

Reason and Caring about in Akratic Action

Ying Xue

Student Number: 16146696

UCL

Mphil Stud

I, Ying Xue, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

I shall focus on the following question: what is the relation between agents and their akratic action, in some complicated cases? I shall compare three routes: (1) akratic agents cannot be active, (2) akratic agents can act for good reasons, and (3) agents might care about their akratic action more than their better judgment. In Chapter 1, I shall explain the concept of akrasia and its irrational factors. In addition, I shall offer some cases to show the difficulties of understanding akrasia and the Irrational-passive Proposal's shortcomings.

In Chapter 2, I shall analyse Nomy Arpaly's (2003) view on "inverse akrasia", taking Huckleberry Finn's case as an example, and argue that: it is plausible to say that deliberation is not necessary for responding to reasons. However, this Reason-responsive Proposal might fail to capture some interesting features of Huckleberry Finn's case, especially the lack of the first-person justification. More details taken from the novel might show us an alternate interpretation.

To set the stage for the alternative, in Chapter 3, I shall focus on two concepts offered by Harry Frankfurt (1988, 1999): caring about and volitional necessity. If their relative independence of desires and reason-responsiveness can be defended, this leaves room for a discrepancy between caring about and reason-responsiveness, which can be reduced to neither one between desires and reason-responsiveness, nor one within reason-responsiveness.

In Chapter 4, I shall offer a proposal based on Chapter 3, and suggest that in some cases, agents act akratically because they do not care about their better judgments as much as they care about what their akratic action benefits. Then I shall use this Out-of-caring-about Proposal to interpret several cases to show its merits and features.

In Chapter 5, I shall reply to some objections and mention one implication that the Out-of-caring-about Proposal might have.

Impact Statement

This thesis presents a new explanation of a kind of akratic action based on the concept of “caring about”. Therefore, this work might contribute to our understanding of human action and practical irrationality, especially the active expression of human agency in irrational action. Moreover, it draws our attention to the complexity of the human mind.

Contents

Acknowledgements	7
Chapter 1 What Akrasia is and the Irrational-passive Proposal	8
1. Some Clarification of the Definition of Akrasia.....	8
2. A Platonic Picture and the Irrational-passive Proposal.....	9
3. What is Wrong with Akrasia? – “Irrationality”.....	13
4. What is Wrong with Akrasia? – “A Divided Self”.....	18
5. Two Cases	19
6. Conclusion.....	22
Chapter 2 Inverse Akrasia and the Reason-responsive Proposal	24
1. Nomy Arpaly’s Analysis of Huck Finn’s Case and the Reason-responsive Proposal	24
2. A Challenge: What Can Count as Responding to or Acting for a Reason	27
3. An Alternate Proposal (First Impression).....	32
4. Acting out of Caring about O	36
5. Conclusion.....	39
Chapter 3 Caring about, Volitional Necessity and Discrepancies	40
1. The First Concept: Caring about	40
2. The Second Concept: Volitional Necessity.....	49
3. Conclusion: Two Discrepancies	52

Chapter 4 The Out-of-caring-about Proposal	54
1. The Framework of the Out-of-caring-about Proposal	54
2. Three Features of the Out-of-caring-about Proposal.....	63
3. Circumstances.....	71
4. Conclusion.....	75
Chapter 5 Objections, Replies and an Implication.....	77
1. An Objection from Reason-responsiveness	77
2. An Objection from the Picture of a Divided Self.....	81
3. An Implication on Self-understanding.....	82
4. Conclusion.....	87
Bibliography	89

Acknowledgements

I would like to thank Ulrike Heuer for her patient guidance and helpful comments. I would like to thank Sebastian Gardner, Lucy O'Brien and Douglas Lavin for their fabulous suggestions and Richard Edwards for his timely assistance. I also would like to thank my first philosophy teacher Liping Wang and my undergraduate supervisor Xiangdong Xu, who drew my attention to the topic of "akrasia". I am grateful to my Mphil peers, especially Diana Craciun, Hongyi Li and Leo Serafini, whose friendship is indispensable, and to my long-time friends Shuyuan Liang and Danfang Li.

I worked on this thesis at home in Linyi, China, so I thank my parents and sisters for offering me a quiet environment. Lastly, I would like to thank my bass named "Axe" and my favourite entertainer Bo Yuan. The song he covers is set as an alarm by me. It always succeeds in waking me up. Whenever I hear it, I feel full of energy and hope (even at 5 a.m.). The following is part of the lyrics:

"After an endless dream, in this world of nothingness,
It seems as if our beloved dreams will lose.
However, even with these unreliable wings,
I'm sure we can fly, on my love."

– Butter-fly

Chapter 1 What Akrasia is and the Irrational-passive Proposal

Despite the difficulties of finding a fully satisfying explanation, akrasia is commonly accepted as a genuine possibility in our daily lives. We can act against our better judgment without compulsion. For example, even if I believe that playing on my phone before falling asleep is not good, and it would be much better if I could leave my phone in the living room, I, nonetheless, play on my phone for about one hour lying on my bed every night. I am not compelled to do that, nor do I change my judgment that I ought not to play on the phone. In this chapter, as a beginning, I shall clarify what akrasia means in this thesis and explain why it is significant and puzzling from the perspective of irrationality and that of the divided self. Several examples would be offered and they might motivate us to use different methods to interpret them.

1. Some Clarification of the Definition of Akrasia

Several features of akrasia make it special and puzzling. Firstly, an agent's better judgment, which is violated by her akratic action, is all-things-considered. It means that the akratic agent has taken all the relevant factors she notices into account¹. Some of them are relevant to morality, while others might not. Therefore, a better judgment merely based on moral considerations might be overridden by other factors (e.g. aesthetic ones), and then the agent, who has an all-things-considered better judgment and acts according to it, is not akratic, even if she acts against the better judgment based on moral considerations. As Donald Davidson says, akrasia (what he calls "incontinence") cannot be reduced "to such special cases as being overcome by the beast in us, or of failing to heed the call of duty, or of succumbing to temptation". (2001, p. 34)

Secondly, if an agent with her better judgment changes it when facing

¹ In this sentence, some words and phrases (e.g. "all the relevant factors" "notices" "takes...into account") are ambiguous, and I do not think I can spell out them at this stage. However, this does not undermine the point highlighted here: akrasia is not a local failure like acting against a better judgment based on a certain subclass of reasons (e.g. moral reasons).

stronger temptation or realising new relevant information which tips the balance of reasons, then her acting against the prior judgment cannot be classified as akrasia. This is because her action aligns with her here-and-now better judgment, which has overridden the prior one. Rather, it might be called “the weakness of will” (Holton, 2009)², which will not be discussed here.

Thirdly, akrasia is distinguished from compulsion. We normally think that akratic agents are not subject to physical disabilities or mental disorder. They are expected to be able to resist what motivates their akratic actions in a certain way. Moreover, when their akratic actions lead to some terrible results, “I acted akratically” might not be regarded as an excuse as justified as “I could not do otherwise because my body malfunctioned / was manipulated” or “I could not do otherwise because I suffered from a certain mental disease”.

Therefore, when an agent acts against her all-things-considered and here-and-now better judgment without compulsion, her action is akratic. However, although we can pin down the phenomenon in this way, it is hard to make sense of how it happens and what it can tell us about ourselves. I shall start with a Platonic picture to accommodate akrasia.

2. A Platonic Picture and the Irrational-passive Proposal

Given this definition, we might be led to a Platonic picture naturally. That is to say, our Reason gives us a rational instruction in the form of a better judgment, while our rebellious Desire disobeys it. Or, in a modern version, we can say that our valuational system offers us a better judgment, which is overlooked or overwhelmed by our blind motivational system. By virtue of the valuational system, an agent with some factual beliefs gets judgments in this form: “the thing for me to do in these circumstances, all things considered, is a”. (Watson, 1975, p. 215) By contrast, the motivational system of an agent is the set of considerations finally

² A challenge to the distinction between “akrasia” and “the weakness of will” made by Holton can be seen in Alfred Mele, 2012, pp. 18-30.

moving her to act.

Apparently, this picture can help us to make sense of some features of akrasia. For one thing, with two relatively independent systems, it accommodates the discrepancy between one's better judgment and action and thus the possibility of akrasia. Our valuational system can make an all-things-considered better judgment and retain it even though a contrary action stems from the motivational system at the same time. For another, akratic actions resulting from the discrepancy can be distinguished from actions resulting from an external force. The former is a product of the discrepancy *within* an agent while the latter is caused by a discrepancy between an external force *and* the agent.

However, this neat picture faces some problems. If we accept this picture, akrasia cannot be distinguished from subtler psychologically compulsive cases. What a person addicted to drugs faces would be a similar discrepancy between her motivational system and valuational system, although her desire for drugs is strengthened to the extent that it functions as if it were an external irresistible force. Although we might hope that the addict could have a will that is strong enough to resist the desire for drugs, we have to admit that the desire is so strong that this is not a case of akrasia.

Then an objection offers itself: why should we bother to find intrinsic differences between the two cases? Can we, as Gary Watson (1977) suggests, claim that akratic actions and compulsive actions are inherently indistinguishable? That is to say, only if we apply a social standard (i.e. "normal" capacities of self-control) to them can we classify them as akrasia and compulsion respectively. Let us call this suggestion the "Irrational-passive Proposal". Its basic idea is that akratic action, like compulsive action, is rooted in the irrational part of human agents. Since this part is not sensitive to an agent's rational control, it, as well as its manifestation in action, is necessarily passive.

However, several points can be made to motivate us to defend the akrasia-compulsion distinction instead of Watson's approach and the Irrational-passive Proposal.

Firstly, the distinction is of first-, second- and third-person significance. We can zoom in on the case of the addict. When she, a teenager, tried drugs offered by her friends for the first time, she might be fully aware that it would be better to reject it all-things-considered (i.e. she ought to reject it.). However, out of curiosity, she took the drugs and changed to an addict gradually. We cannot find the exact cut-off point between the previous akrasia (i.e. she believed she ought not to take the drug again, but she surrendered to her desire akratically) and the current addiction (i.e. now she believes she ought not to take the drug again, but the desire just forces her to take the drug brutally, and she is a hopeless bystander during the process).

Nonetheless, as she became more and more obsessed with drugs, her first-person experience changed. When she was tempted to take her first drug, she did not think she was forced to do so. (Suppose that her friends did not threaten her.) After that night, when she reflected on what she did, she would think that she was weak-willed. "The desire for the drug was not weak," she might think, "but it was not irresistible either." By contrast, when she has become an addict, she might feel she takes the drug pathologically in the sense that she loses control of herself when facing unbearable suffering of the body on the one hand and alleviating it by taking drugs on the other. If she blames herself, she would not blame herself for this particular failure. Instead, she blames herself for her doing nothing to reverse the degradation when it has not been irreversible. In addition, if she wants to change the situation now, she would use some external forces or ask for help from others. Therefore, the akratic agent blames herself and wishes for a stronger will, while the compulsive agent blames what generates the force she faces and resorts to another external force hopefully beating the existing one.

The distinction between compulsion and akrasia manifests in the second-person and the third-person perspectives as well. The more obsessed with drugs the addict is, the less necessary other people feel to directly encourage or motivate her to resist the desire. Instead, they might find only medical treatment

is feasible. In addition, suppose you are her friend (not the one tempting her to try drugs). You might try to tell her what she ought to do and encourage her to stop taking drugs. However, you would not be surprised if you see that she is aware of what she ought to do, but she cannot bring herself to do it. If you put yourself in her shoes, you might concede that: if you were her, you could not resist the desire for the drugs as well. Additionally, you might, on second thoughts, tell yourself: "Anyway, I would not put myself in such a tricky situation in the first place. I have a strong enough will to resist the drugs when I have not become an addict." Therefore, from the second-person and the third-person perspectives, whether the agent acts akratically or compulsively does matter³.

It might be argued that all of these observations (i.e. the difference made by the akrasia-compulsion distinction) are mere illusions, and they can be explained without appealing to the distinction. For example, we make this apparent distinction merely for practical usage, or we have different reactive attitudes towards seeming compulsion and seeming akrasia, which is shaped by conventions. Nonetheless, it seems plausible to be acute when we face these complex phenomena. This enables us to capture the information about human agency that these phenomena convey to us. Here, I have a goal similar to Nomy Arpaly's:

"... I hope to achieve at least one very significant goal, and that is to expand the domain of moral-psychological inquiry, enriching our philosophical discussion of human beings in all their complexity." (2003, p. 8)

Secondly, besides some interesting phenomenological features of akrasia, the bunch of theoretical puzzles involved in akrasia are different from that in compulsion as well. Facing a case of compulsion, we might assume a mechanistic picture of the mental states of the agent with the agent's effort on the one hand and the force from some external origins on the other. Then we do not treat the agent as a genuine agent in the sense that the picture we draw for her

³ A similar attempt to "save the phenomenon" can be found in Thomas Hill (1986). He thinks from the first-person perspective, an akratic agent (say, Amy) does *not* see herself as passive as an addict, and "except for those who draw contrary conclusions from philosophical determinism or particular psychological theories, most other people who know Amy share her view of her weakness." (Ibid., p. 97)

is close to that for a manipulated spider or a puppet. By contrast, when exploring a case of akrasia, we might think we are wading into the realm of irrationality and tackling some questions exclusive to human agency. We wonder how an agent can act against her own better judgment, and whether she could respond to reasons which have been overridden in her deliberation. Therefore, it would be fruitful if we zoom in on cases of akrasia and keep alarmed when a theory leads to the impossibility of an inherent akrasia-compulsion distinction.

To sum up, the Irrational-passive Proposal claims that: akrasia stems from the discrepancy between two relatively independent systems, and the fact that some cases of compulsion share the same origin with akrasia leads us to a kind of scepticism about akrasia (i.e. akrasia and compulsion are not inherently distinguishable). I have argued that this approach is not plausible because of the significance of the akrasia-compulsion distinction in our practices and understanding.

However, even if we set Watson's rationalistic picture and the Irrational-passive Proposal aside, we might find it hard to get rid of some core elements (i.e. irrationality and the divided self) of this picture in our attempt to make sense of akrasia. I shall focus on them in the next two sections.

3. What is Wrong with Akrasia? - "Irrationality"

Normally, akrasia is regarded as a typical case of irrationality. What is irrational in akratic action then? There are two distinct ways to understand rationality. Accordingly, I shall offer two different explanations: First, akrasia implies a salient incoherence of an agent's mental states; Second, the agent fails to respond to some conclusive reasons as she sees them. Apparently, the second explanation implies the first, if we think a discrepancy between seeing and responding to reasons is a signal of incoherent mental states. However, they actually offer distinct diagnoses of akrasia and presuppose different understandings of rationality.

As for the first approach, one of the well-known accounts is John Broome's. He argues that "rationality requires you to intend to *F* whenever you believe you ought to *F*" (2020, p. 15). This requirement of rationality is called "enkrasia" (Ibid.), which is precisely what is violated by an akratic agent: An akratic agent (say, the girl who was tempted to take drugs for the first time) was irrational because she (1) had the belief that she ought to reject the drugs and she (2) had an intention to take the drugs. According to the "enkrasia" principle, in this specific case, rationality requires her to intend to reject the drugs whenever she believes she ought to reject the drugs. Then, given that she did believe she ought to reject the drugs, her intention to take the drugs would be a violation of this "enkrasia" principle. It is worth noting that a "rational requirement" is different from a "reason", and the akratic agent's belief does not give her a "reason" to give up the intention, otherwise even a false or immoral belief could generate reasons.

By contrast, the second explanation would say that an akratic agent does not appropriately respond to conclusive reasons as she sees them. This explains where the normative pressure an agent with the better judgment feels arises from. From the first-person perspective, the akratic agent does *not* fail to conform to a special kind of requirements called "rational requirements" or "enkrasia". In other words, she does not reason in this way: to be rational (or to avoid being irrational), I am required to give up my intention to take the drugs which would be inconsistent with my better judgment. Instead, Kolodny proposes "the transparency account": "The (seeming) normative force of the 'ought' of rationality derives from a (seeming) reason, the reason that the subject believes he has." (Kolodny, 2005, p. 558) That is to say, given that the agent sees some reasons (call these reasons "seeming reasons") from the first-person perspective, it seems to her that she ought to, as these reasons require, respond to them appropriately by forming a corresponding belief or intention. Then, what goes wrong in akrasia is obvious: it seemed to the akratic agent that she ought to reject the drugs given some reasons she had seen (e.g. taking drugs is illegal and harmful; it would ruin her life...), but she failed to do what these reasons required

her to do from her first-person perspective.

However, does it mean that an akratic agent is thus irrational, or at least more irrational than she would be if she acted according to the better judgment? It depends. According to this explanation, what matters would be responding to reasons appropriately, and the reasons have been taken as reasons from the first-person perspective. If we assume that: when an agent forms the better judgment, she responds to conclusive reason as she sees it, then it follows that when the agent acts against the better judgment, she fails to respond to conclusive reason as she sees it. Therefore, the akratic agent is more irrational than she would be if she acted according to the better judgment.

Nonetheless, we are left with a possibility: If we think (1) she can “see” reasons in ways other than deliberation; (2) she can respond to reasons in ways other than forming the better judgment and acting according to it, then it is possible that: when she acts akratically, she does not completely fail to respond to reasons as she sees them. For example, she does “see” some reasons as conclusive, but she underestimates or overlooks them in her deliberation leading to the better judgment. Then if she acts for these reasons (i.e. against the better judgment), she does respond to reasons as she sees them. I shall spell out this possibility in Chapter 2.

Apparently, both of the two explanations are plausible. Is it not the case that what draws our attention to akrasia is the puzzling incoherence within an agent? It seems that the set of conflicting mental states manifesting in akrasia is independent of what the world really is and what the content of these mental states is. For example, we might find it plausible to say: both a dying billionaire who believes she ought to donate all her money to a charity while failing to do it without compulsion, and a businessman who believes he ought to kill his competitor while failing to do it without compulsion are akratic agents.

Nonetheless, the second reason-responsive explanation seems to be plausible as well. As Donald Davidson says, when an agent acts against her better judgment, she normally has reasons to do so in the sense that her akratic

action is not completely undesirable⁴. For example, when a teenager faces a friend offering her illegal drugs, she does have some *prima facie* reasons to take them: Getting an exciting experience, showing her trust towards her friends and maintaining their friendship constitute some understandable reasons supporting her akratic action. However, the problem is that they are overridden by reasons to do otherwise. Therefore, in some cases of akrasia, "... although the agent has a reason for doing what he did, he had better reasons, by his own reckoning, for acting otherwise. What needs explaining is not why the agent acted as he did, but why he didn't act otherwise, given his judgement that all things considered it would be better." (2004, p. 176) Then we might find what is puzzling about akrasia is not the static conflict between a judgment and an intention or action, but how the agent processes the information she collects from the world, especially "what the conditions are that are responsible for the inefficacy of" (Mele, 1987, p. 41) reasons the agent has given an upper hand.

Although both of the explanations capture part of what akrasia puzzles us, I shall focus on the diagnosis offered by the second explanation for two reasons. First, the first explanation leaves little room for a fruitful exploration of akrasia. According to it, akrasia would by definition be irrational for the incoherence manifesting in mental states is a crucial feature of it. If the rationality requirement is an "enkrasia" principle, as Broome argues, it seems that nothing can be more obviously irrational than "akrasia".

Second, following the second explanation, we can see the reason why we feel uneasy about the akratic actions of ourselves and others, which can accommodate our first-, second- and third-person experience. For the akratically drug-taking agent, it seemed to her that reasons required her to reject taking her first drug but she failed to respond to the reasons appropriately in action. In addition, from the third-person perspective, she was expected, given some reasons she had seen, to do what seemed to be her reasons required her to do.

⁴See also Arpaly and Schroeder (2014, p. 74): because of the existence of reasons for the akratic action, the agent's action has "more than zero practical rationality, within conventional rational choice theory".

However, the normative pressure she felt did not generate a new reason, nor did it imply a different source of normativity called rational requirements. Then how could it be relevant to akrasia? I have shown the point that, according to the second explanation:

An agent is rational when she responds to reasons as she sees them. She is irrational when she does not respond to reasons as she sees them⁵.

Then, according to this understanding, if (1) the way an agent sees and weighs the reasons is exhausted by the deliberation leading to a better judgment and (2) her responding to reasons as she sees them implies her following the better judgment (if there is one), then the akratic agent is necessarily irrational.

By contrast, if we think (1) an agent might see and weigh reasons in different ways or (2) she can respond to some reasons she sees without conforming to her better judgment, then it seems too early to say that the akratic agent is more irrational than she would be if she acted according to her better judgment.

For example, a racist shopkeeper might think she ought to treat customers of other races rudely all-things-considered, but she in fact treats them not that bad. There are some conceivable situations motivating us to think: though acting akratically, she is responding to reasons, and this is not a coincidence. Suppose that she is very greedy for money. When she treats customers of other races kindly, it is possible that she acts for earning more money. It might be a conclusive reason for her, although she does not consciously take it as conclusive in her deliberation leading to the better judgment. Or, suppose that she is an overall nice person, despite her racist view caused by her educational background. She believes she ought to be friendly to her fellow creatures, but she tends to dehumanise people of other races to some extent. Then it is possible that when she treats customers of other races kindly, she does so because she does register their humanity implicitly and she acts for the reason that they are humans. In both cases, can we be sure that she would be more rational if she, as her better

⁵ In some cases, an appropriate response to reasons might be “impossible” because of the situations (e.g. the agent is manipulated) or some limits of the agent (e.g. muscular failures), then the agent’s failure to respond is not a failure of rationality, and a mere intention she has might be a sufficient indication of her rationality.

judgment suggests, treated customers of other races rudely?

This is a route opened by the second explanation but denied by Broome's account which focuses on static conflicts among mental states. Although it faces some difficulties (e. g. the relation between reason-responsiveness and the better judgment, and that between deliberation and "seeing" reasons), I will not defend it here and a critical analysis of it is shown in Chapter 2. Before that, one clarification should be made:

When the two diagnoses of the irrationality of akrasia are shown, we can see two different accounts of rationality underlying them. Roughly speaking, Broome's account is "static" in the sense that rational requirement is a set of "state requirements". It is a ban on conflicts among mental states (e.g. believing *p* and believing *-p*), and how the agent gets them or gets rid of them does not matter.

By contrast, the other account proposes "the identity claim" attacked by Broome: "...the property of rationality is identical to the property of responding correctly to reasons". (Broome, 2020, p. 7) A rational agent ought to respond to reasons as she sees them, and the process of responding does matter. The agent should assess the evidence she has in one way or another, and a mere elimination of one of the conflicting states, such as one because of an electric shock (Kolodny, 2004, p. 520) might *not* make her more rational. I will not take sides in this contentious "what-rationality-is" discussion, but when I use "rational" in this thesis, it is more likely that it means "reason-responsive" or "responding to reasons appropriately". For people who feel reluctant to accept this reason-responsive approach to understanding rationality, they can interpret "rational" in this thesis as "reason-responsive" or "normative" in Broome's sense, which would not undermine or trivialise the arguments in this thesis.

4. What is Wrong with Akrasia? – "A Divided Self"

What is wrong with akrasia? The last section shows us a diagnosis from the perspective of reason-responsiveness. In this section, another perspective

relevant to the prior one would be offered.

Although Watson's approach, which might lead us to a sharply split picture of two systems within an agent, is problematic, it is not hard to find evidence of a divided self in some cases of akrasia. In other words, an akratic agent seems to stand by both sides (i.e. her better judgment and her akratic action) in this battle. If she wholeheartedly sided with her better judgment, then why would not she make the "I am forced to do it" excuse? If she wholeheartedly sided with her action, then why would she still label the action as "akrasia" or even "wrongdoing"? The better judgment does have part of the agent for it is the agent who forms the better judgment by her reasoning. However, the remaining part of her seems to stand by her akratic action. This is only a rough description, but it does reflect some feelings akratic agents have.

How can we make sense of the seemingly divided self? One way to understand this ambivalent state is to say that the part standing by the better judgment is explicitly reason-responsive while the other part can be implicitly reason-responsive. In addition, when an agent is responding to reasons, she is active. I will not spell out this possibility here, but it seems that it can accommodate the intuition that the akratic agent is active in two different ways, and that is why she is not a mere bystander to her akratic action. Let us call it the Reason-responsive Proposal and elucidation of it will be offered in Chapter 2.

Another possibility is that the akratic agent is responding to reasons when forming the better judgment, and her akratic action is motivated by something external to the system of reason-responsiveness. However, her akratic action does hold part of her in a way different from that of reason-responsiveness. It is relevant to what an agent cares about, which cannot be reduced to desires or reasons. We will return to this Out-of-caring-about Proposal in Chapters 3 and 4.

5. Two Cases

Let us finish this chapter with two examples of akrasia. Both of them might

cast light on the prior arguments in a different way.

A Sexist Father:

A sexist father living hundreds of years ago believed that girls were not as good as boys, and there was no need to give them educational opportunities. Nonetheless, he still supported his daughters' education. The father might think this was an unwise decision, and he sincerely admired other fathers who did not waste money on girls' education. From the third-person perspective, it is an akratic action because it meets all the three standards, but it is not a "wrong" action. If we accept the reason-responsive understanding of rationality, we might even feel reluctant to label it as "an irrational action". Admittedly, it can be a lucky coincidence showing the father's mental disorder, but there is another possibility: the father, though influenced by the mainstream narratives in the society, felt that his beloved daughters also wanted to know more about the world, get equal opportunities and become independent. However, he, unconsciously or consciously, overlooked or underestimated these factors in the process of deliberation, and believed: the very fact that they were girls has told him what he ought to do.

Given that, we might think this father, though regrettably having some morally wrong ideas, was a better father compared with whom he admired (e.g. a father, with a strong determination, did not support his daughters' education). Furthermore, if we accept the reason-responsive approach to understanding rationality, we can even say that this father was more rational for acting against his better judgment than he would be if he acted in accordance with his better judgment. If we consider the father's action in this way, we might assume that his akratic action's aligning nicely with some good reasons is not a mere coincidence. In other words, he acted for some good reasons which his deliberation failed to capture but he was, at least, dimly aware of. This leads us to the reason-responsive approach to understanding akrasia: When an agent acts akratically, it is possible that she responds to reasons at the same time. That is why she is active. Nonetheless, I will introduce the second case, which might cast suspicion

on this analysis.

24 & Ready to Die

From the outside, Emily was a physically healthy young woman with a loving family and friends. However, she found her life unbearable and constantly harmed herself. She believed that death could bring her peace. Euthanasia is legal in the country she lived in, and her family members and friends respected every decision she made. She was sure that accepting euthanasia was the best choice. However, even if she set a date for euthanasia, she rejected the lethal injection on that day. After that, she said: “When Dr Proot arrived, we talked. I said calmly, ‘I cannot do it’. Because the two previous weeks were relatively bearable. They were free from crises. I don’t know why. Was it because the serenity of death was so close? And that I felt OK? Or had something changed in me? ... I’ve tried not to think about the truth.... Because the truth is: I’d still rather not be here.” (*24 & Ready to Die*, a documentary made by *The Economist*)

Apparently, there is a parallel between the *24 & Ready to Die*’s case and the sexist father’s case. We might think Emily’s rejection of the lethal injection is a good decision, and we could find some reasons for it: Emily is young, so she has a long life to live. Emily has some friends, who will support her. Emily could expect that, in the future, her mental disease will be cured. Last but not least, living life, even an unbearable one, is something worthwhile in itself. We might ignore what Emily thought about it and think these reasons we have seen “touched” her in some ways and changed her mind at the last minute, although she still thought she ought to die. We can even expect that after that she “mysteriously” realises some meanings of life and continues her life “bravely”.

However, there are some facts about the second half of Emily’s case, shown at the end of the documentary: “Emily’s mental ‘crises’ continued. Two years after making this film, she was given a date for euthanasia. On 25th August 2018, Emily was given a lethal injection and died peacefully. Emily (1990-2018).” This ending does render some of the assumptions we made very dubious. It seems that Emily might not respond to any reasons in her acting akratically, and

the “correct” life-saving decision she made could not be rationalised.

Then, can we say that there is no parallel between this case and that of the sexist father (e.g. they are different types of akrasia, or one of them is not akrasia but compulsion)? Or is it possible that our assumptions about the sexist father’s case are also merely wishful thinking and ungrounded rationalisation from a third-person perspective? The aim is not to answer the questions here, but to show tension: on the one hand, we intend to make akratic action intelligible; on the other hand, it might be risky to rationalise it.

6. Conclusion

In this chapter, I have argued that the distinction between akrasia and compulsion is significant because of the peculiar first-, second- and third-person experience in akrasia and its relevance to irrationality. Therefore, when this distinction conflicts with Watson’s account, it might be more plausible to take this distinction seriously and thus doubt the account failing to capture it. Then I explained what is wrong with akrasia. A common answer would be: akrasia is a typical kind of irrationality. This consists of two different diagnoses. One is that akrasia implies some conflicting mental states, while the other says that akratic agents might respond to a set of reasons “seemingly” weaker than those supporting her better judgment. I take the second diagnosis for it might grasp the complexities of akrasia: an akratic agent fails to respond to some seeming reasons though she does feel the normative pressure from them. This might explain why we are upset and puzzled by akrasia. However, it leaves room for a possibility that: though failing to respond to some seeming reasons collected and weighed by her deliberation, the akratic agent might respond to some reasons, even conclusive ones. If this is not a mere coincidence, then the agent can be “active” in the two different ways of reason-responsiveness and the divided self would be explained. In the last section, this theoretical possibility is demonstrated by the sexist father’s case, but then challenged by the *24 & Ready to Die* case.

In the next chapter, I shall focus on “inverse akrasia”, which has been represented by the sexist father’s case, offer a reason-responsive proposal to understand it, and explain a challenge it faces.

Chapter 2 Inverse Akrasia and the Reason-responsive Proposal

In this chapter, I shall begin by reconstructing Nomy Arpaly's argument on "inverse akrasia" in *Unprincipled Virtue* by using the example of Huckleberry Finn (Huck Finn). Arpaly argues that Huck Finn acts rationally when he helps Jim, a black slave, to escape, although he acts against his better judgment. This is because he responds to some right-making features (e.g. Jim's humanity) as reasons. Therefore, he is praiseworthy. Arpaly's Reason-responsive Proposal is more plausible than the Irrational-passive Proposal, because it shows a deeper relation between an agent's internal story and her akratic action in some cases that interest us.

Then I shall analyse a challenge to Arpaly's account. It is argued that we need to know "under what conditions an agent S can count as 'acting on or for (a particular) reason R'" (Stroud, 2007, p. 455) so that we can make sure that a specific agent can count as "acting on or for (a particular) reason R". This is more urgent in Huck Finn's case, because he himself cannot see the justification of his action. Without this first-person justification, Huck Finn might merely be "caused" to help Jim despite his better judgment (Mason, 2007, p. 447). That is to say, he does not respond to reasons in the right way. Therefore, he is not rational.

I shall point out the deeper concerns underlying Arpaly's Reason-responsive Proposal and the challenge it faces. They capture different features of "inverse akrasia", but to fully understand it, we might be motivated to consider an alternate proposal, which will be sketched in this chapter and expounded in the following two chapters.

1. Nomy Arpaly's Analysis of Huck Finn's Case and the Reason-responsive Proposal

I reconstruct Arpaly's argument as follows:

(1) Huck Finn holds "official" racist views (e.g. Black slaves are their owner's

property), though he is a boy with moral common sense.

(2) Huck Finn undergoes a perceptual shift when he gets along with Jim, an escaping slave: Perceiving the similarity between himself and Jim, Huck Finn begins to see Jim as a friend and treats him more equally (e.g. apologising to Jim⁶).

(3) However, Huck Finn does not reflect on what (2) means.

(4) When the opportunity comes to turn Jim in, Huck Finn experiences a strong reluctance to do so. At the same time, it is true that: (4.1) Huck Finn does not change his explicit racism, while (4.2) he has come to see Jim as a person.

Propositions that Arpaly argues for:

(5) To act for a certain reason, the agent does not necessarily know she is acting for that certain reason. However, the reason is not an external reason either for it (5.1) has been registered (though overlooked or underestimated in deliberation) by the agent AND (5.2) it reflects the agent's concern.

(6) Acting for moral reasons means acting for some right-making features.

And

(7) "Praiseworthiness as Responsiveness to Moral Reason (PRMR): For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons." (2003, p. 72)

Therefore:

(8) Huck Finn is acting for reasons (e.g. Jim's humanity), given (1) (2) and (5);

(9) Huck Finn is acting for moral reasons, because his reason is of the right-making kind; i.e. (6) applies;

(10) Huck Finn is morally praiseworthy, given (7) and (9).

In this argument, providing that (1)-(4) are part of the story, and (8)-(10) are derived from (5)-(7), what needs elucidation is (5)-(7). Before that, I shall cite Arpaly's explanation of what reasons, overwhelming reasons and reason-

⁶ For the occasion in the novel, see Chapter 15: "It was fifteen minutes before I could work myself up to go and humble myself to a nigger; but I done it, and I warn't ever sorry for it afterwards, neither. I didn't do him no more mean tricks, and I wouldn't done that one if I'd a knowed it would make him feel that way."

responsiveness mean in this picture:

“...I will assume, for purposes of argument, that one only has a reason to act in a certain way to the extent that the relevant course of action is likely, given one’s belief, to satisfy one’s desires, and that acting rationally, whatever it turns out to be, involves doing one has overwhelming reasons to do, for these reasons. (Ibid., p. 37)”

It is obvious that Arpaly accepts the reason-responsive understanding of rationality (or “acting rationally”), which has been distinguished from the account that takes rationality as static coherence between different mental states in Chapter 1⁷. Arpaly thinks acting rationally means acting for conclusive reasons. We will discuss in what way an action can count as one “for” reasons (rather than one merely “aligning with reasons”) in the next section. The point at issue here is that Arpaly thinks incoherence between a better judgment and action does not give an exhaustive description of the irrationality of an akratic action. Moreover, the incoherence between an irrational all-things-considered judgment and an akratic action might be more rational than a coherence between an irrational all-things-considered judgment and an enkratic action. It is more crucial to figure out (1) whether the agent, one way or another, responds to reasons she picks up implicitly or explicitly from the outside world and (2) whether the reasons themselves are good.

Bearing the reason-responsive view of rationality in mind, we can understand (5). In (5), it is likely that the agent does not tell the strength of the reasons she has and the right way to respond to them, especially when she is not good at thinking. As a result, she may make a mistake on what she ought to do all-things-considered. That does not mean that her mental states are totally opaque to her, but merely that she might overlook or underestimate some of them. This might manifest a failure of her self-reflection, but it does not invalidate the reasons she has, nor does it exclude the possibility that she responds to these reasons

⁷ Arpaly does discuss what her proposal would look like within the static picture which focuses on coherence (Ibid., pp. 45-46). However, given her final attempt to go beyond this coherence, it is plausible to say that she supports the dynamic understanding of rationality.

regardless of the faulty deliberation. Arpaly argues that Huck Finn is an example. Since Huck Finn is reluctant to turn Jim in, and we can find reasons (e.g. Jim's personhood) based on relevant mental states (though not in his deliberation) in his personal history as well as the causal relation between these reasons and the action, it is justified to say that Huck Finn is acting for reasons.

(6) and (7) sound plausible, partly because they do not intend to offer a substantive list of right-making features. Huck Finn responds to good reasons, because it is what motivates him to act (i.e. Jim's humanity) that makes the action praiseworthy. Then Huck Finn is praiseworthy⁸.

We can call Arpaly's interpretation of inverse akrasia the "Reason-responsive Proposal". The basic idea is that even if an agent acts against her better judgment, it is possible that she responds to good reasons. In other words, it denies that a better judgment one makes in a situation is the only channel of one's reason-responsiveness. Therefore, it tells us the internal story of the agents in cases of inverse akrasia: they are not mere bystanders to their action, and thus akrasia, at least inverse akrasia, can be distinguished from compulsion and contribute to our understanding of rationality.

2. A Challenge: What Can Count as Responding to or Acting for a Reason?

When we discuss the way Arpaly understands rationality, we have already seen a problem: Arpaly seems to claim that (1) the reason-responsive mechanism can respond to reasons though sidestepping the mechanism of deliberation (2) the mechanism could correct the faulty deliberation sometimes. However, this might be counterintuitive. Deliberation, of course, can be irrational itself, and correction via another mechanism (e.g. conscience, emotions or "common sense") is not impossible. However, can the sources of correction be seen as reason-responsive? After all, deliberation is the process we intentionally enter to weigh the reasons we have and make a judgment. Then a general standard for

⁸ See also Arpaly, 2006, p. 13.

reason-responsiveness or acting-for-reasons is needed.

The worry raised by Sarah Stroud is: “what are the general truth-conditions for R’s being, not just a reason to Φ , but S’s reason for Φ ing?” (2007, p. 455) There is a distinction between the claim that “Jim’s personhood is *a reason* to help him” and that “Jim’s personhood is *Huck Finn’s reason* to help him”. This would be obvious if Huck Finn can articulate what he takes as reasons. Suppose that he says that “I do not turn Jim in to irritate his owner whom I hate”. It follows that Jim’s personhood is *a reason* to help him nonetheless, but it is not *Huck Finn’s reason* in this case. By contrast, if he says that “I do not turn Jim in because he is a person like me”, then we tend to think “Jim’s personhood is *a reason* to help him” and “Jim’s personhood is *Huck Finn’s reason* to help him” in this case.

In both cases, we might trust the two “alternate” Huck Finns’ first-person narratives. However, in the “real” Huck Finn’s case, tension emerges. On the one hand, the distinction between “R’s being *a reason* to Φ ” and “R’s being *an agent* S’s *reason* to Φ ” must stand. No matter how obvious R’s right-making features and its relation with Φ ing may seem to us, we cannot infer that S is necessarily responding to R when she Φ s. This is because we can conceive mental states, in which R is absent, that actually cause the same S’s Φ ing. Therefore, to ensure that “R is S’s reason to Φ ”, we need to know something about S’s mental states.

On the other hand, in the case of Huck Finn, if we explore his mental states, the first thing manifesting might be a conflict between his racist beliefs and his tendency to act against the beliefs. We cannot find an explicit piece of explanation of his action in the form that “a certain R is Huck Finn’s reason to help Jim to escape”, even if we can find his cumulative changes and his reluctance to turn Jim in in his personal history.

Then, how can we piece together what we find and conclude that a certain R (say, Jim’s personhood) is Huck Finn’s reason to help Jim to escape? If Huck Finn himself sees his acting for R in a certain way, R’s function as Huck Finn’s reason might be ensured. Therefore, we need some general truth-conditions for R’s being S’s reason to Φ , which might be a set of constraints on the way R is

related to S's mental states. For example, Elinor Mason argues that general truth-conditions for "R's being S's reason to Φ " (or "S does Φ for R") would necessarily exclude the case of Huck Finn as a case of reason-responsiveness, if it is an akratic case. The general truth-condition she offers is: S is able to see the justification of her action from her point of view. Mason's argument (2007, pp. 444-45) can be formulated in the following way:

(1) A distinction must be made between a rational response to a reason and a merely causal response to a feature of the situation.

(2) To judge whether an action is a rational response to a reason or a merely causal response to a feature of the situation, we should rely on the "Justification Internalism Constraint": an action is reason-responsive if it is justified from the agent's point of view. That is to say, "the agent must be able to see what the justification is".

(3) Cases of "internal blurring": It is possible that an agent does not see the justification of her better judgment, because the judgement is caused by her idleness.

(4) Cases of "being rational without deliberation": It is possible that an agent sees the justification of her action without deliberation, because reason-responsiveness is broader than deliberation.

(5) The "akratic cases" in question can be interpreted in two ways:

(5.1) Irrational Action Interpretation: An agent sees the justification of her better judgment first-personally, and she does not see that of her akratic action first-personally. Given (2), her akratic action is not reason-responsive. She is akratic and irrational.

(5.2) Pseudo Better Judgment Interpretation: Given (3) and (4), it is possible that an agent does not see the justification of her better judgment first-personally, but she does see that of her action against her better judgment first-personally. However, since the better judgment, in this case, is not reason-responsive, an action against it cannot be called an "akratic action", in the strict sense.

If we apply it to Huck Finn's case, the conclusion would be: Huck Finn is

EITHER akratic *and* irrational, OR he is rational *and* enkratic.

According to the Irrational Action Interpretation, Huck Finn is not responding to Jim's personhood as a reason when he does not turn Jim in, because he could not see the reason as the justification of his action from his point of view. Instead, he does it in a non-reason-responsive way. Therefore, he is akratic in the sense that he acts against his better judgment, but he is not rational given that he violates the "Justification Internalism Constraint".

By contrast, according to the Pseudo Better Judgment Interpretation, Huck Finn's judgment made against the background of the racist narrative of the society is not "a better judgment" but merely what Mason calls "an internal blurring". That is to say, the racist attitude is a mere product of idleness rather than a genuine better judgment. Therefore, an action against it is not "real" akrasia. At the same time, Huck Finn has been dimly aware of the justification of treating Jim equally from his point of view. Then given that he sees the justification of his akratic action internally, his following action can count as responding to a reason, and his better judgment is insignificant.

Then, what is the challenge two interpretations would put to Arpaly's proposal? According to Arpaly, obviously, the Irrational Action Interpretation would be uninteresting. However, the Pseudo Better Judgment Interpretation is unsatisfying either. Huck Finn's better judgment is not a simple internal blurring, which would be corrected if he could think more carefully. Therefore, it would not be a case of a pseudo better judgment. Moreover, we can find little evidence to support that Huck Finn sees the justification of his akratic action: In the novel, Huck Finn has a group of mental states: "he [Jim] had an uncommon level head" (Chapter 14) for a black slave; "He was mighty" (Chapter 23). However, it is true that: (1) Huck Finn does not have a belief that "a black slave with an uncommon head is entitled to be free" or "if a black slave is mighty, he could not be turned in". Therefore, he does not make an inference based on the information he has, nor does he see it as a justification of his action; (2) Huck Finn has a belief that

turning Jim in is the right thing to do, and doing the opposite thing is shameful⁹.

Now, we can see the tension: On the one hand, Arpaly would admit that (1) to respond to reasons, a special kind of relation between a reason and the agent's mental states should be established. We might need a richer internal story of the agent rather than robotic causation. On the other hand, it seems to Mason that (2) the hallmark of a response to reasons is a justification seen from the agent's first-person point of view. If (2) flows from (1), then Arpaly cannot be right when she says that Huck Finn, who "does not have the belief that what he does is right anywhere in his head" (Arpaly, 2003, p. 77), can respond to reasons when he acts akratically.

Arpaly could reject the "Justification Internalism Constraint" and offer a less demanding standard for reason-responsiveness: "If the content of a group of mental states – under the most common story these would be beliefs and desires, but they do not have to be – *adds up* to compelling reasons to do something, and these mental states *motivate* me to do it by virtue of the very feature that makes them good reasons, good premises in a practical inference – then I have responded to a reason." (Arpaly, 2006, p. 70, emphasis mine) An agent can count as responding to reasons if her mental process goes in this way. We can call this the "Right Way Constraint". Unlike the "Justification Internalism Constraint", it does not necessarily exclude akratic action. An agent might notice these right-making factors, see them as conclusive reasons and act for them. This would be a case meeting the standard. Likely, even though an agent merely registers these factors and fails to see the justification of them, she could respond to reasons if her mental states "semi-automatically" process in this way.

The differences between the two standards offered by Mason and Arpaly are significant. Both of them accept that an agent could respond to reasons without deliberation. Nonetheless, Mason argues that an agent who acts against her better judgment can act for reasons only if (1) her better judgment is mere

⁹ In Chapter 31, Huck Finn thinks "it would get all around that Huck Finn helped a nigger to get his freedom; and if I was ever to see anybody from that town again I'd be ready to get down and lick his boots for shame."

“internal blurring” the justification of which she does not see, and (2) she sees the justification of her action. According to the “Justification Internalism Constraint”, an agent cannot simultaneously act against what she gives her first-person justification to *and* respond to reasons. By contrast, Arpaly argues that an agent can do this, because her mental process can function in accordance with the “Right Way Constraint”, although it violates the “Justification Internalism Constraint”.

It is hard to weigh the two different constraints on reason-responsiveness. Instead of doing it, in the next section, I will return to Huck Finn’s case, and argue that the different interpretations in this section focus on two distinct aspects of cases like Huck Finn’s, and an alternate proposal will be raised to grasp both at the same time.

3. An Alternate Proposal (First Impression)

In this chapter, so far, we have discussed:

Arpaly suggests a dynamic reason-responsive view of rationality. This picture can accommodate the cases in which the agent seems to act better when she acts against her better judgment. It is not a mere lucky coincidence. Rather, although akrasia necessarily implies an inconsistency between one’s judgment and action, if the action itself has features of reason-responsiveness (e.g. tracing some right-making features), then the inconsistency would be much less important than the static understanding of rationality might suppose. However, the crux of Arpaly’s proposal is that it might be counterintuitive and theoretically problematic to say akratic agents, like Huck Finn, respond to reasons when acting against their better judgment. According to Mason, a standard for reason-responsiveness is required, which would distinguish it from a causal response; the standard would be a kind of first-person justification, which is different from deliberation; akratic action against one’s first-personally-justified better judgment would lack the first-person justification, so it is not rational.

Therefore, concerning Huck Finn's reason-responsiveness, we might get the opposite interpretations if we accept different standards of "what count as reason-responsiveness". It is significant because they reflect distinct understandings of our rational capacities. In this section, we will slightly deviate from the route we took and consider a relevant question: when they get different interpretations of the case by applying different standards, what are the features they try to grasp? By answering this question, we might find a starting point for an alternate proposal outside the discussion on reason-responsiveness.

Arpaly focuses on the feature that Huck Finn does engage with his action in an active way. This special engagement, she supposes, could be explained only by reason-responsiveness. If what motivates Huck Finn would be "some purely atavistic mechanism – akin, perhaps, to the tendency some animals have not to harm creatures with infantile features" (Arpaly, 2003, p. 76), the case might be uninteresting in two ways. First, it speaks little of what kind of person Huck Finn is. Second, he might not be praiseworthy. There must be a different internal story of Huck Finn.

However, Mason argues that the irrational colour of Huck Finn's action cannot be overlooked. If we think Huck Finn takes his better judgment seriously and sees the justification of it, then his action has a reason-insensitive feature.

Both features are constitutive elements of Huck Finn's case. It seems that Huck Finn's rational capacities do face a real crisis. They are not reliable for either Huck Finn or anyone who wants to interpret the case by the Reason-responsive Proposal. This is not simply because Huck Finn's action is not justified by a deliberation, but because Huck Finn's action is not justified by himself. However, we could not retreat to some purely atavistic mechanisms, which are close to the motivational system in the Irrational-passive Proposal, because we would give up the richness of the case and the praiseworthiness of Huck Finn. Now, it might be helpful to change our direction and find a new option other than the Reason-responsive Proposal (which needs further defence) and the Irrational-passive Proposal (which we feel reluctant to retreat to).

Firstly, let us consider two examples. Huck Finn₂ is similar to the original Huck Finn, except that he is much better at reflecting on what happens to him. He did not get any opportunities to closely observe black men's behaviour, nor did he feel the necessity to reflect on his racism until he met Jim. Therefore, other things being equal, Huck Finn₂ realises that Jim is equal to him. "Then, why should I turn him in? My prior beliefs and attitudes are wrong. He deserves equality and liberty." Huck Finn reckons, and then racist beliefs are overridden.

By contrast, Huck Finn₃ is similar to the original Huck Finn, except that he is more sensitive to any changes within himself. He sees himself as a compassionate boy who deeply cares about the happiness of other people. However, before he met Jim, he did not feel black men are his fellow-creatures and his compassion towards them had not been invoked. Then, other things being equal, Huck Finn₃ realises that he cannot help liking Jim and trying to protect him. "I have good reasons to turn Jim in," Huck Finn₃ thinks, "but I cannot bring myself to do it, just like I cannot help stealing bread in a kitchen to help a tramp. I feel guilty, but sometimes I am even angry at reason's harshness."

By the two cases, I shall make a slow move. Firstly, we can see the distinction between "R's being a *reason* to Φ " and "R's being *S's reason* to Φ " in the case of Huck Finn₃. Huck Finn₃ does react to features newly appearing to him, but he does not respond to them in a reason-responsive way. Therefore, Huck Finn₃'s case is not an implicit version of Huck Finn₂'s case, given that Huck Finn₃ brings his special feelings towards Jim into consciousness and considers it as something in opposition to the reasons at hand. It seems plausible to say that Huck Finn₃ does not respond to reasons.

Secondly, without Huck Finn's explicit first-person justification, we cannot know whether the original Huck Finn's case is an implicit version of Huck Finn₂ or that of Huck Finn₃, then it is problematic to say that Huck Finn responds to reasons without knowing it rather than that he is "motivated" by something other than reasons without noticing it. Admittedly, the former is philosophically clearer to us. Accordingly, it would be easier if we could find a parallel of the former in a

case that we have not made sense of. That is to say, if we are certain that Huck Finn is an implicit version of the explicit reason-responsive Huck Finn₂, it is plausible to merely relocate the mechanism in operation in Huck Finn₂ to somewhere less transparent within Huck Finn. By contrast, Huck Finn₃' case itself is obscure. However, without further investigation, we cannot say in which case what happens to Huck Finn_{2/3} is closer to that to the original Huck Finn.

Furthermore, I suppose that if the original Huck Finn can be closer to the Huck Finn₃, a state quite different from that depicted by Arpaly (i.e. that Huck Finn responds to reasons without knowing it) might manifest itself:

A state "X" where the strength of one's caring about an object is disproportionate to the weight of it as a rationalising reason. That is to say, at this stage, the agent's caring about¹⁰ an object O motivates her to find relevant reasons¹¹, but the reasons in favour of what she cares about are overridden by others. Her caring about O functions in a "bare" way, nonetheless. At this point, Huck Finn, if understood in this way, has *tried* to respond to reasons in the sense that he has traced what could count as reasons, and arrived at a better judgment. However, there is a significant discrepancy between the reasons in favour of his caring about O and the strength of his caring about O. Then Huck Finn's letting Jim escape is genuinely against reasons from his perspective, and even from a third-person perspective, we cannot attribute reason-responsiveness to his action rather than his better judgment of which he sees the justification.

This can help us to see why the akratic action, though seemingly independent of the reason-responsive mechanism, is not alien to the person. It is motivated by the agent's caring about a certain object, which, on certain occasions like that in Huck Finn₃'s case, can function in a quite active way.

¹⁰ What "caring about" means here is not clear, and it will be explained in the next chapter. However, at this stage what we need to think about is merely a pre-philosophical understanding of this phrase, and our intuition that "caring about O" is, without further philosophical reflection, a bit different from "being attracted to O" or "acknowledging O's value".

¹¹ Normally an agent might take the object's value as a reason. Moreover, she could take the fact that she cares about it as a reason (e.g. since she cares about the object, she could not bear its suffering) as well. However, it is not necessarily the case for (1) she might not realise the fact that she cares about the object, or the fact that the extent to which she cares about it could make a difference in her rational judgment; (2) she might think her caring about the object does not reflect the real value of the object, and she does not take her caring about the object as a reason. We will return to this point in Chapter 4.

However, we are left with some questions: How can it happen? Can the “state X” collapse to part of a blind mechanism that has already been considered by Arpaly? The underlying mechanism needs further elucidation, and this will be done in the following two chapters. Here I will end this chapter with several cases in which the mechanism of caring about would be more noticeable. Therefore, it might not be confused with reason-responsiveness, and the first impression it leaves can be vivid.

4. Acting out of Caring about O

Example 1

The Duke of She said to Confucius, “Among my people there is one we call ‘Upright Gong [an upright man].’ When his father stole a sheep, he reported him to the authorities.”

Confucius replied, “Among my people, those who we consider ‘upright’ are different from this: fathers cover up for their sons, and sons cover up for their fathers. ‘Uprightness’ is to be found in this.” (*Analects* 13: 18, Translated by Edward Slingerland).

There are two ways to interpret this case. One is that Confucius thinks the father-son relation should function as the overwhelming reason, and an upright man ought to act according to it. The other is that the word “uprightness” here is used in a merely rhetoric way. It means something right and humane. Accordingly, we can understand Confucius’s view in two different ways: Confucius tries to point out what is correct to do here by emphasising what is the most important thing *qua* the overwhelming reason (the first way)¹² or *qua* what weighs over all reasons (the second way). If we take the second way, we might see a kind of caring function in a bare way.

¹² For example, Qingping Liu (2003, p. 239) thinks, according to the text, “so long as one’s action comply with the highest principles of filial duty and are an attempt to achieve the end of serving one’s parents, they will be naturally and necessarily right and good, no matter whether they violate other principles or standards of human action”.

Example 2 (a story of Mencius who is the most famous follower of Confucius)

Tao Ying asked, “If, while Shun was Son of Heaven and Gao Yao was minister of justice, the Blind Man [Shun’s father] had murdered someone, what would have happened?”

Mencius said, “Gao Yao would have apprehended him; that is all.”

“But wouldn’t Shun have prevented this?”

“How could Shun have prevented it? Gao Yao had received the authority for this.”

“Then what would Shun have done?”

“Shun would have regarded abandoning the realm as he would abandoning an old shoe. Secretly, he would have taken his father on his back and fled, dwelling somewhere along the seacoast. There he would have happily remained to the end of his life, forgetting, in his delight, about the realm.” (*Mencius*, 7a35, Translated by Irene Bloom)

Here the suggestion of helping one’s father to get rid of punishment appears again. This is a case more like Huck Finn’s case. Shun in this case and Huck Finn seem to “follow their heart”. They are motivated by something other than reason-responsiveness, but can we say that they are passive bystanders to their action? Shun is regarded as a role model in Chinese traditional culture, and it is (or “was”) believed that he copes with his situation well, and I suppose that is mainly because he is seen as doing something appropriately, though he might not do it via the mechanism of reason-responsiveness.

Example 3 (from *Southeast the Peacock Flies*. It is a poem written about two thousand years ago.)

The Background: A clerk’s mother hated the clerk’s wife. She forced the clerk to divorce his wife and arranged another marriage for him. During that period, a parent’s authority was absolute, and the clerk, as a servile son, knew it. However, he decided to commit suicide with his ex-wife, whom he loved deeply. Before that, he returned home and talked with his mother:

“The clerk returned to his home,
ascended the hall, bowed to his mother:
‘Today the winds blow fierce and cold,
the cold winds break the tree limbs,
and harsh frost collects on the orchids in the garden.
Your son today goes into darkness,
leaving you behind all alone
I do this bad thing of my own will—
do not rail at the gods or spirits.
May your years be like the rock on the southern mountain,
your four limbs sturdy and straight.’
When his mother heard this,
her tears fell in time to her words:
‘You are the son of a great family
who have served in high government office.
Don’t be foolish and die for this woman,
when she is so far beneath you!
The family to the east has a virtuous daughter,
her beauty the boast of the whole city.
Your mother will arrange for you to have her,
it will be done in the space of a day!’
The clerk bowed once more and withdrew,
in his empty bedroom sighed unendingly,
then made his plan, determined to see it through,
turned his head, looked toward the door,
grief pressing in on him more cruelly than ever.” (Translated by Burton
Watson)

Like Huck Finn, the clerk thinks he is acting wrongly (“I do this bad thing of my own will— do not rail at the gods or spirits.”). Sometimes he even shares the better judgment with his mother and apologises to her. If we follow the Reason-

responsive Proposal, we might think Huck Finn responds to Jim's personhood without explicitly knowing that. But in this case, it is hard to rationalise the clerk's love for his wife. The parallel between this case and the case of Huck Finn might be clearer if we read the following paragraphs in the novel. When Huck Finn helps Jim again by tearing up a letter he planned to send, he thinks:

"All right, then, I'll go to hell—and tore it up. It was awful thoughts and awful words, but they was said. And I let them stay said; and never thought no more about reforming. I shoved the whole thing out of my head, and said I would take up wickedness again, which was in my line, being brung up to it, and the other warn't." (Chapter 31)

When Huck Finn tells Tom that he plans to rescue Jim, he puts it in the following way:

"I know what you'll say. You'll say it's dirty, low-down business; but what if it is? I'm low down..." (Chapter 33)

We can find the similarities between Huck Finn's thoughts and the clerk's.

5. Conclusion

In this chapter, we start by "inverse akrasia" and Arpaly's reason-responsive interpretation of it. It might face a challenge, because akratic agents, even in cases of inverse akrasia, do not see the justification of their akratic action. Then another question we could consider is: what features of the case in question do the different interpretations intend to capture? This leads us to deviate from the discussion of constraints on acting for reasons, and think about an alternative interpretation that relies on a mechanism of "caring about" which we have not considered yet. So far, only examples are offered to give a hint of its existence, and we largely rely on a pre-philosophical understanding of it. The concept of "caring about" will be further explained in Chapter 3, and an interpretation of akratic cases based on it will be given in Chapter 4.

Chapter 3 Caring about, Volitional Necessity and Discrepancies

In the preceding chapter, the tension in some cases of akratic action can be seen. Intuitively, these cases are different from some trivial cases (e.g. an agent, knowing that she ought not to eat sweet food, eats a chocolate bar). This is because they convey some information about the agent's mental states, which might not feature a mere conflict between Reason and Desire. This motivates the reason-responsive interpretation of cases like Huck Finn's. According to it, the fact that an agent acts akratically does not exclude the possibility that she acts for reasons in a non-deliberative way. However, by forming a better judgment, the agent seems to exhaust all her intellectual effort and fully express her willingness to act for good reasons. Moreover, she could see the justification of the judgment. Accordingly, when she acts against her better judgment, her first-person experience is different from a normal response to reasons. Therefore, on the one hand, we try to avoid a trivialisation of these cases; on the other hand, we might feel hesitant to attribute its peculiarity and importance to reason-responsiveness, and see it as over-rationalisation. Whether a Reason-responsive Proposal, with a more flexible understanding of reasons and acting for reasons, can avoid the worry of over-rationalisation will be discussed in section 5.1.

Now let us suppose: it is advisable to give due respect to akratic agents' experience from their first-person perspective (e.g. "I act wrongly!" "I'll go to the hell!"), and have a welcoming attitude towards a wider range of philosophical vocabulary on the condition that it helps us to understand the complicated human life. As a result, an alternative proposal can be offered, but before doing it in the next chapter, I shall focus on two concepts in this chapter to set the stage.

1. The First Concept: Caring about

Harry Frankfurt thinks that the concept of "caring about" "appears to be so fundamental that a satisfactory analysis of it may not be possible at all" (1988, p.

82). However, we might be able to see what a difference it makes in our lives. By caring about some things, we make them important to us and guide ourselves by identifying ourselves with what we care about (2004, p. 23). Caring about implies “a certain consistency or steadiness of behavior; and this presupposes some degree of persistence” (1988, p. 84). As a result, we can organise our lives in a richer way different from a mere “formal relations of sequentiality” (1988, p. 83. See also 2004, p. 43). This is the way caring about functions in the dynamic personal history, and even if we merely take a static specimen from it, we can also see that caring about has a comprehensive influence, which some transient desires and urges would lack.

It might also explain why caring about is not under the direct control of the agent (2004, p. 44). She might intend to change the object she cares about, but sometimes she fails because her caring about it is not a transient desire. However, this is not frustrating. Without caring about anything, we “would not be disposed to make any effort to maintain any of the interests, aims, and ambitions by which we are from time to time moved” (1999, p. 162). In this sense, “what we care about” constitutes the essence of us (2004, p. 17). I shall compare caring about with some familiar concepts to show its irreducibility in spite of its vagueness.

1.1. Caring about Versus Urges and Instincts

It is not difficult to distinguish caring about from some transient urges without any deeper root in one’s self, like an urge to hurt a passer-by. However, when we consider some more stable and noticeable instincts accompanied by strong urges, finding the difference between them and “caring about” would be harder.

Let us start with an example offered by Harry Frankfurt (1988, p. 89, 1999, p. 111). A mother might see a lot of good reasons (e.g. a better future for her child) for abandoning her child for adoption, and she even judges that she ought to do it all-things-considered, but that is what she could not bring herself to do. A parallel case is rejecting euthanasia at the last minute without changing one’s

better judgment that “I ought to die to stop suffering, all-things-considered” in *24 & Ready to Die*. In both cases, two judgments might be made: first, the mother and the patient care about her child and her life respectively, and their caring shapes their lives in a way with which they are so familiar that they might not notice and attach due importance when they deliberate. However, their caring about functions strongly when it is threatened. Second, the mother has a strong parental instinct while the patient has an instinct to live no matter how unbearable her life is. Therefore, when they act, they surrender to strong urges, which render them passive.

However, it does not mean that caring about is equal to instincts and urges, even in cases where it is natural to think of some corresponding instincts. There are two reasons. First, they have different phenomenological features. The brief depiction of the mother’s case which has been offered leaves room for two different interpretations, and more details are needed. Frankfurt (2002, pp. 163-64) suggests that we can imagine two mothers in this situation. Facing her failure to give up her child, mother₁ “may continue to wish that she could give the child away; and she may regret, and even chastise herself for, her inability to do so”, while mother₂ “may recognize her discovery as a revelation not just of the fact that keeping the child is what is most important to her, but also of the deeper fact that it is what she truly wants to be most important to her”. In mother₁’s case, we might attribute her action to her parental instinct, while in mother₂’s case caring about appears.

Second, the way caring about functions is different from that of instinct or urges. Arpaly (2006, pp. 62-63) makes a distinction among “robot causation”, “content-efficacious causation” and “reason-responsiveness”. “Robot causation” functions at a physical or chemical level, while “content-efficacious causation” functions at the psychological level. In mother₁’s case, it is more likely that she is subject to robot causation, and her physical (or neurological) states make it impossible to give up her child whenever she tries. After that, she might try to overcome this urge by establishing or strengthening the opposite robot-causal

chain. However, in mother₂'s case, it is more likely that she would be immersed in a more comprehensive and vivid psychological atmosphere, in which the time she spent with her child, her child's innocent face and lovely eyes flood her and (at least, partially) cause her to act against her better judgment. This process is content-efficacious, because it is the content that makes a difference, so it can be distinguished from the robot causation in mere instincts and urges.

Therefore, even stable instincts with irresistible urges are different from "caring about" at phenomenological and causal levels.

1.2. Caring about Versus Desires

Apparently, desires are more promising candidates, because they have more intimate relations to an agent than those of instincts and urges, which might lead to a richer first-person experience. Moreover, desires have a variety of contents, and their channel of causation is not as simple as "robot causation".

Arpaly and Schroeder (2014, pp. 105-10) offer a minimal theory of caring: for ordinary and healthy people, "one who intrinsically desires the wellbeing of another a great deal is always a person who experiences a broad range of emotions surrounding that person, and is always someone who would be said to care about that other person". They argue that this theory stands, unless we can find a case (which should be nonclinical) in which a person intrinsically desires the wellbeing of another person to a large extent but does *not* care about that person. If counterexamples cannot be found, it follows that X's caring about O can be reduced to a large-quantity and high-intensity version of X's intrinsic desire for O's wellbeing, and no qualitative difference is necessary.

Nonetheless, as far as I am concerned, this is not satisfying, given that the concept of caring about is designed to capture "the most distinctive feature of human life (Frankfurt, 1988, p. viii)". Let us consider an example from *The Master and Margarita*. Pontius Pilate presided over the trial of Jesus and his crucifixion. In this novel, Pilate secretly admired Jesus's philosophy and spirit, and thus he

resented himself for his cowardice and suffered from guilt. Here is an impressive description of how Pilate, and his loyal dog, suffer:

“For about two thousand years he [Pilate] has been sitting on this platform and sleeping, but when the full moon comes, as you see, he is tormented by insomnia. It torments not only him, but also his faithful guardian, the dog. If it is true that cowardice is the most grievous vice, then the dog at least is not guilty of it. Storms were the only thing the brave dog feared. Well, he who loves must share the lot of the one he loves. (Chapter 32)”

Why do both Pilate and his dog suffer? Pilate seems to suffer from his betrayal of what he cares about, and his dog from its owner’s suffering. If we accept the minimal theory of caring, it follows that:

A. Pilate intrinsically desires the wellbeing of Jesus and his philosophy a great deal. = B. Pilate cares about Jesus and his philosophy.

C. Pilate’s dog intrinsically desires the wellbeing of Pilate a great deal. = D. Pilate’s dog cares about Pilate.

A and C are plausible, given that the concept of desires is applicable to both human and non-human animals without difficulties. To say the dog intrinsically desires its owner’s wellbeing does not imply that it could understand what “wellbeing” means. Instead, it suffers when its owner seems miserable to it.

B is acceptable as well: from the way Pilate suffers from his wrongdoing, and the way the suffering shapes his mental life and behaviour, we might think that is the very thing the concept of caring about is meant to grasp. However, D might sound a bit strange. The dog’s reaction might lead us to think it “loves” Pilate and it is willing to share Pilate’s suffering. Nonetheless, if we accept Frankfurt’s way of using the concept of “caring about”, it might be unjustified to say that the dog “cares about” Pilate, because the dog cannot (1) orient itself and arrange its life in a richer way because of caring about; (2) be conscious of that it cares about Pilate more than it cares about its own happiness; (3) feel inner conflict or doubt

if what it cares about is worthy of caring.... Therefore, it is not suitable to say the dog cares about Pilate, at least in Frankfurt's sense, in spite of its intrinsic desires for Pilate's wellbeing.

Two objections should be considered. The first objection is: the seeming unsuitability is due to an impoverished understanding of desires and confusion between "what caring about means" and "how caring about interacts with something else against the background of human life". It is possible that human desires are different from non-human desires in two aspects: first, the former has more contents (e.g. religion, philosophy); second, it shows special richness in the context of human life. Therefore, the equation between a strong intrinsic desire for O's wellbeing and caring about O would stand within the scope of human beings. However, I suppose this does not undermine the argument in this section, the basic idea of which is: The concept of "caring about" is raised to capture something peculiar to human life, which is not intrinsically written in the concept of desires (A similar emphasis on this feature of caring can be seen in Agnieszka Jaworska, 2007a, p. 480); We might overlook it if we accept the minimal theory of caring.

Second, it might sound emphatic to say that non-human animals lack certain capacities, for we face many epistemological difficulties to understand them. Therefore, it is possible that animals like gorillas are "carers" like us (but, according to Jaworska, dogs are not. *Ibid.*, p. 496). However, given what we have known so far, it is plausible to say that caring about is a capacity more complicated than desiring. Even if some animals share it with human beings, it tells us that the capacity might be continuous in the biological world without simplifying it.

1.3. Caring about Versus Conventions and Traditions

It has been shown that caring about can be distinguished from some mental components we are more familiar with. However, is it possible that caring about

is closer to some external factors, like social norms an agent takes for granted or traditions and habits an agent feels constitutive to what she is? The answer would be: no matter how influential the role social background plays in our personal history and mental constitution, it cannot be equated to caring about. We can imagine an agent treats traditions like authoritative orders or physical laws. She does not care about the law of gravity and she might not care about praying before eating food even though she takes them for granted. In my opinion, to say an agent cares about a tradition rather than she conforms to it, it needs to meet two standards. First, the object brings the agent a life which only an agent with certain objects of caring about could enjoy; Second, though the importance of a tradition for an agent might not be sensitive to reasons, it can be sensitive to and surrender to other objects the agent cares about. Let me elucidate it by an example from the movie *Fiddler on the Roof*.

Tevye lives in a conservative village, and at the beginning of the movie, he is proud of the traditions he has. He says “because of traditions, everyone knows who he is, and what God expects him to do.” He makes judgments according to the traditions and arranges a marriage for his daughter. Therefore, when a poor tailor tells Tevye that he wants to marry Tevye’s daughter, and they have already given each other a pledge, Tevye gets irritated. He says:

“You give each other a pledge? They give each other a pledge! Unheard of! Absurd!... Unthinkable! Where they think they are? In Moscow? In Paris? ... Who do you think you are? You are nothing! Something I cannot allow! TRADITION! Marriages must be arranged by the papa. They should never be changed!”

However, when Tevye sees the tailor’s determination, he begins to consider this marriage more carefully.

“He’s beginning to talk like a man. On the other hand, what kind of marriage it would be with a poor tailor. On the other hand, he is an honest

hard worker. But on the other hand, he has absolutely nothing. On the other hand, things could not get worse for him but only gets better. They give each other a pledge, unheard of! Absurd!... But look at my daughter's face. She loves him. She wants him. And look at my daughter's eyes, so hopeful! TRADITION!"

Then he tenderly asks the lovers: "Well, children, when shall we make the wedding?"

In this case, Tevye cares about the tradition, not because he was taught to strictly conform to it at an early age, but because the relation between him and tradition meets the two standards: First, Tevye not only respects the tradition, but thinks tradition infuses his life with meaning and orientation; Second, facing his caring about his daughter, which is a counterpart, his caring about traditions surrenders.

1.4. Caring about Versus Pro-attitudes

Now let us consider an idea more attractive than those we have discussed: an agent's caring about O implies her pro-attitudes toward O to some extent. In other words, given that an agent cares about O, it is necessary that she sees O or some properties of O as good, desirable, suitable, caring-worthy. She might notice it and takes it as a reason to act. Or, she dimly holds a pro-attitude towards it and thus she is disposed to benefit it. In the former case, her caring about would be part of her practical reasoning, while in the latter, she has a favourable cognitive state on it, which might also be integral to a non-deliberative process of reason-responsiveness.

However, caring about is relatively independent of the evaluative system empirically and theoretically. At the empirical level, according to Frankfurt, an agent caring about O could hold a variety of attitudes towards O: sometimes, she holds an "altogether neutral attitude of acceptance" (2002, p. 161) (e.g. "alright, I

love fame and money”), sometimes “a welcoming approval” (e.g. “that is who I am, an ambitious young person seeking reputation and fortune!”), but sometimes it may come with a “weary resignation” (2006, p. 8) (e.g. “I confess: I am a vulgar person, who merely cares about fame and money”). A worry is: is it possible that they are not sincere when they make these claims? For example, the “vulgar person” holds a pro-attitude to “fame and money” indeed, but he realises that other people would not think highly of him, so he puts it in this way. Therefore, it is risky to completely rely on these examples, so we need to further expound the relation theoretically.

On the one hand, if we suppose an agent cares about O, it is hard to imagine that she could not see any value in O (otherwise she might be pathological). Moreover, she might hold a pro-attitude towards that fact that she cares about O. For example, she feels that this relation gives her life purpose and meaning (e.g. she cares about studying philosophy), helps her to build her character (e.g. she cares about the happiness of non-human animals) and guides her to act. When we reflect on what we care about, we could find some good reasons to act in a certain way because of it. Even if we do not reflect, we might follow it naturally.

On the other hand, an agent’s caring about O is not always sensitive to the attitude she has towards O and her relation to O. This becomes obvious when another object Q appears. She could not pursue O and Q at the same time, and she sees more value in Q than O, or she sees more value in pursuing Q than O. Therefore, in this situation, her prior pro-attitude towards O is offset, withdrawn or turned to be a con-attitude towards O, when she has to decide to pursue either O or Q. She might decide to pursue Q, but it is possible that she could not do it, because she could not bring herself to care about Q. As Frankfurt puts it:

“The fact that someone cares about a certain thing is constituted by a complex set of cognitive, affective, and volitional dispositions and states. It may sometimes be possible for a person, by making a certain choice or decision, effectively to bring it about that he cares about a certain thing or that he cares about one thing more than about another. But that depends upon conditions

which do not always prevail. It certainly cannot be assumed that what a person cares about is generally under his immediate voluntary control.” (1988, p. 85)

This point is significant, when we try to develop an interpretation of some cases of *akrasia* later.¹³

After comparing caring about with instincts, urges, desires, traditions, conventions and pro-attitudes, we can see some features of it: first, unlike instincts and urges, it does not influence people at a mere physical and chemical level; second, compared with desires, it is peculiar to human life; third, it might be coloured by traditions or conventions, but it means more than them; and fourth, it might imply a pro-attitude towards the object of caring about (say, O). However, caring about and pro-attitude can be relatively independent, when a stronger pro-attitude towards another object (say, Q) defeats that towards O but fails to change the fact that the agent cares about O. That is not to say, without instincts, urges, etc., an agent could care about a certain object nonetheless. However, given all the distinctions made in this section, we cannot find easy ways to reduce caring about to concepts we are more familiar with. In the next section, I shall focus on a relevant concept – volitional necessity, to show how “caring about” functions.

2. The Second Concept: Volitional Necessity

“Volitional necessity” is a genuine “necessity” because it does reduce the number of alternates an agent faces. This kind of necessity “renders certain actions impossible for him – not by depriving him of the capacity to perform them but by making it impossible for him to use that capacity.” (Frankfurt, 1999, p. 111) It is “volitional necessity” because where it originates from is an agent’s own volitional essence (i.e. what she cares about).

¹³ Here is a stronger version of the argument (Jaworska, 2007b), which I will not discuss further. The basic idea is that “caring about O” means the agent “is identified with O”. This is ontological identification in the sense that “caring about O” would constitute a structure which “plays a vital role in creating and maintaining the unity of agency over time” (p. 568). By contrast, “a pro-attitude toward O” means an agent identifies herself with O, which is a kind of “subjective identification”. The ontological identification is more fundamental and it can be non-reflective and non-judgmental.

It is because we care about certain things that some options are not genuinely open to us. We can think about the case in the preceding section. Since the mother cares about her child so much, abandoning her child is not a genuine option available to her, though she could do it easily if she did not care about her child. “When someone is tending to be distracted from caring about what he cares about most, the force of volitional necessity may constrain him to do what he really wants to do.” (Frankfurt, 1988, p. 87). The exclusion of these options is not caused by external forces but by what we care about. Here, some contrasts can be made.

There are two other kinds of necessity. One is the necessity of desires: in this case, we might feel a strong aversion when we think about abandoning someone in our family, which renders it impossible. The other is the necessity of reason (Frankfurt, 2004, p. 46): we might find that the reason against abandoning a family member is overwhelming, and we cannot act against that reason. We can draw a parallel between theoretical reasoning and practical reasoning. Given the scientific knowledge an agent has already known, it seems impossible for him to believe that the age of the Earth is only several thousand. Likely, “I am powerless to form the intention to lop off my leg, given my other (non-destructive) doxastic and practical commitment”. (Moran, 2003, p. 404)

Both of the two necessities might coexist with the volitional necessity and overdetermine an action. Take Pilate in *The Master and Margarita* for example. This can be a case of overdetermination, because Pilate’s self-blame is necessitated by i. his strongest desires (e.g. his desires for being moral and shouldering his responsibility render stopping blaming himself “impossible”) ii. conclusive reasons (e.g. he sees himself as blameworthy for wrongdoing, and takes it as a conclusive reason to blame himself) and iii. what he cares about (e.g. he could not stop caring what he cares about and blaming himself for betraying it, so he could not be indifferent towards what he did).

However, they are not the same thing. We can see their differences, if we

return to the mother's case¹⁴. She might see a lot of good reasons (e.g. a better future for her child) for abandoning her child for adoption, but she could not bring herself to do it. In this example, she faces a kind of necessity rendering a certain act impossible. It is not the necessity of acting for conclusive reasons, because she might think she has conclusive reasons to abandon her child. It is not the necessity of desires either. If what motivates the mother were a mere desire for keeping her child, the mother would have two equally plausible ways to cope with it: either to satisfy the desire or to give up the desire. If this desire is, in fact, a kind of obsession or infatuation, the agent might favour the latter way over the former way. We can imagine such a mother. However, there is another possibility: the mother might find that the latter is not available to her, because this motivation itself is what she feels reluctant to give up, despite that it leads to *akrasia*. (2004, p. 12).

The divergences are not surprising, given that we have discussed the differences between caring about and pro-attitudes, and those between caring about and desires, instincts and urges. Volitional necessity flows from caring about, while the other two kinds of necessity are linked to pro-attitudes and desires respectively. Volitional necessity is not a stream that necessarily flows into either the branch of the necessity of desires or that of the necessity of reasons. Therefore, it is *not* the case that when a person acknowledges the internality of a "passion" which she disapproves of in judgments, only two versions of interpretation are available: one is that "the person is simply resigned to the passion of which he disapproves" while the other is that "the person goes on, perhaps out of resignation, to decide to treat that passion as reason-giving". (Bratman, 1996, p. 18). An alternate interpretation is that the person does not simply resign to the passion, and she could treat it as necessitated by her caring rather than reason-giving.

Despite these differences, Frankfurt also points out that we can draw a

¹⁴ A similar case is offered by Mele (1987, p. 19): A man "after careful deliberation, judges it best to put the horse out of its misery and intends to shoot it in the head here and now. While aiming at its head, he notices the sad expression in its eyes and, still thinking that *it would be best to shoot it*, he intentionally refrains from doing so."

parallel between the volitional necessity and the necessity of reason in some aspects, because in both cases “the inability to forbear is not a simple matter of deficient capacity on the part of the agent” (1988, p. 66). Rather, the two kinds of necessity constitute important capacities (i.e. to reason and to care about respectively) peculiar to human agents. This distinguishes both of them from the necessity of desires, which seems to be more atavistic. Moreover, a parallel can also be drawn between volitional necessity and the necessity of desires, for in both cases, “the agent experiences himself as having no choice but to accede to the force by which he is constituted even if he thinks it might be better not to do so” (Ibid.) This distinguishes both of them from the necessity of reason, which is always sensitive to evaluations.

3. Conclusion: Two Discrepancies

Since caring about is distinguished from mere desires, there can be a discrepancy between the strongest desires motivating an agent’s action and what the agent cares about the most, which is clearly depicted in Frankfurt’s early proposal. When an unwilling addict identifies herself with the desire to resist the drug while being motivated by the desire to take it (1988, p. 18), there is a discrepancy between her strongest desire and her higher-order volition, which can be traced back to what she cares about.

Likely, caring about cannot be reduced to evaluations (e.g. taking a certain object as good), because it is not necessarily sensitive to judgment, decisions, deliberation and other reason-responsive activities. “To care about something differs not only from wanting it and from preferring it but also from judging it to be valuable. A person who acknowledges that something has considerable intrinsic value does not thereby commit himself to caring about it. (1999, p. 158)” However, this discrepancy is vaguer than that between a desire and caring about. This is because, unlike desires which might be alien to a person, evaluations and better judgment usually imply an agent’s commitments and a first-person justification. It

follows that the relation between reason-responsive activities and caring about might be one necessarily within the human agency.

If we can leave room for a discrepancy between reason-responsiveness and caring about, we might find a way to cope with some complicated cases of akrasia, the cause of which is neither the discrepancy between desires and reason-responsiveness (as the Irrational-passive Proposal supposes) nor one within reason-responsiveness (as the Reason-responsive Proposal suggests) but one between the one's volitional necessity and reason-responsiveness. To elucidate this discrepancy is the aim of the next chapter.

Chapter 4 The Out-of-caring-about Proposal

In Chapter 3, I explain what caring about and volitional necessity mean, the roles they play in human agency and the peculiar contour of agency they contribute to. In this chapter, I shall use these concepts to construct a framework for some cases of akratic action.

I shall analyse what the better judgment is in some akratic action in section 1. In addition, I shall argue that since there can be a mismatch between the degree to which we care about O and O's weight as rationalising reasons, then it is possible that caring about independently motivates an action sidestepping the reason-responsive process. Following this approach, we can make sense of some puzzling cases of akrasia, in which the agents are not overwhelmed by urges for some objects, but act out of their caring about the objects. Let us call it the "Out-of-caring-about Proposal". In section 2, it will be applied to some cases we have met. In section 3, three features of it will be shown. Then, it will be fit into the big picture in which caring about interacts with inner circumstances, such as attention (Mele, 1987), and emotions (Tappolet, 2003), and external circumstances such as social expectations (Rorty, 1997).

1. The Framework of the Out-of-caring-about Proposal

1.1. Three Motivations

There are three motivations to consider the Out-of-caring-about Proposal:

First, with a fine-grained understanding of akrasia, we might find a puzzling type in which the agent, though acting against her all-things-considered better judgment, stands a more intimate relation towards her action than that in another type, in which an agent is overwhelmed by an urge. Correspondingly, the information conveyed by the action is more philosophically interesting and it might demand further interpretation. For example, in the cases taken from *Fiddler on*

the Roof and the *Adventures of Huckleberry Finn*, we might have the intuitions that: it is *not* the case that Tevye is merely spoiling his daughter and, luckily, he does something good from an objective perspective; it is *not* the case that Huck Finn only reveals himself as a weak-willed child though his action is good; they are cases different from those such as eating chocolate akratically. Call the former “interesting cases of akrasia” and the latter “trivial cases of akrasia”. However, if we accept the divided “Reason-Desire” picture, the “Irrational-passive Proposal” flows from it: at a deeper level, both the “interesting” and the “trivial” cases of akrasia share the same “uninteresting” origin (e.g. the blind Desire’s triumph over Reason) and the same outcome (e.g. an agent’s passive state). By contrast, with a richer understanding of human agency, we might find differences between the agent’s relation to her akratic action in the “interesting cases of akrasia” and that in the “trivial cases of akrasia”, then our intuitions might be sharpened and defended. This motivates us to set the Irrational-passive Proposal aside and consider the Reason-responsive Proposal.

Second, the Reason-responsive Proposal shows us the relation between an agent’s mental states and her akratic action in some cases. For example, Huck Finn’s mental states that motivate his akratic action are sensitive to reasons, and his akratic action is an action for reasons. Deliberation, combined with better judgment, does not exhaust reason-responsiveness. However, a question would be: is this an over-rationalisation, which might overlook some significant features of this case? After all, Huck Finn does put all of his effort to respond to reasons in deliberation; Huck Finn thinks helping Jim is absolutely wrong; Huck Finn feels that his reluctance to turn Jim in and his motivation to help Jim are somehow insensitive to reasons, impenetrable and recalcitrant. They are phenomenological features as important as a defence of Huck Finn’s praiseworthiness, and to accommodate them might motivate us to leave the Reason-responsive Proposal for a while.

Third, if the explanation of the concepts of caring about and volitional necessity in the preceding chapter is on the right track, this might give us a hint

of the plausibility of a proposal based on them. The two concepts are different from desires and their blind mechanism, and they can be distinguished from reason-responsiveness and its reason-sensitive mechanism. Therefore, if a proposal to make sense of interesting cases of akratic action can be developed from them, it might have two advantages: First, an action out of caring about, no matter what its relation to the better judgment is, would be different from one motivated by “desires”; Second, seeing an akratic action as one out of caring about, due to its relative independence of reason-responsiveness, might help us to better grasp the first-person phenomenological features. The advantages are absent in the Irrational-passive Proposal and the Reason-responsive Proposal respectively, and it is their absence that motivates us to make the move.

Here is a point of clarification. Some readers may find a certain move motivated by a certain concern misleading or unnecessary. For example, they might think the Irrational-passive Proposal is satisfying, and the distinction between “interesting” and “trivial” cases are superficial; they might find the Reason-responsive Proposal plausible and believe that there is no need to take the first-person experience as seriously as I do; or, they might argue that caring about is reducible to either desires or reason-responsiveness, so the Out-of-caring-about Proposal would be hopeless. Nonetheless, the plausibility of a more developed version of the Out-of-caring-about Proposal might in turn prove that taking this route is worthwhile.

1.2. A Better Judgment and What One Cares about

An agent might get an all-things-considered better judgment by deliberating in the following way: I think Φ -ing is better than ϕ -ing all-things-considered, and I know to Φ I shall γ at t ; Now it is t , I ought to γ . It is possible that she arrives at a better judgment immediately, and holds it unless new evidence motivates her to reconsider it or change it. In this section, I shall focus on two features of the better judgment: on the one hand, to get a better judgment (not a pseudo one discussed

in Chapter 2), an agent could see conclusive reasons for it, which override reasons against them. This process is sensitive to reasons, and the agent gives her first-person justification to it. On the other hand, despite the sincerity of her endorsement of the better judgment, the agent cannot take her better judgment as the only way she engages the situation.

One cannot be a mere bystander to her better judgment. Admittedly, we might find an idea out of the blue in our mind, and sometimes it is in the form of a judgment (e.g. I shall move to another city). However, since it is not located within a web of reasons and considerations, we tend to take its form as a mere guise. By contrast, when an agent deliberates to get a better judgment, she does take factors she sees as relevant into account and monitors how the process goes. She pays special attention to the terminus of the process, which is supposed to inform her of “what is better”. This degree of the engagement is sufficient for us to think that she is active.

In some other cases, the better judgment is not so obvious, but it would be revealed when it is challenged. For example, K does not make an explicit all-things-considered judgment in the form that “I ought to keep my promise to go to my friend’s birthday party tonight, though staying home and watching my favourite TV show would be enjoyable”. Without any conscious deliberation and self-command, she puts her friend’s birthday party into her schedule in the morning and buys a gift in the afternoon. However, before leaving her home for the party at night, it occurs to her that she will meet Q at the party, whom she dislikes, and thus she feels reluctant to go. Then she tells herself: “You should go, otherwise your friend would be disappointed and you would break a promise!” Then due to the new information, she forms an explicit better judgement after a fluctuation.

In the cases in which the justification of the better judgment can be seen by the agent, we can see that the agent is active in collecting relevant information and making inferences. As for the cases in which the better judgment becomes obvious only after new factors are input, the agent does not lose her rational monitor. Both are possible in akratic action. An agent might explicitly hold a better

judgment before she finds that she is motivated to act against it. It is possible that the better judgment is noticed by her only when the opposite motivation is felt. What is of significance here is that in akratic action, better judgment necessarily implies engagement from the akratic agent's rational capacities. Arpaly thinks, in cases of inverse akrasia, the agents are "incompetent abstract thinkers" (2003, p. 78). However, the process of getting a better judgment is *not* that an agent engages in a situation at an abstract level. Since the agent does take some substantive commitments, interests of him and the outcomes of his action into consideration, any "bad" but "real" better judgment is not a mere mistake in "abstract thinking".

However, it does not follow that forming a better judgment is the only way the agent can be seen as active¹⁵. The agent might be aware of an overriding motivation to act against her better judgment, which could interrupt her at different stages: (1) even if she has a better judgment, she cannot form an intention to carry out it; (2) even if she has an intention to carry out her better judgment, she cannot bring herself to try to realise her intention; (3) even if she brings herself to try to realise the intention, she fails. What matters here is not the exact point at which the motivation against the better judgment enters, but the first-person experience the agent has.

In the preceding chapter, I have argued that the degree to which an agent cares about a certain object does not necessarily reflect how much she values it. This might be used to explain what happens here. An agent could value the object to a large extent while not caring much about it. Then she might accept this discrepancy as given or feel frustrated about it. For example, though admiring a self-disciplined friend who always studies hard, resists any temptation, and makes progress steadily, J might admit that she could never do the same thing, not because she does not see the value of her friend's way of doing things, but

¹⁵ An example of the argument for the privilege of better judgment is David Pugmire's (1982, pp. 188-89). He argues that by forming a better judgment an agent "put himself through a deliberation that opened him, as much as anything in his power could, to what he was doing, and he reached a dissuasive all-things-considered value-judgment, resolved and set himself against what he then did anyway". These are his best efforts, if they are defeated by a desire then the agent cannot do anything and the desire is irresistible.

because she does not care about her career or love the subject as much as her friend does. Likely, an agent might take an action or object as bad while noticing that she cares about it to a large extent.

This discrepancy does not imply that caring about and reason-responsiveness are irrelevant. First, it does *not* mean that: the way an agent values different objects reflects nothing about whether she cares about these objects. J does care about studying hard, a self-disciplined life, and a brilliant career, and that is why she could appreciate their value rather than see them as some standards for a good student written on a Student Manual. However, J personally cares more about some transient pleasure, socialising, her leisure time and so on.

Second, it does not mean that: J cannot develop reasons from the self-knowledge of how much she cares about different objects. Before an exam, J might think “well my friend is studying hard. How diligent she is! She will get a high mark as usual. As for me... Let me enjoy my party!” She does not need to deliberate in this way, but she is acting for some good reasons, and they are good reasons to her because they fit what she cares about well.

However, the discrepancy still exists. It would manifest when an agent finds that: from caring about O , she cannot form a rationalising reason to Φ as strong as some reasons not to Φ . For example, we can imagine that the exam J faces is very significant. J thinks, all-things-considered, she ought to study hard rather than go to a party. She finds her caring about pleasure would not rationalise her going to the party. If she does it akratically nonetheless, we might have two interpretations. The first is trivial: she cannot resist temptation. The second is out-of-caring-about: she is more careless about her future and more short-sighted than she supposes, and the degree to which she cares about pleasure is higher than that she would be comfortable with. She might feel frustrated about herself, but the object to which she feels frustrated is still “herself”, which is more accurately depicted by her action than her judgment in this case.

1.3. The Proposal

According to the alternate proposal, what Arpaly calls inverse akrasia (e.g. the racist shopkeeper's and the sexist father's cases in Chapter 1, Huck Finn's case in Chapter 2, and Tevye's case in *Fiddler on the Roof* in Chapter 3) can be interpreted as akratic action out of caring about. That is to say, agents might not act for reasons better than those supporting their better judgment. An alternate interpretation is:

First, an agent X cares about an object O (the Fact of Caring);

Second, an action Φ -ing is beneficial to O or the relation between X and O (the Fact of a Beneficial Effect of Φ -ing);

Third, X sees some conclusive reasons to avoid Φ -ing despite its benefits to the object O or the relation between X and the object O (the All-things-considered Better Judgment against Φ -ing);

Fourth, X cares about the object O more than her commitments (e.g. the appreciation of the value of the corresponding objects or principles) underlying the conclusive reasons (the Relevant Degree of Caring);

Fifth, X acts against the conclusive reasons due to volitional necessity, which is the channel through which her caring about functions (the Akratic Action)¹⁶.

The advantage of the Out-of-caring-about Proposal is that it could explain some phenomenological features of this kind of akrasia. For one thing, we can see X's appreciation of the value of her better judgment at the third stage. Therefore, the judgment she makes is not "pseudo" because it shares a similar process with a better judgment in an enkratic action so far. That is to say, what happens in an enkratic action might be a similar series of the first, second and third stages. Both the akratic agent and the enkratic agent¹⁷ take the beneficial effects of Φ -ing as a reason for Φ -ing, but it is overridden by other considerations

¹⁶ It does not mean that this process is linear. An agent might experience one stage repeatedly in the real temporal sequence.

¹⁷ By an "akratic agent" and an "enkratic agent", I am not supposing that they have different characters or habits of being "akratic" or "enkratic" in various situations. Instead, an akratic agent is called in this way because of one akratic action she does in question, and an "enkratic agent" would mean her counterpart in a similar situation.

to avoid Φ -ing. The overriding considerations constitute their better judgment, the justification of which is seen by both the akratic and the enkratic first-personally. At the reason-responsive level, the difference might be insignificant.

For another, it might explain the first-person experience of an akratic agent. The difference between an akratic agent and an enkratic agent occurs when the transparent moments of the better judgment slide at the fourth stage. The better judgment then falls into the opaque mental background of the agent as a whole, which is, however, the indispensable context for the judgment to serve its practical function. Then the better judgment's efficacy is limited. This is because it might face either a hospitable context or the opposite, and the resources that it is able to use are not always sufficient. It is likely that an agent's caring about has shaped the agent by volitional necessity and precluded some options seemingly open to her. This is not always transparent to the agent's reason-responsive capacity, nor would it conform to her reason-responsive capacity.

Nonetheless, this does not mean that the two capacities (i.e. one of responding to reasons and the other of caring about) of the human mind would always disagree with each other. Rather, two points should be made: first, in non-clinical cases, an agent can be aware of her caring about O to some extent and she could take it as a reason. Therefore, her caring about O does not lie at the unconscious level that interests psychoanalysis.

However, secondly, the two mechanisms are not necessarily in lockstep. At the third stage, the reason-responsive mechanism might misinterpret the mechanism of caring about. Or, it could interpret the mechanism of caring about correctly but it attempts to calibrate it instead of mirroring it. Both of them might end up with either an akratic action or a struggle for the strength of will.

Therefore, we could interpret Huck Finn's case in the following way:

First, Huck Finn cares about Jim or something relevant, such as the fact that he does not harm anyone (the Fact of Caring);

Second, helping Jim is beneficial to the object he cares about (the Fact of a Beneficial Effect of Helping Jim);

Third, Huck Finn sees some conclusive reasons to not help Jim, despite that by helping Jim what he cares about could benefit (the All-things-considered Better Judgment against Helping Jim);

Fourth, Huck Finn cares about Jim or something relevant more than his commitments (e.g. protecting Jim's owner's "property", avoiding being blamed or despised by others) underlying the conclusive reasons (the Relevant Degree of Caring);

Fifth, Huck Finn acts against his better judgment due to volitional necessity, which is the channel through which his caring about functions (the Akratic Action).

In addition, there is another kind of akratic action, which can be called "action out of carelessness" in the sense that an agent does not care about her better judgment enough. This might lead to a different first-person experience: "cool akrasia".

Christine Tappolet (2003) thinks akratic action can be made intelligible partially by reasons supporting the akratic action (although they are overridden in deliberation) and partially by emotions motivating akratic action. Then, the concept of "cool akrasia" is used to describe akratic action in which an agent is not influenced by any especially strong emotions. For example, although an agent sees some conclusive reasons to avoid smoking, she smokes without any obvious struggles or emotional fluctuation. An interpretation offered by the Out-of-caring-about Proposal would be: the degree to which the agent cares about her better judgment is low. Then, following what she cares about slightly more would not be an especially bitter choice, although it implies akratic action.

Can we call akrasia out of caring about "hot akrasia" and akrasia out of carelessness "cool akrasia"? It seems that in akrasia out of caring about, an agent cares about both her better judgment and the object benefiting from her akratic action. Therefore, although she finally acts out of what she cares about more against her better judgment, many bitter emotions and feelings might be involved in the process. Therefore, it is "hot akrasia". By contrast, akratic action out of carelessness is "cool akrasia", because although an agent's better judgment is

supported by conclusive reasons, an agent does not care about it to the extent that giving up it would make a significant difference on her feelings.

Therefore, I think this distinction might be plausible if we leave room for adjustments according to different social contexts. Let us reconsider the example taken from *Mencius* in Chapter 2. Shun's father murdered a man. As a ruler, Shun lets Gao Yao, a judge, apprehend his father. However, after it, "Shun would have regarded abandoning the realm as he would abandoning an old shoe. Secretly, he would have taken his father on his back and fled, dwelling somewhere along the seacoast. There he would have happily remained to the end of his life, forgetting, in his delight, about the realm." In this case, we can see a feature of hot akrasia: Shun cares about both law and his father; and a feature of cool akrasia: Shun does not feel especially guilty or ambivalent about acting against the law. I do *not* mean that in this case, Shun takes what the law tells him to do as his better judgment, which renders his action akratic. Rather, I intend to draw attention to a possibility (which will not be further discussed here): in a background where people do not necessarily give priority to reason-responsiveness, it might have more tolerance towards acting out of caring about though it cannot be fully rationalised. Therefore, even if an agent cares about both sides (i.e. her better judgment and the object which would benefit from her akratic action), this does not necessarily place her into a "hot" state. Rather, she might be cool, because she thinks it is quite natural to benefit what she cares more about, regardless of what her better judgment tells her, and this is tolerated or even implicitly supported by society.

2. Three Features of the Out-of-caring-about Proposal

This section will discuss three features of the Out-of-caring-about Proposal.

2.1. The Symmetry between Akrasia and Enkrasia

As shown in Chapter 1, the normal way of putting the puzzle of akratic action is to ask: why does an agent fail to follow her better judgment? By formulating the question in this way, we usually presuppose that there is something absent or excessive in akratic cases, compared with enkratic cases. An enkratic action becomes intelligible when we see conclusive reasons an agent responds to by forming and following a judgment. However, an explanatory “gap” appears in akratic action. Even if we see the conclusive reasons an akratic agent tries to respond to, this could not explain her action. In this sense, they are asymmetric.

However, in the Out-of-caring-about Proposal, we can see a symmetry between out-of-caring-about/carelessness akrasia and enkrasia. That is to say, the basic stages and the factors involved in them are similar. It is what happens at these stages and how these factors interact that make their outcomes (i.e. an akratic action and an enkratic action) different, and no explanatory “gap” is needed. Here, I agree with Michael Stocker (1979). He thinks that an agent has a complex array of psychic structures, which could influence whether an agent could act for the believed good. In cases of akrasia, the agent’s better judgment, showing the believed good, might be undermined when it interacts with the complex array of psychic structures. By contrast, in the cases of enkrasia, the better judgment could be maintained or even strengthened by the structures. Therefore, he suggests that “if weakness of will, desiring the (believed) bad, is problematic, so is strength of will, desiring the (believed) good. We must replace those moral psychologies which generate the traditional philosophical problems about weakness of will. We need moral psychologies that recognize, in general, the complexities of the psyche and, in particular, those complex arrays of psychic structures of mood, interest, energy, . . . and also the complex mediating roles played by these arrays between motivation and evaluation”. (Ibid., p. 753) Therefore, our concerns in explaining enkrasia and akrasia are asymmetric, rather than the phenomena themselves. When we try to explain the former, we normally focus on reasons. Only when we find the insufficiency of this method in akratic cases do we resort to something else.

The Out-of-caring-about Proposal could be seen as an attempt on this track. It fits “caring about” into the psychic structures, and classifies a kind of akratic action in which caring about changes the result as out-of-caring-about/carelessness akratic action. Admittedly, the symmetric structure needs further defence. The aim here is not to do this work. Rather, the point is that: The Out-of-caring-about Proposal seems to reject the asymmetry between akrasia and enkrasia, which might be the core of the puzzle of akrasia at first glance. However, it does not mean the proposal gets lost, because it tries to cope with the phenomena without presupposing the asymmetry, and the asymmetry is not inherent in the phenomena.

2.2. The Praiseworthiness of Out-of-caring-about Akratic Action

So far, we try to use the Out-of-caring-about Proposal to understand akrasia at the descriptive level. Now, a series of questions might be asked: does the Out-of-caring-about Proposal change anything at the normative level? Can people who act out of caring about be rational? Are they praiseworthy?

Inverse akrasia attracts Arpaly mainly because agents, like Huck Finn, seem to be akratic and praiseworthy. According to her account, three points can be made:

(1) If an agent acts for moral reasons, which are some right-making features of her action, then she is praiseworthy;

(2) A further explanation of why she acts in this way: she has the right kind of moral concern;

(3) Even if an agent acts against her better judgment, she could (3.1) have the right kind of moral concern and (3.2) engage in this good-moral-reason-responsive process. Therefore, (3.3) she is praiseworthy.

The Out-of-caring-about Proposal would reject (3.2). According to it, an agent who acts akratically out of caring about is not rational in Arpaly’s sense. Although she has some reasons to act in this way, these reasons are overridden

and they do not offer the main motivation to her akratic action. The fact that the agent cares about something makes a difference. However, the Out-of-caring-about Proposal would *not* reject (3.1) and (3.3). That is to say, if an agent has the right kind of moral concern (or, she cares about something with moral value), then she could be praiseworthy for acting out of caring about it. (3.2) is not necessary for her praiseworthiness.

To explain this point, let us consider an example taken from Amelie Rorty (1997, p. 651): a stockbroker believes that fleecing her infirm and aged client is the best thing to do. “The client is at death’s door.... Her client’s family would give him better care if he were to become indigent. Still, for all her reasoning to the contrary, Angela might be akratically unable to profit from an easy chicanery... she is, as she sees it, irrational - and moral.” It is not sufficient to say that she acts for moral reasons, which are overridden when she makes a better judgment. It is likely that she acts out of her caring about morality, without changing the justification (she sees) of acting immorally in this case. Nonetheless, we tend to think the stockbroker is praiseworthy.

This might lead us to think: if an agent cares about something with moral value, and this is reflected in her action, then she is praiseworthy. Whether the channel is reason-responsive or not is not significant.

On the one hand, a similarity can be found between this account and Arpaly’s. She offers a version of interpretation of Huck Finn closer to that suggested by the Out-of-caring-about Proposal: “Huck, not a sophisticated deliberator, does not draw the conclusion so clear to the contemporary reader – that he morally ought to help Jim just as he would help any other friend, perhaps any other innocent person – but when he cannot bear to turn Jim in, it is at least partially a response to Jim’s humanity; *Huck cares about all humans and thus does not want some things done to them or happening to them.*” (Arpaly, 2015, p. 142, emphasis mine) According to this interpretation, it is Huck’s caring about all humans that makes him sensitive to moral reasons even if his better judgment gets it wrong. Since the praiseworthiness of responding to moral reasons is more certain, Huck’s

praiseworthiness could be defended.

On the other hand, this could be a risky move. Huck Finn is not like the stockbroker who does something moral, because it is, as she sees it, moral. Three questions can be asked to see the differences: first, does what they care about have moral value? In the stockbroker's case, she cares about what she treats as moral principles, such as "do not cheat". This has moral value. In Huck Finn's case, what he cares about is not certain (as we have seen in the interpretation offered by the Out-of-caring-about Proposal). We could say that he cares about Jim (as a friend or a fellow creature), and could not bear Jim's broken hope for freedom. Or, he might care about that he does not deprive the freedom of someone or make someone miserable. Even if we could not find a specific object to fix Huck Finn's caring about, we could see that all of these options have moral value to some extent.

The second question is: Can they see moral value in what they care about, which is not overridden by any other moral considerations they have? The stockbroker could see moral value in her moral principles. This is not overridden by any other moral considerations, although it is overridden by financial considerations. By contrast, Huck Finn could see moral value in his caring about Jim, or his not depriving the freedom of anyone. However, this is overridden by what he sees as morally better. Compared with it, the moral value of what he cares about is little. For Huck Finn, it even turns out to be "evil".

Now we can consider the trickiest question: is Huck Finn's acting out of caring about something with moral value praiseworthy, despite the fact that (1) he could not find any non-overridden moral value in it and (2) he even sees it as evil? It is parallel to the question we meet when we take the reason-responsive route. That question is: is Huck Finn acting for moral reasons and is he praiseworthy for doing so despite the fact that (1) he does not take them as moral reasons and (2) he even sees his action as evil? Now the Out-of-caring-about Proposal rejects that Huck Finn's action is basically one for moral reasons, so it could not appeal to Huck Finn's responding to moral reasons as a defence of his praiseworthiness.

However, one answer is available to the Out-of-caring-about Proposal. It can say: Huck Finn is praiseworthy because he has the kind of caring about that is rooted in a morality-promoting living foundation and contributes to a “civilised” moral picture. Huck Finn’s caring about what he takes as a moral principle (i.e. “Return a black slave to his owner”) might not have moral value for two reasons: (1) we do not think the mere fact that one cares about what she takes as “moral principles” has any moral value, because the principles she holds can be terrible or immoral; (2) we see something as having a-historical moral value, like equality and freedom, which could override some historical conventions.

By acting akratically, Huck Finn cultivates the morality-promoting living foundation. Although in this specific case and for the specific agent, the act is not morality-orienting, it contributes to the ideal moral picture. It is not the case that, to get moral reasons, we can completely rely on what we care about and trust its mysterious tendency to cause moral enhancement. The point is that: if an agent does *not* see the rationalising power and the moral good of an action, she might act out of caring about something which we could see as the “bright” side of humanity and a positive signal of the human future. It is likely that we praise Huck Finn for the same reasons why we praise morality.

However, this might be an uninteresting point. Even if we do not take the long route and consider the three proposals of understanding the case, we could get a similar judgment based on common sense: Huck Finn is praiseworthy because what he does is morally good from a modern perspective, although he does not act for moral reasons (because he fails to see them as moral reasons), nor does he act out of caring about morality (because he even sees evil in his caring about Jim). Nonetheless, the defence offered by the Out-of-caring-about Proposal is different from this pre-theoretical view, because it is historical: it leaves room for the historical process of exploring moral reasons, the possible interaction between morality and caring about, and a more complicated explanation of Huck Finn’s action, as he sees it rather than as we see it.

2.3. The Capacity of an Akratic Agent

Two features of the Out-of-caring-about Proposal have been discussed. The first is that it supports a symmetric understanding of the explanation of akratic and enkratic actions, and the second is that its implication on an agent's praiseworthiness is more complicated and subtler than that of the Reason-responsive Proposal. Its third feature, which is shared by the Reason-responsive Proposal, is: according to it, the locus of akrasia could be a kind of capacity rather than inability. We cannot offer a general diagnosis of akratic action by merely saying that the agents are "weak" in certain aspects, or they fail and get nothing. Instead, we need to interpret the rich information it conveys. This point is nicely made by Frankfurt and Arpaly:

"If a person who is constrained by volitional necessity is for that reason unable to pursue a certain course of action, the explanation is not that he is in any straightforward way too weak to overcome the constraint. ... Unlike the addict, he does not accede to the constraining force because he lacks sufficient strength of will to defeat it. He accedes to it because he is unwilling to oppose it and because, furthermore, his unwillingness is itself something which he is unwilling to alter." (Frankfurt, 1988, p. 87)

"Similarly, Huckleberry Finn need not be particularly lacking in self-control – it may simply be the case that his reluctance to treat a human being like escaped livestock is stronger than any inhibitions he may have, which does speak well of him." (Arpaly, 2003, p. 99)

If we could interpret akratic action in this fine-grained and neutral way, we might find more methods to overcome it or live with it. To be enkratic, normally, an agent is suggested to try hard to conform to her better judgment and eliminate any disruptive factors. Now more options are available:

(1) Change the better judgment. Instead of focusing on this local inconsistency, the akratic agent could try to get more knowledge of herself. She could ask herself: "Is this better judgment really better? Given who I am, is this

better judgment really better / practical for me?" As a result, she might get some reasons to change her better judgment. For example, she realises that although she could find more value in her better judgment, it would not bring her happiness because she would betray what she cares about deeply. Or, she might find that carrying out the better judgment is impossible for her, and, if she believes "ought implies can", she should give up the better judgment which implies an impossible "ought".

(2) Act out of caring about. In some cases (especially those having little moral impact), keeping the inconsistency between a better judgment and an action is not that worrying. For example, a person might think she ought to stick to a low-calorie diet to be very slim, all-things-considered. However, she usually fails to do it. Her roommate, who shares the better judgment as well as reasons for it with her, can do it for several years. The difference between them can be: the akratic person does not care about her appearance as much as her roommate does. Instead, she cares more about being energetic, eating foods she likes and so on. She might not take them as reasons better than those to be very slim, and she blames herself for being weak-willed. However, she is, generally, "happier" and more comfortable.

(3) Try to care about a better judgment. The Out-of-caring-about Proposal argues that what one cares about can be out of her direct control, and that is why it can be insensitive to a better judgment. However, it does not mean that we cannot change what we care about and the extent to which we care about certain things gradually. Think about a student who is working on her philosophy thesis. She knows she ought to spend five hours on her thesis to submit it on time, but she fails to do it akratically. After observing her classmate, who works on her thesis for more than seven hours without any tiredness, she thinks: "I suppose my classmate really loves philosophy. Three years ago, I decided to study philosophy because I was interested in it. However, I found it is not what I expected and I gradually lost my interest..." Instead of directly "strengthening her will-power" by self-command, she might try something else. She can read a

biography of a great philosopher, or she can contact her first philosophy teacher who evoked her interest in philosophy. In this process, she might not have new reasons to support her better judgment, which has already had the upper hand in her reasoning. Instead, she might care more about philosophy, and this does help her to overcome akrasia.

3. Circumstances

Since so far we have only focused on the relation between an agent's better judgment and what she cares about, it might result in a misunderstanding that akratic action is caused by two mental items separated from the larger psychological and social background. This is not the case, and many other mental items and social factors do have impacts to some extent, but this is compatible with the Out-of-caring-about Proposal. The aim of this section is to fit the proposal into the background.

3.1. Internal Circumstances

The first psychological item is attention. Mele argues that attention plays an important role in akratic action: "a person's focusing on and vividly representing the prospective pleasant results of an action, A, that is contrary to his better judgment may increase his motivation to do A even though these mental events are not reasons for him to be more motivated to do A. (1987, p. 79)" I think it is correct, and it can also serve an ancillary function in the Out-of-caring-about Proposal.

The impact of attention is neutral in the sense that it does not have an intrinsic inclination to support either a better judgment or an akratic action. However, when an akratic action does occur, it is likely that more attention is paid to the attractive aspects of the akratic action. For example, when Huck Finn focuses his attention on what he takes as a moral principle, he feels the

motivation to follow his better judgment and write a letter to report Jim. However, when he completes the letter, his attention is drawn to something else: “[Jim] would always call me honey, and pet me, and do everything he could think of for me, and how good he always was...” (Chapter 31) Then he cannot help tearing the letter up.

If we interpret Huck Finn’s case by the Out-of-caring-about Proposal, it is easy to see that: the attention Huck Finn pays to what he cares about strengthens the power of it to motivate the akratic action. If we further trace his attention, we might find it is because he cares about Jim or something relevant that he pays special attention to it, which in turn amplifies the power of his caring about in this specific situation. Therefore, attention and caring about can be mutually stimulating.

The second kind of psychological items is emotions. There are two views of the role emotions play in akratic action. First, the role is causal in the sense that it is one of the disruptive forces contributing to an akratic action. The Out-of-caring-about Proposal can accommodate it, because emotions can fuel the power of caring about and help volitional necessity. Likely, it can help one’s better judgment. Then its function is similar to that of attention¹⁸.

By contrast, the second role emotions might play is more comprehensive. Christine Tappolet (2003, p. 111) argues that “cases of akrasia caused by emotions involve a conflict between a value perception and an evaluative judgement that can be compared to perceptual illusions...” Let us call it the Emotion-based Proposal. According to it, emotions consist of certain ways of seeing an object. For example, fear might imply that an agent sees an object or a situation as dangerous, although it might not be supported by sufficient evidence noticed by the agent. It is because of the discrepancy between value perceptions contained in emotions and evaluative judgments based on explicit evidence that an agent can act against the latter due to the former. This makes

¹⁸ The relation between emotions and attention itself has significance independent of their similar roles in akratic action, which will not be discussed here. See Christine Tappolet (2003, p. 105).

some akratic action intelligible. Moreover, “it need not necessarily be the emotion that gets things wrong” (Ibid., p. 112), and sometimes akratic action can lead to a better result.

With regard to the Emotion-based Proposal, two points can be made. First, the way the concept of emotions is postulated and the explanatory power it is supposed to have are akin to those of the concept of non-deliberative reason-responsiveness in the Reason-responsive Proposal. Emotions’ sensitivity to situations can be integral to the non-deliberative reason-responsive process, as emotions might “make us more rational, in the sense of allowing us to track reasons which we have but which we have neglected in our deliberation” (Ibid., p. 115). Second, if emotions are linked to an implicit way of reason-responsiveness, their relation to caring about might be ignored¹⁹. However, the function of emotions is ambiguous and neutral, because they could stimulate both reason-responsiveness and caring about. If we take emotions as a kind of perception, they might be sensitive to both.

Therefore, we can fit the Out-of-caring-about Proposal into a bigger picture of the human mind, and see its compatibility with other attempts, based on different psychological items, to make sense of akrasia.

3.2. External Circumstances

Now let us move on to external circumstances consisting of personal histories and social expectations.

An agent’s personal history is external in the sense that it is unchangeable for an agent but it plays an important role in her life. A worry might be: if what an agent cares about as well as the degree to which she cares about certain things largely depend on her experience, does it mean that akratic action out of caring about is, in the deepest sense, a kind of compulsion or the domination of the past?

¹⁹ A demonstration of the relation between “caring about” and the corresponding “complex emotional structures characteristic of caring” can be seen in Agnieszka Jaworska, 2007a, p. 485.

Furthermore, if we think an agent's better judgment is subject to similar influences from her education and experience, this seems to undermine all the distinctions we make.

However, we can think of it differently. We have a variety of experiences because we stand in different relations to the past, and that is why we need distinctions among blind desires, caring about and reason-responsiveness. All of them might collapse if determinism is presupposed. However, first, we do not need to presuppose it and, second, the different relations are more obvious to us and shape our lives efficiently. Moreover, personal history has a facet benign to "caring about". That is to say, it is because personal histories shape our relations to certain objects to some extent that our caring about them can achieve its full complexity and richness. For example, as Joseph Raz points out, in *Little Prince*, when the little prince faces a crisis that his rose is not unique in appearance, he solves it by recognising that the rose is "made unique by the history of their love" (2001, p. 22). We will return to this point in Chapter 5.

As for social expectations, they can be part of personal histories. However, it needs special attention because in nearly all of the cases we have discussed, some "distorted" or "unreasonable" social expectations lurk. For example, Huck Finn is expected to turn Jim in, the son in the poem *Southeast the Peacock Flies* is expected to do what his mother tells him to do, and Tevye in *Fiddler on the Roof* is expected to strictly conform to traditions. The background of all these cases features closed societies, unfair social structures and biases from a modern point of view.

This might challenge the Out-of-caring-about Proposal. Is the Out-of-caring-about Proposal merely applicable to cases in which agents' better judgments are caused by distorted social expectations, which always provoke revolt from something rooted deeply in humanity? If this is the case, then the diagnosis offered by the Out-of-caring-about Proposal would be secondary, and the crux is the mismatch between social expectations or "bad morality" (Bennett, 1974) and universal humanity, which is demonstrated in individuals who have not been

thoroughly poisoned.

Admittedly, it is plausible to say that social expectations can increase or decrease the frequency of a certain type of akratic action. For example, in modern society, much fewer people would face the dilemmas that Huck Finn, the son and Tevye faced respectively. However, two points can be made to show the Out-of-caring-about Proposal, though compatible with the influence from social background, has significance independent of the “society-versus-individuals” picture.

First, it is worth noting that social expectations not only appear as disciplines shaping one’s better judgment, but also penetrate one’s way of life mildly and change what she cares about. Second, the significance of the social impact is not a phenomenon peculiar to conservative social backgrounds. As Amelie Rorty observes, some akratic action is rooted in self-contradictory social expectations, which the modern society is based on: “Social institutions and economic systems encourage and foster the very actions that they also condemn. While promoting habits of cooperation, they also reward radical independence; while condemning aggression, they also praise ‘aggressive initiative.’ While admiring selfless devotion, they also reward canny self-interest.” (1997, p. 652) Different notions and stimulating objects penetrate the members of the society, and it is not necessarily the case that these external factors mould members’ rational judgments in a more efficient way than that they influence what they care about. Therefore, it is plausible to say that Huck Finn’s caring about Jim is also influenced by a social expectation that “be nice to your friend”, and the son in the poem is moved by a social expectation that “marriage was meant to last”. Therefore, what the Out-of-caring-about Proposal emphasises is not conflicts between individuals and the social order in historical contexts, but an inconsistency within an agent, in which her intake of social expectations is not fixed.

4. Conclusion

In this chapter, a detailed explanation of the Out-of-caring-about Proposal is offered. The basic idea is that an agent cares more about what could benefit from her akratic action. It has three features. First, it implies a symmetry between akrasia and enkrasia. Second, its relation to moral praiseworthiness is not as direct as that of the Reason-responsive Proposal. Third, it draws our attention to a different capacity we have, and offers some suggestions to overcome or live with akratic action. In the end, it is fitted into the bigger picture of the human mind and society.

Chapter 5 Objections, Replies and an Implication

In this chapter, I shall consider two objections. The basic idea of the first objection is that the Reason-responsive Proposal actually leaves room for the partiality which caring about focuses on. Therefore, the Out-of-caring-about Proposal can be reduced to the Reason-responsive Proposal. However, I shall argue that it is not plausible. Second, adding a new mechanism of caring about seems to add a sub-agent, which might lead to a divided picture of the self and make normal cases of self-control “miracles”. However, I shall argue that, at least, this problem is not especially challenging to the Out-of-caring-about Proposal. Even for the Reason-responsive Proposal, the mechanism of “caring about” underlies reason-responsiveness. What the Out-of-caring-about Proposal differs from the Reason-responsive Proposal is *not* that the former requires a new system that is absent in the latter, but that the former relies on a new way of function of a system that has already been manifest in the latter. Then, a possible implication of the Out-of-caring-about Proposal will be shown.

1. An Objection from Reason-responsiveness

In Chapter 2 and Chapter 3, one of the motivations to put the Reason-responsive Proposal aside and consider the Out-of-caring-about Proposal is that: the former’s interpretation of cases like Huck Finn’s might be over-rationalised to the extent that it fails to capture some first-person features in akratic action.

However, this charge might imply a misunderstanding of the Reason-responsive Proposal. For example, facing a similar charge of being “excessively rationalistic”, Richard Moran says (2007, p. 53): “I think I see the aspects of my view that give rise to this charge, but I hope I can show that my account only looks that way if we start with an excessively rationalistic picture of reason.” Likely, if the Reason-responsive Proposal can offer an understanding of reason and reason-responsiveness which can accommodate the first-person features in

cases like Huck Finn's, it might be unjustified to see it as "over-rationalised", and it would be unnecessary to step out of the Reason-responsive Proposal and seek an alternative.

I shall take Joseph Raz's account (2001) as an example, though the question his arguments target is not akrasia but the relation between impartiality and partiality. I shall explain Raz's arguments, and point out how it could support the Reason-responsive Proposal, and what the Out-of-caring-about Proposal might reply to it.

Raz offers the concept of "attachments": "People derive a sense of purpose and of value in their life from their engagements with pursuits and relationships that they regard, implicitly or explicitly, as worthwhile, that is, from their attachment." (Ibid., p. 16) It is not difficult to find the similarities between it and the concept of "caring about". However, they are different because a fuller engagement of an agent's rational capacities is required by "attachments". First, the establishment of X's attachment to O normally implies X's recognition of O's impersonal value, otherwise this attachment would not be appropriate. In other words, the concept of attachment "refers only to ties of which we are conscious, and it excludes those we would rather be without but from which we cannot, at least, not without much effort shake ourselves free (Ibid., p. 17)". Second, if an attachment is appropriate, it could be a source of responsibilities: special responsibilities of "a citizen, a parent, a lover, an academic (Ibid., p. 34)". Therefore, the concept of "attachments" is more sensitive to reasons, because the attachments themselves are subject to the norms of rationality – they should be based on values, and the attachments can generate agent-relative reasons, for which an agent ought to act in a certain way.

If this argument is plausible, it might "update" the Reason-responsive Proposal. For example, we can explain Huck Finn's case in the following way: Huck Finn has an attachment to Jim, and this is based on impersonal value, say, that Jim is a nice man and he is kind to Huck Finn. Therefore, this attachment is appropriate. As a result, Huck Finn has a set of special responsibilities because

of his personal attachment, and he could take them as reasons. However, when he deliberates, he overlooks the set of good reasons which are tailored by his attachment and rooted in impersonal value. Nonetheless, when he helps Jim against his better judgment, he actually responds to these value-based reasons which are applicable to him because of his personal attachment.

Then what would the Out-of-caring-about Proposal say? A reply would be: yes, this would be a nice case if Huck Finn himself could think it in this way. However, it is worth noting that even if Huck Finn takes his personal attachment into account, it does not efficiently rationalise his akratic action. Huck Finn could see the value in his personal attachment, but he decides to set it aside to act morally (as he sees it) in this situation, all things considered. We have little evidence that, in this case, the attachment functions in a reason-sensitive way, although it could function via the reason-responsive channel, as described by the updated Reason-responsive Proposal. However, it does not mean that this attachment would be “pathological, and self-destructive” (Ibid, p. 17). Rather, it shows us the broader meaning and an independent function of attachments, which is what the Out-of-caring-about Proposal focuses on.

Let me elucidate the difference between “one’s attachment to O that tailors reasons within reason-responsiveness” and “one’s attachment to O as a non-rationalising motivation” by an example. Now electronic sports are popular in many countries. We can imagine 20 years ago, when the mainstream narrative tended to demean electronic sports, a teenager called Bobo was a huge fan of electronic sports and she could play some games very well. She did care about playing these games and she might claim “playing them is my passion”. However, she thought that her parents were right when they said that she ought to stop playing games to study hard, or at least largely reduce the time she spent on it, otherwise she could not have a bright future. Nobody could earn money or have a good social status by playing electronic games then. In this case, she did have a personal attachment to playing electronic games, and we could tell some impersonal value in it. However, whenever she played games, it is likely that she

could not develop reasons sufficient to rationalise her action.

By contrast, we can imagine a similar teenager called Coco, who lives in a society where the industry of electronic sports is developing and many people take it as their career. Happily, Coco thinks she ought to take playing electronic games as her career, all things considered. Then her personal attachment might be “appropriate”. However, I suppose, in Huck Finn’s case, the role of his attachment to Jim is similar to Bobo’s attachment to electronic games rather than the lucky Coco’s. In both Bobo’s case and Huck Finn’s case, the agents can develop reasons from the personal attachments but the reasons are not sufficient to rationalise their action, because they have been overridden by the opposite considerations. However, if we treat the attachment as a “caring about-cared about” relation, we cannot say that: the cases are mere ones of infatuation or obsession, given the irreducible role we have conferred to the concept of “caring about” in Chapter 3.

This is compatible with the rationalising role of one’s attachment to a certain object, but it would serve a counterexample to two views: the first is that attachment is always sensitive to reasons at a descriptive level; the second is that, at a normative level, attachments should be sensitive to reasons and they should be appropriate in the sense that they are in lockstep with the value of the objects seen by an agent. Raz compares the establishment of attachments to the concept of “taming” explained by the Fox in the *Little Prince*. By taming a valuable object voluntarily, an agent establishes a personal attachment to it voluntarily. However, what is worth noting in cases like Huck Finn’s is not how an agent can voluntarily establish attachments to “appropriate (impersonal) value, and make it meaningful” for the agent (Ibid., p. 19), but the facet of an agent’s “being tamed” when “taming”. In a more “romantic” tone, we can say that sometimes the attachments cannot be fitted into either a picture of rationality or one of madness, but it renders a quite different experience of “being tamed”, which is available to us: “After all, it is quite possible for someone to anticipate realistically that he would be unable to bring himself to pursue a certain course of action even if he

were to recognize it as the best. It may be clear to him that there are matters with respect to which he is incapable of acting rationally....” (Frankfurt, 1988, p. 184)

To sum up, the main idea of this objection is that “caring about” can be accommodated to a fine-grained picture consisting of reasons, value and attachments. However, the response of the Out-of-caring-about Proposal would be that the object of one’s caring about might not always be the one which is tamed by an agent, but the one by which an agent is tamed. It can be seen as a personal attachment, but its emphasis would be different from that in Raz’s account.

2. An Objection from the Picture of a Divided Self

The basic point throughout the thesis is that X’s caring about O cannot be reduced to either (1) X’s strong desire for O or (2) that X takes some value in O or the relation between X and O as reasons to act. However, it might lead us to a worry that: would this imply an independent mechanism or channel of the human mind?

First, I agree with Alfred Mele’s opinion (1987, p. 83) that when we add a “new system” to the picture of the human mind, we do not artificially create any sub-agent but try to accommodate certain results as a given: “To say that these elements constitute a system is (roughly) to say that together they generate (or are capable of generating) a certain result. The explanatory work is done by the elements and their interrelationships, not by the supposition that they are somehow partitioned off from other mental items.”

Second, if any explanatory attempt aiming at making sense of akrasia cannot avoid introducing different systems within an agent, then what the Out-of-caring-about Proposal has to prove is that: the picture of the human mind implied by it is no more fragmented than that implied by other proposals, such as the Reason-responsive Proposal. Apparently, the Out-of-caring-about Proposal has a new layer called “caring about” which is absent in the Reason-responsive Proposal.

However, this is not the case. Even in the Reason-responsive Proposal, a layer of this volitional structure is necessary. Arpaly uses the concept of “concern” to grasp it. She thinks one’s concern makes a difference in her motivational set, emotional makeup and cognitive features (2003, pp. 85-86). For example, if “one cares about morality, moral facts matter to one emotionally, and they are salient to one. As a result, other things being equal, a person of more moral concern will be more sensitive to moral features of situations...” (Ibid., p. 87). In this sentence, “caring about morality” and “having a moral concern” would be interchangeable. Accordingly, one’s sensitivity to a kind of reasons depends on the content of her concern.

For example, Huck Finn acts out of deep concern (Ibid., p. 92). Both the Reason-responsive Proposal and the Out-of-caring-about Proposal would accept this point. However, what the Out-of-caring-about Proposal emphasises is that caring about can function in a bare way. That is to say, it does not necessarily manifest as a certain kind of sensitivity to reasons. Rather, it could manifest as a kind of insensitivity to deliberation and reasons, which is caused by volitional necessity. This process would sidestep the explicit deliberation as well as any implicit alternate reason-responsiveness. Nonetheless, it is still an expression of caring about rather than blind urges. Therefore, we can see the Out-of-caring-about Proposal’s raw materials are no more than those of the Reason-responsive Proposal, but it argues that the relations among them are more complicated, and they can explain some discrepancies in our lives.

3. An Implication on Self-understanding

If we leave room for akratic action out of caring about and the framework underlying it, then it might change the way we cope with the self-knowledge we get and the self-understanding we could achieve. I shall start with an account of self-understanding offered by Richard Moran, and then point out how the Out-of-caring-about Proposal might challenge it.

Moran thinks people have two perspectives, from which they can get knowledge of themselves. The first is a deliberative perspective. From it, an agent does *not* have exclusive access (e.g. an inner perception) to her mental states. Rather, it “involves a stance of agency beyond that of being a kind of expert witness” (2001, p. 4). It implies a kind of “transparency condition”. When an agent is asked “do you believe p?”, normally she would not check her mental states to find a corresponding piece of belief p or -p, but she would consider the possibility of p directly by collecting evidence from the world and making inferences. Although by this deliberative process, she arrives at a judgment on a certain feature of the world (say, “yes, it will be sunny tomorrow!” or “I will go shopping tonight”), normally, this also serves as an avowal of a belief she has (say, “I believe it will be sunny!” or “I believe I will go shopping tonight”). Then we can see what “transparency” means: “a first-person present-tense question about one’s belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world” (Ibid., p. 62).

Moreover, in Moran’s opinion, human agents also have a theoretical perspective, from which they can observe themselves in a way similar to that they observe each other third-personally. They are able to discover a desire, a belief, an emotion or an intention within themselves, and sometimes they are impenetrable in the deliberative perspective. For example, a person might have some biases they do not take as reasons in their deliberation, and, in this sense, the biases (and their power) look “invisible” from the deliberative perspective. However, if she sees herself from the theoretical perspective, collects evidence (e.g. her actual attitudes and behaviour towards a certain group) and makes inferences, she may admit that she has these biases and she even predicts that she cannot change these attitudes in the near future, given her failures before. Then she achieves an attribution of a belief to her rather than an avowal. She can “report” an intention or an attitude of her, which can explain her action. However, in this case, the person is alienated from herself, for she loses the peculiarly first-

person perspective (i.e. deliberative perspective) and sees herself from a perspective she might share with other people.

This picture can throw light on akrasia, which exemplifies the clash between the two perspectives. Moran interprets an example taken from Sartre's *Being and Nothingness*. A gambler decides to avoid the gaming tables. This decision is one he makes rather than one he finds through evidence of his mental states. By this decision, he avows an intention he has. However, at the same time, he knows himself empirically. From his personal history consisting of many failures to carry through this kind of resolution, the resolution "appears to him then as an ungrounded, inconstant thing on which to base any confidence about what he will in fact do" (Ibid., p. 79). This is a report of the mental states he attributes to himself. Furthermore, Moran offers a diagnosis of akrasia: it is corrosive because "it often begins with the tactical substitution of the theoretical point of view for the practical one." (Ibid., p. 81)

Moran thinks the deliberative perspective is more philosophically interesting. For one thing, it helps us to understand what is special about self-knowledge. "It is self-knowledge through avowal, I claim, that is the form of self-knowledge that has attracted philosophical interest, and which exhibits a difference between first and third-person cases." (2007, p. 68). For another, it also throws light on a relevant puzzle that "what is human agency?" The sense of "agency" Moran has in mind is one engaged by "evidence, reasons and justification" (2003, p. 403), and orienting "oneself toward the question of one's beliefs by reflecting on what's true", or orienting "oneself toward the question of desires by reflecting on what's worthwhile or diverting or satisfying" (2001, p. 64). Therefore, we can see that: the deliberative perspective is not only an exclusively first-person way of acquiring self-knowledge but fundamentally constitutive to us as agents. It has a rationalistic colour. As Sebastian Gardner (2004, p. 266) points out: "Transparency is connected directly and firmly to rational personhood, and when Moran gives examples of its failure, these are all cases that, everyone would agree, call for its restoration". Akrasia is an example of its failure, for an akratic

agent is supposed to rely on the theoretical perspective to acquire self-knowledge.

We can draw a parallel between the Out-of-caring-about Proposal and Moran's distinction. The deliberative process underlying the better judgment can be seen as produced by the deliberative perspective in Moran's view. However, the capacity of caring about underlying the akratic action is significantly different from the theoretical perspective, but before revealing the difference, I shall show a similarity between them.

Both of them go beyond the picture of a rational agent and attempt to wrestle with some bare "facts" about the agent. The bare facts might flow from an agent's personal history and contribute to the way she normally orients herself in different situations. These facts might be taken as reasons, but it is not necessary. In Moran's account, from the theoretical perspective, an agent can see some sediments of her personal history, which are still active in shaping her life and even future. Likely, according to the Out-of-caring-about Proposal, when an agent cannot help caring a certain thing despite that her reasons direct her to the opposite way, she might also see how the past influences her and constitutes some undeniable facts of her. Therefore, in both Moran's way to cope with akrasia and the Out-of-caring-about Proposal, these facts serve as the rival of the better judgment.

Intuitively, the findings of the theoretical perspective are more like a daunting ghost from the past because they tend to cloud rational agents' present-tense judgments. However, the function of "caring about" in the Out-of-caring-about Proposal is more neutral, and it could be like a burst of benign sunshine emitting throughout the temporal flow sometimes. However, they are not mere two different ways to describe the same object (i.e. the former is more pessimistic but sensible while the latter is more optimistic but wishful), but show a substantial difference between the theoretical perspective in Moran's account and the concept of caring about in the Out-of-caring-about Proposal. I shall argue that the theoretical perspective described by Moran is either (1) not as purely "theoretical" as its third-person counterpart or (2) so extreme that it cannot serve the role

Moran gives it in explaining akratic action.

Let me amplify the first possibility. Moran emphasises that when an agent gets self-knowledge from the theoretical perspective, he relies on “the same kind of considerations available to others” (2001, p. xxx). However, it is, as far as I am concerned, hard to imagine such a theoretical perspective first-personally. This is because we are familiar with our situation as agents with a past or a personal history, no less than we are familiar with our “privileged” position as rational agents. That is to say, when we see the personal facts which are recalcitrant to our deliberation, we cannot be disinterested or neutral towards them. Instead, we might recognise ourselves embedded in them first-personally, in a way different, but no less significant, from taking them into deliberation. For example, in the case of the akratic gambler, he finds that although his decision to avoid gaming table is so vivid and undeniable from his deliberative point of view, he, from another point of view, is dimly or clearly aware of the facts that (1) his motivation to gamble is so strong (2) his hedonic disposition is entrenched and (3) he is not the kind of people who can take control of their own destinies, and he always disappoints himself and people who love him by not caring about certain things enough. Calling this process one from a theoretical perspective might be misleading to some extent, given that the agent (e.g. the akratic gambler in this case) normally does not treat what he sees as neutral facts. The bare way in which these facts appear to the gambler does not lead him to hold a third-person-like perspective. Rather, the gambler might find another undeniable fact: in a non-trivial way, his present-tense self is better reflected by what the facts show him rather than the decision that he has happily embraced.

Therefore, the perspective of “caring about” might depict the second perspective an agent has towards herself better. It does not mean that what Moran refers to by “the theoretical perspective” is impossible, but it is almost pathological. At least, it is doubtful that in the case of the akratic gambler, such a perspective can appear without rendering this common case far stranger than Moran supposes it to be. By contrast, the Out-of-caring-about Proposal draws our

attention to the complexity of the way we understand ourselves and blur the boundary between the two perspectives to some extent. The non-transparent part of human agency seen from the deliberative perspective might also play an important role, and people are not mere observers to this part who hold only a third-person-like theoretical attitude and expect empirical findings.

4. Conclusion

In this thesis, we focus on an interesting kind of akrasia, in which the agent acts against her better judgment but it does not mean that she is subject to compulsion. Three proposals have been considered. According to the Irrational-passive Proposal, the agent is motivated by irrational factors, and thus she is passive. However, this would be futile to make sense of the cases we are interested in. According to the Reason-responsive Proposal, the agent can respond to reasons although she acts against her better judgment, because reason-responsiveness is not exhausted by her deliberation. It is plausible to say that reason-responsiveness can be non-deliberative. However, in the specific cases of akratic action that interest us, the agents seem to lie in a situation in which they act “in spite of reasons”. Accordingly, they seem to get a cluster of first-person experience which is equal to neither that after a normal response to reason nor that caused by brute forces. To accommodate this peculiar first-person experience, the concepts of “caring about” and “volitional necessity” might be helpful. The general idea of the Out-of-caring-about Proposal based on the two concepts is that the akratic agent cares about an object to the extent that its motivating power cannot be fully rationalised by any reason she develops from her caring about the object. It might lead to two objections, and one implication has been discussed.

Two kinds of questions remain unanswered: one is “upstream” while the other is “downstream”. The “upstream” question is: what is the metapsychological or ontological status of caring about? We have tried to pin down it by

distinguishing it from other psychological items. However, the method is negative for we largely rely on acknowledging “what it is not”. Moreover, the method is horizontal in the sense that it might have an unknown deeper cause and require complementation from a detailed account of metapsychology. Only if we could figure out it, can we get the full picture of caring about and a complete genesis of akratic action out-of-caring-about/carelessness. The “downstream” questions are: is out-of-caring-about/carelessness action autonomous? Should akratic agents be responsible for their actions in these cases?²⁰ Although some hints of it have been given when we discuss the praiseworthiness of “inverse akrasia” as an action out of caring about, a more elaborate account is needed. This thesis aims to demonstrate the link between caring about and some cases in which agents act against their better judgments. It is because of the establishment of this link by the Out-of-caring-about Proposal that the two kinds of questions could be raised in this way, but the link itself does not favour any answers to them, and it does not even guarantee that they are answerable.

²⁰ Frankfurt’s answer is “yes”, and Agnieszka Jaworska argues that “carings function as the most elemental building blocks of the capacity for autonomy” (2007a, p. 494)

Bibliography:

Arpaly, Nomy. 2003. *Unprincipled Virtue*. Oxford University Press.

_____. 2006. *Merit, Meaning, and Human Bondage*. Princeton University Press.

_____. 2007. 'Reply to Harman, Stroud and Mason'. *Philos Stud.* 134:457-65.

_____. 2014. 'Duty, Desire and the Good Person towards a Non-Aristotelian Account of Virtue'. *Philosophical Perspectives*, Vol. 28, Ethics, pp. 59-74.

_____. 2015. 'Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance' in *The Nature of Moral Responsibility*. Oxford University Press. pp. 141-46.

Arpaly, Nomy. and Schroeder, Timothy. 2014. *In Praise of Desire*. Oxford University Press.

Bennett, Jonathan. 1974. 'The Conscience of Huckleberry Finn'. *Philosophy*. Vol. 49, No. 188, pp. 123-34.

Bloom, Irene translated. 2009, *Mencius*. Columbia University Press.

Bratman, Michael E. 1996, 'Identification, Decision, and Treating as a Reason'. *Philosophical Topics*, Vol. 24. No. 2, Free Will, pp. 1-18.

Broome, John. 2020 (forthcoming). 'Reasons and Rationality'.

Bulgakov, Mikhail. 2008. *Master and Margarita* (Translated by Richard Pevear and Larissa Volokhonsky).

Davidson, Donald. 2001. *Essays on Actions and Events*. Oxford University Press.

_____. 2004. *Problems of Rationality*. Oxford University Press.

Frankfurt, Harry. G. 1988. *The Importance of What We Care about*. Cambridge University Press.

_____. 1999. *Necessity, Volition and Love*. Cambridge University Press.

_____. 2002. "Reply to Gary Watson", in *Contours of Agency*, Edited by Sarah Buss and Lee Overton, The MIT Press, pp. 160-64.

_____. 2004. *The Reasons of Love*. Princeton University Press.

_____. 2006. *Taking Ourselves Seriously & Getting It Right*. The Tanner Lectures in Moral Philosophy. Stanford University Press.

Gardner, Sebastian. 2004. 'Critical Notice of Richard Moran, Authority and Estrangement: An Essay on Self-Knowledge'. *The Philosophical Review*, Vol. 113, No. 2, pp, 249-47.

Hill, Thomas. 1986. 'Weakness of Will and Character', *Philosophical Topics*, Vol. 14, No. 2: pp. 93-115.

Holton, Richard. 2009. *Willing, Wanting, Waiting*. Oxford University Press

Jaworska, Agnieszka. 2007a. 'Caring and Full Moral Standing'. *Ethics*. Vol. 117, No.3 (April 2007), pp. 460-97.

_____. 2007b. 'Caring and Internality'. *Philosophy and Phenomenological Research*. May. 2007. Vol. 74. (May, 2007), pp.529-68.

Kolodny, Niko. 2005. 'Why be rational?'. *Mind* 114, pp. 509–63.

Liu, Qingping. 2003. 'Filiality Versus Sociality and Individuality: On Confucianism as "Consanguinitism"'. *Philosophy East & West*. Volume 53. Number 2, pp. 234-50.

Mason, Elinor. 2007. 'Rationality and morality: thoughts on *Unprincipled Virtue*'. *Philos Stud*. 134, pp. 441–48.

Mele, Alfred. R. 1987. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. Oxford University Press.

_____. 2012. *Backsliding: Understanding Weakness of Will*. Oxford

University Press.

Moran, Richard. 2001. *Authority and Estrangement*. Princeton University Press.

_____. 2003. 'Responses to O'Brien and Shoemaker'. *European Journal of Philosophy* 11:3. pp. 402–19.

_____. 2007. 'Replies to Critics'. *An International Journal for Theory, History and Foundations of Science*, Enero 2007, SEGUNDA EPOCA, Vol. 22, No. 1(58), pp. 53-77.

Pugmire, David. 1982. 'Motivated Irrationality'. *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 56. pp.179-96.

Raz, Joseph. 2001. *Value, Respect, and Attachment*. Cambridge University Press.

Rorty, Amelie Oksenberg. 1997. 'The Social and Political Sources of Akrasia'. *Ethics*. Jul. Vol 107. No. 4. pp. 644-57.

Slingerland, Edward translated. 2003. *Confucius: Analects with Selections from Traditional Commentaries*. Hackett Publishing Company.

Stocker, Michael. 1979. 'Desiring the Bad: An Essay in Moral Psychology'. *The Journal of Philosophy*, Vol. 76, No. 12: pp. 738-53.

Stroud, Sarah. 2007. 'Moral worth and rationality as acting on good reasons'. *Philos Stud* 134:449–56.

Tappolet, Christine. 2003. 'Emotions and the Intelligibility of Akratic Action'. in *Weakness of Will and Practical Irrationality*. Edited by Sarah Stroud and Christine Tappolet. Clarendon Press. pp. 121-46.

Twain, Mark. 2001. *Adventures of Huckleberry Finn*. Hungry Minds, Inc.

Watson, Gary. 1975. 'Free Agency'. *The Journal of Philosophy*. Apr. 24. Vol. 72. No. 8, pp. 205-20.

_____. 1977. 'Skepticism about Weakness of Will'. *The Philosophical*

Review. Vol. 86, No. 3 (Jul., 1977) pp. 316-39.

The Economist. 24 & Ready to Die.

<https://www.youtube.com/watch?v=SWWkUzJ4M>

Southeast the Peacock Flies. Translated by Burton Watson.

<https://www.poetrynook.com/poem/southeast-peacock-flies>