [1] However, our answer to the first question may have implications on the second as it will involve investigating the role of the agent in the process of character change. In a nutshell, my answer to the first question is: it is the individual's deliberative reflections on their own character, for one thing, and alterations in their salience patterns, for another, that make character change possible. Its implications for the second question is: the individual can play an active role in either way of reaching character change, and so is responsible for their character either way. Notice that we will be concerned with metaphysics rather than epistemology of character traits in this thesis. So, I think we can safely put aside the epistemic issues and possible sceptical worries, and assume that manifestation is a reliable (though not

[1] See Frankfurt (1988) and Callard (2018) for relevant discussions.

infallible) sign of possession. [2]

Here is a roadmap of the thesis. In Chapter 1, we will clarify our use of the term "character," differentiating it from several other uses in folk talk, philosophy, and psychology. We will analyze character in terms of character traits, and conceive of character traits as relatively stable, multi-track dispositions. We will then introduce several principal features of character traits, including their ontological dependence on manifestations, social accountability, and graduality. In Section 1.3, we will introduce the debate over situationism and the challenge situationism raises to virtue theory. We will see that it shall not worry us in our enquiry. In Chapter 2, we will explain what character change consists in. In Section 2.1, we will determine the ontological category to which character change belongs. In Section 2.2, we will distinguish character change from three other phenomena which may look like but are not character change. They are: mental disorders, personality mirroring, and mere changes of the observer's perspective. In Chapter 3, we will discuss what makes character change possible. I will put forward two different approaches to the possibility of character change. The first is called the "top-down" approach, as it requires the agent's reflections on their own character. The idea is based on Sartre's view that one cannot truly grasp one's own character independently of their deliberative choice about it (Section 3.1). The second is called the "bottom-up" approach, as it does not require the agent's reflections on their own character. It appeals to the notion of salience and claims that one's character can change as their salience patterns change, the latter of which can be reached via environmental changes or deliberate practice of attending to certain aspects of a situation (Section 3.2).

[2] Blackburn (1998: 124) claims: "there is reason to say that the way in which we have to *interpret* people is the way that, psychologically, they actually *are*." See his Ch. 3 for more details.

CHAPTER 1 WHAT IS CHARACTER?

The term "character" has many usages in everyday discourse: e.g., to say that someone "is a character" is to say that they have some distinctive eccentric features and to say that someone is a "man of character" is to say that they have good character. It also has different usages in different philosophers' work. In his lecture notes, *Anthropology from a Pragmatic Standpoint* (1798), Kant distinguishes between different kinds of character: temperament *(Temperament*), "individual nature" *(Naturell)*, and "character purely and simply" (*Character schlechthin*). According to Kant, temperament refers to the basis of an individual's inclinations which is related to their bodily constitution; individual nature comprises human beings' mental powers (*Gemüthskräfie)*, which underpin their natural abilities; and the "character purely and simply" of a person is moral and reflects what a person is prepared to make of themselves.[3] More recently, some philosophers distinguish character from personality. Among others, Fileva (2016), Miller (2013, 2014), and Goldie (2004) all take character traits to be a subset of personality traits for different reasons. Fileva and Miller draws the distinction by appealing to the idea of responsibility and norms, whereas Goldie thinks character, as opposed to personality, is a purely moral matter.

Thus, for simplicity, it is necessary to clarify our usage of the term in this thesis. By "character," we mean a relatively enduring feature of an individual—in contrast with those that can change much more often, e.g., belief, desire, and intention, etc.—manifested through one's behaviors, occurrent thoughts (e.g., talking to oneself), and feelings (e.g., feeling surprised). In other words, I am using the term "character" interchangeably with the term "personality" as is used in empirical psychology. My reason for not following them in drawing the line is simply

[3] I rely on Wood's (1999) and Frierson's (2006) accounts on this and for further references.

that there is no obvious reason for a conceptual dichotomy between traits given our current purposes. On the one hand, the grounds of change with respect to different character traits may be too divergent to be captured by a dichotomic analysis anyway. On the other hand, we may discover them to be of a uniform nature so that the grounds of their change can be given a uniform analysis after all (see Alvarez 2017). [4]

1.1 CHARACTER TRAITS AS DISPOSITIONS

In this subsection, we will first introduce character traits via different examples. We will treat character traits as a specific kind of dispositions. Then, we will provide alternative analyses of such kind of trait terms. Lastly, we will remark that one can have higher-order dispositions with respect to a particular trait disposition.

Usually, we describe someone's character by using trait terms, such as the adjectives "gentle," "agitated," "impulsive," and so on. Each character trait has its typical manifestations (Schwitzgebel (2002) uses the term "stereotypical"). Some character traits (e.g., compassion or selfishness) seem to involve both cognitive and conative manifestations as well as intentional and nonintentional behaviors. Others seem to involve no cognitive manifestations (e.g., irritability or general anxiety) or no conative manifestations (e.g., foresight or closed-mindedness). Still others seem to involve no bodily actions, e.g., the traits of being analytical and logical, which pertain to a person's reasoning capacities, though they plausibly involve verbal behaviors.

Arguably, each manifestation has a distinctive conscious character. Hence, each character trait may be associated with a distinctive conscious character through its typical manifestations. The relevant conscious character may have not only qualitative elements but also involve a

[4] Thank Iskra Fileva for help me think through this.

spatio-temporal, causal, and conceptual structure. For instance, procrastination and punctuality are each associated with a distinctive temporal awareness. Talkativeness and uncommunicativeness may each be associated with a distinctive awareness of other persons. In our example, high spirits, impulsiveness, spontaneity, and grace are character traits belonging to the younger Félicie. In contrast, agitation and ungentleness are character traits belonging to the elder Félicie. These character traits have different somatosensory bases and are associated with the kinaesthetic awareness of one's movement.

From a metaphysical point of view, character traits are generally regarded as a kind of disposition. Dispositions are what Vetter (2015) calls "easy possibilities": e.g., "almost everyone can be provoked to anger, but the mark of an irascible person is that they get angry *easily*" (71). (Vetter notes that Bird (2007: 19) and Mumford (1998: 5) also employ the term "easily" in their explications of dispositions.) Anjum and Mumford (2018) make a similar point, contending that dispositions are "considerably more than pure contingency, but considerably less than pure necessity" (149). Moreover, many philosophers follow Ryle (1949) in claiming that each character trait is a multi-track disposition, "the actualizations of which can take a wide and perhaps unlimited variety of shapes": [5]

> When Jane Austen wished to show the specific kind of pride which characterized the heroine of 'Pride and Prejudice', she had to represent her actions, words, thoughts and feelings in a thousand different situations. There is no one standard type of action or reaction such that Jane Austen could say 'My heroine's kind of pride was just the tendency to do this, whenever a situation of that sort arose.' (32)

We could apply the standard, simple conditional analysis of dispositions to multi-track

[5] For a criticism see Lyons 1973.

dispositions, analyzing a trait disposition in terms of *a set of* conditionals. The antecedent of each conditional indicates a stimulus condition, which triggers the manifestation as presented in the consequent. For instance, the elder Félicie is agitated in that "at odd moments of the day, when the need for effusion seized her, she would suddenly overwhelm us with the floods of her affection," and so on and so forth. Moreover, Manley and Wasserman (2007, 2008) suggest that the set of conditionals associated with a multi-track disposition should be uncountably infinite. They argue that each conditional's antecedent should describe "a fully specific scenario that settles everything causally relevant to the manifestation of the disposition" (2007: 72). According to their view, if a trait disposition, such as loquaciousness or irascibility, could be triggered by any stimulus condition, then the corresponding conditionals would consider any scenario relevant in their antecedents.

The simple conditional analysis distinguishes stimulus conditions of manifestations from background factors which interfere in the operation of a trait disposition. For example, even when there are relevant stimuli such as recognizing someone in need, one might nevertheless fail to manifest compassion because of severe emotional disturbance, serious depression or other mental illness, extreme fatigue, being intoxicated, or being under the influence of tranquilizers, and so on. However, the boundary between background factors and stimuli can be blurry. We can describe Félicie as acting "out of character" if she does not display "floods of affection" at "odd moments of the day." However, we can also think that sometimes behaving otherwise than we expect is part of her character, too. Due to this reason, some philosophers propose to cancel the distinction. Vetter (2015), for example, proposes to analyze dispositions in terms of their manifestations alone, while specifying the factors intervening with the manifestations (for similar ideas, see Mumford and Anjum 2011: 22, Hüttemann 2013). Whether we choose to dismiss the blurry boundary or not, it follows that a person's character can be interpreted in many equally valid ways, each matching a familiar pattern. When we

cannot identify a recognizable profile in someone's character, we tend to describe that person as insane (Schwitzgebel 2002: 263).

A last remark is that one can have higher-order dispositions about character traits, e.g., one can be disposed to become more patient. In that case, the strengthening or weakening of a particular trait disposition will be a manifestation of the relevant higher-order disposition. For instance, a mild-tempered person might become irascible as they manifest the higher-order disposition to become irascible (genetically, for instance). Of course, one can also strengthen or weaken a trait disposition without being disposed to strengthen or weaken it. For example, a mild-tempered person who is not disposed to become irascible can become irascible after a traumatic event.

So far, we have introduced character traits as multi-track dispositions and candidate analyses of trait terms. In the next subsection, we will introduce several main features of trait dispositions, in particular those which distinguish them from other kinds of dispositions.

1.2 FOUR FEATURES OF TRAIT DISPOSITIONS

In this subsection, I introduce four main features of character traits for our purposes. A first is that the existence of character traits depends on previous manifestations, which may or may not be frequent or regular. The second is that character traits are more than mere patterns or statistical regularities; they are a *sui generis* kind of dispositions. The third is that character traits come in degrees. The fourth is that (certain) character traits may have a maximal and/or minimum degree.

First, let us start with the idea that character traits are manifestation-dependent. As we conceive of character traits as a specific kind of dispositions, it is worth asking whether (and if yes, how) they differ from paradigmatic physical dispositions. Some philosophers maintain that,

above all, someone possesses a given character trait *only if* they have manifested that character trait, whereas physical dispositions are ontologically independent of their manifestations (see Alvarez 2017, Alston 1970, Hampshire 1953: 6). In this view, someone who has never had a timid thought, feeling, or behavior, for example, will not count as a timid person. By contrast, a glass is fragile whether or it has broken before. According to Alvarez (2017), we can at most say that such a person merely *potentially* or *counterfactually* possesses a given character trait. Take bravery for example: someone who has never exhibited bravery can only be considered *capable* of being brave or *would* be brave *if* the circumstances allowed (see also Mumford 1998: 8, Wright 1988). This perspective on trait dispositions aligns with the commonsense view of habits (see Ryle 1949: 85, 119). Conversely, the opposing view argues that a timid person, for example, may never have shown timidness simply due to the lack of relevant stimuli (see Miller 2014: 19ff., Brandt 1970; note that Brandt's argument is restricted to manifesting actions). Alvarez counters that the stimulus conditions for most character traits (e.g., generosity, malice) are not so special that it should be challenging for them to be exhibited:

> For even someone in solitary confinement can have malicious thoughts, generous intentions or mean reactions even if only to imagined scenarios; moreover, failure to have certain thoughts, images, etc. may also, given certain conditions, constitute manifestation of a character trait. It seems that being conscious and having basic mental abilities is all that is required to be able to manifest one's character traits. (Alvarez 2017: 85)

For our purposes, it is convenient to assume character traits as manifestation-dependent dispositions, as we are primarily concerned with ordinary examples of trait attributions such as the Félicie case. Moreover, whether character traits should be manifestation-dependent is presumably independent of our main question of what makes character change possible.

However, this immediately raises the question of whether possessing a character trait requires that the person manifests it at least relatively frequently or regularly. Brandt (1970) provides a negative answer, arguing that we do not necessarily consider all of a person's past behaviors and manifested thoughts when making trait attributions. Similarly, Alvarez (2015) contends that, depending on the nature of the action or the circumstance, we may attribute character traits to a person even if there has been only one manifestation. I find these arguments persuasive. Everyday character trait attributions are not based on evidence of statistical regularities, nor should trait attributions in philosophy.

This leads us to consider the second feature of character traits, namely that character traits are a *sui generis* kind of dispositions rather than mere patterns or statistical regularities. Denying that character traits rely on regular or frequent manifestations allows us to reject the so-called summary theory of character traits, which claims that a person's character traits are simply patterns of their actual behavior, thoughts, and/or feelings (see Miller and Knobel 2015, Miller 2014). In addition to the argument from independence, another major consideration is the normativity of character traits, which Schwitzgebel (2002: 262) refers to as "social accountability." The basic idea is that when we attribute a character trait to someone, we expect that their actions and thoughts not only align with—but also ought to align with—the stereotypes associated with that trait. According to Schwitzgebel, we explain and predict their behavior, thoughts, and feelings in ways that fit those stereotypes, as demonstrated by the common saying, "she is that kind of person." Moreover, we tend to be bothered if someone fails to meet our expectations in a trait-relevant situation. Conversely, individuals to whom a character trait is ascribed often behave in ways that conform to these stereotypes (262). This can be explained by the psychological tendency of human beings to avoid disappointing others, leading them to fit into others' interpretations.

Another theory that might be useful here is the theory of affordances. We can view character trait dispositions as a special kind of affordance in social interactions, indicating possibilities for social behaviors among social agents. For example, traits such as generosity, kindness, and honesty are positive traits that encourage trust and cooperation. On the other hand, traits like narrow-mindedness, arrogance, and deceitfulness are negative traits that can lead to conflict and mistrust. Neutral traits, such as reservedness, routine-orientedness, and neutrality, do not inherently provoke positive or negative interactions but can influence the dynamics of social engagement depending on the context. To take an interesting example from Haslanger (2018: 236), while gentlemanliness is conventionally seen as a virtue, it may not be viewed as such from a feminist perspective.

From either perspective, considering character traits as a *sui generis* kind of disposition rather than mere patterns or statistical regularities clarifies that "our common-sense psychology is a normative practice we learn to engage in for the purpose of becoming responsible, understandable agents" (McGeer 1996: 512). Furthermore, as Anjum and Mumford (2018: 7) point out, a disposition might ground a probability, but the two are not equivalent.

The third feature of character traits is that they come in degrees. [6] To illustrate this, it is useful to conceive of each multi-track trait disposition as a set of single-track dispositions as Schwitzgebel (2002) does. I find the theory very useful and will represent it here. Schwitzgebel refers to each set of single-track disposition as a "stereotype" of the trait:

> A dispositional stereotype is a stereotype whose elements are dispositional properties. An example is the stereotype for being hot-tempered. This stereotype will include the disposition to respond angrily to minor provocations, the disposition to be slow in

[6] For discussions on this idea, see Miller 2013: Ch. 1, fn. 26 for further references.

> cooling off after a fight, the disposition to feel and express frustration quickly when one's will is thwarted, and so forth. Personality traits, such as being reliable, affable, or tenacious, are arguably all characterizable by means of such dispositional stereotypes. To have these personality traits is really nothing more than to match these stereotypes. (251)

Which stereotypes we specify will not matter because no single stereotype will be necessary or sufficient for possessing a character trait (Schwitzgebel discusses the case of belief at p. 252, but he thinks the same applies in the theory of character traits). So, for instance, the disposition to consciously judge that timing matters is neither necessary nor sufficient for being punctual (see Miller 2014 on the relationship between possessing character traits and possessing normative judgments). Nevertheless, Schwitzgebel acknowledges that some stereotypes will be central and others marginal (265). His account implies that the degree to which someone possesses a given character trait positively correlates with the proportion of their matched stereotypes and the centrality of those stereotypes (250). Regarding this matter, Miller (2014) suggests two alternative ways of evaluating the degrees of trait possession: by correlating them with the frequency of manifestation in trait-relevant situations and with the agent's willingness to perform trait-relevant actions, especially those that are demanding (see Alston 1970: 75–6, 78 and Upton 2009: 60).

The continuity of character immediately raises the question concerning the fourth feature of character traits, namely whether the boundaries of a trait category are vague. This centers on whether character traits (or some of them) have minimal or maximal degrees. The question of minimality is sometimes framed as categoricity, i.e., whether there is a "minimal threshold" for possessing a particular character trait (see Miller 2013, 2014). Some philosophers express doubts on this matter. Alvarez (2017), for instance, argues that there might not be a universal

answer to this question. Russell (2009: 114) suggests that it can be unclear whether someone truly possesses a particular trait disposition. Similarly, Schwitzgebel argues that whether an individual counts as possessing a particular trait disposition depends on the context (270). It follows from his account that it is context-dependent whether individuals described by Ryle (78) as "sentimentalists," who experience "lively and frequent" "patriotist" feelings but exhibit no patriotist actions, are genuine patriotists. A good consequence, Schwitzgebel points out, is that this vagueness and context-dependency afford trait attributions with flexible utility (253).

Regarding the question of maximality, I have no affirmative arguments to offer here, but I shall clarify the issue. Vetter (2015) discusses two conceptions of the maximal degree of a disposition. One is that something possesses a disposition to the maximal degree just in case it can do nothing other than manifest it. For example, being maximally honest means being capable of only honest actions. Similarly, being impulsive means being capable of only impulsive behaviors. The other conception proposes that maximally possessing a disposition means exercising it; furthermore, the closer a disposition is to its exercise, the greater its degree. To take Vetter's example, when an irascible person becomes angry, their irascibility peaks, but then recedes to a more moderate level once they calm down. The first conception aligns maximality with necessity, while the second with actuality (86–7). In the case of character traits, conceived as relatively stable features, it is the necessity condition that should be relevant.

Fortunately, for our purposes, it is unnecessary to determine the conditions for gaining or losing a given character trait. For instance, we do not have to establish whether Félicie completely ceases to manifest high spirits, impulsiveness, spontaneity, and grace. It suffices for us to discuss the strengthening and weakening of character traits. That is, we only need to acknowledge that Félicie possesses certain character traits to a lesser degree as she gets older.

Thus far, I have discussed four main concerns about the main features of character traits. In the following section, I will delve into an important issue concerning the robustness of character traits.

1.3 SITUATIONISM AND GLOBAL TRAITS

Gilbert Harman (1999, 2000, 2001, 2003, 2009) and John Doris (1998, 2002, 2010) are the leading philosophers who brought situationism in psychology into philosophy (for further references see Miller 2023). [7] Situationism, based on experimental findings, denies the existence of what they call global (or robust) character traits, i.e., character traits as manifested in behavior both across trait-relevant situations and over time in the same situations. According to the situationist, the relevant experimental evidence does not prove that there are such things as cross-situational consistent character traits, even if it is compatible with the existence of what they call "local" character traits. Situationists argue that this constitutes a big challenge to virtue theory (including virtue ethics, virtue epistemology, virtue aesthetics, and so on), as the tradition of virtue theory going back to Aristotle is generally interpreted as identifying virtue and vice with robust character trait. In this thesis, we are not concerned with normative ethics; however, the challenge from situationism does bring the reality of global character traits into question, and thus bring the substantivity of character change into question. The worry is that if situationism is true, then individuals never undergo substantive character change, but only have varying reactions to different situations. To put it another way, there is no character change, but only changes of reactive patterns.

In this subsection, I will argue that this shall not worry us given our present purposes. Even if situationism is true, it does not follow that there is no substantive character change. I will first

[7] For more discussions, see, e.g., Athanassoulis 2000, Badhwar 2003.

introduce one experiment whose interpretation is controversial in the debate.

Let us consider one experiment which is interpreted in support of situationism. In this experiment, individuals in the experiment group found a dime in the coin return slot of a phone booth, whereas those in the control group did not. All individuals in both groups are supposed to see a passenger dropping a stack of papers by accident. In the end, 88% in the experiment group helped pick up the dropped papers, whereas only 4% in the control group did. Here comes the situationist's argument. If we conceive of compassion as a global character trait manifested in behavior cross trait-relevant situations, then a compassionate person will help whether they found a coin or not. On reasonable albeit controversial statistical assumptions, the situationist concludes that the result shows that the average individual is not compassionate in the sense we just specified. It follows that human behavior is more influenced by trivial situational factors rather than an internal, relatively reliable character.

Interpreting the results of the experiments means explaining the human behavior involved in the experiments. There are several alternative ways of explaining the distributions of helping behavior in the two groups. The key point is that the situationist argues that it is implausible to say that the helpers are all compassionate people, whereas none of the non-helpers are. By contrast, it is more plausible to explain the behavioral divergences in terms of the seemingly correlating situational factor of having discovered a dime in the coin return slot of a phone booth.

However, it is unclear to me how this interpretation, even if plausible, could undermine the possibility of substantive character change. As discussed in Section 1.1, we consider character traits as multi-track dispositions. This fortunately leaves room for there being local character traits only. To put it another way, the possibility of substantive character change does not rely on the reality of global character traits. Moreover, we allow the possibility of acting out of

character. It means that an individual can act contrary to the relevant behavioral dispositions in situations associated with a certain character trait. For example, when a compassionate person ignores the need to help pick up dropped papers, they are acting out of character. Explanation human behavior requires us to consider many related issues, e.g., whether to interpret the individual's behavior as an instance of acting out of character or the manifestation of a local trait. Moreover, an individual can be acting out of a character trait while manifesting another relevant one; for instance, we cannot simply assume that the individuals in the experiment either has compassion or no character traits at all. Furthermore, this underdetermination is to be expected given our understanding that a character trait may not be fully captured by an explicit specification of its dispositional stereotypes (see Section 1.1).

All in all, given the above considerations, I do not think the situationist account straightforwardly challenges our enquiry into the possibility of substantive character change.

CHAPTER 2 WHAT IS CHARACTER CHANGE?

In this section, I present a conception of character change. In Chapter 2.1, I propose that we are concerned with the kind of character change that is a continuous, open-ended process. In Section 2.2, I distinguish character change from three other kinds of phenomena: changes in mental health condition, personality mirroring, and changes in trait attributions merely due to shifts in the observer's perspective.

2.1 AN ONTOLOGICAL RESOLUTION

Given what we have said about character in Chapter 1, character change consists in the strengthening or weakening of certain character traits. Therefore, character change is change of dispositional states, and is itself not a state. What ontological category does character change belong to? In this subsection, I propose that the kind of character change we are interested in is a continuous and open-ended process as opposed to an event.

Let us begin with the ontological trichotomy frequently referenced by philosophers of mind: process, event, and state. It is useful to illustrate this by associating it with Vendler's (1957) classic fourfold classification in linguistics. Vendler categorizes verbs into four types: state, achievement, activity, and accomplishment. For simplicity, we can refer to Rothstein's (2004) approximation of Vendler's classification as follows: [8]

> Crudely, states are non-dynamic situations, such as be happy or believe; activities are open-ended processes, such as run; achievements are near-instantaneous events which are over as soon as they have begun, such as notice; and accomplishments are processes

[8] Mourelatos (1978) also discusses ways of associating the two classifications in detail.

which have a natural endpoint, such as read the book. (2004: 6)

With the fourfold distinction in mind, we can differentiate between instantaneous change and continuous change. An instantaneous change is an achievement, whereas a continuous change is an activity or an accomplishment. Presumably, a person's character can change instantaneously. However, such a change of character is plausibly unobservable (at least) at the time it occurs. For example, imagine a young mother who becomes strong and decides to live boldly after witnessing her 4-year-old child hide his tears to spare her further grief during a violent moment in her abusive marriage. It could be after years that she would recognize the change that occurred in that moment. Similarly, in the case of Félicie, we often identify a person's character change by comparing the traits they possess at different stages of life rather than different moments in a day. Félicie's character change is not drastic; we can still see the younger Félicie's qualities in the elder Félicie, even if they have faded with age. However, long-term character changes can also be drastic. Consider an innocent young man who becomes indifferent and selfish after devoting his heart to his lover and being betrayed. Another example from literature is the transformation of Xiangzi, the protagonist of the Chinese modern novel *Rickshaw Boy* by Lau Shaw. Initially simple, upright, and kind, Xiangzi becomes corrupt, pedantic, bored, and numb after various life twists.

Diverse forms of character change can occur throughout a person's life. For convenience, we will focus on continuous, observable character change in the remainder of this thesis. Such character change is a process, which can be classified as either an activity or an accomplishment. For example, becoming humorous is a process with an endpoint, namely reaching the minimal threshold of humor. In contrast, the process of becoming more humorousis open-ended, assuming there is no maximum level of humor or the maximum level is indefinite. Given that the minimal threshold of a character trait can be vague (as noted in Section 1.3), any form of

character change—whether becoming humorous or becoming more humorous—can be conceived as as an open-ended process. This means it is a process without a definite endpoint or anticipated result, encompassing an indefinite length of time, and therefore is not an event. Thus, when we inquire about the possibility of character change, we are concerned with the unfolding of a process rather than its culmination.

In the next subsection, we will differentiate three phenomena which assimilate character change but should not be conflated with it.

2.2 CHARACTER CHANGE: WHAT IT IS NOT

In this subsection, I discern three phenomena from character change: changes in mental health condition, personality mirroring, and mere changes in trait attributions.

First, changes in mental health condition may lead to character changes, but they themselves do not constitute character changes. For example, consider depression. Depression is a mood and, in some cases, diagnosed as a mental disorder. According to DSM-5, a person has depression if they have persistent feelings of sadness, low mood, and loss of interest in their usual activities for a relatively long period of time, while demonstrating a number of other common symptoms, including appetite change, sleep disturbance, decreased energy, psychomotor changes (e.g., agitation or retardation), impaired ability to concentrate, thoughts of death, etc. (163–4). Obviously, depression involves changes in one's patterns of behaviors, thoughts, and feelings. However, it is not character change, for it is a pathological process. Character change may occur due to pathological reasons, but the process itself is not pathological. Changes in mental health condition might amount to character change, but that will depend on multiple factors, such as the severity of depression, ranging from mild to severe. It also depends on the person's present character. A resilient, strong and adaptable person, for

instance, may respond to depression in more positive ways and continue to exhibit their character even if they are struggling with depressive symptoms. Furthermore, it depends on the patient's access to social support, including supportive relationships and therapy resources. Finally, mitigating the impact relies on addressing the underlying genetic, biological, environmental, and psychological issues, such as trauma and stressors.

Consider the following case of depression which involves no character change:

> Sarah, a 32-year-old marketing manager, has always been known among her friends and colleagues for her outgoing and sociable character. However, in recent months, she has been struggling with feelings of sadness, hopelessness, and fatigue. Despite her depression, Sarah continues to maintain her usual friendly and optimistic demeanor when interacting with others. She still attends social gatherings and engages in conversations with her friends and coworkers, although she may occasionally feel emotionally drained or disinterested. At work, Sarah continues to fulfill her professional responsibilities to the best of her ability. However, her concentration and productivity have decreased slightly, and she occasionally struggles to meet deadlines due to feelings of lethargy and disinterest. Despite these challenges, Sarah's coworkers still view her as a reliable and competent team member. In her spare time, Sarah spends more time alone than usual, not because she wants to isolate herself, but because she lacks the energy to socialize or participate in activities she once enjoyed. While alone, she also finds it challenging to muster the motivation to pursue hobbies or engage in self-care activities.

In this case, Sarah is experiencing depression, but her outgoing character and professionalism remain largely unchanged. We can construct similar cases in which the subject is experiencing Post-Traumatic Stress Disorder (PTSD) without experiencing a character change. For instance, consider someone who has experienced a traumatic event (e.g., a car accident) and developed

a heightened fear response (which typically involves physiological changes such as increased heart rate and sweating, as well as avoidance behavior) to triggers associated with the trauma. Plausibly, we do not say they have become a fearful person. The reason is not that they tend to have fear responses only in restricted kinds of situations. Rather, it is because their fear responses are not part of their intrinsic nature at all, but a symptom of mental illness.

The second phenomenon which we might mistake for character change is called "personality mirroring" (sometimes also called the "chameleon effect," see Chartrand & Bargh (1999) for an introductory reading). Personality mirroring is the social phenomenon in which individuals unconsciously mimic the behaviors, gestures, speech patterns, or attitudes of those around them (or even fictional characters). It can occur in various types of social interactions, including conversations, group settings, or even observing each other from a distance. Such mirroring behavior is temporary and serves several social and psychological functions. As to social functions, it can create a sense of similarity, leading to increased comfort and trust in the interaction. Moreover, it can enhance communication effectiveness and mutual understanding, as synchronized body language, facial expressions, or verbal tone can contribute to nonverbal communication. Regarding psychological functions, individuals might mirror each other to fit into a group, either by conforming to group norms or by signaling empathy with others' experiences. Mirroring behavior, even if it deviates from one's previous patterns of behavior, does not indicate a character change. It remains a further question whether (and under what conditions) such situational adaptations can lead to genuine character change.

The third phenomenon I am concerned with here involves cases in which it is only the observer's attitude toward a person, not the person's own character, that has changed. Murdoch (1997) vividly describes such a case in detail:

> A mother, whom I shall call *M*, feels hostility to her daughter-in-law, whom I shall call *D*. *M* finds *D* quite a good-hearted girl, but while not exactly common yet certainly unpolished and lacking in dignity and refinement. *D* is inclined to be pert and familiar, insufficiently ceremonious, brusque, sometimes positively rude, always tiresomely juvenile.... Thus much for *M*'s first thoughts about *D*. Time passes, and it could be that M settles down with a hardened sense of grievance and a fixed picture of *D*, imprisoned (if I may use a question-begging word) by the cliché: my poor son has married a silly vulgar girl. However, the *M* of the example is an intelligent and well-intentioned person, capable of self-criticism, capable of giving careful and just attention to an object which confronts her. *M* tells herself: 'I am old-fashioned and conventional. I may be prejudiced and narrow-minded. I may be snobbish. I am certainly jealous. Let me look again'. Here, I assume that *M* observes *D* or at least reflects deliberately about *D*, until gradually her vision of *D* alters. If we take *D* to be now absent or dead this can make it clear that the change is not in *D*'s behavior but in *M*'s mind. *D* is discovered not to be vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful, and so on. (Murdoch 1997: 312–3)

Some theorists, such as Hayes et al (2016), makes the bold claim that character just is reputation. However, this contradicts our preliminary assumption in this thesis. In our view, character is a real mental feature of the individual. Therefore, we must distinguish changes in trait attributions from changes in one's character itself. That said, we can reconcile Hayes et al's position with our realist position by conceiving of character traits as a kind of affordances for social interaction (see Section 1.2), that is, possibilities for actions in social interactions.

So far, I conclude my discussion of the three phenomena which we should not conflate with character change. In the next section, we will start to investigate the main question of this thesis,

namely what makes character change possible.

CHAPTER 3 WHAT MAKES CHARACTER CHANGE POSSIBLE?

In this section, I present two proposals concerning the possibility of character change. The top-down approach claims that a person's character can change as they reflect on their character (Section 3.1). The bottom-up approach claims that a person's character can change as their salience patterns change (Section 3.2). At the end of Section 3.2, I will briefly remark that both approaches leave room for agential freedom in the process of character change.

3.1 THE TOP-DOWN APPROACH

Sometimes people desire a different character for themselves. Often, they have more or less a sense of what their present character is like, and what it ideally should be. It is relatively easy to imagine how one might come to desire a different kind of character. For instance, we might hope that we could become more extraverted, more confident, and less procrastinating, either for these character traits' own sake or for their instrumental value. In contrast, it is relatively unclear how one could come to know about their own character. It is a variant of the issue of self-knowledge, but its subject matter, i.e., character traits, has received relatively little attention in the literature. To my knowledge, Sartre is almost the only exception. In the book *Being and Nothingness* (2003[1956]), Sartre develops a theory of self-consciousness. According to this theory, self-consciousness asserts the subject's freedom of self-interpretation with respect to whatever mental state as *given* within consciousness. In his words, "Consciousness is a pure and simple negation of the given, and it exists as the disengagement from a certain existing given and as an engagement toward a certain not yet existing end" (500). Sartre applies this theory to self-assessment of character, contending that it is inseparable from deliberative reflections on one's own character.

In what follows, I will introduce Sartre's view about self-assessment of character. Compared to employing more of Sartre's jargons to explain his thought, I think it would be more helpful to offer an interpretation within our dispositional framework of character traits. The key notions I will rely on throughout my interpretation are willingness and choice.

Basically, Sartre thinks that any purported self-ascription of character traits (whether on the basis of self-observation or testimony, and so forth) or even raising the question "Do I have a particular character trait X?" will ultimately collapse for me into (or, to make it weaklier, imply) the question "Shall I be X?" In other words, for Sartre, discovering one's character traits is not independent of choosing one's character traits. For example, when I ask myself "Am I a punctual person?" I am also deliberating whether to be a punctual person. This differs from the view that one can first identify their character traits and then forming a second-order inclination to strengthen some of them and weaken others (see Ryle 1949: 97, Alvarez 2017: 91). This latter stance is well illustrated by Ryle's example, "A person may find that he is too fond of gossip, or not attentive enough to other people's comfort" (75). Moreover, there are many reasons for which one may find their character too strong or too weak in one aspect or another: either for its own sake or for its instrumental value. Ryle describes a hotel-proprietor who tries to possess "equability, considerateness, and probity" only due to his strong desire to get rich (97). However, notice that these motivating reasons are external reasons (which is to be contrasted with the internal reasons which Sartre's account focuses on; see below).

Sartre's view may first sound implausible to some of us, or at least far less intuitive than the second-order inclination view. For the reasons I gave earlier, now I will offer an interpretation of his account within our dispositional framework of character traits. The interpretation may or may not respect Sartre's original thought, but for the time being we can

just focus on the simplified Sartrean idea that the "Am I…?" and "Shall I…?" questions cannot be separated.

In our dispositional framework of character traits, when I ask myself "Am I a punctual person?" what I am really asking is whether I am *disposed* to be on time when I have an appointment (for instance). In my interpretation, the crux of the Sartrean view is this: since I am a self-conscious being, the question about my disposition would ultimately amount to a question about my *willingness*. That is, I would actually be asking myself: am I *willing* to be on time when I have an appointment?

Second, the question "Shall I be punctual?"—asked from a first-person point of view—is not asking for considerations on social norms or external reasons for being a punctual person, nor concerned with predicting my future behavior based on past evidence (which differentiates the Sartrean view from the second-order inclination view, as anticipated earlier). Rather, it asks for a *resolution* with regard to acting in a certain way, based on my internal (or internalized) reasons for or against that way of acting. In other words, this second question puts one *in a position* to make choices. In this sense, the "Am I…?" question will ultimately collapse into (or imply) the "Shall I…?" question, because one cannot be willing to act *X*-wise without being in a position to choose to act *X*-wise. For example, being willing to act punctually puts me in a position to choose to act punctually. Similarly, being unwilling to act punctually also puts me in a position to choose not to act punctually.

What is worth emphasizing here is that when we adopt the weaker thesis—i.e., that of inseparability rather than collapsibility—it is the two questions, not the answers to them each, that are inseparable. I could have been willing to act punctually, yet chosen not (or not chosen) to act punctually for other reasons. For example, suppose that I have regularly or frequently been late for appointments, not because I would like to do so but because I am a disorganized

or forgetful person, with a disordered temporality. In that case, although I am willing to be on time for appointments, I am equally or even more willing to do things in a relatively disorganized way or pay more attention to things I consider to be more important than remembering to bring my umbrella, for instance. In this relation, it is worth emphasizing that at least within our dispositional framework of character traits, no assessment of character is supposed to be assessment of past patterns of behavior, thought, and feelings. In other words, remind that we must not conflate dispositions with statistical regularities. This can never be overemphasized when the assessment in question is self-assessment, in particular if (but not only if) we assume a self-other asymmetry when it comes to the epistemology of mind, meaning that the subject has a special access to their willingness.

The freedom of self-interpretation is not unlimited. Sartre does recognize a limit to it, namely others' interpretation of oneself. My freedom is limited in that others have a form of interpretation of me from the third-person point of view—based on empirical evidence—and unavailable to me from the first-person perspective. They ascribe characteristics to me—which Sartre calls "unrealizables"—such as my race, my job, my physical look, my character traits, and so on (548 ff.). With respect to character traits in particular, Sartre writes:

> Such qualities as 'evil', 'jealous', 'sympathetic' or 'antipathetic' and the like are not empty imaginings; when I use them to qualify the Other, I am well aware that I want to touch him in his being. Yet I cannot live them as my own realities. If the Other confers them on me, they are admitted by what I am for-myself; when the Other describes my character, I do not 'recognize' myself and yet I know that 'it is me'. I accept the responsibility for this stranger who is presented to me, but he does not cease to be a stranger. (298)

This limitation is important, and we will analyze it in two steps. First, we will consider the unavailability thesis. I will first introduce one interpretation provided by Moran (2001: 186). Then, I will consider a worry about this interpretation, and propose a solution to this problem, again, by appealing to the notion of willingness. Second, we will consider the limitation thesis. I will provide my interpretation of it, too.

Moran (2001: §5.5, fn. 19) suggests one way of understanding this unavailability thesis, with the idea sourced from discussions in moral psychology (Moran discusses quotes from Williams (1985: 10–11)). According to Moran, my own awareness of a certain character trait can undermine the claim that I possess it (170). For example, suppose that I am wicked. The fact that I myself am aware of my wickedness to a degree undermines the claim that I am wicked. Because we do not expect a wicked person to be aware of their own wickedness. Or, at least, a wicked person who is aware of their own wickedness is not so wicked after all. This interpretation is based on the general observation that our attitude toward someone to some extent depends on their attitude toward themselves (169).

Now, I consider one possible objection to Moran's interpretation when it is applied to character traits in general, and propose a response to it. The objection is that this interpretation may only apply to moral character traits such as kindness, and will not apply to character traits such as punctuality and shyness. For, the objection goes, it seems unlikely that a person is less punctual or shy just because they are aware of their own punctuality or shyness. However, I think we can respond to this objection again by appealing to the notion of willingness, in a way echoing the above interpretation of Sartrean freedom. The idea is that a punctual person who considers themselves as punctual is considering themselves as willing to act punctually. So, according to our previous discussion, such a person would consider themselves as being in a position to choose to act punctually. However, it is implausible that a punctual person is ever

in a position to choose to act punctually, though they do sometimes choose to act unpunctually. This is because when they act punctually, it is never a choice but a natural inclination.

Next, let us consider how the fact that others can interpret me in a way that is unavailable to me should pose a limit to my freedom of self-interpretation. Here, I offer an account based on Schwitzgebel's (2002: 262) notion of social accountability (see Section 1.2). As Schwitzgebel points out, character traits, unlike beliefs, are not thoroughly reason-responsive. For instance, while a dishonest person can think that there is good reason for being honest, an atheist cannot consistently think that there is good reason for believing in God. Nonetheless, he says, character traits are socially accountable, meaning that the individual has to make sense of themselves by ensuring their behavioral patterns are interpretable by others. As Schwitzgebel reminds us, we tend to describe someone as insane if we fail to identify a tendency that aligns with their patterns of behavior, thought, and feelings (263). So, for example, we would consider someone who has been late for every single date in life as an unpunctual person, and if such a person who understands the meaning of punctuality takes themselves to be punctual, we would consider them as insane. In this sense, an interpretation from the third-person point of view will limit the individual's freedom of self-interpretation, in the sense that it sets the "situation" in which the agent exercises their freedom. The notion of willingness, again, can help us better understand this idea of situational freedom. To take the above example, I can freely interpret myself as punctual, i.e., as willing to act punctually. However, the fact that I have never acted punctually puts me in a situation in which I should exercise this freedom. This situation may be my past record of unpunctuality, or be my incapability of acting punctually due to the neurodiverse nature of my brain. In Sartre's own terminologies, the situation conditions the self-conscious being, and their situational freedom is "the small movement which makes of a totally conditioned social being someone who does not render back completely what his conditioning has given him" (Sartre 1969: 45).

To sum up, in this subsection, I represented and interpreted Sartre's account that discovering one's character traits is inseparable from deliberative reflections on one's own character. This account implies that one's character gets specified or renewed every time the individual reflects upon it. We can read this as allowing a form of instantaneous (and probably unobservable) character change. Of course, we can raise further questions concerning this account. First, we can compare Sartre's more general theory of self-consciousness with other existing accounts of self-knowledge, where self-knowledge means knowledge about one's own standing mental states such as beliefs and desires. In particular, we can compare Sartre's view with different "transparency" approaches to self-knowledge, according to which self-knowledge is obtained by "looking through" the (transparent) mental state directly to the aspect of the world it represents (see, e.g., Davidson 1987, Moran 2001, Byrne 2018, etc.; for a general survey, see Gertler 2024). In particular, it would be interesting to see which approach would be most advantageous in providing a uniform analysis of knowledge about one's own standing mental states, including character traits as dispositional states. The second question asks to what extent the agent is free in balancing the weight of empirical evidence such as observations or reports of one's regular or frequent behavior and deliberative reflections, i.e., their facticity and transcendence as a self-conscious being in self-enquiry. [9] Alvarez (2015) has discussed interesting cases in which one's inclinations diverge from their actual behavior, so that the latter is suppressed by oneself to a certain degree. Alvarez takes courage for example: in Aristotle and his followers' view, a person who feels fear in the face of danger but tries their best not to run away counts as courageous. Alvarez did not discuss this kind of higher-order freedom in self-enquiry (nor many other philosophers in the field, to my knowledge), but it seems to me that further work needs to be done in this direction.

[9] Crane and Farkas (2022) recently recognized both sources as complementary means of attaining self-knowledge, particularly concerning typical standing states like beliefs and desires. They stress that the agent has the freedom to balance potentially incompatible empirical evidence with deliberative choice during this process.

3.2 THE BOTTOM-UP APPROACH

People do not always consciously reflect on their own character. In many cases (e.g., possibly in the Félicie case), one's character changes without any form of reflection on character. In this subsection, I explore a bottom-up approach to character change, which centers on the notion of salience. I will first explain the notion of salience as in recent usages, illustrate it by examples, and then connect it with character change.

Salience is construed subjectively (as opposed to objectively or intersubjectively), meaning that something's salience is relative to an individual (Archer 2022b: 114). Moreover, it is an occurrent, personal-level phenomenon involving conscious awareness in either perceptual experience or thought (114, 118). For our purposes, we work with Archer's minimal account of salience (also see Watzl 2017, 2022). According to Archer, "for something to be salient to you is for it to 'stand out' to you" (115). In this account, the notion of salience presupposes the notion of attention.

> ***The Minimal Account*** something's salience to you is its commanding your attention.

To illustrate the idea of "commanding," let us consider the following scenarios of salience (for more similar examples, see Archer 2022b and Watzl 2017: 135):

> ***Interview*** Imagine you are at a job interview for your dream position. The interviewer walks in and is extremely charismatic and charming. Throughout the interview, you are so captivated by their charisma and charm that you fail to notice their inappropriate questions and comments. They frequently interrupt you, ask questions that seem irrelevant or too personal, and make several condescending remarks. Despite these red flags, you remain focused on their charm, interpreting their behavior as confidence and

assertiveness. It’s not that you notice their inappropriate behavior and choose to ignore it; their charm simply overshadows everything else.

Keynote Talk Imagine you are attending a conference on climate change, listening to a keynote presentation by a renowned scientist whose work you greatly admire. Their research aligns perfectly with your views on the urgency of addressing climate change. During the presentation, they make a compelling argument but include a significant factual error. You are so engrossed in their passionate delivery and so aligned with their overall message that you completely miss the mistake. It's not that you notice the error and choose to overlook it; you are so focused on the larger truth of their message that the error goes entirely unnoticed by you.

Family Reunion Imagine you are at a family reunion, and you have an important announcement to make about your recent engagement. You've been looking forward to sharing this news with everyone for weeks. During the gathering, you decide that now is the perfect time to make your announcement. You stand up and start speaking excitedly about your engagement, completely oblivious to the fact that a close relative has just shared some distressing news about a serious health issue. Despite the somber mood in the room and the concerned faces of your family members, you continue with your announcement, absorbed in your own joy and excitement. It's not that you notice the family's mood and choose to overlook it; you are so focused on your happy news that you fail to recognize the emotional atmosphere around you.

Interview focuses more on perceptual experience, whereas ***Keynote Talk*** and ***Family Reunion*** focus more on one’s cognitive and affective attitudes. In those cases, you have a degree of control (whether voluntary or involuntary, direct or indirect) over what is salient to you (Archer 2022b: §6.3). You could have attended to what is not salient to you, or withhold your attention

from what is salient to you (Mole 2022: 140). This is the sense in which we say that salience commands your attention. However, what is salient to you sometimes seems to be not commanding your attention but enforcing it (Archer 2022b: 126). To take Archer's example, someone suffering from obsessive-compulsive disorder (OCD) can find it "maximally salient" that they have not washed their hands in the last half hour. The experience of salience can persist even though they understand that it is not important, in fact, far less important than listening to their child tell them about their day at school at that moment. Despite this, they cannot attend to their child's stories (121). Archer points out that in a pathological example like this one, it seems that the mental health patient could not have withhold their attention from what is salient to them at all. It is akin to a case in which you find it so difficult to not attend to the fire alarm, even if you understanding that it is only a part of the system test and does not matter in that moment. In such a case, what is salient to you is "biomechanical," in that attending to the fire alarm, like the behavior of a lioness defending her cubs, is an instinctual reaction (125–6). In both of these two cases, the agent experiences something as *demanding* attention without experiencing it as *deserving* attention (see Siegel 2014).

There is an important difference between the first three cases and the case of OCD and fire alarm that is relevant to our discussion of character. Namely, it is in the former kind of cases that what is salient to you can reflect what Archer calls your "standing evaluative worldview about what matter in general" (120; also see Smith 2005: 245). That you find the interviewer saliently charming may tell us something about your type. The fact that you fail to notice their inappropriate questions could even suggest that you are a lookist—the kind of person to whom others' physical look is maximally salient—without ever being aware that you are. In the same vein, the fact that you fail to notice the fallacy of the presenter's argument might tell us you are more a sympathetic person than a logical one. But these are mere possibilities. You may just have acted out of character for good reasons. Notice that we do not have to equate character

with one's standing evaluative worldview as Archer does. It is sufficient for us to accept that the two aspects of one's mind are associated in an important way (see also Crane and Farkas (2022) for another view of the connection between one's character and "worldview").

The association between salience and character goes in two directions: the first is that salience can reflect character, and the second is that salience can influence character. In what follows, we can follow Katsafanas (2016, also see his 2011) in focusing on how salience might *confirm* and *reinforce* the corresponding character traits, assuming that our theory will equally apply when salience *disconfirms* or *attenuates* certain character traits. To take Katsafanas' example:

> The irascible and sanguine persons both see the faces of the audience, but the former finds frowns and grimaces salient and interprets them as marks of hostility and disdain, whereas the latter finds nods and thoughtful expressions salient and interprets them as approval. And these appearances confirm the very traits that generate them, thereby reinforcing the traits. (2016: 138)

In the following, I will examine the two directions in turn. Importantly, we can anticipate that the direction that salience can impact our character is crucial to our exploration of the possibility of character change.

The aspect that salience is representative of one's character has been emphasized in moral psychology for decades. As Chappell (2022: 138) recently asserts, various theories of moral perception and practical wisdom reveal the psychological mechanisms that make the right choices salient to the person of good character. For instance, McDowell has written on kindness, not in terms of salience or attention but in terms of sensitivity:

> A kind person has a reliable sensitivity to a certain sort of requirement that situations impose on behavior. The deliverances of a reliable sensitivity are cases of knowledge, and there are idioms according to which the sensitivity itself can appropriately be described as knowledge: a kind person knows what it is like to be confronted with a requirement of kindness. The sensitivity is, we might say, a sort of perceptual capacity. (McDowell 1998, 51)

McDowell's idea of the requirement of kindness assimilates Siegel's (2014: 56–7) and Cavedon-Taylor's (2022: 21) idea of a "normative constraint" on acting in a specific way as posed by an experience of salience. For example, stereotypically speaking, an extravert will experience a party at which they know nobody as a solicitation with accompanying motivational elements, talking it as an opportunity to make new friends. By contrast, an introvert will not feel the same solicitation in the same situation. They may feel the social norms to chat with people, however, they do not feel an internal solicitation like the extravert does. In other words, the extravert will feel for themselves that talking to strangers is *the thing to be done*, whereas the introvert will feel that talking to strangers is *the thing that one should done*.

Archer generalizes the idea beyond the moral realm when she proposes the following account of salience, which associates one's character with a "standing evaluative worldview about what matter in general" (also see Smith 2005: 245):

> ***The Sophisticated Constitutive View*** something's salience to you is constituted by your occurrent evaluation that it matters in the context of the situation you are in, which in turn emerges from your standing evaluative worldview about what matters in general (Archer, §6.7).

Archer argues that at least in the cases we are interested in (such as the first three cases we

considered above), something's salience must be connected with a particular instance of occurrent evaluation that it matters to *you* in *that* context and, moreover, with standing evaluation that it matters in general.

The other direction that the association between salience and character goes is that salience is not simply an upshot of one's character, but can conversely shape it (Archer 2022b: 123). Unfortunately, it has received relatively less attention so far than the first direction we discussed. However, this direction can be derived from the first. I find either of the two virtue-theoretically inspired views discussed by Watzl's (2022: 91) useful in this aspect. The stronger view says that salience patterns partially constitute one's character, whereas the weaker view says that good character is not constitutive of some patterns of salience but nevertheless requires having the right motivational susceptibilities (see McDowell's quote above for an illustration). Both views contend that possessing a certain salience pattern is necessary for maintaining a certain character trait. It follows that, without possessing the necessary pattern of salience, one would no longer be able to maintain a certain character trait. It is in this sense we say that salience can influence character. Notice that the impact of salience over character does not imply that any variation in salience will bring about variations in character.

More importantly, this second direction is crucial to our enquiry, as it suggests an answer to the question concerning the possibility of character change. As Watzl (2022) points out, the salience patterns associated with meditation and mindfulness might be instrumental in cultivating the moral virtue of compassion in an individual (91). Conversely, reducing experiences of certain saliences can also prevent reinforcing certain character traits. There are many ways of altering one's salience patterns. First, when the situation remains the same, we can direct the individual's attention toward or away from certain aspects of a situation. For instance, we often attempt to cultivate a character trait in a child (say, optimism) by training

them to attend to the aspects of a situation which they can make efforts to improve. Alternatively, an individual can practice attending to certain aspects of the phenomenon to bring about intended changes. Second, changes in the external environment might reduce chances of experiencing certain saliences, thereby preventing reinforcing certain character traits. It should be emphasized that, usually, changes in the external environment can bring about changes in the agent's own physical and mental states. To take Félicie's increased neuroticism for example, we can plausibly imagine that certain negative life events have resulted in mental and physical health deteriorations, thereby influencing her perspectives of the same kinds of situation.

This brings us back to the second question that we put forward in Introduction, concerning the role of the agent in character change. Watzl (2022: 97) describes salience as a "passive force," which is not under one's direct voluntary control as raising one's own arm is (also see Watzl 2017: 135–7). Nevertheless, when the agent practices attending to what is previously not salient to them, they are exercising a form of indirect voluntary control over what is salient to them (Archer 2022b, Hieronymi 2009; for psychological evidence, see Mole 2022: 151–7). This assumes that you initially assess what should be important to you in a particular type of situation and recognize that repeatedly attending to something can increase its salience. By deciding to direct your attention to a specific aspect of a situation, you can immediately alter what is salient to you. In this manner, your character, which is based on your patterns of salience, may change. From this perspective, the bottom-up approach, like the top-down approach, allows for agential freedom.

In this subsection, I present some preliminaries about the notion of salience, attention, and their connections to character. I suggest that the individual's character might change if their salience patterns changes, and the latter might change if their attention patterns changes as

intended or the external environment changes, in a way that brings about changes in their physical or mental states.

CONCLUSION

It is time to take stock. In this thesis, I addressed the question of what makes character change possible. While character change is an ordinary phenomenon, the question has barely received any attention in philosophy (Alvarez 2015 does raise this question without discussing it much). More generally, the philosophy of character has not received enough attention. However, it is an important aspect of our mental life and contributes to self-understanding and mutual understanding in interpersonal relationships. Moreover, the empirical study of character (or personality) faces many difficulties, such as the difficulty of conducting long-term studies of character change (see Kupperman (1995, Appendix A) for relevant discussions). This makes the philosophical study of character valuable for contributing its unique insights and perspectives. I hope this thesis, with its selected perspective, will contribute to the development of the philosophy of character. Furthermore, I hope the development of the philosophy of character will contribute to the broader picture of philosophy of mind and action, metaphysics and ethics. Many further interesting questions awaits exploration. Just to mention one possibility: one may be interested in the relationship between one's self and one's character, and, in this connection, the relationship between self-change and character change.

My project has developed as follows. In Chapter 1, I mapped out the preliminaries for our enquiry. In Section 1.1, I analyzed character in terms of character traits, which I conceived of as multi-track dispositions. In Section 1.2, I emphasized several aspects to character trait dispositions that should differentiate them from other kinds of dispositions. First, for the sake of arguments, I proposed to treat character traits as dispositions that ontologically depend on previous manifestations. However, I denied that they have to be manifested relatively frequently or regularly before. Second, following Schwitzgebel (2002), I acknowledged a kind of normativity to character traits, called social accountability, according to which we not only tend

to understand others as acting in character but also try to make sense of ourselves by acting in character. Finally, I maintained that character traits come in degrees, but left it open whether they (or some of them) have minimal or maximal degrees, which shall not concern us here. In Section 1.3, I introduced the debate over situationism in philosophy. My concern was not with the existence of robust character traits and hence the legitimacy of virtue theory, but rather with whether situationism challenges the existence of substantive character change. I argued that it does not.

In Chapter 2, I conceived of character change as the strengthening or weaking of specific character traits. In Section 2.1, I discussed which ontological category character change belongs. I claimed that there is instantaneous character change as well as continuous character change, both of which is open-ended. In Section 2.2, I distinguished the phenomenon of character change from three other phenomena that might be mistaken for character change: that is, changes in mental health condition, personality mirroring, and changes in trait attributions merely due to shifts in the observer's perspective. These phenomena are themselves not character change, but may give rise to character change under appropriate circumstances.

In Chapter 3, I proposed two approaches to the main question concerning the possibility of character change. In Section 3.1, I proposed the top-down approach, which is built on my own interpretation of Sartre's theory of self-consciousness within the dispositional framework of character traits. It claims that when one attempts to discover one's character traits, one is in a position to choose their character traits. This makes instantaneous character change possible, as one's character gets specified or renewed whenever they reflect on it. At the end of the subsection, I raised two further questions that awaits future investigations. The first asks about the similarities and differences between Sartre's theory of self-consciousness and other transparency approached to self-knowledge. In particular, it would be interesting to see which

approach would be most advantageous in providing a uniform analysis of knowledge about one's own standing mental states, including character traits as dispositional states. The second question inquiries about the extent of agential freedom involved in balancing empirical evidence and deliberative reflections in self-enquiry. In Section 3.2, I proposed the bottom-up approach, which draws on the idea of salience. Roughly, salience is what demands the subject's attention, and one can attend to what is not salient and withdraw attention from what is salient. It claims that there is a bi-directional relationship between salience and character: namely, salience can reflect character, and it can also influence character. The latter direction can be accounted for by the modal claim that maintaining a given character trait requires possessing a certain pattern of salience. Because, without possessing the necessary pattern of salience, one would no longer be able to maintain a certain character trait. Notice that the impact of salience over character is not such that any variation in salience will bring about variations in character.

REFERENCES

Alston, W. (1970). Toward a Logical Geography of Personality: Traits and Deeper Lying Personality Characteristics. In *Mind, Science, and History*. Ed. H. Kiefer and M. Munitz. Albany: State University of New York Press, 59–92.

Alvarez, M. (2015). Ryle on Motives and Dispositions. In *Ryle on Mind and Language*. Edited by David Dolby, 74–96. Basingstoke: Palgrave Macmillan.

Alvarez, M. (2017). Are Character Traits Dispositions? *Royal Institute of Philosophy Supplement* 80:69–86.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* [DSM-5], 5th edition, Arlington, VA: American Psychiatric Association.

Anjum, R. L. and Mumford, S. (2018). *What Tends to Be: The Philosophy of Dispositional Modality*. Routledge.

Archer, S. (ed.). (2022a). *Salience: A Philosophical Inquiry*. Routledge.

Archer, S. (2022b). Salience and What Matters. In Archer (2022a), pp. 113–129.

Athanassoulis, N. (2000). A Response to Harman: Virtue Ethics and Character Traits. *Proceedings of the Aristotelian Society* 100: 215–21.

Badhwar, N. (2009). The Milgram Experiments, Learned Helplessness, and Character Traits. *The Journal of Ethics* 13: 257–89. 10.1007/s10892-009-9052-4

Bird, A. (2007). *Nature's Metaphysics*. Oxford: Oxford University Press.

Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.

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

Byrne, A. (2018). *Transparency and Self-Knowledge*. Oxford University Press.

Callard, A. (2018). *Aspiration: The Agency of Becoming*. New York: Oup Usa.

Cavedon-Taylor, D. (2022). Life Through a Lens: Aesthetic Virtue and Salience vs Kantian Disinterest. In S. Archer (ed.) (2022a), pp. 10–23.

Chappell, S. G. (2022). Salience, Choice, and Vulnerability. In S. Archer (ed.) (2022a), pp. 130–139.

Chartrand, T. L. & Bargh, J. A. (1999). The Chameleon Effect: The Perception–Behavior Link and Social Interaction. *Journal of personality and social psychology*, *76*(6), 893–910. doi:10.1037/0022-3514.76.6.893.

Crane, T. & Farkas, K. (2022). Mental Fact and Mental Fiction. In T. Demeter, T. Parent & A. Toon (eds.), *Mental Fictionalism: Philosophical Explorations*. New York & London: Routledge. pp. 303–319.

Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association* 60: 441–58. Reprinted in his (2001). *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.

Doris, J. (1998). Persons, Situations, and Virtue Ethics. *Noûs* 32: 504–30.

Doris, J. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Doris, J. (2010). Heated Agreement: *Lack of Character* as Being for the Good. *Philosophical Studies* 148: 135–46.

Fileva, I. (ed.) (2016). *Questions of Character*. New York, US: Oxford University Press USA.

Frankfurt, H. (1988). Identification and Wholeheartedness. In his *The Importance of What We Care About*. Cambridge: Cambridge University Press, pp. 159–176.

Frierson, P. R. (2006). Character and Evil in Kant's Moral Anthropology. *Journal of the History of Philosophy*, *44*(4), 623–634.

Gertler, B. (2024). Self-Knowledge. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2024/entries/self-knowledge/.
Gide, A. (1924[1909]). *Strait is the Gate* (*La Porte Étroite*), D. Bussy (trans.). New York: Knopf.
Goldie, P. (2004). *On Personality*. Routledge.
Hampshire, S. (1953). Dispositions. *Analysis* 14: 5–11. 10.1093/analys/14.1.5
Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society* 99: 315–31.
Harman, G. (2000). The Nonexistence of Character Traits. *Proceedings of the Aristotelian Society* 100: 223–6.
Harman, G. (2001). Virtue Ethics without Character Traits. In *Fact and Value*, A. Byrne, R. Stalnaker, and R. Wedgwood (eds). Cambridge: MIT Press, 117–127.
Harman, G. (2003). No Character or Personality. *Business Ethics Quarterly* 13: 87–94. 10.5840/beq20031316
Harman, G. (2009). Skepticism about Character Traits. *The Journal of Ethics* 13: 235–42.
Haslanger, S. (2018). What is a Social Practice? *Royal Institute of Philosophy Supplement*. 82:231-247. doi:10.1017/S1358246118000085
Hayes, T. L., Hogan, R., and Emler, N. (2016). The Psychology of Character, Reputation, and Gossip, in I. Fileva (ed.), (2016), pp. 268–282.
Hieronymi, P. (2009). Two Kinds of Agency, in L. O'Brien, and M. Soteriou (eds), *Mental Actions*, Oxford University Press.
Hüttemann, A. (2013). A Disposition-based Process Theory of Causation. In Stephen Mumford and Matt Tugby (eds.), *Metaphysics and Science*. Oxford: Oxford University Press, pp. 101–22.
Kant, I. (1978) *Anthropology from a Pragmatic Point of View*. Translated by Victor Lyle Dowdell. Carbondale: Southern Illinois University Press.
Katsafanas, P. (2011). Activity and Passivity in Reflective Agency, in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 6*, Oxford University Press.
Katsafanas, P. (2016). Autonomy, Character, and Self-Understanding, in I. Fileva (ed.) (2016).
Kupperman, J. (1995). *Character*. New York: Oxford Univeristy Press.
Lyons, W. (1973). Ryle and Dispositions. *Philosophical Studies*, 24, (5): 326–34.
Manley, D. and Wasserman, R. (2007). A Gradable Approach to Dispositions. *The Philosophical Quarterly* 57:68–75.
Manley, D. and Wasserman, R. (2008). On Linking Dispositions and Conditionals. *Mind* 117:59–84.
McDowell, J. (1998). Virtue and Reason. In *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, pp. 50–76.
McGeer, V. (1996). Is "Self-Knowledge" An Empirical Problem? Renegotiating the Space of Philosophical Explanation. *The Journal of Philosophy, 93*(10), 483–515. https://doi.org/10.2307/2940837
Miller, C. (2013). *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.
Miller, C. (2014). *Character and Moral Psychology*. Oxford University Press.
Miller, C. (2023). Empirical Approaches to Moral Character. In E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2023/entries/moral-character-empirical/.
Miller, C. and Knobel, A. (2015). Some Foundational Questions in Philosophy about Character. In Christian B. Miller et al (eds.), *Character: New Directions from Philosophy, Psychology, and Theology*. Oxford University Press. pp. 19–40.
Mole, C. (2022). The Moral Psychology of Salience. In S. Archer (ed.) (2022a), pp. 140–158.

Moran, R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
Mourelatos, A. P. D. (1978). Events, Processes and States, *Linguistics and Philosophy* 2:415–34.
Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.
Murdoch, I. (1997). *Existentialists and Mystics: Writings on Philosophy and Literature*. New York: Penguin.
O'Shaughnessy, B. (2000). *Consciousness and the World*. Oxford: Oxford University Press.
Rothstein, S. (2004). *Structuring Events: A Study in the Semantics of Lexical Aspect* (Explorations in Semantics 2). Malden, MA & Oxford: Blackwell.
Russell, D. (2009). *Practical Intelligence and the Virtues*. Oxford: Clarendon Press.
Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
Sartre, J. P. (2003[1956]). *Being and Nothingness: An Essay on Phenomenological Ontology*, Hazel E. Barnes (trans.), New York: Philosophical Library.
Sartre, J. P. (1969). Itinerary of a Thought. *New Left Review*, (58), 43.
Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Noûs*, 36: 249–275. https://doi.org/10.1111/1468-0068.00370
Shaw, L. (1945) *Rickshaw Boy* (luo tuo xiang zi). Trans. by Evan King. New York: Reynal & Hitchcock.
Siegel, S. (2014). Affordances and the Contents of Perception. In B. Brogaard (Ed.), *Does Perception Have Content*? Oxford University Press, pp. 1–28.
Smith, A. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics* 115: 236–271.
Upton, C. (2009). *Situational Traits of Character*. Lanham: Rowman & Littlefield.
Vendler, Z. (1957). Verbs and Times. *Philosophical Review* LXVI: 143–60. Reprinted in a revised version in Vendler (1967), *Linguistics in Philosophy*. Ithaca, NY: Cornell.
Vetter, B. (2015). *Potentiality: From Dispositions to Modality*. Oxford Univeristy Press.
Watzl, S. (2017). *Structuring Mind: The Nature of Attention and How It Shapes Consciousness*. Oxford: Oxford University Press.
Watzl, S. (2022). The Ethics of Attention: An Argument and a Framework. In Archer (ed.) (2022a), pp. 89–112.
Williams, B. (1985). *Ethics and the Limits of Philosophy*. Harvard University Press.
Wood, A. W. (1999). *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
Wright, A. (1988). Dispositions, Anti-realism and Empiricism. *Proceedings of the Aristotelian Society* 91: 39–59.